

| | |
|--|---|
| <p>Dagobert Soergel Room 526 Baldy 716-645-1474 (O), 716-645-3775 (OFax) 703-823-2840 (H), 703-823-6427 (HFax) 703-585-2840 (cell) dsoergel@buffalo.edu www.dsoergel.com <i>Office hours by appointment</i></p> | <p style="text-align: right;">Amy Miller Room 549 Baldy Amy Miller <anmiller@buffalo.edu> <i>Office hours TBA</i></p> |
| <p style="text-align: center;">Web site: www.dsoergel.com/571/ (All materials searchable, advanced materials)</p> | |

Department of Library and Information Studies. University at Buffalo

LIS 571

Organization and Control of Recorded Knowledge

Summer 2012

| | |
|--|--|
| <p>Catalog Description</p> | <p>Knowledge analysis and representation; information presentation and assimilation; bibliographic and record control.</p> |
| <p>Purpose of the course: prepare for <i>searching</i> and <i>cataloging</i></p> | <p>The course will make you a better searcher and start you off on cataloging by giving you a fundamental understanding of modern principles of knowledge organization. It will prepare you for advanced courses and for selecting areas to pursue. It will prepare you for lifelong learning. The emphasis is not on specific skills but on concepts that give you the foundation for acquiring a wide range of skills as required by the tasks at hand over a life-time career. You will be prepared to become a leader in the field.</p> |
| <p>General learning objectives</p> | <p>Theoretical foundation of all types of information systems You will acquire the theoretical foundation needed to understand and apply a range of concepts and techniques to create and fully use traditional, modern, and future information systems:</p> <ul style="list-style-type: none"> • paper and digital libraries, • the Web, • management information systems, content management, intranets, and enterprise portals in organizations, • substantive databases, • artificial intelligence and expert systems (e.g. , for diagnosis). The Semantic Web. <p>Put differently, you will understand the primary conceptual structures that organize data, information, knowledge, language, and text. Often the course will not create full knowledge but open a window into more advanced work.</p> <hr/> <p>User orientation as an essential principle You will acquire a user-oriented (problem-oriented, request-oriented) approach to the design and operation of information systems.</p> |

LIS program objectives met by the course

- 1.1 **Graduates understand the nature of information** and its role in learning, research and scholarship, business, society, and culture.
- 2.1 Graduates are able to analyze information needs and to design, promote, and assess information services. (related to 4.1 Management)
- 2.3 **Graduates understand and are able to apply principles of knowledge organization for a wide range of applications, from organizing a collection to expert searching to support for learning.**
- 2.5 Graduates understand and are able to apply the principles of information seeking and of reference and user services in different modes of communication (face-to-face, telephone, computer communication) for users of all ages and groups, including those with diverse styles of information use and diverse physical and intellectual abilities and needs. **Graduates understand and are able to apply a wide range of advanced search techniques.**
- 4.3.1 Graduates are equipped and motivated to use information to transform lives and to promote equity, mutual respect, and a rich social fabric in a diverse society.
- 4.3.4 Graduates will demonstrate understanding of, respect for, and sensitivity to the diversity in society, including age, culture, economic means, ethnicity, language, physical and mental ability, race, and sexual orientation.
- 4.6 Graduates are ready to serve as leaders and advance the field.
- 5 Graduates understand the importance of personal qualities conducive to professional success. The program fosters the development of professionals with such qualities. Especially Assertiveness, Innovativeness.

Information about the instructor (UBlearns)

GSE Faculty spotlight
Short CV
Full CV

Note 1 Small Groups 1, 2, 3. In the seated class, these lectures/class meetings are done in small groups with much interaction and work on the board. For the online class this designation signals that there is a PowerPoint presentation in which you need to work on solving small problems before going to the solution presented on the next slide

Note 2 Summer. In the regular semester there are 14 weeks, and the course is divided accordingly with lectures numbered 1.1 is Week 1, Lecture 1, 1.2 is Week 1, Lecture 2, ... In the summer the work is condensed into 12 weeks, but there is still the same number of lectures. It would be impractical to change the lecture numbers, so weeks are renamed units; there are still 14 units with two lectures each.

Teaching Statement

"There are those who look at things the way they are, and ask why?
I dream of things that never were, and ask why not?"
George Bernhard Shaw as paraphrased by Robert F. Kennedy.

Guiding principles. I endeavor to inspire students to look beyond present practice – which, of course, they need to understand – to what could be, so they can become true leaders in the field. Second, I guide students to an understanding of basic, often timeless, concepts and theories – always illustrated by examples and assignments that connect to practice – so that they have a firm foundation both for lifelong learning and for critically examining present practices and reinventing them as needed to better serve users. Thirdly, I imbue students with a spirit of user orientation, so that they bring an understanding of user tasks, sense-making processes, and resulting requirements to everything they do, from structuring classifications and designing Web sites to indexing to formulating queries to presenting results. In all of this I make students think.

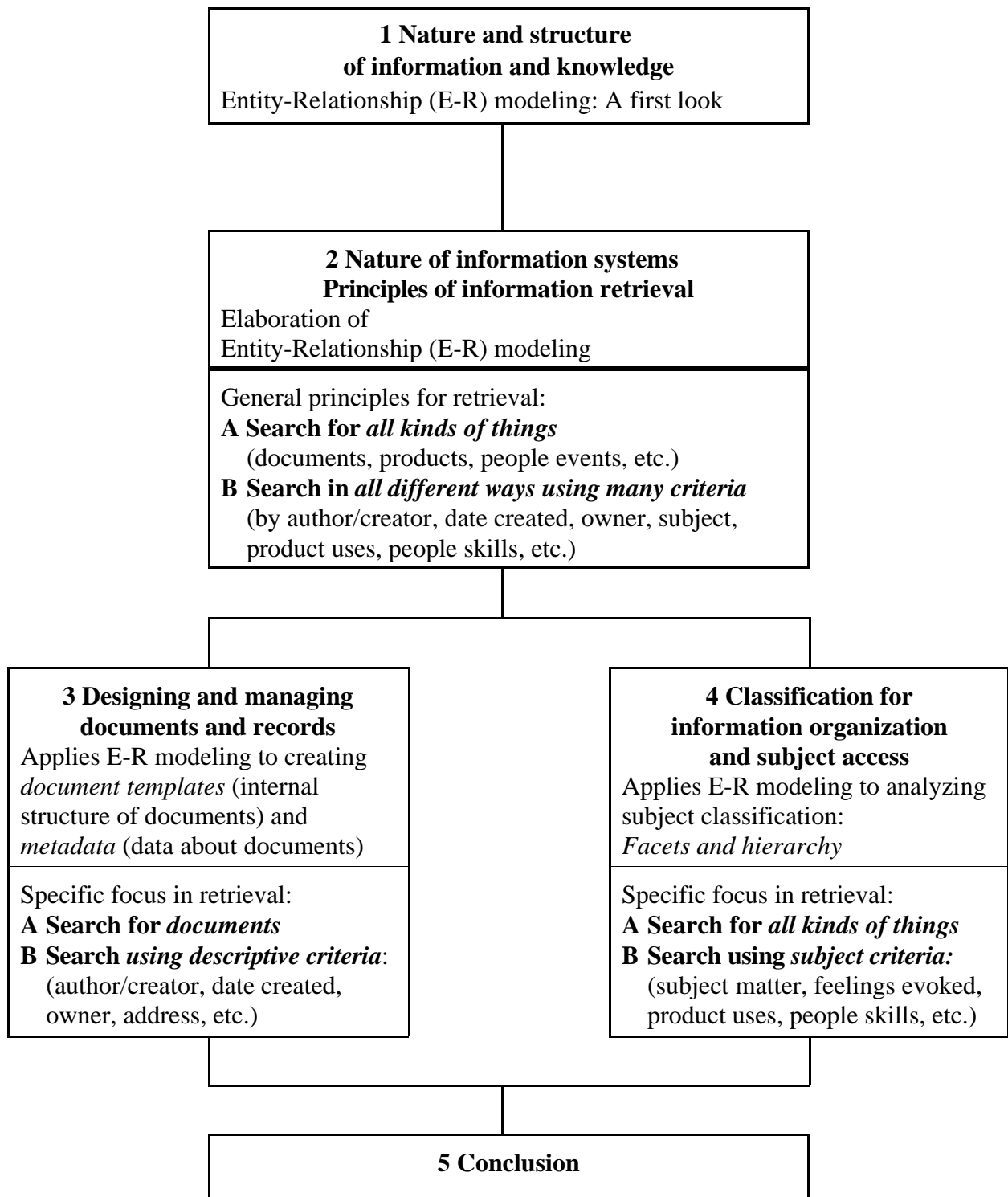
Implementing these principles requires careful structuring of course materials – often creating a new conceptual framework – and good document design. It also requires guiding students to their own discovery of ideas. For this purpose I often conduct small group sessions in front of a blackboard on which a framework evolves step by step from student contributions. Face-to-face class sessions with lecture, discussion, and in-class exercises supported by extensive lecture notes on paper have served me well in pursuing my teaching goals.

I pay great attention to choosing good examples students can relate to, especially examples that illustrate several concepts and principles, so that students do not need to get familiar with a new example every time a new concept is introduced and, perhaps more importantly, so students can see how several concepts work together in practice.

The nature of the course

| | |
|--|---|
| <p>Broad course, prepares for wide range of jobs</p> | <p>The course considers all kinds of information systems and a wide range of concepts and techniques. Emphasis is on concepts rather than specific skills to provide a solid basis for life-long learning as required by ever-changing practice. The course introduces some topics as windows into new areas so you can discover interests you want to pursue in more depth later.</p> <p>The course prepares you for pursuing a wide range of interests. It generalizes insights gained in the library world over decades or even centuries and makes them useful for other contexts, such as expert systems, digital libraries, content management systems, or the World Wide Web, where they are desperately needed. Library cataloging is included as an example of general principles.</p> |
| <p>Content: structure & representation of information</p> | <p>Information and knowledge structure serves two purposes:</p> <ol style="list-style-type: none"> (1) finding and applying information (as in the Semantic Web); often involves logical inference or statistical processing to derive an answer or have a system take action. (2) presenting information in a structure and format the user can assimilate and understand so that she can put the information to use. Thus, we must consider the structure of information in the user's mind. This is crucial for learning and instruction. <p>Information Architecture and Instructional Design, each from its own perspective, use the study of Information Organization and contribute to it.</p> |
| <p>Four parts: From principles to applications</p> | <ol style="list-style-type: none"> 1 Basic nature and structure of information and knowledge 2 The nature of information systems & principles of information retrieval, which are applied and made more concrete in Parts 3 and 4 3 Designing and managing documents and records 4 Classification for information organization and subject access |
| <p>Theory for improved practice</p> | <p>The course presents theory in the service of improved practice. You will</p> <ul style="list-style-type: none"> • be able to apply course concepts to practical problems; • understand the principles that underlie current practice and thus be able to do a job intelligently; • have a sound basis for changing current practice for new requirements. |

The structure of the course (use with Lecture 1)



Materials for the course

| | |
|--|---|
| Syllabus, lecture notes and assignments | <p>Calendar (pink). “Information Central”, gives for each lecture:</p> <ul style="list-style-type: none"> - the theme(s) of the lecture - the readings (do beforehand to prepare for the lecture) - the assignments (do afterwards to apply and reinforce the concepts). |
| | <p>Lecture Notes (pink dividers)</p> <ul style="list-style-type: none"> - Augmented, as appropriate, by recorded lectures, PowerPoint presentations, and guides to the lecture notes |
| | <p>Assignments (gold dividers)</p> <ul style="list-style-type: none"> - Descriptions of tasks, worksheets, and deliverables to be handed in - Assignment materials (explanations, examples, materials to work with) - Word templates for assignment deliverables (on Ublearns) |
| Text | <p>Soergel, Dagobert Organizing information. Principles of database and retrieval systems. Orlando, FL.:Academic Press; 1985. 450 p. ISBN 012-654261-9 Used copies may be available (\$20-40), new copies from instructor (\$80)</p> |
| Readings | <p>Reading packet (white). Required and some optional readings, arranged by lecture, green dividers. Model Catalog (first reading): MARC records illustrating descriptive and subject cataloging. Used throughout, examples for Assignments 8 and 13 Electronic version on Ublearns.</p> <p>Other useful book: Jonassen, David H.; Beissner, K.; Yacci, M. Structural knowledge. Techniques for representing, conveying, and acquiring structural knowledge. Hillsdale, NJ: Erlbaum; 1993. 265p.</p> <p>Optional readings go deeper or cover advanced topics. Some are included in the reading packet and the lecture notes+assignments. Available on Ublearns as pdf.</p> |

The materials come in three three-ring binders:

- Syllabus and lectures (pink)
- Assignments (gold)
- Readings (green)

For each binder, there is a bookmark. Move the bookmark along as you go to always keep your place.

Conduct of the course

| | |
|---|--|
| <p>For each unit</p> <p>Presentations</p> | <ul style="list-style-type: none"> • Do the readings or you will get lost in presentations. Note: You will figure out which readings require careful study and for which you just need to get the gist; partially that depends on your special interests. Optional readings can be done after the presentations. • In the Lecture Notes, read the objectives, notes on practical significance, and discussion questions. • Listen to / view the presentation(s) for the unit (PowerPoints with narration and/or voice recordings to go with the lecture notes on paper). • Think about the discussion questions. There is a thread for each question but likely not enough time for thorough discussion of each. • Read the assignment(s) and post questions if you need clarification. • Do the assignments |
| <p>Assignments</p> | <p>13 assignments (typically 2 -3 hrs) to practice the course concepts. Unless stated otherwise, assignments can be done in teams of 2 -3 with one copy submitted, except individual essays.</p> <p>Assignments are for learning; they must be submitted, and you will get feedback. But what really counts (for taking further courses and ultimately in the work place) is what you have learned through doing the assignment and reading the answer sheet (see Grading below.) Do not work too hard on the assignments; when you go more than 50% over the suggested time, you are doing something wrong, ask the TA. The answer sheets are generally more complete than what is expected from your answers.</p> <p>Assignments 8 and 13 are required to receive a grade in the course; you can work on them in teams but you must hand them in individually. These assignments are required because they are not covered in the final.</p> <p>Students who skip the assignments will not master the concepts or integrate them in their knowledge and thus will fall behind.</p> |
| <p>Class interaction</p> | <p>You are encouraged to form study groups and arrange virtual meetings with the instructor or the TA.</p> <p>Postings to the Ublearns discussion board may be useful, but this class depends less on discussion than other classes</p> <p>You are welcome to post or email questions to the TA or the instructor.</p> <p>Before you spend hours on figuring something out, ask.</p> |

| | |
|---|--|
| <p>Midterm, Term paper, Final</p> | <p>Midterm (take-home, <u>closed-book</u>, approx. 90 min) for feedback. Closed book so you can test what is in your head. Sample questions on pages following Lecture 7.2 in the lecture notes.</p> <p>Term paper / end-of-term essay. (See form following Assignment 6) A reflective essay on a theme of your choice, guideline 4,000-5,000 words:</p> <ul style="list-style-type: none"> - <u>make connections</u> between ideas gleaned from the readings, assignments, and class discussions; - consider implications and applications; - identify issues, questions, problems for further study. <p>This can be a personal essay that reflects your personal experience and views. Starting with Lecture 1 and continuing throughout the course, jot down ideas and connections as they occur to you. Short description due June 11.</p> <p>Final (open-book, 5 hours) covering concepts from all lectures, readings, and assignments, except specifics from Assignments 8 and 13. Sample questions in Lectures 14.1 and 14.2. Open book. You may find it useful to compile review notes.</p> |
| <p>Course requirements All must be fulfilled</p> | <p>Final Term paper / end-of-term essay Satisfactory completion of Assignment 8 (Descriptive cataloging practice) and Assignment 13 (Subject cataloging and searching practice).</p> |
| <p>Grading</p> | <p>Grading is based on the student's understanding and ability to apply course concepts to practical problems. The instructor will consider all available evidence.</p> <p>The final exam and the end-of-term essay are the primary sources of evidence for grading, with more emphasis on the better of the two.</p> <ul style="list-style-type: none"> • If the final and paper result in a grade of A, that is the grade. • If final and paper result in a grade below A, the following will be considered as subsidiary evidence that might improve the grade: <ul style="list-style-type: none"> • midterm (whose main purpose is feedback), • the assignments (whose main purpose is learning), • comments and questions on the discussion board, or in other interactions with the instructor, • (at the student's option) the student's learning blog/diary (see p. 9). <p>Since grading is based on understanding achieved and not points earned, there is some uncertainty. However, there is safety net. If you have done good work all semester, you should not be anxious about the grade. If a student has put forth adequate effort but still has not achieved a grade of B, I will provide guidance for further study which is intended (but not guaranteed) to lead to the understanding needed for a satisfactory grade. It is very rare that a student does not get a B.</p> |

Learning blog / diary

Keep a learning blog/diary in which you enter for each unit

- what have I learned, what was most important, what was most interesting, what was extraneous; what helps me in my (future) work? How?
- How does a course idea support better service to users, directly or indirectly?
- How does a course idea relate to other course ideas? For example, how is a course idea a manifestation of the E-R approach?
- Comments on readings – what did it contribute, how hard was it, ...
- What did I not understand? How does my not understanding this affect my (future) work?
- What questions do I have?
- Course **critique and suggestions**;

Feel free to post any of your observations, especially questions, to the "free write" forum for the unit; you can do so anonymously. All free write posts will be compiled and questions answered in a document prepared by the instructor..

Academic Integrity

"When an instance of suspected or alleged academic dishonesty by a student arises, it shall be resolved according to the procedures set forth herein. These procedures assume that many questions of academic dishonesty will be resolved through consultative resolution between the student and the instructor."
Full statement at www.grad.buffalo.edu/policies/academicintegrity.php

Religious Holidays

It is a policy throughout the State University system that: "on those religious holidays when members of a faith typically observe the expectation of church or synagogue that they be absent from school or work, campuses will avoid the scheduling of such events as registration, the first day of classes, or student convocations, and individual students will be excused from class without penalty if expressly requested." (From SUNY Policy Manual, 1975, Section No. 091.3.) If such a requested absence results in a student's inability to fulfill an academic requirement of a course scheduled on that particular day, the instructor should provide an opportunity for the student to make up the requirement without penalty.

Accommodations for students with disability

Please work with UB Disability Services to receive the support that enables you get the most out of your education. I am also looking forward to working with you to make the course a good experience. "Government regulations require that university policies, practices, and procedures not discriminate on the basis of disability. Disability Services coordinates reasonable modifications so that individuals with disabilities can access and benefit from all programs, services, and activities of the university."
www.ub-disability.buffalo.edu/servc.php

Reference list of detailed objectives and practical applications Skip for now but

Refer to these objectives throughout the class and read again towards the end of the class.

| | Main theme | Objectives: You should | Information professional task supported |
|---|---|---|---|
| 0 | User orientation | .1 have the spirit of user-orientation. | - everything information professionals do |
| 1 | Types of knowledge Types of concepts Lect. 2.1-2.2, 5.2-6.1 | .1 understand the characteristics and facets of different types of knowledge; .2 be able to apply this understanding to an analysis of information needs, to the organization of information, and to the evaluation of information found. | - understanding information needs (as in a reference interview); - analyzing and assessing info. found; - organizing different kinds of knowledge in an information system. - understanding how people think; thus - understanding how people ask questions and how they process information; thus - presenting the right content in the right form. |
| 2 | Knowledge representation Document design Lect. 2.1-2.2, 5.2-6.1 | .1 understand the main issues in and approaches to knowledge representation; apply this understanding in the analysis and design of information systems. | - adapting answers to users' mental structures; - searching effectively; - organizing a body of knowledge for simple retrieval, and for more complex inference. (Inference enables a system to draw conclusions from the knowledge stored, creating new knowledge.) |
| 3 | Information structure Lectures 1.1-3.1, 4.1-5.1 | .1 be able to design the conceptual data schema for a new info. system; .2 be able to analyze the conceptual data schema of an existing information system; .3 be able to apply this understanding to indexing and query formulation. | |
| 4 | Retrieval performance measures Lecture 3.2 | .1 understand retrieval performance measures and be able to apply them to: the specification of individual search requirements; the determination of optimal search effort; the evaluation of search results; and the evaluation of an information system as a whole. | - establishing and meeting the requirements for individual searches; - selecting information systems to be acquired; - designing information systems to meet performance criteria. |
| 5 | File organization and search algorithms Lecture 5.1 | .1 understand basic principles of file organization, storage structures, and index structures .2 understand search algorithms, including ranked retrieval. | - searching: formulating simple Boolean queries; using knowledge of search algorithms for effective searching; - designing efficient storage structures, incl. the use of hierarchical inheritance. |
| 6 | Information structure and usability. Document design Lectures 5.2-6.1 | .1 understand the relationship between information structure/representation and usability; .2 be able to apply principles of document design and text structure to the creation and selection of good documents. | - designing systems that can generate tailor-made documents; - preparing customized answers; - delivering search results in a useful layout; - evaluating documents. |

| | Main theme | Objectives: You should | Information professional task supported |
|----|--|--|--|
| 7 | Linguistic techniques: syntactic and semantic analysis Lecture 5.2 | .1 be aware of linguistic techniques and their applications in info. retrieval (IR). .2 [understand the basics of parsing sentences and semantic analysis, including word sense disambiguation.] .3 understand anaphora and its effect on retrieval and fact extraction. | - replacing labor-intensive human processing of vast amounts of text with automated techniques for text processing based on linguistics. Many commercial products available. - understanding free-text search systems; - being aware of systems that extract from text just the data the user needs; - being aware of automated translation - increasingly important with globalization, and very important in the Web. |
| 8 | Descriptive cataloging of documents Lectures 6.2-7.2 | .1 understand the application of general information structure principles to the descriptive cataloging of documents; .2 be aware of the variety of codes for bibliographic description; and .3 be able to catalog consulting AACR2. | - understanding and applying traditional descriptive cataloging; - describing and organizing electronic documents — in an organization's collection of documents/records or in the huge document collection on the World Wide Web (metadata). |
| 9 | Vocabulary control Lectures 8.1-8.2 | .1 understand the problems and principles of vocabulary control and be able to apply these principles to indexing and searching. | - designing systems that help users cope with vocabulary confusion through controlled vocabulary or query term expansion; - doing good searches without such support. |
| 10 | Functions of classification Lectures 8.2, 9.2, 11.2 | .1 understand the functions of classification in information retrieval systems, especially request-oriented indexing and inclusive searching. .2 Understand the functions and importance of classification for a wide range of other tasks. | - applying request-oriented indexing for improved system performance; - exploiting the vast intellectual capital available in classifications for functions beyond bibliographic retrieval, at the same time increasing compatibility and synergy between different information functions. |
| 11 | Structure of subject classification: Facet structure and hierarchy Lectures 8.1-13.2 | .1 understand the principles of the structure of subject classification, in particular facet organization and hierarchy; .2 be able to apply these principles to the analysis of existing schemes and to indexing and query formulation. | - indexing (cataloging); - analyzing a search topic and formulating the query, using hierarchic expansion; - designing systems that assist users in search topic clarification and in query formulation; - developing classifications / thesauri. See also functions of classification above. |
| 12 | Familiarity with specific subject classification schemes Lectures 12.1-13.2 | .1 be aware of the variety of classification schemes, thesauri, etc. and be acquainted with major American schemes: Yahoo (or DMOZ) Classification, Dewey Decimal Classification, Library of Congress Classification, Library of Congress Subject Headings. | - provides a door into a vast array of sources of knowledge about concepts and terminology, knowledge that lets you apply the general principles of facet structure and hierarchy to high-quality indexing (cataloging) and searching. |

Outline and calendar. Overview

Part 1. Foundations. Knowledge and knowledge representation

| | | |
|-----------|--------|---|
| * 1.1-1.2 | May 21 | Intro. and overview. Information systems and information structure. |
| 2.1 | May 21 | The nature of knowledge. |
| 2.2 | May 28 | Knowledge representation |

Part 2. Information retrieval: General principles and methods

| | | |
|-----|--------|--|
| 3.1 | May 28 | The structure of information systems. |
| 3.2 | May 28 | Objectives and performance measures for information systems |
| 4.1 | June 4 | An integrated information structure model |
| 4.2 | June 4 | Data schemas and formats |
| 5.1 | June 4 | Access to information: data structure & search modes. Retrieval as prediction. Ranking |

Part 3. The nature, design, and management of documents and records

| | | |
|---------|---------|--|
| 5.2-6.1 | June 11 | Document function, structure, analysis, and design. 5.2a Knowledge (re)presentation in text and images. Text linguistics. 5.2b Text analysis. [5.2c Parsing, in Supplement] 5.2d Document design 6.1a Document macrostructure. Document templates 6.1b Document markup languages |
| 6.2-7.2 | June 18 | Cataloging and metadata. Bibliographic control: description, entries and access 6.2a-c Metadata & description 7.1a Entries 7.1b RDF & Dublin Core 7.2 Descr. exercise |
| | June 18 | <i>Take home midterm exam distributed. Due M June 25</i> |

Part 4. Classification and subject access

| | | |
|------|---------|---|
| 8.1 | June 25 | Small Groups 1. Explorations in subject access Continuing Ass.10. (to be scheduled) |
| 8.2 | | Vocabulary control. Lexical relationships. Index language functions |
| 9.1 | July 2 | Small Groups 2. Index language structure 1: conceptual (to be scheduled) |
| 9.2 | | Small Groups 2. Application of index language structure to searching |
| 10.1 | July 9 | Small Groups 3. On constructing a hierarchy from facet combination (to be scheduled) |
| 10.2 | | 10.2a Brief introduction to Assignments 13.1 - 13.4 10.2b Introduction and in-class exercise: Ass. 13.1 Dewey Decimal Classification (DDC) |
| 11.1 | July 16 | Introduction and in-class exercise: Assignment 13.2 Yahoo |
| 11.2 | | Introduction and in-class exercise: Assignment 13.2 LCC |
| 12.1 | July 23 | Index language structure 2: database organization |
| 12.2 | | Media Streams Demo (indexing movie scenes using icons as descriptors) (short) |
| 13.1 | July 30 | Exploration of Knowledge Organization Systems (KOS) |
| 13.2 | | Indexing and system performance (conceptually also belongs to Part 2) |

Part 5. Conclusion

| | | |
|------------|----------------|--|
| 14.1- 14.2 | Aug. 6 | Final review |
| | Aug. 10 | Final exam Posted F Aug. 10, due M Aug. 13 |
| Wednesday | Aug. 15 | Term paper due. Last day for Assignments 8 and 13 (email if you need an extension) |

* Lecture number: 1.1 is Unit 1, Lecture 1, 1.2 is Unit 1, Lecture 2, ... | Each block = one sheet in the calendar

List of Assignments

| No | Assignment | Assigned | Due |
|--------|--|--|--|
| 1 | Hypermedia exploration: Perseus (Lecture 1.2; 2.5 hours) | May 21 | May 28 |
| 2 | Bibliographic retrieval system exploration: MEDLINE (Lect. 1.2; 3 h) | May 21 | May 28 |
| 3 | Online catalog search exercise (Lecture 1.2; 1.5 hours) | May 21 | June 4 |
| 4 | Restructuring a semantic network (Lecture 2.2; 1 hour) | May 28 | June 4 |
| 5 | Analytical description of an information system (Lecture 3.1; 3 h) | May 28 | June 4 |
| 6 | Developing a conceptual data schema (Lecture 4.2; 1.5 hrs) | June 4 | June 11 |
| | Short description of term paper (Lecture 5.1) | June 4 | June 11 |
| 7 | Apply linguistic techniques to retrieval problems (Lect. 5.2b; 2 h) | June 11 | June 18 |
| 8 | Descriptive cataloging practice (Lect. 6.2c; 4 hrs) (flex. due date) | June 18 | June 25+ |
| 9 | Problems of entry (Lecture 7.1a; 1.5 hours) (flex. due date) | June 18 | June 25+ |
| 10 | Indexing of three documents and prep for Lecture 8.1 (2 hrs) (do before Lecture 8.1) | June 18 | June 25 |
| *** | Take-home midterm, covers Units 1 - 7 (1.5 hrs) | June 18 | June 25 |
| 11 | Request-oriented indexing (Lecture 8.2b, 2 hours) | June 25 | July 2 |
| 12.1-3 | Conceptual analysis and synthesis (Lecture 9.1- 10.1) (total 7 h) | | |
| 12.1 | Semantic factoring (Lecture 9.1) (1.5 hours) | July 2 | July 9 |
| 12.2 | Building a hierarchy of elemental concepts (Lecture 9.1) (1.5 h) | July 2 | July 9 |
| 12.3 | 12.3a Practice Hierarchy from facet combination with education concepts (Lecture 10.1) (2 hours) 12.3b Real Hierarchy from facet combination with concepts from 12.1 / 12.2 (Lecture 10.1) (2 hrs) | July 9 | July 16 |
| 13.1-4 | Subject cataloging and searching practice (Lect. 10.2 - 11.2) | see each | see each |
| | 13.1 Dewey Decimal Classification DDC (Lect, 10.2)(4 hrs) | July 9 | July 16 |
| | 13.2 Yahoo: Yahoo (or DMOZ) classification (L.11.1) (6 h) OR LCC: Library of Congress Classification (L. 11.2)(6 h) OR DDC 2 More practice with DDC (no Lect.) (6 hours) OR Choice in consultation with instructor | July 16 July 16 July 16 July 16 | July 23 July 23 July 23 July 23 |
| | 13.3 Libr. of Congress/Sears Subject Headings (LCSH) (no Lecture) (3 h) | July 23 | July 30 |
| | 13.4 ERIC Thesaurus (no lecture) (3 hours) | July 23 | July 30 |
| | Term paper | May 21 | Aug. 15 |

Outline and Calendar

Units 1 - 14

Model catalog

Required. Refer to this throughout the course.

- 1 Soergel, Dagobert.
Model Catalog for LIS 571. Including a summary of the MARC Format.
August 2005. 46 p.

The model catalog gives many examples of cataloging documents, including a Web site, using AACR2R (Anglo-American Cataloguing Rules. 2. edition, revised) and the MARC (MACHINE Readable Cataloging) format.

It includes an **outline of the MARC format** for study and provides **examples** for

Lecture 4.2. Data schemas and formats,

Lecture 5.1. Access to information: data structure & search modes. Retrieval as prediction. Ranking (through the searching capabilities with Library Master).

Lectures 6.2-7.2. Bibliographic and record control. General issues, description, entries and access
and

Assignment 8. Descriptive cataloging of three documents

Lectures 12.1-13.2 Subject cataloging and searching practice
and

Assignment 13. Subject cataloging and searching practice.

Electronic version for Library Master on UBlerns.

Needs to be installed following the instructions in the paper/pdf version

Outline and Calendar

Unit 1. May 21

Lecture **Part 1. Foundations. Knowledge and knowledge representation.**

| | |
|------------|---|
| 1.1 | <p>Overview of the course and course materials (35 min)</p> <p>Introduction: Information Professionals in the 21st century (30 min)</p> <p>www.dsoergel.com/571/UBLIS571Lecture01.1Introduction.mp3 www.dsoergel.com/571/UBLIS571Lecture01.1Introduction.docx (text) www.dsoergel.com/571/UBLIS571Lecture01.1OverviewSlides.zip</p> |
| 1.2 | <p>Information systems and information structure (70 min)</p> <p>www.dsoergel.com/571/UBLIS571Lecture01.2Slides.zip www.dsoergel.com/571/UBLIS571Lecture01.2ExampleSlides.zip</p> |

Readings and assignments see on back

Unit 1. May 21, continued

To prepare

Read beforehand. Optional. If you have trouble early in the course, come back to these.

General introductory readings (UBlearns))

Soergel, Dagobert.

Information retrieval

Information organization

Berkshire Encyclopedia on Human-Computer Interaction. 2004 UBlearns, **Optional**
Overview articles directed at a general audience. Overview of many course concepts

Berners-Lee, Tim; Hendler, James; Lassila, Ora

The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities

Scientific American. 2001 May (get from the Web) A famous article
www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21 **Optional**

Lecture 1.1 Introduction: Information Professionals in the 21st century

Special Libraries Association

SPECIAL LIBRARIANS Putting Knowledge to Work

www.sla.org/content/SLA/professional/meaning/what/index.cfm UBlearns, **Required**

Competencies for Information Professionals of the 21st Century

www.sla.org/content/learn/comp2003/index.cfm (electronic copy of excerpts on UBlearns)

Excerpts **Required**, full document **Optional**

U.S. Department of Labor. Bureau of Labor Statistics

Occupational Outlook Handbook. Librarians (a quite traditional view, limited)

<http://stats.bls.gov/oco/ocos068.htm> **Optional**

Lecture 1.2 Information systems and information structure

No readings

Assignments assigned

Expanding on Lecture 1.2 Information systems and information structure

- ▶ **Assignment 1, Hypermedia exploration: Perseus and Freebase** (due May 28) (2.5 hours)
- ▶ **Assignment 2, Bibliographic retrieval system exploration: MEDLINE** (due May 28) (3 hours)
- ▶ **Assignment 3, Online catalog search exercise** (due June 4) (1.5 hours)

Part 1. Foundations. Knowledge and knowledge representation, continued

| | |
|------------|--|
| 2.1 | The nature of knowledge www.dsoergel.com/571/UBLIS571Lecture02.1Guide.pdf |
| 2.2 | Knowledge representation www.dsoergel.com/571/UBLIS571Lecture02.2Guide.pdf |

Unit 2. May 21- May 28

To prepare, read beforehand:

Lecture 2.1 The nature of knowledge Lecture objectives etc. (pink sheet)

Text Chapter 1. **Introduction: information systems for problem solving**
 Text Chapter 2. **The nature of information**
 Text Section 9.3 **Criteria for the design and evaluation of data schemas** (p. 150-152)

- 1 Skemp, Richard R. *The psychology of learning mathematics*. Expanded American Edition (Also: 2. ed. 1986, 1.ed. 1971, page numbers vary), In packet, multiple copies in Baldy 14A
 Chapter 2 **The formation of mathematical concepts**, p. 9 -21 **Required**
 Chapter 3 **The idea of a schema**, only p. 22-29 **Required**
 Chapter 5 **Symbols**, p. 46-55 **Required**

A general introduction to the structure of knowledge and its representation, applicable to any subject, not just mathematics.

Lecture 2.2 Knowledge representation Lecture objectives etc. (pink sheet)

- 1 Lindsay and Norman. *Human information processing. Intro to psychology*. NY: Ac. Press, 1972.
 Chapter 10. **The structure of memory (semantic networks, DS)**, p. 374-401 **Required**
 Chapter 11. **Memory processes (restructuring semantic. networks, DS)**, p. 402-434 **Required**
- 2 Cohen and Kjeldsen, **Information retrieval by constrained spreading activation in semantic networks**. Information Processing and Management. 1987; 23 (4): 255-260. In packet, **Optional**
- 3 Jonassen, David H.; Beissner, Katherine; Yacci, Michael. *Structural knowledge: Techniques for representing, conveying and acquiring structural knowledge*. Hillsdale, NJ: Lawrence Erlbaum, 1993.
 Ch. 12. **Implicit methods for conveying structural knowledge through frames and slots**, p. 125-133. In packet, **Optional, highly Recommended for LMS**)
- 4 Parsaye, Kamran; Chignall, Mark.. *Expert systems for experts. New York: John Wiley and Sons, 1988.*
 Section 2.2.3. **Frames: Packaged Structures**, p. 48-57 In packet, **Optional**
- 5 Fikes, Richard and Kehler, Tom. **The role of frame-based representation in reasoning**. Communications of the ACM. 1985; 28(9): 904-920. In packet XXX, **Optional**

Assignments Due: Assignment 1, Hypermedia exploration: Perseus and Freebase (2.5 hours)

Assignment 2, Bibliographic retrieval system exploration: MEDLINE (3 hours)

Assignments assigned Based on example in Lecture 2.2

► **Assignment 4, Restructuring a semantic network** (Due June 4) (1 hour)

Unit 3. May 28

Part 2. Information retrieval: General principles and methods

| | |
|-------------------|--|
| <p>3.1</p> | <p>The structure of information systems In-class exercise: Analytical description of an information system www.dsoergel.com/571/UBLIS571Lecture03.1.mp3</p> |
| <p>3.2</p> | <p>Objectives and performance measures for information systems (60 min) Lecture www.dsoergel.com/571/UBLIS571Lecture03.2.mp3 (Subtract 4 from Lecture Notes page nos)</p> |

To prepare, read beforehand

Lecture 3.1 The structure of information systems

Lecture objectives etc. (pink sheet)

Text Chapter 5. **The structure of information systems** (for discussion)

Lecture 3.2 Objectives and performance measures for information systems

Lecture objectives etc. (pink sheet)

Text Chapter 6. **Systems analysis**

Text Chapter 7. **Assessment of users' problems and needs**

Text Chapter 8. **Objectives and performance measures for ISAR systems** (for discussion)

Assignments due

Assignment 1, Hypermedia exploration: Perseus and Freebase (2.5 hours)

Assignment 2, Bibliographic retrieval system exploration: MEDLINE (3 hours)

Assignments assigned

Based on in-class exercise in Lecture 3.1

► **Assignment 5, Analytical description of an information system** (due June 4) (3 hours)

Unit 4. June 4

Part 2. Information retrieval: General principles and methods, continued

| | |
|-------------------|--|
| <p>4.1</p> | <p>An integrated information structure model www.dsoergel.com/571/UBLIS571Lecture04.1.mp3 Lecture explaining p. 14 - 47 of the reading; these pages serve as lecture notes Ideas from this lecture/reading will be used and clarified throughout the course www.dsoergel.com/571/UBLIS571Lecture04.1.mp3</p> |
| <p>4.2</p> | <p>Data schemas and formats www.dsoergel.com/571/UBLIS571SoergelLecture04.2InClassExercise.mp3 In-class exercise: Developing a conceptual data schema (45 minutes) Questions and answers / discussion of Chapter 9 www.dsoergel.com/571/UBLIS571Lecture04.2ConceptualDataSchemaExerciseReview.docx www.dsoergel.com/571/UBLIS571Lecture04.2ConceptualDataSchemaExerciseReview.mp3 www.dsoergel.com/571/UBLIS571Lecture04.2.mp3 (for Lecture Note pages 72-76)</p> |

To prepare, read beforehand

Lecture 4.1 An integrated information structure model

Lecture objectives etc. (pink sheet)

- 1 Soergel, Dagobert. **A general model for searching linked data OR Design of an integrated information structure interface.** A unified framework for indexing and searching in database, expert, information retrieval, and hypermedia systems. January 1999, edited with a new title December 2011. 78 p. Prologue and Part 1 (p. 1-48) In packet, whole paper on UBlerns.
 Read Prologue and p. 1 - 17 **Required** The rest of this reading serves as the notes for the lecture.
- 2 Soergel, Dagobert. **A language for the description of foods.** Manuscript. 1992. In Packet, **Optional**
 More examples for the Entity-Relationship approach and hierarchical inheritance

Lecture 4.2 Data schemas and formats

Lecture objectives etc. (pink sheet)

Text Ch. 3. **The structure of information** and Ch. 9. **Data schemas and formats** (including Appendix)

Text Chapter 9. **Data schemas and formats** (including Appendix)

Model Catalog (very first reading in packet) (useful to look at)

Assignments due

Assignment 3, Online catalog search exercise (1.5 hours)

Assignment 4, Restructuring a semantic network (1 hour)

Assignment 5, Analytical description of an information system (3 hours)

Assignment assigned

► **Assignment 6, Developing a conceptual data schema** (due June 11) (1.5 hours)

Unit 5. June 4, Lecture 5.1

Part 2. Information retrieval: General principles and methods, continued

| | |
|------------|---|
| 5.1 | <p>Access to information: data structure and search modes (70 min.)</p> <p>Retrieval as prediction (probabilistic retrieval)</p> <p>Review of Boolean retrieval (Text Chapter 10)</p> <p>In-class exercise: Ranking of retrieved objects</p> <p>Review of search modes and data structures (Text Chapter 11)</p> <p>www.dsoergel.com/571/UBLIS571Lecture05.1RankingSpreadsheet.xlsx</p> |
|------------|---|

To prepare, read beforehand

Lecture 5.1 Access to information: data structure and search modes

Lecture objectives etc. (pink sheet)

Text Chapter 10. **Elementary query formulation**

Chapter 11. **Data structures and access**

Model Catalog (useful to look at, search in Library Master)

For Lecture 5.1 see Part 3, p. 27-28.

Assignments due: Assignment 3, Online catalog search exercise (1.5 hours)

Assignment 4, Restructuring a semantic network (1 hour)

Assignment 5, Analytical description of an information system (3 hours)

Assignment assigned

Based on in-class exercise in Lecture 5.2b

► **Assignment 6, Developing a conceptual data schema** (due June 11) (1.5 hours)

► **Prepare description of term paper** using the form found after Assignment 6 (due June 11)

Unit 5. June 11, continued under Part 3

Unit 5 June 11, Lecture 5.2 - Unit 6 June 11, Lecture 6.1

Part 3. Nature, design, and management of documents and records

Lectures 5.2 and 6.1. Document function, structure, analysis, and design

www.dsoergel.com/571/UBLIS571Lecture05.2.mp3 (5.2a-d)

www.dsoergel.com/571/UBLIS571Lecture06.1.mp3

| | |
|-------------|--|
| 5.2a | Knowledge (re)presentation in text and images. Text linguistics (30 min) |
| 5.2b | Text analysis overview and examples (30 min) In-class exercise: Extracting data from text, especially resolving anaphoric references |
| 5.2c | [Natural language processing. Syntactic and semantic parsing] (0 min) Supplement |
| 5.2d | Document design. Formatting documents for understanding by people |
| 6.1a | Document macrostructure. Document templates. Inter-document relationships (40 min) |
| 6.1b | Formatting documents for interpretation by computers. Markup languages. (20 min) Very brief introduction; covered in 506. Some materials in the supplement |

Lecture 5.2a Knowledge (re)presentation in text and images. Text linguistics

Lecture objectives etc. (pink sheets) for Lectures 5.2 - 6.1 and for Lecture 5.2a

- 1 Soergel, Dagobert. **The nature of texts.** 1999. 12 p. In packet, **Optional**
- 2 Crombie, Winifred, original author ; Soergel, Dagobert, adapter. **Semantic relations between propositions.** Original article 1985, adaptation 1998. 7 p. In packet, **Required**
Shows how universal the entity-relationship approach is.

Lecture 5.2b Text analysis overview and examples Lecture objectives etc. (pink sheet)

Lecture 5.2c [Natural language processing. Syntactic and semantic parsing] [In Supplement]

- 1 **Xerox linguistic software** (Web announcement) and **Temis / Luxid** www.luxid.com In packet, **Required**
Illustrates practical importance of text analysis and NLP. Look over, get the gist, look at examples
- 2 Etzioni, Ore; Banko, Michele; Soderland, Steven; Weld, Daniel S.
Open Information Extraction from the Web. Comm.of the ACM December 2008 51(12): p.68-74
- 3 Shuldberg, Kelly H.; MacPherson, Melissa; Humphrey, Pete; Corley, Jamie.
Distilling information from text: The EDS TemplateFiller system.
Journal of the American Society of Information Science. 1993.10; 44(9): 493-507. Ublearns, **Optional**
- 4 Allen, James. **Natural language understanding.** 2nd ed. Redwood City, Calif.: Benjamin/Cummings, 1995.
Table of contents and Chapter 1. Introduction to natural language understanding, p. 1-17. In packet, **Optional**
- 5 Feldman, Susan **NLP Meets the Jabberwocky: Natural Language Processing in Information Retrieval.**
ONLINE, May 1999. 23(3): 62-64,66-68,70-72. www.scism.sbu.ac.uk/inmandw/ir/jaberwocky.htm. Ublearns,
Optional, Recommended **over for 5.2d, 6.1a, 6.1b, assignments**

Lecture 5.2d Document design. Formatting documents for understanding by people

Lecture objectives etc. (pink sheet)

- 1 Mayer, Richard E. **The Balloons Passage: Understanding requires a schema.** From Mayer, R. E. Thinking, Problem-Solving, Cognition. NY: Freeman, 1983, p. 207- 208.) In packet, **Required**
- 2 Novak, J. D. & A. J. Cañas, **The Theory Underlying Concept Maps and How to Construct Them.** Tech. Report IHMC CmapTools, Florida Institute for Human and Machine Cognition, 2006-01, rev. 2008-01., 33 p. Retr. 2012-04-29 from <http://cmap.ihmc.us/publications/researchpapers/theorycmaps/theoryunderlyingconceptmaps.htm>. In packet, **Required p. 1-12, LMS all**
- 3 Keyes, Elizabeth. **Information design: Maximizing the power and potential of electronic publishing equipment.** IEEE Transactions on Professional Communication 30(1) (1987): 32-37. In packet, **Required**
- 4 Soergel, D., comp. **Some useful document design guidelines. & Supplemental style rules.** 2011 In packet, **Requ.**
- 5 Lynch, Patrick J.; Horton, Sarah. **Web style guide. Basic design principles for creating Web sites.** 3. ed. New Haven, CN: Yale Univ. Press; 2009. 352 p \$19 at Amazon. Full text at www.webstyleguide.com/wsg3/index.html Ch. 3. Site design. p. 23 - 31. Ch. 4. Page design. p. 53 - 56 Ch.. Typography. p. 79 - 92. Find on Web,
- 6 Meyer, Bonnie J. F. **Following the author's top-level organization: an important skill for reading comprehension.** In: Understanding Readers' Understanding: Theory and Practice. Tierney, R.J.; Anders, P.L.; Mitchell, J.M., eds., p. Hillsdale, N.J.: Erlbaum, 1986. p. 59 - 76. In packet, **Optional, Rec. LMS**
- 7 Rumelhart, David E. and Norman, Donald H. **Accretion, tuning, and restructuring.** In: Semantic Factors in Cognition. Cotton, J.W. & Klatzky, R.L., eds. Hillsdale, N.J.: Erlbaum, 1978, p. 37 - 53. In packet, **Opt., Rec.LMS**

Lecture 6.1a Document macrostructure. Document templates. Inter-document relationships

Lecture objectives etc. (pink sheet)

Lecture 6.1b Formatting documents for interpretation by computers. Markup languages.

Lecture objectives etc. (pink sheet)

Lectures 5.2 - 6.1

- 1 **Content management suite from Interwoven** [Included for the ideas, no endorsement of product or company.] Data sheets from Web site www.interwoven.com/products XXX Autonomy In packet, more on Web, **Optional** Download the .pdf Data sheets and use the zooming function of Adobe Acrobat to actually see the small pictures.
- 2 Kamps, Thomas; Hüser, Christoph; Möhr, Wiebke; Schmidt, Ingrid. **Knowledge-based information access for hypermedia reference works: Exploring the spread of the Bauhaus movement.** In Agosti, M.; Smeaton, A. *Information retrieval and hypertext.* Boston: Kluwer; 1996. Chapter 10, p. 225-256. **Required**
A very nice and easy to understand example that uses all the technologies discussed here
- 3 Noemie Elhadad, Min-Yen Kan, Judith Klavans, and Kathleen McKeown. **Customization in a unified framework for summarizing medical literature.** Journal of Artificial Intelligence in Medicine, 33(2):179-198, 2005 www.cs.columbia.edu/nlp/papers/2005/elhadad_al_05a.pdf. UBLearns, **Optional**
Find documents based on a patient record, then extract relevant pieces and present in a multi-document summary

Assignments due Unit 6: Assignment 6, Developing a conceptual data schema (1.5 hours)

Description of term paper using the form found after Assignment 6

Note: Assignment 3, Online catalog search (assigned May 21, due June) prepares for Lectures 6.1 - 7.2

Assignment assigned in Unit 6, June 11 See p. 30

Unit 6 June 18, Lecture 6.2 - Unit 7, June 18, Lectures 7.1 - 7.2

Part 3. Nature, design, and management of documents and records, cont.

Lectures 6.2a - 7.2. Metadata. Bibliographic and record control

| | |
|-------------|---|
| 6.2a | General introduction to metadata (10 min.) Lecture |
| 6.2b | Bibliographic and record control. General issues (40 min) Lecture www.dsoergel.com/571Lecture6.2.mp3 |
| 6.2c | Bibliographic and record control. Description. (20 min) Describing texts and documents in a more general context Lecture |
| 7.1a | Bibliographic and record control: Entries and access (40 min) In-class exercise: Problems of determining author entry www.dsoergel.com/571/UBLIS571Lecture7.1aSlides.zip |
| 7.1b | Metadata, Resource Description Framework (RDF), Dublin Core (DC) (30 min) Lecture www.dsoergel.com/571/UBLIS571Lecture7.1b.mp3 |
| 7.2 | In-class exercise on descriptive cataloging. Done individually or in pairs |

To prepare, read beforehand

Lectures 6.2a - 7.2 Metadata. Bibliographic and record control. Lecture objectives etc. (pink)

Lectures 6.2 b - c. General issues. Description

The Model Catalog is helpful as a source of examples.

1 Tillett, Barbara What is FRBR?: A Conceptual Model for the Bibliographic Universe.

www.loc.gov/cds/FRBR.html. In packet, **Required**

2 **Functional Requirements for Bibliographic Records: Final Report.** 1998

www.ifla.org/VII/s13/frbr/frbr.pdf. UBlerns, **Optional**

3 **Statement of international cataloguing principles.** IFLA 2009. In packet, **Required**

Retrieved on Sept. 18, 2010 from www.ifla.org/files/cataloguing/icp/icp_2009-en.pdf

4 Oliver, Chris. **Introducing RDA: A Guide to the Basics.** Chicago, IL: ALA Editions; 2010.

128 p. ISBN-13: 978-0-8389-3594-1. **Chapter 1 Required, other Optional**

Read at http://books.google.com/books?id=WIBFVbU_ozYC, 1 HC in Baldy 14A

5 **Descriptive Cataloging. Sample codes.** In packet, **Required**

Includes **Excerpts from AACR2, Part 1**; *Brief introduction to APA rules*; and

Sample list of citation styles supported by EndNote

over

Get a general idea. Do not read the whole document in detail; know where things are so you can refer to the document in later cataloging exercises.
AACR2 is accessible through cataloger's desktop

Lecture 7.1a Bibliographic and record control: Entries and access

Read afterwards:

- 1 **Lubetzky's conditions for author entry** (from Needham, *Organizing knowledge in libraries*, 1971), rearranged by D. Soergel
- 2 **Excerpts from AACR2, Part 2** (Familiarize yourself with the general layout and some major rules) (The reading package contains merely a list of sections to look at; use the AACR2 book in Baldy 14A.)
- 3 Corresponding pages from RDA XXX

Assignments due Unit 6, June 11

Assignment 6, Developing a conceptual data schema (1.5 hours)

Assignments due Unit 7, June 18

- ▶ **Assignment 7. Apply linguistic techniques to retrieval problems** (due June 18) (2 hours)

Assignments assigned Unit 6, June 18

- ▶ **Assignment 8, Descriptive cataloging practice** (complete when you can, Aug. 15) (1.5 hours)
If you are unsure, start this assignment after Lecture 7.2, Descriptive cataloging exercise
- ▶ **Assignment 9, Problems of entry** (complete when you can, no later than Aug. 15) (4 hours)
- ▶ **Assignment 10, Index three documents, prepare for Lecture 8.1 (due June 25)** (2 hours)
(This assignment is preparation for Small Groups 1, Lecture 8.1)

MIDTERM

Take-home, closed-book midterm exam handed out (due June 25) (1.5 hours)
May type or handwrite

Unit 8. June 25 (Small Groups 1, 4 hours)

Part 4. Classification and subject access

| | |
|-------------|---|
| 8.1 | Explorations in subject access Continuing Assignment 10. (2 hrs.) www.dsoergel.com/571/UBLIS571Lecture08.1Slides.zip |
| 8.2a | Vocabulary control. A special case of authority control (20 min.) Text Chapter 12 review Lexical relationships: Paradigmatic relationships (synonymy, antonymy, hyponymy) and homonymy/polysemy Lecture (10 min.) www.dsoergel.com/571/UBLIS571Lecture08.2.mp3 |
| 8.2b | Index language functions Lecture (60 min.) www.dsoergel.com/571/UBLIS571Lecture08.2.mp3 |

To prepare, read beforehand

Part 4 Objectives etc. (pink sheet)

Lecture 8.1 Explorations in subject access

Lecture 8.1 objectives etc. (pink sheet). Look over all pages for Lecture 8.1 in the Lecture notes

Lecture 8.2a Vocabulary control & Lexical relationships

Lecture objectives etc. (pink sheet)

Text Chapter 12. **Terminological control** (for brief discussion)

Lecture 8.2b Index language functions

Lecture objectives etc. (pink sheet)

Text Chapter 13. **Index language functions** (for discussion)

- 1 Mooers, Calvin; Brenner, Claude W. **A case history of a Zatacoding information retrieval system.** In: Punched Cards: Their Applications to Science and History, 2nd ed., Robert S. Casey; James W. Perry; Madeline M. Berry; and Allen Kent, eds., New York: Reinhold, 1958. p. 346-352. In packet, **Required**

This is the **seminal article on request-oriented indexing** (called *filtering technique* in the article. Mooers coined the terms *information retrieval* and *descriptor*)

- 2 Soergel, **Functions of a thesaurus / classification / ontological KB.** In packet, **Required**
From Soergel, Dagobert. **Knowledge Organization Systems. Overview.** 2009.
UBlearns, **Optional**

Over

Assignments due

Assignment 8, Descriptive cataloging practice (4 hours) (can be handed in later)

Assignment 9, Problems of entry (1.5 hours) (can be handed in later)

Assignment 10, Indexing of three documents (2 hours)

This assignment is preparation for Small Groups 1, Lecture 8.1.

Midterm Exam**Assignment assigned**

Assignment based on request-oriented indexing exercise in Lecture 8.2b

► **Assignment 11, Request-oriented indexing** (due July 2) (2 hours)

Unit 9. July 2 (Small Groups 2)

Part 4. Classification and subject access, continued

| | |
|------------|---|
| 9.1 | <p>Index language structure 1: conceptual</p> <p>In-class exercise</p> <p>Conceptual analysis and synthesis: Semantic factoring and hierarchy building</p> <p>www.dsoergel.com/571/UBLIS571Lecture9.1SlidesA.zip</p> <p>www.dsoergel.com/571/UBLIS571Lecture9.1SlidesB.zip</p> |
| 9.2 | <p>Application of index language structure to searching</p> <p>In-class exercise: Retrieval of documents in a sample collection</p> <p>In-class exercise: Retrieval access to the documents from Assignment 11</p> <p>http://www.dsoergel.com/571/UBLIS571Lecture9.2SlidesA.zip</p> <p>http://www.dsoergel.com/571/UBLIS571Lecture9.2SlidesB.zip</p> |

To prepare, read beforehand

Lecture 9.1 Index language structure 1: conceptual

Lecture objectives, etc. (pink sheet)

Text Chapter 14. **Index language structure 1: conceptual** (for discussion)

Lecture 9.2. Application of index language structure to searching

Lecture objectives, etc. (pink sheet)

Assignments due

Assignment 11, Request-oriented indexing (2 hours)

Midterm Exam

Assignment assigned

Based on Small Groups 2, Lecture 9.1

► Assignment 12, Conceptual analysis and synthesis, first part

(Note: There is no Assignment 12 as such, just Assignments 12.1, 12.2, 12.3a, and 12.3b)

► **Assignment 12.1, Semantic factoring** (due July 9) (1.5 hours)

► **Assignment 12.2, Building a hierarchy of elemental concepts** (due July 9) (1.5 hours)

Unit 10. July 9, Small Groups 3**Part 4. Classification and subject access, continued**

| | |
|--------------|--|
| 10.1 | Constructing a hierarchy from facet combination www.dsoergel.com/571/UBLIS571Lecture10.1Slides.zip |
| 10.2a | Brief introduction to Assignments 13.1 - 13.4 www.dsoergel.com/571/UBLIS571Soergel-20120328-Lec10.2a.mp3 |
| 10.2b | Introduction and in-class exercise: Assignment 13.1 Dewey Decimal Classification |

To prepare,**Lecture 10.1, Small Groups 3 Constructing a hierarchy from facet combination**

Complete Assignment 12.1 and 12.2

Lecture 10.2a, Brief introduction to Assignments 13.1 - 13.4

Lecture objectives, etc. (pink sheet)

General introduction to Assignments 13.1 - 13.4, p. XXX in the Assignments (gold)

Lect. 10.2b Introduction and in-class exercise: Ass. 13.1 Dewey Decimal Classification

Look over Assignment 13.1 materials.

In a study group: Work on the DDC Worksheet, skip what you do not get quickly, 2 hrs

1 Needham, Christopher D. **Organizing knowledge in libraries: An introduction to information retrieval.** 2nd ed. New York: Seminar Press, 1971.

Ch. 7 **Review of classification principles**, p. 109-131

Ch. 8 **DDC**, p. 133, 140-152

Still the clearest exposition of timeless classification principles

In Packet, **Required**

For further study

2 Chan, Lois Mai; Comaromi; Mitchell, Joan

Dewey Decimal Classification: Principles and Application by Lois Mai Chan and Joan S. Mitchell . 3. ed.

Dublin, Ohio : OCLC, 2003. xi, 216 p.

This version not found in Lockwood **Optional**

On the new ed. 23: www.oclc.org/dewey/versions/print/default.htm

Find on Web, **Optional**

over

Unit 10. July 9, continued

Assignments due

- Assignment 12.1, Semantic factoring (1.5 hours)
- Assignment 12.2, Building a hierarchy of elemental concepts (1.5 hours)

Assignments assigned

Based on Small Groups 3, Lecture 10.1

- ▶12.3a Practice Hierarchy from facet combination with education concepts (due July 16) (2 hours)
- ▶12.3b Real Hierarchy from facet combination with concepts from 12.1 / 12.2 (due July 16) (2 hours)

Assignments 13.1-13.4 Subject cataloging and searching practice

- ▶Assignment 13.1, Dewey Decimal Classification (DDC) (4 hours)
Start July with work in study group, then listen to presentation, complete by July 16
- ▶Assignment 13.2 Yahoo: Yahoo (or DMOZ) classification (a semi-faceted classification) (6 hours)
 - OR LCC: Library of Congress Classification (LCC) (6 hours)
 - OR DDC2. More practice with DDC
 - OR Your choice in consultation with instructor

For Yahoo and LCC, we will start going through the worksheet, index a document, and formulate a query so that everyone has at least some idea of these schemes. July 16
Due on July 23 (whichever you choose)
- ▶Assignment 13.3, Library of Congress/Sears Subject Headings (LCSH) (5 hours)
Start July 23, complete by July 30, not covered in class
- ▶Assignment 13.4, ERIC Thesaurus (3 hours)
Start July 23, complete by July 30, not covered in class

Absolute deadline for Assignments 13.1-13.4 to receive a grade in the course is Aug. 15.

Assignment 13.2 LCC needs Classification Web (logon information will be communicated; it is useful to have a look at the printed volumes; a few copies in Baldy 14A or from some library.

You are encouraged to meet with the instructor in study groups to ask questions on these assignments.

Unit 11. July 16

Part 4. Classification and subject access, continued

| | |
|-------------|--|
| 11.1 | <p>Introduction and in-class exercise on Assignment 13.2 Yahoo</p> <p>We will start going through the worksheet, index a document, and formulate a query. www.dsoergel.com/571/UBLIS571Lecture11.1aFirstExampleSlides.zip www.dsoergel.com/571/UBLIS571Soergel-20120404-Lec11.1b.mp3 www.dsoergel.com/571/UBLIS571Lecture11.1cInClassExercisesSlides.zip</p> |
| 11.2 | <p>Introduction and in-class exercise Assignment 13.2: LCC</p> <p>We will start going through the worksheet, index a document, and formulate a query www.dsoergel.com/571/UBLIS571Soergel-20120404-Lec11.2a.mp3 www.dsoergel.com/571/UBLIS571Soergel-20120404-Lec11.2b.mp3</p> |

To prepare

11.1 Introduction and in-class exercise on Assignment 13.2 Yahoo

Optional: Look at <http://dir.yahoo.com/>

11.2 Introduction and in-class exercise Assignment 13.2: LCC

1 Needham, Chapter 8, **Schemes of classification**, p.163-168 LCC (In Packet, **Required**)

2 Chan, Lois Mai

A guide to the Library of Congress Classification. 5th ed.

Englewood, Colo.: Libraries Unlimited, 1999.

Z696.U4C47 1999 in Lockwood Library Ublearns, All sections **Optional**

p. 1-14 The history of the classification

p. 14-19 Focus and use

p. 23-47 Principles, structure, and format (skim the examples).

For further study, read more from this book

Assignments due

12.3a **Practice** Hierarchy from facet combination with education concepts (2 hours)

12.3b **Real** Hierarchy from facet combination with concepts from 12.1 / 12.2 (2 hours)

13.1 Dewey Decimal Classification

Assignments assigned

▶ Assignment 13.2 **Yahoo** (due July 23) **OR**

▶ Assignment 13.2 **LCC** (due July 23) **OR**

▶ Assignment 13.2 **DDC2. More practice with DDC** (due July 23)

Unit 12. July 23**Part 4. Classification and subject access, continued**

| | |
|-------------|--|
| 12.1 | Index language structure 2: database organization Text Chapter 15 review Concluding in-class exercise: vocabulary control and hierarchical structure Concluding in-class exercise: conceptual analysis and synthesis |
| 12.2 | Media Streams Demo (scheme for indexing movie scenes using icons as descriptors) http://www.dsoergel.com/571/UBLIS571Soergel-20120411-Lec12.2.mp3 |

To prepare**Lecture 12.1 Index language structure 2: database organization read beforehand**

Lecture objectives, etc. (pink sheet)

Text Chapter 15. **Index language structure 2: database organization** (for discussion)

Lecture 12.2 Media Streams Demo

Optional readings if you are interested in the Media Streams classification, using icons as descriptors

1 Davis, Marc. **Media Streams: An Iconic Visual Language for Video Representation.**

In: Readings in Human-Computer Interaction: Toward the Year 2000, eds. Ronald M. Baecker, Jonathan Grudin, William A. S. Buxton, and Saul Greenberg. 854-866. 2nd ed., San Francisco: Morgan Kaufmann Publishers, Inc., 1995.

www.dsoergel.com/571/MarcDavisMediaStreamsChapter.pdf UBlerns, **Optional**

2 Davis, Marc. **Media Streams: An Iconic Visual Language for Video Annotation.**

Telektronikk 4.93 (1993): 59-71

www.dsoergel.com/571/MarcDavisMediaStreamsJournalPaper.pdf UBlerns, **Optional**

Over

Assignments due

Assignment 13.2, Yahoo, OR LCC, OR DDC 2, OR Own choice (6 hours)

Assignment assigned

► Assignment 13.3, **Library of Congress/Sears Subject Headings** (LCSH) (5 hours) (due July 30)

► Assignment 13.4, **ERIC Thesaurus** (3 hours) (due July 30)

Prepare requests for topics to be included in the final review.

(Can be submitted by e-mail to dsoergel@buffalo.edu until Aug. 6, 10 am)

For Assignment 13.3, Library of Congress/Sears Subject Headings

1 Needham, Ch. 10, *The alphabetic subject catalog*, p. 199-223 UBLearns, **Optional**

2 Chan, Lois Mai 2007

Cataloging and classification: An introduction

Lanham, Md. : Scarecrow Press, 2007

Lockwood Z693.5 .U6 C48 2007, Silverman Library [Capen] Reserve Z693.5 .U6 C48 2007

Chapter 8 on LCSH UBLearns, **Optional**

For further study

Chan, Lois Mai 2005

Library of Congress Subject Headings. principles of structure and application. 4. ed.

Englewood, CO: Libraries Unlimited; 2005.

This newest version not found at UB

Perreault, Jean M. 1979

Library of Congress Subject Headings: A New Manual.

International Classification 1979 Nov.; 6(3):158-169.

Extensive review of an earlier version of Chan's book. Gives a good feel for some of the problems in LCSH. UBLearns, **Optional**

Unit 13. July 30**Part 4. Classification and subject access, continued**

| | |
|--------------|---|
| 13.1 | <p>Exploration of Knowledge Organization Systems (ontologies, classification schemes, thesauri)</p> <p>The big stack of readings for Lecture 12.2 (End of Reading Packet, blue dividers) will serve as lecture notes. www.dsoergel.com/571/UBLIS571Lecture13.1ExplorationKOS.mp3 Have a look at them beforehand. not meant for word-by-word reading You should explore two of the schemes in more depth after the lecture</p> |
| 13.2a | Questions on Assignment 13.3 LCSH / Sears and 13.4 ERIC or anything else (15 min) |
| 13.2b | <p>Indexing and system performance (50 min.)</p> <p>Text Chapter 16 review</p> |

To prepare, read beforehand**13.1 Exploration of Knowledge Organization Systems**

The big stack of readings for Lecture 12.2 (End of Reading Packet, blue dividers)
 Have a look at them beforehand. not meant for word-by-word reading

13.2b Indexing and system performance

Text Chapter 16. **Indexing and system performance** (for discussion)

1 Soergel, Indexing and retrieval performance: The logical evidence.

Journal of the American Society for Information Science, 1994.9; 4(8): 589-599.

In Packet, **Required**

Note: Some students found it easier to read this article before Chapter 16.

Prepare requests for topics to be included in the final review and email to instructor (dsoergel@buffalo.edu) or post to UBLearn

Assignments due

Assignment 13.3, Library of Congress/Sears Subject Headings (LCSH) (5 hours)

(Nothing to be handed in now, but you should have indexing and query formulation done)

Assignment 13.4, ERIC Thesaurus (3 hours)

Unit 14. Aug. 6
Part 5. Conclusion

| | |
|--|---|
| 14.1 and 14.2 | Final review http://www.dsoergel.com/571/lectures/UBLIS571Soergel_20110504_Lec15.1.mp3 http://www.dsoergel.com/571/lectures/UBLIS571Soergel_20110504_Lec15.2.mp3 |
|--|---|

To prepare

Read through lecture notes for Lectures 14.1 and 14.2:

Sample final questions to be discussed

Final review: Natural language processing [Supplement]

Final review: Precombination vs. postcombination [Supplement]

Think of questions, post to UBLearn or email them to the instructor ahead of time,

Dagobert Soergel

LIS 571
Organization and Control
of Recorded Information

Summer 2012

Lecture Notes

Lecture 1.1

May 21

Overview of the course and course materials

Information professionals in the 21st century

www.dsoergel.com/571/UBLIS571Lecture01.1Introduction.mp3
www.dsoergel.com/571/UBLIS571Lecture01.1Introduction.docx (text)
www.dsoergel.com/571/UBLIS571Lecture01.1OverviewSlides.zip

| | |
|-------------------|---|
| Objectives | <ol style="list-style-type: none"> 1 You should have an appreciation for the wide variety of information tasks that an education in information studies enables you to undertake and the wide variety of information environments you can work in. Put differently, you should gain a sense of the breadth of functions and the breadth of environments you can work in, the breadth of careers. 2 You should have an appreciation for the wide variety of information systems that exist, including expert systems. 3 You should have an idea of Organization of Information concepts and skills that are needed in practice 4 You should have an overview of the course and know what to expect and what is expected of you. |
|-------------------|---|

Outline

Introduction to the course

www.dsoergel.com/571/UBLIS571Lecture01.1Introduction.mp3

Introduction of students

Overview of the course and course materials

www.dsoergel.com/571/UBLIS571Lecture01.1OverviewSlides.zip For the rest of 1.1

What do information professionals do Read p.3

See sample CVs in Supplement

Types of information systems Read p. 4 - 5

On teaching and organization of information Read p. 6

Salaries of information professionals See supplement

What do information professionals do?

| | |
|---|--|
| <p>Answer questions, find things</p> | <ul style="list-style-type: none"> • Explore the information need with the user: <ul style="list-style-type: none"> Understand the user's problem, understand what the user knows already, understand how the user thinks • Find answers in external and internal sources, such as <ul style="list-style-type: none"> N Library catalogs, bookstore catalogs (mostly online now), N Reference tools (bibliographies, biographical tools, almanacs, encyclopedias, etc.), print or online N Numeric databases, such as census databases, N Maps N The Web at large, intranets N Archives (find records on a given subject even though they are not indexed by subject) N A repository of instructional materials • Make a report that draws on several sources (extensive example: Congressional Research Service reports for Congress) • Organize the answer for quick perusal |
| <p>Organize things so they can be found</p> | <ul style="list-style-type: none"> • Catalog books using the MACHine-Readable Cataloging (MARC) format and the Anglo-American Cataloging Rules (AACR2-2002) or soon RDA • Catalog Web pages using Dublin Core • Catalog learning materials using educational metadata standards • Format documents using XML • Write abstracts for and index journal articles |
| <p>Help people produce information</p> | <ul style="list-style-type: none"> • Assist in editing and formatting documents • Help teachers in creating lesson plans (find instructional materials or learning objects, help format the lesson plan, help format materials for students, for example using graphic organizers) • Create Web pages (for the library or school media center on the organization) |
| <p>Teach</p> | <p>Teach people</p> <ul style="list-style-type: none"> • how to find information. Requires teaching them about information organization • how to assess and evaluate information • how to use and integrate information • how to present information |
| <p>Develop and set up systems for all of the above</p> | <ul style="list-style-type: none"> • Set up bibliographic and other databases, including library catalogs • Set up an intranet or a more ambitious enterprise portal that supports the work of all people in the organization • Set up document templates for easy creation of documents • Develop classification schemes, thesauri, taxonomies for special user groups (Each US agency must have a taxonomy to present its material to the public) • Help users with setting up their own personal information systems |

Types of information systems and information environments

Information systems can be classified along many dimensions or facets. Any specific information system can be characterized by a combination, one concept from each dimension, for example

A system

- dealing with loosely structured information
- dealing with published or semi-published information
- serving a government agency
- information used for research and patient care
- dealing with the medical domain
- using paper technology for storage and accessing digital information

= **a traditional medical government library**

Sample dimensions (facets) for characterizing information systems

Types of information (such as bibliographic data, text and images, multimedia, numerical and other primary data, organization data and records);

Degree of structure of the information (unstructured or loosely structured information as in text vs. tightly structured information as in numeric databases)

Processing to create answers: plain retrieval vs. drawing conclusions

Origin of information (such as generally published information - paper or online, government information, organizational information, information. about customers or patients);

internal vs. external information

Users of the information, audience or organization served (groups - such as children, farmers, scholars, urban communities - or organizations - such as schools; universities or colleges; government agencies; businesses);

Uses of information (such as research, learning, problem solving, decision making, collaboration, day-to-day transactions);

Subject field (such as physics, medicine, or anthropology);

Technical means of providing access (such as paper vs digital).

The combinations are many, illustrating the flexibility and diversity of the information field

The table on the next page gives some examples for these dimensions. The information systems listed have characteristics for the other dimensions too but we do not list these.

Sample systems illustrating selected dimensions

| | |
|--|--|
| <p>Information & processing</p> | <p>Expert systems (medical diagnosis, computer configuration, detecting mineral deposits from satellite images, loan approval, etc.)</p> <p>Software libraries / databases for ease of access and reuse</p> <p>Employment service databases</p> <p>Personnel information system (usual personnel data plus skills and assignments to manage an organization's workforce)</p> <p>Geographical information system (GIS)</p> |
| <p>Users</p> | <p>Information systems in organizations</p> <p>Knowledge management: Make sure all applicable information is used to best advantage by organizing all types of internal and external sources of information – paper files and computer files no matter who keeps them, people and the knowledge they have in their heads – for access and usability.</p> <p>Information resources management</p> <p>Day-to-day transaction systems (order, inventory, etc.)</p> <p>Management information systems (MIS), Decision support systems</p> <p>Records management, archives (especially electronic records)</p> <p>A personal information system managing Web bookmarks, bibliographic references, downloaded Web pages, computer files, paper documents in personal collection, all kinds of notes, addresses, appointments</p> |
| <p>Use</p> | <p>Instructional information systems matching learner's needs with instructional materials</p> <p>In formal educational institutions</p> <p>In organizations for training (this is big business! Coordinate with personnel information system)</p> <p>For both: long-distance learning</p> |
| <p>Technology</p> | <p>Paper libraries of all kinds (public, academic and school, special)</p> <p>Online information systems</p> <p>Digital libraries</p> <p>Intranets</p> <p>An organizations Web site</p> <p>Any kind of computer database</p> <p>Bibliographic databases (e.g., Medline or OCLC's WorldCat)</p> <p>OCLC = Online Computer Library Center, the world's largest cooperative cataloging agency)</p> <p>Full-text databases (e.g., Westlaw or Lexis for law)</p> <p>Multimedia databases. Problem of retrieving still and moving images</p> <p>Substantive databases (directories, statistical data, material properties data)</p> |

On teaching and organization of information

Implications for curriculum and instruction

The findings from this study suggest that today's students need to learn – in a way that transcends their learning of specific content – a good deal about the structure of knowledge and about the importance of that structure. In order to learn in an information-rich environment, they need to learn

- (1) that knowledge is indeed structured in meaningful ways;
- (2) that various structures can be applied to various kinds of knowledge; and
- (3) that a key part of learning is learning how to create personal structures that organize their own learning accurately and coherently.

They must learn that knowledge is an organized, systematically related set of ideas and that they need to work at building an understanding of that organization as well as learning the individual ideas. They must learn the nature and uses of various kinds of structures—for example, time lines, maps, and hyperlinks as well as traditional narrative structures—that they can use as tools for building their own knowledge. They must learn criteria and procedures for building appropriate and coherent structures that will allow them to integrate and communicate their thoughts. A curricular emphasis on teaching students how to structure information is, I believe, the most important implication for learning and teaching that stems from the presence of the information-rich environment in which we and our students live.

Learning theorists tell us that learning consists of constructing mental models or schemas, structures that are comprised of ideas and patterns or frameworks that organize and link those ideas. At some basic level, then, learning is the equivalent of organizing information. And no one in a school knows more about organizing information than the library media specialist best. Helping both teachers and students understand and learn to create a variety of ways to structure information is the key task for our profession in these best and worst of times.

From

Delia Neuman

Learning in an information-rich environment: Preliminary results

Treasure Mountain/Elms Research Retreat

Elms Resort and Spa

Excelsior Springs, MO

May 31, 2002

Part 1. *May 21- May 28*
Foundations. Knowledge and knowledge representation

Lecture 1.2 *May 21*
Information systems and information structure
www.dsoergel.com/571/UBLIS571Lecture01.2Slides.zip
www.dsoergel.com/571/UBLIS571Lecture01.2ExampleSlides.zip

| | |
|-------------------------------|--|
| Objectives | <ol style="list-style-type: none"> 1 Gain an appreciation for the variety of information systems that exist, including knowledge-based systems / expert systems. 2 Understand the importance of information structure / knowledge representation as the heart of an information system. 3 Have a first idea of the entity-relationship approach to knowledge representation. |
| Practical significance | <p>Being knowledgeable about databases is a requirement for every executive assistant, let alone information professionals. Databases are the key to dealing efficiently with many types of information.</p> <p>Knowing about many types of information systems makes your skills more widely applicable and thus increases career opportunities. Expert systems are now widely used in many subject areas, for example, medicine, computer system configuration, and processing of loan applications; see the list at the end of Lecture 1.2 for some examples</p> <p>Designing or understanding the information structure of a system is key to building or using the system. The entity-relationship approach is the most natural and at the same time most general way for representing information.</p> |

Note on terminology: The Artificial Intelligence (AI) community speaks about *knowledge representation*, the database community speaks about *data modeling*.

Key idea:

Combining different kinds of facts to find an answer. (Inference. Chaining)

Done by people - reference librarian or user consults different databases as needed to find all the facts needed to construct an answer.

Done by systems - all facts must be accessible to the system

Relates to → LIS 518 Reference Sources and Services (concepts important for how to search)

Introduction

Purpose of an information system generally: Answer questions by either finding an answer that exists ready-made in the database or deducing an answer from multiple statements in the database.

Answering a question means **going from something known to something unknown.**

The lecture will show through examples **how information structure is used to find answers.**

We will look at three examples of information systems:

- 1 An expert system for medical prescriptions
- 2 A database that supports the operation of a university (Organizing Info., Chapter 3)
- 3 Medline, a bibliographic information system (Supplement)

Example 1. An expert system for medical prescriptions

| | |
|------------------|---|
| Purpose | From the data in the patient record, including new diagnoses, find drugs the patient should take. |
| Questions | <ol style="list-style-type: none"> 1 What new disease does patient Fred have 2 What drugs are used to treat asthma? <i>Known:</i> Disease asthma, <i>unknown:</i> Drug 3 What drugs should patient Fred take? <i>Known:</i> Patient Fred, <i>unknown:</i> Drug |

You will see examples of data that are needed to come up with a prescription for a patient. A physician deciding on a prescription combines these data through a reasoning process.

To prescribe a drug, a physician proceeds as follows

- A Find the disease newly added the patient record (patient data, specific to each patient).
- B Find out what drugs are available to treat that disease (drug treatment data, med. textbook)
- C Check for each of these drugs what, if any, harmful effects they may have on the patient.

In most medical environments, there are three simple databases each of which • contains one kind of facts and • can be queried to retrieve facts as they are stored in the database

The physician is left with the task of combining these facts to decide what drug to prescribe

We will show these databases, and then develop an expert system that has access to all three kinds of facts and can combine them to deduce a suggestion for a drug to be prescribed.

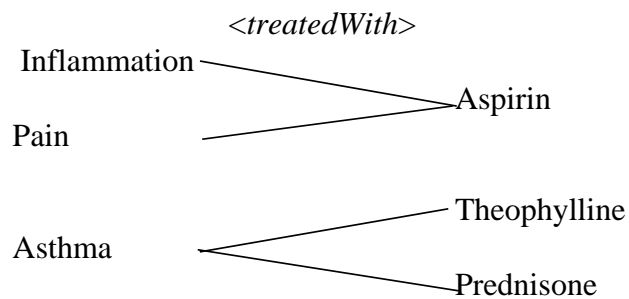
Note: Customarily, the term "fact" is used in this context, but "statement" or "assertion" would be better. "Fact" implies that an assertion is true, but not all so-called "facts" in a database are true.

A Patient facts

| | | |
|----------------------|------------------------------------|-----------------------------------|
| Patient facts | A1 Fred <hasDisease> Liver disease | |
| | A2 Fred <hasDisease> Asthma | (New disease, basis for question) |
| | A3 Phil <hasDisease> Inflammation | |

B Treatment facts

| | | |
|------------------------|---------------------------------------|-----------------------------|
| Question | What drugs are used to treat asthma? | Asthma <treatedWith> Drug X |
| Treatment facts | B1 Inflammation <treatedWith> Aspirin | |
| | B2 Pain <treatedWith> Aspirin | |
| | B3 Asthma <treatedWith> Theophylline | |
| | B4 Asthma <treatedWith> Prednisone | |
| Answer | Theophylline, Prednisone | |

Graphical representation of facts**C Contraindication facts** (one consideration in harmful side effects)

| | |
|---|---|
| Contra- indication facts | C1 Aspirin <contraIndicatedWith> Peptic ulcers |
| | C2 Theophylline <contraIndicatedWith> Peptic ulcers |
| | C3 Theophylline <contraIndicatedWith> Arrhythmia |
| | C5 Prednisone <contraIndicatedWith> Liver disease |

The physician uses contraindication facts to filter out drugs that treat the disease in question but would do harm to the patient and retain only drugs the patient tolerates.

The general rules for the drug prescription reasoning process are given on p. 11 – 12. These rules must be applied to a specific patient, in the example Fred. How this is done is explained step by step in UBLIS571Lecture1.2ExampleSlides.zip

Developing a system that can deduce answers by combining facts

Saves the physicians time. Need to give instructions (rules on how to combine facts in a language a computer can understand). We use *Prolog* (Programming in logic).

Overall strategy: Divide the problem into two sub-problems (as the physician does)

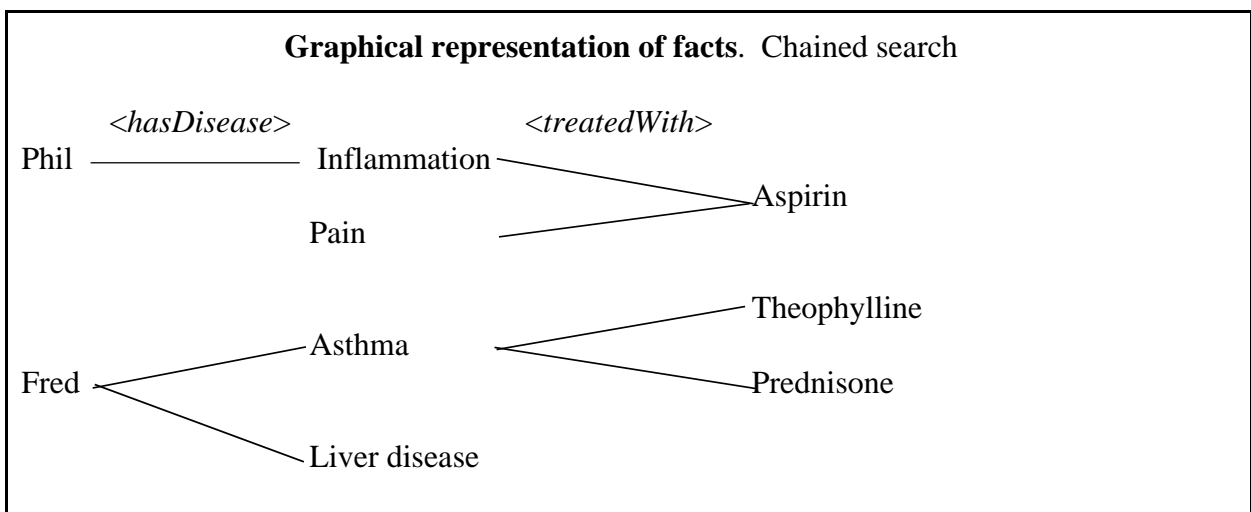
| | |
|-----------------|---|
| Question | What drug(s) should Fred take? Fred <i><should take></i> Drug X |
| Rule | Person X <i><shouldTake></i> Drug Z IF Person X <i><hasCandidateDrug></i> Drug Z AND Person X <i><tolerates></i> Drug Z /* filter for: drug does no harm */ |

First sub-problem. Combine patient facts with treatment facts to find candidate drugs

| | |
|-----------------|---|
| Question | What are candidate drugs for Fred Fred <i><hasCandidateDrug></i> Drug Z |
| Rule | Person X <i><hasCandidateDrug></i> Drug Z IF Person X <i><hasDisease></i> Disease Y AND /* from patient facts, A */ Disease Y <i><treatedWith></i> Drug Z /* from treatment facts, B */ |
| Answer | Theophylline, Prednisone |

This rule does a chained search, as further illustrated in the graph:

1. The first condition starts from a person and finds a disease.
2. The second condition starts from the disease found and finds candidate drugs



Note: We could build a drug prescription expert system that considers just the newest patient disease to be treated and treatment facts. But such a system would not be good for patients; more knowledge is needed to avoid drugs that would harm the patient.

Second subproblem: Combine patient facts with contraindication facts to retain only drugs the patient tolerates.

For example, Prednisone must not be taken by a person with liver disease, that is:

Prednisone <contraIndicatedWith> Liver disease

| | |
|-----------------|---|
| Question | What drugs does Fred tolerate. Fred <tolerates> Drug Z In the example, we need to apply this check to Theophylline and Prednisone |
| Rules | <p>Person X <tolerates> Drug Z IF Drug Z <contraIndicatedWith> Disease W AND Person X <freeOf> Disease W</p> <p>Person X <freeOf> Disease W IF NOT (Person X <hasDisease> Disease W)</p> <p>Person X <tolerates> Drug Z IF NOT (Drug Z <contraIndicatedWith> any Disease W) /* If a drug has no contraindications at all, it can be prescribed without checking the patients diseases. */</p> |
| Answer | Theophylline |

Note: If a drug is not ruled out based on the diseases seen from the patient record, a good system would **alert the physician to all contraindications** so that the patient can be checked out for these conditions.

| | |
|----------------------------|---|
| Further refinements | <ul style="list-style-type: none"> • drug effectiveness • drug side effects and their severity • drug interactions and incompatibilities • drug cost <p>A system containing all these data for a large number of drugs can prescribe as well as a human expert and would be called an expert system.</p> |
|----------------------------|---|

Some additional types of data used:

Drug effectiveness data (effectiveness may depend on several factors):

Disease <treatedWith> (Drug, Effectiveness, Ethnicity, Gender, Age)

Drug incompatibility data (bad effects that can happen if a person takes two drugs)

Drug <incompatibleWith> (Drug, Effect)

Example 2. **A database that supports the operation of a university**

Discussion of the example presented in Organizing Information, Chapter 3

Partial conceptual schema and some illustrative data for a university database

From Soergel, Organizing Information, Chapter 3

See next page

We will discuss in class how these interconnected data can be used to find answers by combining different kinds of facts; in the graphical representation this can be visualized as "chaining"

In supplement

Types of information systems from simple to complex (and more useful)

Characteristics of a good information system

Advanced ideas to ponder

Lecture 2.1*January 25***The nature of knowledge**

www.dsoergel.com/571/UBLIS571Lecture02.1Guide.pdf

| | |
|-------------------------------|--|
| Objectives | <ol style="list-style-type: none"> 1 Understand the characteristics and facets of different types of knowledge and be able to apply this understanding to an analysis of information needs, the organization of information, and the evaluation of information found. 2 Understand findings from cognitive psychology on the way people form and deal with concepts and to apply these findings to a better understanding of information needs, to the design of classifications, and to information presentation. |
| Practical significance | <p>Knowing about types of knowledge is important for</p> <ul style="list-style-type: none"> ⌋ understanding information needs (as in interviewing a library patron before doing a search – reference interview); ⌋ analyzing and assessing information found; ⌋ determining how to organize and process information /knowledge in accordance with its type; ⌋ matching documents to the needs of the patron according to the type of information they contain. <p>Knowing about types of concepts is important for understanding how people think and, therefore, how they ask questions, how they determine relevance, and how they process information. The answers, in turn, determine how information should be retrieved (retrieval should approximate human relevance judgment) and what information should be presented to a user in what form.</p> |

Outline

1 Types of knowledge: characteristics/facets/dimensions

- 1.1 Types of knowledge by content
- 1.2 Types of knowledge by scope of applicability
 - 1.2.1 Knowledge about regularities (laws, rules) vs. knowledge about individual detail from which the regularities can be derived
 - 1.2.2 Scope of applicability to natural or social phenomena. Scope of validity of a statement in space and time.
- 1.3 Types of knowledge by degree of "vagueness" of knowledge
- 1.4 Types of knowledge, other aspects

2 The nature of concepts / categories

- 2.1 Types of concepts. Individual concepts and class concepts
 - 2.1.1 Individual concepts – individual entities. Persistence over time
 - 2.1.2 Class concepts / categories. Simplified account
 - 2.1.3 Mass concepts (oil, flour, sugar) vs. count concepts (sugar cubes, books).
 - 2.1.4 Abstract concepts (freedom, justice). Can define the concrete class of all countries in which freedom prevails.
- 2.2 Objectivist vs. organism-centered view of categories
- 2.3 Explicit definition of categories vs. prototypes and fuzzy membership. Radial categories
- 2.4 Basic level categories (Eleanor Rosch)

1 Types of knowledge: characteristics/facets/dimensions

→ LIS 518 Reference Sources and Services: Selection of reference tools

1.1 Types of knowledge by content

Definitional knowledge (dictionary) vs. **assertive knowledge** (encyclopedia, world almanac).
Essential vs accidental attributes

(These are relative distinctions, see discussion of concepts in Section 2.1 below.)

Knowledge about **static relationships** ("what is the area of Nigeria") vs.

knowledge about **events and actions** ("what has happened in the religious strife in Nigeria")

Knowledge by subject area or by relationship type used in a statement

1.2 Types of knowledge by scope of applicability

The more widely applicable an item of knowledge, the more important to obtain it, validate it, and store it in an easily accessible form. There are several aspects or facets of scope.

1.2.1 Knowledge about regularities (general laws, rules) vs. knowledge about individual detail from which the regularities can be derived

Knowledge about regularities that apply to many cases or throughout a system (e.g. medical text book knowledge) vs. knowledge about individual detail that applies only to the individual case.

Examples:

| | | |
|---------------------------|---|---|
| Regularity: | Asthma < <i>treatedWith</i> > Theophylline | [Applies to all asthma patients] |
| Individual detail: | Fred < <i>hasDisease</i> > Asthma | [Applies just to one patient.] |
| Regularity: | FDST 257 < <i>hasPrerequisite</i> > FDST 101 FDST 257 < <i>isOfferedAs</i> > COF02 | [Applies to all students in any section] [Applies to all students in this section] |
| Individual detail: | COF02 < <i>hasStudent</i> > R. Smith | [Applies to this student] |
| Regularity: | Hearing tests < <i>hasNarrowerTerm</i> > Audiometry | [Applies to all searches for these concepts.] |
| Individual detail: | Document 4 < <i>dealsWith</i> > Audiometry | [Applies to retrieval of just one document.] |
| Regularity: | Kepler's laws of planetary motion | [Applies to all planets at all times] |
| Individual detail: | The observational data about planet positions | [Each observation applies to the position of one planet at one time] |
| Regularity: | Burglary is punishable with 3 - 10 years of prison. | [Applies to all burglaries] |
| Individual detail: | Weaver broke a large Window and entered the house. He took . . . | [Applies to this particular burglary] |

To generalize from two of these examples.

| Domain | Type of fact | Examples |
|-----------------------------|-----------------------------|---|
| Medical | Regularities, general facts | Facts about symptoms and diseases and treatments are broad; they apply to many cases. |
| | Individual detail facts | Facts about an individual patient are narrow; they apply only to one case. |
| Subject access to documents | Regularities, general facts | Facts about concept relationships are broad; they apply to all searches for the concepts involved and affect the retrieval of all documents indexed by one of the concepts. |
| | Individual detail facts | Facts linking a document to a concept (indexing facts) are narrow; they affect only the retrieval of this one document. |

| Ways to reason from past experience | |
|---|---|
| Regularities or laws are known | Reasoning from general laws (deductive): Draw conclusions on specific cases to which the laws can be matched. |
| Regularities or laws are <u>not</u> known | Case-based reasoning (inductive): Find similar past cases and assume the new case will have similar outcomes. Examples: Weather forecast Decide a legal case where the law is inconclusive (precedents) Decide on patient treatment based on past experience with similar cases |

| Two important specific kinds of knowledge about regularities | |
|--|---|
| Type of knowledge | Examples |
| Knowledge about restrictions on data values | <i>A male individual of a mammal species cannot be pregnant.</i> <i>A two-year-old human cannot weigh more than 30 pounds.</i> Used for checking data for errors |
| Default knowledge | Default knowledge: <i>A car has four wheels.</i> Specific knowledge about an individual case / knowledge about exceptions: <i>The Runabout has three wheels.</i> Default knowledge used all the time in daily life. Default values in date entry forms |

1.2.2 Scope of applicability to natural or social phenomena. Scope of validity of a statement in space and time

Regularities can differ in the scope in which they apply:

Examples

| Narrow scope | Broad scope |
|---|--|
| A law describing the free fall of objects towards the earth applies only on the earth (strictly speaking, only on a given point on the earth). Kepler's laws apply only to the movement of objects moving in an orbit around another object (originally they were conceived as applying only to the movement of the planets around the sun). | The general law of gravity applies to many phenomena throughout the universe; many more specific laws, like the two mentioned, can be derived from it. |
| A property value for a specific material (such as the electrical conductivity of copper) applies only to phenomena involving that material. | The gravitational constant holds through the entire universe (or so physicists think) and is involved in many phenomena. |
| A social rule, custom, or etiquette rule that applies only in one country | A social rule, custom, or etiquette rule that applies world-wide |
| Many rules of grammar apply to only one language family (such as all Indo-European languages) or to only a single language. | Some linguists believe that some principles apply to all languages (language universals). |

| Related distinction | |
|---|--|
| Domain-specific knowledge | Common sense knowledge |
| Examples: Effects of a drug; how to teach math to fourth-graders | Example: Use cost-benefit analysis general principles of management |

Example:

A World Bank project aimed at improving schools by training education ministry staff in
(1) general management procedures, especially procurement,
(2) domain-specific (in World Bank speak: sector-specific) knowledge in education.
The country staff found (1) just as useful or more useful because it could be transferred to other domains.

How are scope and usefulness of knowledge related?

1.3 Types of knowledge by degree of "vagueness" of knowledge

"Vagueness" is a vague umbrella term for the more well-defined distinctions listed below.

Concepts of "vagueness" of knowledge can be applied, for example, to knowledge of document relevance to a user's request (see below).

Rarely is knowledge "hard and fast"; it is important to assess a given knowledge item along the vagueness dimension to see how sound a basis it is for decision making

1.3.1 Precise vs. imprecise knowledge

This has to do with the error range or rounding with which variable values are given.

Examples

Knowing that a certain Shakespeare quote is from Macbeth is less precise than knowing it is from Macbeth, Act 2, Scene 3.

Persons' names with increasing precision: last name with initial; full first name; plus middle initial; full middle name; plus date of birth (important in cataloging)

Numeric values derived from empirical data, such as physical measurements or poll results, are subject to error; they have an error range. In reporting such numbers, give only significant digits and preferably indicate the error range to avoid conveying unwarranted precision.

Do not give data with more precision than is known (warranted by the measurement) (e.g., polling data).

Do not give data with more precision than is required even if more precision is known. (In a financial report to the board a \$10 million organization, round numbers to the nearest thousand, but for balancing the books use numbers to the penny.)

1.3.2 Certain vs. uncertain knowledge

Linked with risk. Combined with precision: Confidence intervals

Yes/no statements (such as facts or rules in an experts system) vs. probabilistic statements.

We can assume (erroneously) that a document either is relevant to a user's question or it isn't, with no shades of gray in between (see next point for a different stance), yet still say that document X has a probability of 0.7 of being relevant. Our knowledge about relevance is uncertain.

1.3.3 Graded assertions

For example, a document can be highly relevant or somewhat relevant. This can be expressed by a numerical score between 0 and 1. In other words, the set of relevant documents had no sharp boundary but rather is a fuzzy set. We cannot say that a document is a member of the set or that it is not a member of the set; rather, membership in a fuzzy set is a matter of degree.

1.3.4 Unambiguous vs. ambiguous statements (including intentional ambiguity)

Oder-Neisse line as border between Poland and Germany after WW2.

1.3.5 Facts (“true”, “objective”) or statements asserted as facts **vs. opinion**

"Hard" statements vs. judgment statements. News page vs. editorial page

1.3.6 Knowledge about the accuracy, certainty, or trustworthiness of facts or rules

1.4 Types of knowledge, other aspects

1.4.1 **Beliefs.** Need to indicate whose belief.

1.4.2 **Modality of knowledge items** (descriptive, prescriptive, statement of possibility)

Descriptive statement: The car is going 50 miles per hour (what is)

Prescriptive statement: The speed limit is 45 miles per hour (what should be)
(a prescription for drivers as to speed of their cars)

Possibility statement: The car can go 100 miles per hour (what could be)

More examples

Descriptive knowledge: Knowledge about the effects of calorie intake, specific nutrients (such as vitamin E), and exercise.

Prescriptive knowledge: Guidelines on nutrition and exercise.

People writing to an advice columnist report the facts of the case as they see them – descriptive knowledge. The advice columnist tells them what to do – prescriptive knowledge.

The law is prescriptive knowledge

Politicians and planners deal with the art of the possible; they need knowledge of what is possible. For example, some people claim to know that it is not possible to change Social Security because public opinion is against it and the votes to pass legislation are not there. On the other hand, visionary politicians may defy conventional knowledge of what is possible and make things possible. Time horizon of statements about possibility.

1.4.3 **Knowledge about what kinds of knowledge are important:** Conceptual data schema (introduced in Lecture 1.2)

2 The nature of concepts / categories / classes

Importance: The nature of concepts is fundamental to information processing in people and in machines (see readings, particularly Skemp). Another way of looking at types of knowledge.

2.1 Types of concepts. Individual concepts and class concepts

2.1.1 **Individual concepts – individual entities.** Persistence over time

2.1.2 **Class concepts / categories.** Simplified account

See Sections 2.2 - 2.4 for a discussion of the complexities of the structure of categories.

| Concepts have | |
|---|--|
| Intension, intensional definition, "meaning" | <p>Definitional knowledge as opposed to world knowledge (empirical knowledge) A concept or class defined in terms of attributes or characteristics all entities must possess in order to be members of the class. These are called essential attributes or characteristics. A characteristic of an individual entity can expressed in several ways:</p> <ol style="list-style-type: none"> (1) the entity possesses an attribute (2) the entity is capable of entering a given relationship (occupy a comparable place in a network of relationships) <p>A query formulation is a definition. It defines what it means for a document (or a person, or a computer program) to be relevant for the user. It encapsulates the user's intention.</p> <p>Description logic provides a formal way for defining concepts in a classification or ontology used in the semantic Web.</p> |
| Extension | <p>The set of individual entities belonging to the category</p> <p>For example, the set of all relevant documents</p> <p>In some cases it is possible to define a category by exhaustively listing all its members. This is called an extensional definition.</p> <p>Some defined concepts have empty extensions in reality, see below.</p> |

Example definitions (isa = is a type of):

| Class | Definition |
|---------------------------------|---|
| red balls | All objects that meet two conditions: Object <i><hasType></i> Ball and Object <i><hasColor></i> Red |
| pews | All objects that meet two conditions: Object <i><isa></i> Bench and [Object <i><locatedIn></i> Building, Building <i><isa></i> Church] |
| government documents | All documents that meet the conditions: Document <i><publishedBy></i> Organization, Organization <i><hasType></i> Government agency. |
| water-soluble substances | All chemical substances that meet the condition Substance <i><solubleIn></i> Water. |
| ? | All English words that meet the condition Word <i><canServeAsObjectFor></i> Refine. |

| | |
|------------------------------|---|
| Essential attributes | Attributes that are used in the definition of a class |
| Accidental attributes | Any other attribute that one or more members of a class may possess |

It often happens that all members of a class share an accidental attribute, that is an attributes that is not defining but nevertheless present in each member of the class. Such a general law can be determined by observation.

Example: Assume it is true that all government documents are authoritative.

Thus, if we have ascertained that a given document meets the definition for government document, we can conclude that the document is authoritative (knowledge of a regularity).

It is this ability to predict the behavior of an entity once it has been identified as belonging to a concept/category that makes for the usefulness of concepts; concepts are essential for economy of mental operations.

In law:

Fit facts of the case to a legal concept, for example determine that the facts of a case meet the definition of burglary.

Then apply the legal rule applicable to that concept.

Erroneous generalization: Stereotypes.

Relationship of definition to empirical data. Examples: One can define an animal species through a list of attributes such that no animal existing in nature fits the definition, for example, "an animal that looks like a horse and has a horn in the middle of the forehead". There are no unicorns in the real physical world, but there are plenty in fiction. Such fictitious animals are written about and depicted and become objects of searches. *Imaginary animals* is a very popular search topic in the International Children's Digital Library (ICDL).

2.1.3 **Mass concepts** (oil, flour, sugar) vs. **count concepts** (sugar cubes, books)

2.1.4 **Abstract concepts** (freedom, justice). Can define the concrete class of all countries in which freedom prevails.

2.2 Objectivist vs. organism-centered view of categories

(Quotes from Lakoff, *Women, fire, and dangerous things*. U. of Chicago Pr.; 1987)

Next page

2.2 Objectivist vs. organism-centered view of categories

Important: Information is not just transmitted but needs to be actively processed and assimilated by the receiver or learner (see the last paragraph of this section).

Objectivist view of categories (as characterized by George Lakoff)

- Symbols that correspond to the external world are *internal representations of external reality*.
- Abstract symbols may stand in correspondence to things in the world independent of the particular properties of the organism that holds the symbols.
- Since the human mind makes use of internal representations of external reality, the mind is *a mirror of nature*, and correct reason mirrors the logic of the external world. (p. XIII)

Organism-centered view (DS term) of categories (George Lakoff)

- Do meaningful thought and reason concern merely the manipulation of abstract symbols and their correspondence to an objective reality, independent of any embodiment (except, perhaps, for limitations imposed by the organism)?
(Summary of the objectivist view for contrast)
- Or do meaningful thought and reason essentially concern the nature of the organism doing the thinking – including the nature of its body, its interactions in its environment, its social character, and so on? (p. XV - XVI) (Organism-centered view)
Embodied cognition versus symbolic representation

A balanced view (D. Soergel)

- Interacting with the physical, social, and intellectual world around us as well as with our own selves, we form complex mental models which allow us to better understand the world around us and better understand ourselves and thus help us to take actions in the world towards achieving our objectives. This formation of mental models has a social dimension; it is often done in interaction with or building on the models of others – as in group learning.
- These mental models, which include concepts / categories, reflect a structuring of experience in ways useful to the person. A person's experience is shaped by perceptions of the world (within the limitations of the person's faculties for perception and thought) and by the modes of interaction with objects in the world. Thus, **a mental model is not simply a mirror image of the outside world but rather an actively shaped image, adapted to the person's needs**, often distorted, often enriched (or contaminated, depending on one's point of view) with elements that have no counterpart in the “real” world (but might well be realized as the person shapes the world).
See Supplement

The sense-making approach to information service

The view of mental models, concepts, and categories presented above is important for an understanding of how people use information and what information should be provided to people. In this view, a person must assimilate information into her mental model; a person **must make personal sense of the information**. Different people may get different things out of the same document. In that sense, one might say that information does not exist objectively, but only as it gives rise to a change in a person's mind. Or that a book does not convey information as much as it is a stimulus for the reader to create and elaborate her own information in her own mind.

The sense-making approach in related disciplines

In literary theory

This is the position taken by **reader response theory**. The expression "I did not get much out of it" is in tune with this "active reader" position. The art of giving the reader a "relevant" book, then, is to find a book that allows this reader with his background and mental models to "get something out of" the book, to construct his own knowledge, updating his mental models in a way that will help him to find better solutions to the problems he faces.

In education

The **constructivist theory of learning** holds that we learn best by constructing or reconstructing knowledge for ourselves.

Discovery learning is a closely related approach. It holds that a students learn best when they explore a subject and discover facts and relationships for themselves. In science this means that students discover scientific laws through their own experiments.

The **job of the teacher or information specialist** then is to create an environment, including access to information, that enables students or users to do their own discovery and knowledge construction with guidance provided only to the extent necessary ("scaffolding").

2.3 Explicit definition of categories vs. prototypes and fuzzy membership

Radial categories

Prototype. Example *Chair*:

Chair, living room chair, kitchen chair, lawn chair, easy chair, rocker, armchair, chaise longue, bar stool, stool?, footstool??. ottoman??

Necessary attributes vs. sharing a sufficient number of attributes.

Knowledge of concepts stored in memory as explicit definitions or prototypes?

In reality a mixture of both?

Importance of examples in thesaurus scope notes

Radial category. Example: *Mother*

(a category that has a central case but then many cases deviating more or less in different directions)

There are many "models" of what a *mother* is (Lakoff 1987, p. 83).

"The central case, where all the models converge, includes a mother who is and always has been female, and who gave birth to the child, supplied her half of the child's genes, nurtured the child, is married to the father, is one generation older than the child, and is the child's legal guardian."

The following cases share some, but not all, of these features. The first four emphasize a biological perspective, the others a social perspective.

- Biological mother (also called natural mother, but that term was abandoned)
- Birth mother (term for biological mother in the context of adoption)
- Surrogate mother
- Genetic mother
- Rearing mother
- Stepmother
- Adoptive mother
- Foster mother
- Unwed mother

Importance for reference interview and searching.

What would you search for if the user says he wants documents about *mother*?

Understanding radial categories helps you ask the user questions to better pinpoint the information need.

Application to retrieval:

The documents (or statements, such as statements of fact or hypotheses, etc.) relevant to a query constitute a category (a subset of all documents). We can define such a category in two ways:

- through a query formulation that explicitly specifies the features that make a document relevant (expressing the intent of the user, intensional definition). (This query formulation could be applied in a Boolean search (to be retrieved, a document needs to meet all conditions) or in a ranked retrieval search (documents are retrieved even if they do not meet all conditions exactly, and are ranked by how closely they meet the conditions);
- through a sample document that serves as a prototype of relevant documents (“more like this”) or several documents that serve as examples.

The category “relevant documents” can be a radial category when there are different ways in which a document can be relevant to the query. Needs several query formulations.

2.4 Basic level categories: application of categories or concepts to action (Eleanor Rosch)

From this perspective, what categories are most useful and worth the effort to learn?

Example. Gain in knowledge of what to do as concepts get more specific

- If somebody tells you that there is a piece of *furniture* in a room you have been assigned, that does not tell you much. You still do not know what you can do in the room.
- If somebody tells you there is a *chair* in the room, that tells you a lot more; you know you can sit.
- If somebody tells you there is a *easy chair* in the room, you know a little more, but not much more; you still know only that you can sit(perhaps a bit more comfortably).

There is a **big information gain from *furniture* to *chair***, but a **small gain from *chair* to *easy chair***. So it is worthwhile to learn about the category of *chair*, but the added benefit of knowing all the many specific types of chair would be low and the learning effort would be very high.

chair is at the optimal level in the hierarchy, it is a **basic level category**.

More examples

| Superordinate | Basic level | Subordinates | |
|---------------|-------------|-------------------|-------------------|
| Furniture | Chair | Kitchen chair | Living room chair |
| | | Lawn chair | Bar stool? |
| | | Easy chair | Footstool?? |
| | Table | Dining room table | Kitchen table |
| | | Card table | Pool table |

The idea of basic level categories is important for information services and learning and instruction in many ways, for example:

- Users tend to ask questions using basic level categories even if their information need is more specific. The reference librarian (or a computer system trying to replace the reference librarian) needs to recognize when it is necessary to probe for the real specific information need.
- Basic level categories are learned first. A classification scheme for children must be built from basic level categories.

Empirical results of studies in cognitive psychology

- Subjects were given words that name categories of objects, such as *furniture*, *chair*, *lawn chair*, and were asked to list **attributes** of that category.

For **superordinate categories**, such as *furniture*, subjects listed **few attributes**.

For **intermediate categories**, such as *chair*, subjects listed **many attributes**.

Basic level

For **subordinate categories**, such as *easy chair*, subjects listed **a few additional attributes** beyond those for *chair*.

"Basic level categories are the most inclusive level of classification at which **objects have a significant number of attributes in common**." (p. 214)

- Basic level categories are the most inclusive level of classification at which **objects share substantive functionality**. (p. 217). Example:

furniture (*table*, *chair*, *cabinet*) (**superordinate category**):

Few, if any movements or other things you do are **in common** to all types of furniture.

chair (*any type*) (**basic level**):

People make the **same kind of movements** (sitting down) for all types of chair

easy chair (**subordinate category**):

movements are hardly different from any other type of chair

- Basic level categories are **learned first**.

| Level | Number of attributes | Number of instances | Number of categories | Usefulness for action |
|---------------|--|---------------------|---------------------------------|-----------------------|
| Superordinate | Few attributes | Gazillion instances | Few categories | Low |
| Basic level | Many attributes (high information gain) | Many instances | Medium number of categories | High |
| Subordinate | Only a few more attributes (low information gain) | Few instances | Very large number of categories | Only slightly higher |

Note: Basic level may depend on group - culture and subculture.

Elaboration in supplement

Lecture 2.2**January 25****Knowledge representation**

→ LIS 506 IT, LIS 569 Data Management

www.dsoergel.com/571/UBLIS571Lecture02.2Guide.pdf

| | |
|-------------------------------|--|
| Objectives | <ol style="list-style-type: none"> 1 The student should understand the different approaches to knowledge representation. 2 The student should understand different mechanisms of knowledge processing. 3 The student should be able to apply this understanding to analyze the knowledge structure of existing systems and designing the knowledge structure for a system to be built. 4 The student should be able to apply this understanding to finding answers to a user's question by combining knowledge from one or more systems. |
| Practical significance | Knowing about system-internal knowledge representation and conceptual data schemas is important for organizing a body of knowledge for retrieval and beyond that, for inference, that is, for a system that can draw conclusions from the knowledge stored (and thus create new knowledge), rather than simply retrieving what is there. |

Key idea:**Hierarchical inheritance**

Using knowledge about a class to answer questions about any member of the class.

Using knowledge about a broad class to answer questions about any subordinate class or any member of such a subordinate class.

Reorganize a set of data so it requires less storage space and can be communicated more efficiently.

Outline

- 0 Forming categories in a set of entities to create a more efficient data structure using hierarchical inheritance – introductory exercise
- 1 Definition of knowledge representation (in the mind, on paper, for computers)
Examples
- 2 Approaches to knowledge representation
- 3 Some mechanisms in knowledge representation
- 4 Some criteria for describing and evaluating knowledge representations (Supplement)

0 Forming categories in a set of entities to create a more efficient data structure using hierarchical inheritance. In-class exercise

Consider the menu listing below (the actual menu as it appears on the restaurant's website is on the back of this page).

How could this menu be presented so it takes less space and is easier to read

Hint: Try to figure out what is actually different for each choice

| Third Course | | |
|---|--|--|
| Fondue Feast | Fondue Fusion | Lobster Indulgence |
| Filet Mignon, Sauerbraten NY Strip, Nueske's Applewood Smoked Bratwurst, Hefeweizen Marinated Shrimp, Roasted Garlic Crusted Chicken, Sun Dried Tomato Ravioli and Fresh Vegetables. | Lobster Tail, Filet Mignon, Sauerbraten NY Strip, Nueske's Applewood Smoked Bratwurst, Hefeweizen Marinated Shrimp, Roasted Garlic Crusted Chicken, Sun Dried Tomato Ravioli and Fresh Vegetables. | Lobster Tail(s), Sauerbraten NY Strip, Nueske's Applewood Smoked Bratwurst, Hefeweizen Marinated Shrimp, Roasted Garlic Crusted Chicken, Sun Dried Tomato Ravioli and Fresh Vegetables. |

| Revised Presentation |
|-----------------------------|
| |

ALPINE BIG NIGHT OUT



• FOR A LIMITED TIME ONLY •

TAKE A CULINARY JOURNEY TO THE PLACE WHERE FONDUE ORIGINATED...

We are taking you back to the birthplace of fondue, the Alps.
This feature menu highlights the delicious recipes of oldeworld Switzerland, Germany and France.

FIRST COURSE

Alp and Dell Cheese Fondue

A delicious combination of white wine, garlic, whole grain mustard and nutmeg, with a smooth blend of award-winning Gruyère, Raclette and Fontina cheeses.

SECOND COURSE

Alpine Ridge Salad

A bed of mixed greens topped with oven roasted tomatoes, award-winning Gruyère, Raclette and Fontina cheeses, hard-boiled eggs and honey roasted almonds, topped with a robust shallot vinaigrette.

THIRD COURSE

Fondue Feast*

Filet Mignon, Sauerbraten NY Strip, Nueske's Applewood Smoked Bratwurst, Hefeweizen Marinated Shrimp, Roasted Garlic Crusted Chicken, Sun Dried Tomato Ravioli and Fresh Vegetables.

\$42.00 per person

Fondue Fusion*

Lobster Tail, Filet Mignon, Sauerbraten NY Strip, Nueske's Applewood Smoked Bratwurst, Hefeweizen Marinated Shrimp, Roasted Garlic Crusted Chicken, Sun Dried Tomato Ravioli and Fresh Vegetables.

\$88.00 per couple

Lobster Indulgence*

Lobster Tail(s), Sauerbraten NY Strip, Nueske's Applewood Smoked Bratwurst, Hefeweizen Marinated Shrimp, Roasted Garlic Crusted Chicken, Sun Dried Tomato Ravioli and Fresh Vegetables.

\$48.00 per person

ENTRÉE COOKING STYLES

Coq au Vin

Flavors of fresh herbs, mushrooms, garlic, spices and burgundy wine.

Mojo Style

Caribbean-seasoned bouillon with distinctive flavors of fresh garlic with a citrus flair!

Court Bouillon

Homemade, seasoned vegetable broth.

Bourguignonne

European-style fondue in cholesterol-free canola oil.
0g trans-fat Oil

FOURTH COURSE

White Chocolate Apple Cobbler Fondue

White chocolate with seasoned apples, streusel topping and spices.

All entrées are cooked tableside in any of our four flavorful cooking styles located above.

Regular menu item substitutions allowed. Ask your server for details.



STIR THINGS UP.

*Our Fondue-style service may result in the undercooking of certain ingredients.

1388.09 Consuming raw or undercooked MEATS, POULTRY, seafood, shellfish or EGGS may increase your RISK for foodborne

Reorganize these bibliographic records, using hierarchical inheritance for efficient storage

Document 1

100 1 Mager, Robert Frank, \$d 1923-
 245 10 Developing attitude toward learning /
 \$c Robert F. Mager.
 260 Belmont, Calif. : \$b Fearon/Pitman Publishers,
 \$c c1968.
 300 vii, 104 p. ; \$c 22 cm.
 650 0 Interaction analysis in education.
 650 0 Learning, Psychology of.
 650 0 Group work in education.
 650 0 Classroom management.

Document 4

100 1 Conant, James Bryant, \$d 1893-1978
 245 10 The comprehensive high school; \$b a second
 report to interested citizens \$c by James B.
 Conant.
 260 New York, \$b McGraw-Hill \$c [1967]
 300 vi, 95 p. \$c 21 cm.
 650 0 Education, Secondary
 650 0 Education \$z U.S. \$y 1945-
 650 0 Comprehensive High Schools \$z U.S. \$y 1945

Document 2

100 1 Candelora, D[eborah] M.
 245 10 Hands-on technology program \$h [computer
 file]
 246 HOT program
 260 [Ramsey, NJ]: \$b [Galaxy Networks], \$c 1996-
 500 Title from the home page HTML title
 500 Material copyrighted by D. M. Candelora
 500 Accessed 1998 Feb. 2
 650 0 Science - Study and teaching (Elementary) -
 Aids and devices
 650 0 Science - Experiments
 650 0 Computers - Study and teaching (Elementary) -
 Aids and devices
 650 0 Mathematics - Study and teaching (Elementary)
 - Aids and devices
 650 0 Learning by discovery
 650 0 Active learning
 856 4 \$u www.galaxy.net/~k12/ \$n Ramsey, NJ

Document 5

100 1 Mager, Robert Frank, \$d 1923-
 245 10 Developing attitude toward learning, \$b or,
 SMATS "n" SMUTS / \$c Robert F. Mager
 250 2nd ed.
 260 Belmont, Calif. : \$b David S. Lake, \$c c1984.
 300 x, 116 p. : \$b ill. ; \$c 24 cm.
 490 1 The Mager library
 500 Rev. ed. of: Developing attitude toward
 learning. 1968.
 650 0 Interaction analysis in education.
 650 0 Learning, Psychology of.
 650 0 Group work in education.
 650 0 Classroom management.

Document 3

100 1 Mager, Robert Frank, \$d 1923-
 245 10 Developing attitude toward learning : \$b or
 SMATs 'n' SMUTS / \$c Robert F. Mager.
 250 2nd ed.
 260 London : \$b Kogan Page, \$c 1991, c1990.
 300 116 p. ; \$c 23 cm.
 650 0 Interaction analysis in education.
 650 0 Learning, Psychology of.
 650 0 Group work in education.
 650 0 Classroom management.
 650 0 Students \$a Motivation

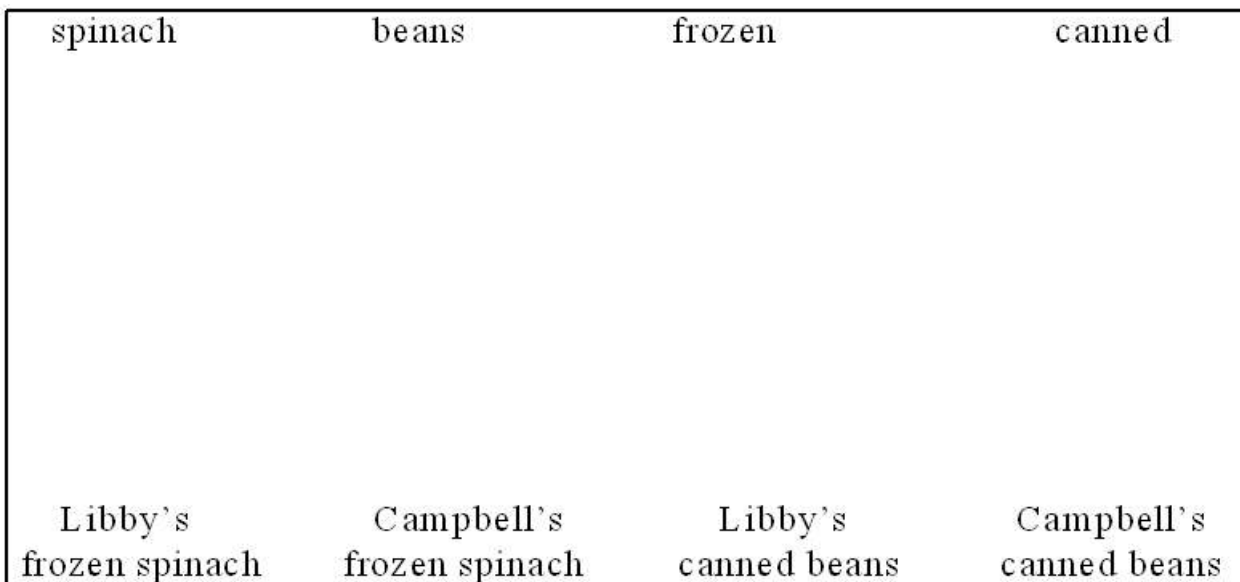
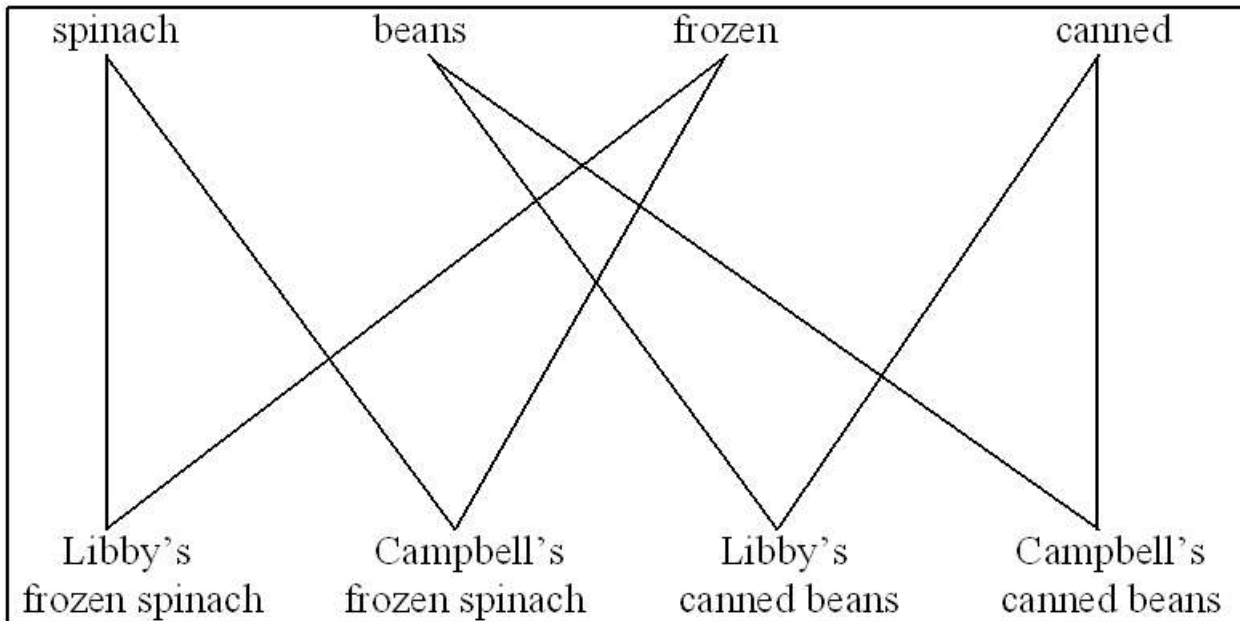
Document 6

100 1 Mager, Robert Frank, \$d 1923-
 240 10 Developing attitude toward learning. \$l Spanish
 245 10 Desarrollo de actitudes hacia la ensenanza /
 \$c Robert F. Mager.
 260 Barcelona : \$b Martacinez Roca, \$c c1985.
 300 158 p. : \$b ill. ; \$c 19 cm.
 650 0 Interaction analysis in education.
 650 0 Learning, Psychology of.
 650 0 Group work in education.
 650 0 Classroom management.

In the following example, there are just lots of data but no classes. Our task is to find groups of

entities that have common characteristics so we can form a class (a new node in a semantic network); the common characteristics are stored only for the new class (node) and inherit down to all the members. Finding entities with common characteristics is an important element in the process of concept formation.

Look at the semantic network below. How can it be restructured for more efficient storage? Complete the second copy of the network to show your restructuring.



Original database**Food product 1. Libby's frozen spinach**

Food: spinach

Preservation: frozen

Manufacturer: Libby

Food product 2. Campbell's frozen spinach

Food: spinach

Preservation: frozen

Manufacturer: Campbell

Food product 3. Libby's canned beans

Food: beans

Preservation: canned

Manufacturer: Libby

Food product 4. Campbell's canned beans

Food: beans

Preservation: canned

Manufacturer: Campbell

Restructured database

Note: The restructured database has more records, but they are much shorter

A good example for hierarchical inheritance is a cookbook which may have a basic recipe for potato soup and then many variations that say,

Rosemary potato soup

“Use the recipe for potato soup but add 1 ½ teaspoons minced fresh rosemary .”

1 **Definition of knowledge representation**

(in the mind, on paper, for computers)

Knowledge representation is the expression of knowledge through a system of symbols or signs, such as words, Dewey numbers, or icons. A knowledge representation scheme must provide

- symbols that refer to objects in the world or objects in the mind (put differently, symbols that refer to entity values, roughly, nouns);
- symbols that refer to relationship types (roughly, verbs);
- a syntax that allows for the expression of statements consisting of entity identifiers linked through relationship symbols.

Natural language is a very expressive knowledge representation system, but it is hard for a computer system to figure out what a natural language text means and then act on this knowledge. Need simpler KR systems for useful computer systems.

Approaches Entity-relationship representation (very common in the database field)

Semantic network representation

Frame representation (artificial intelligence & object-oriented programming)

Note: In 571 we talk about knowledge representation in the abstract. Implementation in databases is treated briefly in 506 Information Technology and extensively in 569 Data management. In the supplement there is an assignment that takes you through creating and querying a simple Microsoft Access implementation of the University Database.

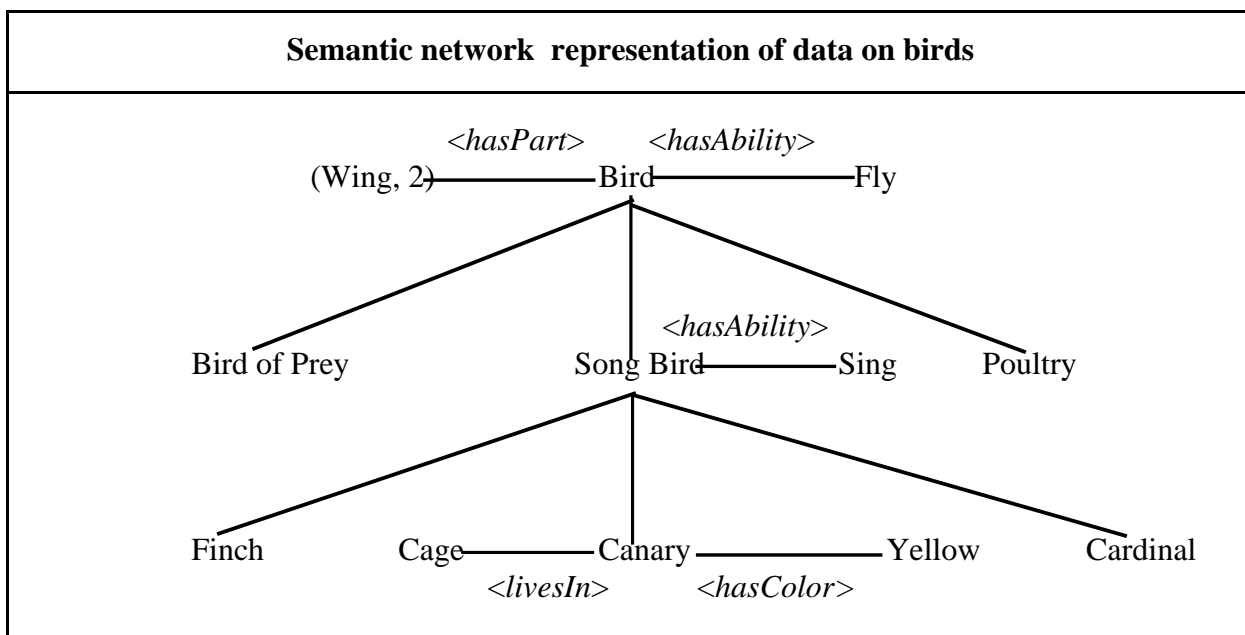
Knowledge representation examples (for computer systems)

A small example with data on birds

A large example with data on food products

Simple example: Representation of data on birds

| Entity-relationship representation of data on birds | | | | | |
|---|-------|-----------|-----------|--------------|-----------|
| Canary | <isa> | Song Bird | Canary | <hasColor> | Yellow |
| Finch | <isa> | Song Bird | Song Bird | <hasAbility> | Sing |
| Cardinal | <isa> | Song Bird | Bird | <hasAbility> | Fly |
| Songbird | <isa> | Bird | Bird | <hasPart> | (Wing, 2) |
| Bird of Prey | <isa> | Bird | Canary | <livesIn> | Cage |
| Poultry | <isa> | Bird | Bird | <isa> | Animal |



| | |
|--|---|
| <p>Key idea:</p> <p>Hierarchical inheritance</p> | <p>A node lower in the hierarchy inherits the characteristics of nodes above it. For example, <u>Song Bird</u> <hasAbility> <u>Fly</u> and <u>Song Bird</u> <has part> <u>(Wing, 2)</u> are both true; <u>Song Bird</u> inherits these characteristics from <u>Bird</u>.</p> <p>Note: The existence of a hierarchy alone does not make hierarchical inheritance. Only when the hierarchy is used to pass down characteristics from higher nodes to lower nodes is there hierarchical inheritance</p> |
| <p>Spreading activation</p> | <p>Activation (or attention) may spread from a node to a neighboring node: A person thinking about <u>yellow</u> (<u>yellow</u> is activated) may be reminded of a <u>canary</u> (<u>canary</u> is activated), and then of <u>song bird</u> and then of a bird singing (<u>sing</u> is activated) and then, again starting from <u>song bird</u>, of any <u>bird</u> and thus of a bird <u>flying</u>)</p> |

Frame representation of data on birds with hierarchical inheritance

| | |
|-------------------------------|--|
| <i>Frame for:</i> Bird | |
| <i>isa:</i> | Animal |
| <i>includesSpecific:</i> | Song Bird; Bird of Prey, Poultry /* This slot does not inherit down */ |
| <i>hasColor:</i> | |
| <i>hasPart:</i> | (Wing, 2) |
| <i>hasAbility:</i> | Fly |
| <i>livesIn:</i> | |

| | | |
|------------------------------------|-------------------------|----------------------------------|
| <i>Frame for:</i> Song Bird | | [inherited] |
| <i>isa:</i> | Bird; [Animal] | |
| <i>includesSpecific:</i> | Canary; Finch; Cardinal | |
| <i>hasColor:</i> | | |
| <i>hasPart:</i> | [(Wing, 2)] | Anything that is true for a bird |
| <i>hasAbility:</i> | Sing; [Fly] | is true for a songbird through |
| <i>livesIn:</i> | | inheritance |

| | | |
|---------------------------------|-------------------------------|-------------------------------------|
| <i>Frame for:</i> Canary | | [[inherited from two levels above]] |
| <i>isa:</i> | Song Bird; [Bird]; [[Animal]] | Anything that is true for a bird or |
| <i>includesSpecific:</i> | | a song bird is true for a canary |
| <i>hasColor:</i> | Yellow | through inheritance |
| <i>hasPart:</i> | [(Wing, 2)] | |
| <i>hasAbility:</i> | [Sing]; [Fly] | Anything that is true for a bird or |
| <i>livesIn:</i> | Cage | a song bird is true for a canary |
| | | through inheritance |

| | |
|---|---|
| <i>Frame for:</i> Penguin /* added to illustrate inheritance override */ | |
| <i>isa:</i> | Bird; [Animal] |
| <i>includesSpecific:</i> | |
| <i>hasColor:</i> | White; Black |
| <i>hasPart:</i> | (Wing, 2) |
| <i>hasAbility:</i> | Swim; NOT Fly; [Fly](example of overriding an inherited piece of data) |
| <i>livesIn:</i> | Antarctica |

See <http://percevia.duncraft.com/> for a bird database using more bird attributes

This page is intentionally blank

More elaborate example: Representation of data on food products

| | |
|---|--|
| <p>Purposes of the food information system</p> | <p>Government agency: Determine the safety of a food product Consumer: Find food products to be avoided with a given allergy Cook: Prepare a food product Food manufacturer: Produce the ingredient label</p> |
| <p>Sample questions</p> | <p>Find all products to be avoided by people allergic to eggs. I have cauliflower, onions, and tomatoes I need to use up. Find a good recipe.</p> |

Entity-relationship (E-R) representation

Conceptual data schema (entity types and relationship types covered; *isa* is short for *is a*)

| Entity types | Relationship types | |
|-------------------|---|-------------------------------------|
| Name | Food product <hasName> | Name |
| Food product (FP) | Food product <isa> | Food product |
| Organism | Food product <comesFromSource> | Organism |
| Person | | |
| Organism part | Food product <comesFromPart> | Organism part |
| Substance | Food product <isExtractedSubstance> | Substance |
| Form | Food product <isMadeFrom> | Food product ¹ |
| Process | Food product <hasIngredient> | Food product |
| Purpose | Food product <withPurpose> | Purpose |
| Container | | |
| Good, commodity | Food product <contains> | Substance (omitted in the examples) |
| Portion | | |
| Legal entity | Food product <processedBy> | Process |
| Person | Food product <withIntensity> | Intensity |
| Money number | Food product <withPurpose> | Purpose |
| | Food product <hasForm> | Form |
| | Food product <packedIn> | (container, volume) |
| | Organism <eat> | Portion or Substance or FP |
| | <buy/sell> (Legal entity [seller], Legal entity [buyer], Good, Money no.) | |

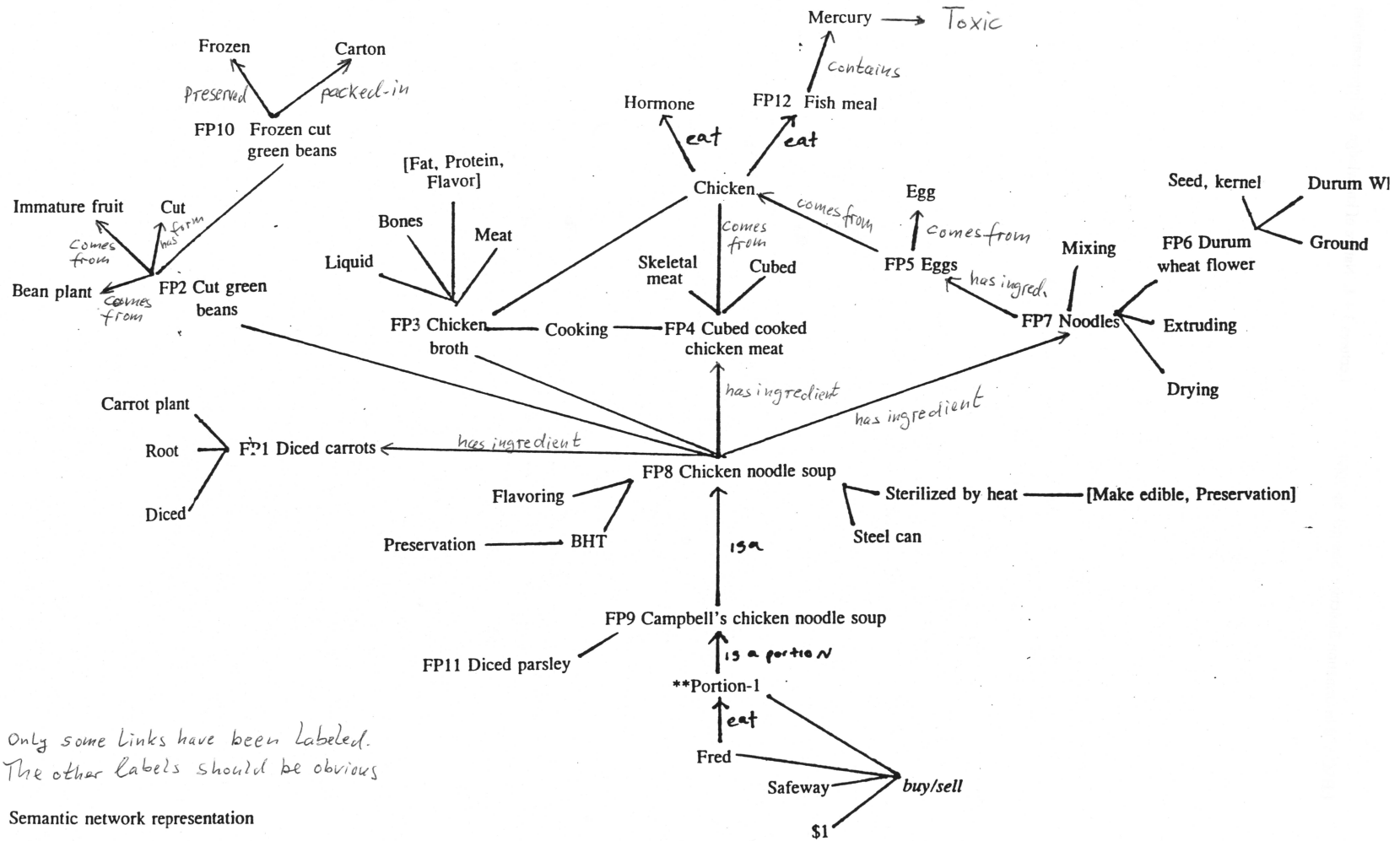
¹Not used in the lecture examples, but in the reading on a food description language

Entity values can be seen from the E-R statements (FP0, FP1, etc, Plant, . . .) FP = Food Product

E-R statements [] inherited from one level above [[]] inherited from two levels above

| | | | |
|-----------------------------|--------------------------|------------------------|-----------------------------------|
| FP0 <hasName> | Food product | FP14 <hasName> | Cubed cooked chicken |
| FP1 <hasName> | Vegetable product | FP14 <isa> | FP2 Meat product |
| FP1 <isa> | FP0 Food product | FP14 <comesFromSource> | Chicken |
| FP1 <comesFromSource> | Plant | FP14 <comesFromPart> | Skeletal meat |
| FP2 <hasName> | Meat product | FP14 <processedBy> | Cooking |
| FP2 <isa> | FP0 Food product | FP14 <hasForm> | Cubed |
| FP2 <comesFromSource> | Animal | FP15 <hasName> | Eggs |
| FP2 <comesFromPart> | Carcass | FP15 <isa> | FP3 Egg product |
| FP3 <hasName> | Egg product | FP15 <comesFromSource> | Chicken |
| FP3 <isa> | FP0 Food product | [FP15 <comesFromPart> | Egg] |
| FP3 <comesFromSource> | Animal | FP16 <hasName> | Durum wheat flour |
| FP3 <comesFromPart> | Egg | FP16 <isa> | FP1 Vegetable product |
| FP4 <hasName> | Prepared food | FP16 <comesFromSource> | Durum Wheat |
| FP4 <isa> | FP0 Food product | FP16 <comesFromPart> | Seed, kernel |
| FP4 <processedBy> | * | FP16 <hasForm> | Ground |
| FP5 <hasName> | Soup | FP17 <hasName> | Noodles |
| FP5 <isa> | FP0 Prepared food | FP17 <isa> | FP4 Prepared food |
| FP5 <processedBy> | * | FP17 <hasIngredient> | FP16 Durum wheat flour |
| FP5 <hasForm> | Liquid OR Semiliquid | FP17 <hasIngredient> | FP15 Eggs |
| FP11 <hasName> | Diced carrots | FP17 <processedBy> | Mixing |
| FP11 <isa> | FP1 Vegetable product | FP17 <processedBy> | Extruding |
| FP11 <comesFromSource> | Carrot plant | FP17 <processedBy> | Drying |
| FP11 <comesFromPart> | Root | FP18 <hasName> | Flavoring (detail omitted) |
| FP11 <hasForm> | Diced | FP19 <hasName> | BHT (detail omitted) |
| FP12 <hasName> | Cut green beans | FP20 <hasName> | Chicken noodle soup |
| FP12 <isa> | FP1 Vegetable product | FP20 <isa> | FP5 Soup |
| FP12 <comesFromSource> | Bean plant | FP20 <hasIngredient> | FP13 Chicken broth |
| FP12 <comesFromPart> | Immature fruit | FP20 <hasIngredient> | FP14 Cubed cooked chicken |
| FP12 <hasForm> | Cut | FP20 <hasIngredient> | FP11 Diced carrots |
| FP13 <hasName> | Chicken broth | FP20 <hasIngredient> | FP12 Cut green beans |
| FP13 <isa> | FP2 Meat product | FP20 <hasIngredient> | FP17 Noodles |
| FP13 <comesFromSource> | Chicken | FP20 <hasIngredient> | FP18 Flavoring |
| FP13 <comesFromPart> | Meat | FP20 <hasIngredient> | FP19 BHT |
| FP13 <comesFromPart> | Bones | <w/ purpose> | Preservation |
| FP13 <isExtractedSubstance> | {Fat, Protein, Flavor} | FP20 <processedBy> | Boiling |
| FP13 <processedBy> | Cooking | <w/ intensity> | Fully cooked |
| FP13 <hasForm> | Liquid | <w/ purpose> | Make edible, |
| | | | Preservation |
| | | FP20 <hasForm> | Liquid with solid chunks |

| | | |
|---|--------------------------------|---|
| FP21 | has name | Diced parsley (statements not shown) |
| FP22 | <hasName> | Campbell's Chicken Noodle Soup |
| FP22 | <isa> | FP20 Chicken noodle soup |
| [FP22 | <hasIngredient> | FP13 Chicken broth] |
| [FP22 | <hasIngredient> | FP14 Cubed cooked chicken meat] |
| [FP22 | <hasIngredient> | FP11 Diced carrots] |
| [FP22 | <hasIngredient> | FP12 Cut green beans] |
| FP22 | <hasIngredient> | FP21 Diced parsley |
| [FP22 | <hasIngredient> | FP17 Noodles] |
| [FP22 | <hasIngredient> | FP 18 Flavoring] |
| [FP22 | <hasIngredient> | FP 19 BHT |
| | <withPurpose > | Preservation] |
| [FP22 | <processedBy> | Boiling |
| | <w/ intensity> | Fully cooked |
| | <w/ purpose> | Make edible, Preservation] |
| FP22 | <packedIn> | Steel can |
| [] inherited from one level above, [[]] inherited from two levels above | | |
| Portion-1 | <isaPortionOf> | FP22 Campbell's chicken noodle soup |
| [[Portion-1 | <hasIngredient> | FP13 Chicken broth]] |
| [[Portion-1 | <hasIngredient> | FP14 Cubed cooked chicken meat]] |
| [[Portion-1 | <hasIngredient> | FP11 Diced carrots]] |
| [[Portion-1 | <hasIngredient> | FP12 Cut green beans]] |
| [Portion-1 | <hasIngredient> | FP21 Diced parsley] |
| [[Portion-1 | <hasIngredient> | FP17 Noodles]] |
| [[Portion-1 | <hasIngredient> | FP18 Flavoring]] |
| [[Portion-1 | <hasIngredient> | FP19 BHT <i>purpose</i> Preservation]] |
| [[Portion-1 | <processedBy> | Sterilized by heat |
| | <WithPurpose > | {Make edible, Preservation}]] |
| [Portion-1 | <packedIn> | (Steel can, 10 fl oz)] |
| FP23 | <hasName> | Frozen cut green beans |
| FP23 | <isa> | FP12 Cut green beans |
| [FP23 | <comesFromSource> | Bean plant] |
| [FP23 | <comesFromPart> | Immature fruit] |
| [FP23 | <hasForm> | Cut] |
| FP23 | <processedBy> | Freezing |
| FP23 | <packedIn> | Carton |
| buy/sell (Safeway, Fred, Portion-1, \$1) | | |
| Fred | <eats> | Portion-1 |
| Chicken | <eats> | Hormone |
| Chicken | <eats> | FP24 Fish meal |



Only some links have been labeled.
The other labels should be obvious

Semantic network representation

Some sample frames (not all data represented in frames)

A minimal frame: Instance of a frame for buy/sell (a relationship with four arguments)

A minimal frame gives information for one relationship.

| buy/sell | |
|---|--|
| <u>SourceOfGoodOrService</u> / <u>ReceiverOfMoney</u> : | Safeway |
| <u>ReceiverOfGoodOrService</u> / <u>SourceOfMoney</u> : | Fred |
| <u>GoodOrService</u> : | Portion-1 (a particular can of Campbell's chicken noodle soup) |
| <u>MoneyAmount</u> : | \$1 |

All slots are essential; each value depends on all the others. The same information cannot be expressed in two separate frames.

Relationships can have two, three, or more pieces of information (called arguments) needed to make a complete statement. In the entity-relationship version, we wrote the same information as:

buy/sell (Safeway, Fred, Portion-1, \$1)

The frame is just a different way of writing this, with specification of the role each piece of information plays. None of the pieces of information can be separated out and given separately. That is why the frame is called minimal.

Linguists specify for each verb or group of verbs the slots that must be filled in order to make a complete statement with the verb; they call this specification a **case frame**. So the above is a case frame for the verb buy and for the verb sell. (Both verbs describe the same transaction, just from different perspectives.)

An extended frame: Instance of the food product frame for FP20

An extended frame packages information from several relationships. Many of the pieces of information in an extended frame could be stored separately.

Frame slots are defined through relationship types.

Corresponding slot (facet) codes and names from the paper on food description language in []

| An extended frame: Instance of the food product frame for FP20 | | Facet from the food classification in Reading |
|--|-------------------------------|---|
| Slot | Value | |
| FP20 <hasName>: | Chicken noodle soup | |
| <isa>: | FP4 Soup | <i>A Product type</i> |
| Slots dealing with food origin | | |
| <comesFromSource>: | N/A | <i>B1 Food source</i> |
| <comesFromPart>: | N/A | <i>B2 Part</i> |
| <isExtractedSubstance>: | N/A | |
| <isMadeFrom>: | N/A | |
| <hasIngredient>: | FP13 Chicken broth | <i>B3 Ingredient</i> |
| ... | | |
| <hasIngredient> | FP19 Preservative BHT | <i>D4 Method of preservation</i> |
| <withPurpose> Preservation | | |
| End food origin | | |
| <processedBy>: | Boiling | <i>D2 Cooking method</i> |
| <processedBy>, <withIntensity>: | Fully cooked | <i>D1 Degree of preparation</i> |
| <processedBy> | Sterilizing by heat (Boiling) | <i>D4 Method of preservation</i> |
| <withPurpose> Preservation: | Liquid with solid particles | <i>C Phys. state, shape, form</i> |
| <hasForm>: | | <i>E2 Container, wrapping</i> |
| <packedIn>: | | |

Another instance of the food product frame, FP22

(inherits most of its information from FP20; inherited slots are not repeated, saving lots of space)

| An extended frame: Instance of the food product frame for FP22 | |
|--|---------------------------------------|
| Slot | Value |
| FP22 <hasName>: | Campbell's Chicken Noodle Soup |
| <isa>: | FP20 Chicken noodle soup |
| <hasIngredient>: | FP21 Diced parsley |
| <packedIn>: | Steel can |

Think of this type of inheritance in the context of recipes in a cookbook.

2 Approaches to knowledge representation

Summary of concepts covered in examples

Entity-relationship approach

Semantic networks

Frames

Role of frames

Grid for data acquisition

Template for data output (for example, city data frame in Wikipedia)

Activation of all frame elements when one element is activated
(Seeing *parsley* may activate in a person's mind the whole frame for
Campbell's Chicken Noodle Soup)

Types of frames

Minimal frames (DS term)

A minimal frame represents an n-ary relationship – each slot corresponds to one argument of the relation. No slot could be omitted without making the frame incomplete, that is, making at least one other slot value indeterminate.

Extended frame (DS term)

An extended frame includes additional slots that represent further relationships, usually binary relationships from the focal entity to other entities.

3 Mechanisms in knowledge representation

Spreading activation

Hierarchical inheritance

Restrictions on values

Default values (for example, telephone area code in the database of a local charity)

Procedural attachments (procedures to be called when data are entered in the slot)

4 Criteria for describing and evaluating knowledge representations

(advanced) See supplement

Part 2.*February 1 - February 15***Information retrieval: General principles and methods****Lecture 3.1***February 1***The structure of information systems** (Organizing Information, Section 5.1)www.dsoergel.com/571/UBLIS571Lecture03.1.mp3

| | |
|-------------------------------|---|
| Objectives | <ol style="list-style-type: none"> 1 Know and understand the functional components of information systems and be able to use this framework <ul style="list-style-type: none"> • to analyze information systems; and • to integrate the subject matter from this and other courses. 2 Understand the wide variety of information systems . |
| Practical significance | <ul style="list-style-type: none"> • To design, operate, or use an information system or a specific function in it, you must understand the information system components, their inputs, output, and functioning. • To take advantage of all available career opportunities, you must understand the multitude of information systems and information environments in which the knowledge and skills acquired in LIS can be applied. • The information system diagram provides a framework for organizing information from many courses. |

Additional question: Determine the percentage of resources allocated to each of the following functions (Refer to Section 2.6 of the text).

- (1a) Assist the user in identifying relevant documents (intellectual access). (A user is given a list of references to documents relevant for her problem or topic. She must then consult these documents and extract the information needed.)
- (1b) Make available known documents (physical access). (A user requests specific documents, often documents found through 1a, and is loaned or given copies.)
- (2) Provide tailor-made packages of substantive data. (A user needs information on a certain topic and is given a report that contains just the information she needs, no more nor less. This report may be prepared by information center staff or a computer program by extracting information from documents or it may be the result of a search of a substantive database.) (See Section 2.5 of the text for an explanation of "substantive data".)

What kind of data do you need to answer this question?

In-class exercises (in preparation for Assignment 5, where you analyze an information system)

Refer to the figure on the following page (which conveniently integrates Text Figures 5.1c, Text p. 47 and 5.6, Text p. 58). Using this figure, analyze the following:

- 1 A **Web search system** (like Google or Yahoo) as an information system.
- 2 A **special library** as an information system.
Question: How does interlibrary loan fit into the information system framework?
- 3 The **production and use of a textbook** in the information system framework.

To analyze Google as an information system using the structure on p. 58, look at each box and determine what Google does using what files (data) and rules

Same for Yahoo Directory, <http://dir.yahoo.com/>

Short description: <http://help.yahoo.com/l/us/yahoo/directory/basics/basics-03.html>

Post your results to the discussion board for Lecture 3.2. This analysis is a class project.

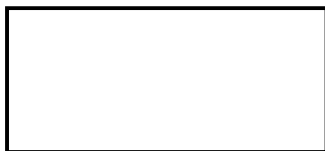
The recording will take you through some of this analysis

Additional question: put on p. 56 to save space.

Note on diagram conventions



Denotes a file, data, inputs and outputs.
Could be a group of people and their problems as an input



Denotes a process that transforms inputs into outputs



Sequence of processes and files, flow of data
Control of processes or file organization

Combined information systems diagram here

Lecture 3.2*February 1***Objectives and performance measures for information systems**

(Organizing. Information, Ch. 8)

www.dsoergel.com/571/UBLIS571Lecture03.2.mp3 (Subtract 4 from Lecture Notes page nos)

| | |
|-------------------------------|---|
| Objectives | <ol style="list-style-type: none"> 1 Understand the purpose and objectives of IR systems so that you can examine the functioning of the individual IR system components in light of these objectives. 2 Understand both the importance and the difficulty of defining suitable measures of information system performance and of applying such measures to actual systems. |
| Practical significance | <p>A clear understanding of objectives and evaluation criteria for both individual searches and for an information system as a whole is important for the following:</p> <ol style="list-style-type: none"> 1 Conducting individual searches. <ul style="list-style-type: none"> • Determining user requirements. • Selecting an information system (database and search system) that can be expected to meet these requirements. • Evaluating search results and determining when to stop searching. • Determining the amount of resources that should be allocated to a search. 2 Selecting information systems to be acquired, including reference tools and online databases (see Lecture 1.1 for a list of types). (Acquisition includes leasing or contracting for use, as well as training in the use of the system.) 3 Designing information systems or communicating requirements to systems analysts. <p>See Text Chapter 8, Introduction, and Section 8.5 for elaboration.</p> |

Discussion questions: see next page

Discussion questions

- 1 Consider the definition of relevance and of performance measures in general in the context of an information system with data on the structure of a nuclear power plant to be used in case of malfunctions. The system gives detailed information about all components down to the last pipe and valve, their functions and interrelationships.
- 2 Consider performance measures for the following information system. The purpose of the information system is to assist in solving crimes. The system stores reports of past crimes — both solved and unsolved — and indexes them by various features of the modus operandi. To use the system, the detective formulates a query based on the features of the unsolved crime. The system provides reports of similar crimes; these might assist in solving the crime in question.
- 3 Consider a Web search service (such as AltaVista or Lycos) that produces ranked retrieval output. Picture two users. User 1 needs a quick answer to a question, and user 2 needs a comprehensive list of materials (for example, a listing of all classifications schemes and thesauri available on the Web). What performance measures would be appropriate for each type of user?

Organizing Information, Chapter 8 review, especially

The need for performance measures for guiding system design, Figure 8.1, p. 111

Deriving performance measures, Figures 8.2 and 8.3.

Figure 8.3 updated on Lecture Notes p. 62 (next page)

The concept of relevance, Text Section 8.4

Relates to → LIS 518 Reference Sources and Services, important for how to search

Have the text open to p. 111

Have the lecture notes open to p. 62

Listen to recording

www.dsoergel.com/571/UBLIS571Lecture03.2.mp3 (Subtract 4 from Lecture Notes page nos)

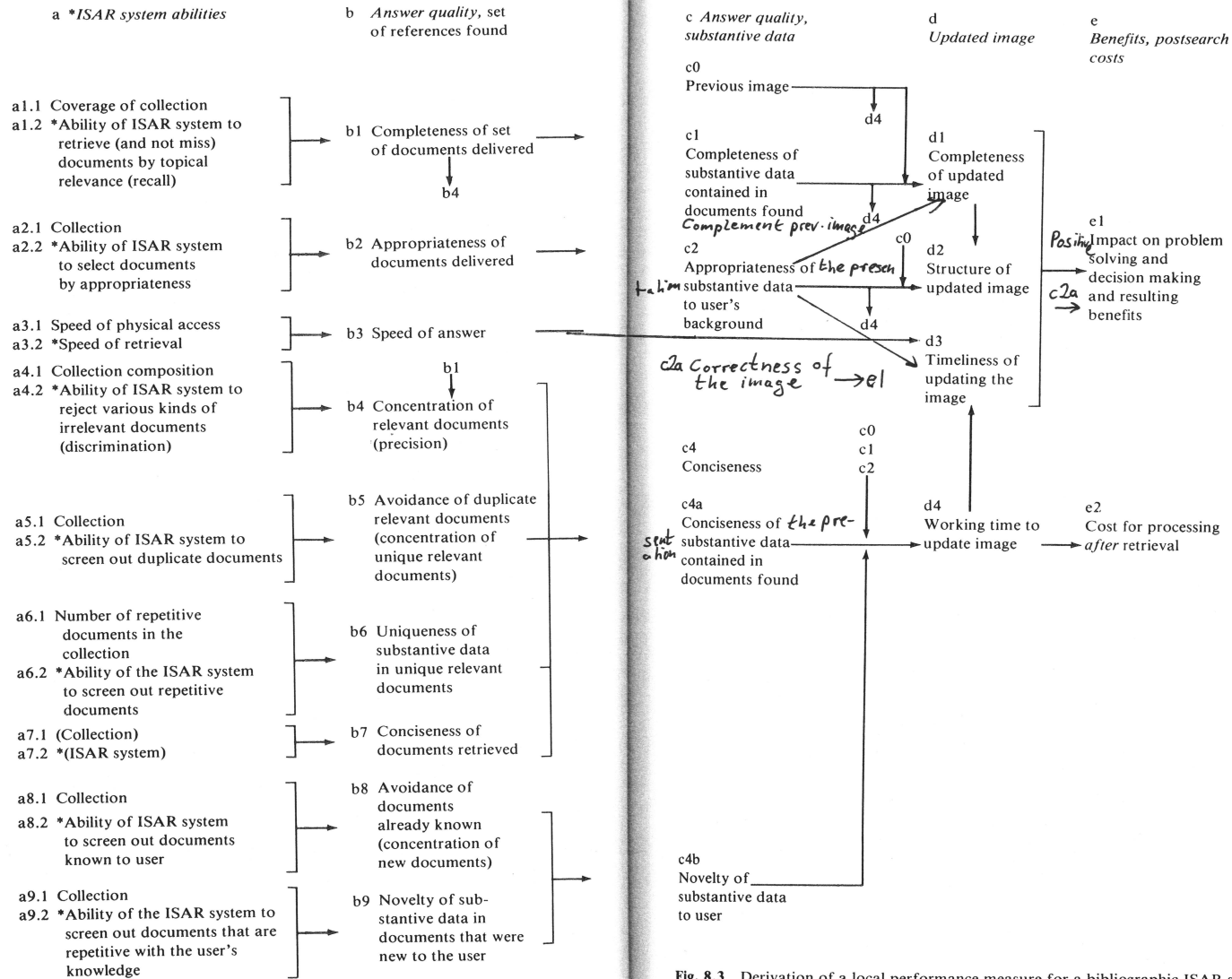


Fig. 8.3 Derivation of a local performance measure for a bibliographic ISAR system.

Relevance criteria for different purposes

Relevance criteria of teachers selecting oral history materials

Relevant to teaching content and method

Relationship to theme

As part of broader curriculum

Relates to other schoolwork

Variety for the classroom

Vocabulary

Characteristics of oral history interviews

Flow of interview

Expressive power

Language & verbal expression

Diction

Nonverbal communication

Characteristics of the story

Positive message for students

Role of interviewee in Holocaust events

Relationship of story to student

Students connect with passage

Students identify with interviewee

Radical difference from students' reality

Represents different populations

Race

Age of interviewee during Holocaust events

Appropriateness

Developmental appropriateness

Acceptability to stakeholders

Technical production quality

Length-to-contribution ratio

Topical relevance for scholars

| Types of topical relevance | |
|---|--|
| Topic: Food in Auschwitz | |
| Relevance type | Example |
| “Classical relevance” (TREC definition) | |
| • Provides direct evidence | Describes types of food and portions served |
| • Provides indirect/circumstantial evidence | Describes undernourished people |
| Additional relevance types | |
| • Provides context | (1) Reports on intensity of manual labor (2) Availability of food in the area around the camp |
| • Useful as a basis for comparison | Food situation in a different camp |
| Pointer relevance | |
| • Provides pointer to a source of information (The information pointed to can be relevant in any of the ways listed above) | Mention of a study on the subject |

TREC = Text REtrieval Conference

A yearly competition of information retrieval systems performing specified retrieval tasks on a given test collection held at NIST (National Institute of Standards and Technology)

CLEF = Cross-Language Evaluation Forum

The corresponding activity in Europe. Uses the MALACH speech retrieval test collection for one task.

Lecture 4.1*February 8***An integrated information structure model**

www.dsoergel.com/571/UBLIS571Lecture04.1.mp3

| | |
|-------------------------------|--|
| Objectives | <ol style="list-style-type: none"> 1 Understand a general model of information retrieval; 2 Be able to analyze specific systems and information retrieval operations in terms of this general model; 3 Integrate knowledge across types of information systems and developing an overall vision of retrieval possibilities. |
| Practical significance | <p>This knowledge will enable you</p> <ul style="list-style-type: none"> • to use existing systems in new and imaginative ways, in particular, to use several different systems in synergistic ways; • to design new systems with increased power, for example to search for Linked Open Data on the Web <p>Note: Pay attention to the first bullet; it will make you a better searcher. You may never be able to use a unified integrated system of the kind described. But you can use existing systems in combination to achieve improved search results enabled by the way of thinking presented here. In other words, you can build your own “virtual” integrated information structure whenever a search requires it.</p> |

Key idea:**Combining different kinds of facts to find an answer. (Inference. Chaining)**

Done by people - reference librarian or user consults different databases as needed to find all the facts needed to construct an answer.

Done by systems - all facts must be accessible to the system

Relates to and elaborates on **Lecture 1.2 Information systems and information structure**

Design of an integrated information structure interface

Part 1. Basic structure and search commands

- 0 **Prolog: Finding answers. The nature of search**
- 1 **Introduction. Scope, purpose, and organization of the paper**
 - 1.1 General introduction
 - 1.2 Organization of the paper
 - 1.3 Introductory example: Concepts, projects, texts, organizations, persons
- 2 **A unified view of systems or The multidimensional design space for information systems**
- 3 **Elements of information structure**
 - 3.1 Objects
 - 3.2 Relationships (links) and connections
 - 3.3 Neighborhoods and queries
 - 3.3.1 "Offspring neighborhood". Example: Modeling documents as a tree of smaller and smaller units
 - 3.4 Links to, from, and between neighborhoods
- 4 **Search**
 - 4.1 Definition of search
 - 4.2 Specification of a search based on relationships
 - 4.3 Single criterion search starting from a single object
 - 4.3.1 Single-criterion search starting from a single object with single object as targets
 - 4.3.2 Single-criterion search starting from a single object with neighborhoods as targets
 - 4.4 Single-criterion search starting from a neighborhood
 - with single object as target
 - with neighborhood as target
 - 4.5 Combination search (Boolean AND or weighted search) with single objects as targets
 - 4.5.1 Combination search with single objects as targets
 - 4.5.2 Combination search with neighborhoods as targets
 - 4.6 Offspring neighborhoods and ancestor neighborhoods in searching
 - 4.6.1 Offspring neighborhoods and searching. Review
 - 4.6.2 Ancestor neighborhoods and searching. Hierarchical inheritance
 - 4.6.3 Indexing with hierarchical inheritance
- 5 **Indexing**

This recorded lecture will present the material from the reading

**A general model for searching linked data
OR Design of an integrated information structure interface**

as if it was a presentation at a conference. Please read the Prologue and p. 1 - 17 beforehand; these sections give examples. The lecture does not assume that you have read the remainder of this reading; rather, the reading is a back-up reference.

Restatement of the objectives: Through this lecture you should

- Get a better understanding of entity-relationship data modeling
- Improve your skills as a reference librarian through better understanding of chained searching, often using multiple data sources
- Acquire a general framework for understanding retrieval system features such as inclusive (hierarchically expanded) searching.
- Get a forward-looking sense what retrieval systems could do in the new world of Linked Open Data (LOD)

Read parts of the assigned reading:

A general model of linked data Or Design of an integrated information structure interface
Prologue (p. VII - X) and p. 1 - 17

Recommended exercise before listening to the lecture

As an example for the dimensions given in Figure 7, p. 17 (and also examples of the kinds on the searches discussed in the reading), you could do search in the UB Libraries catalog:

Type *Library instruction* in the search box

Starting point: a search key, entered by the user

Scroll down to

Bibliographic instruction : a handbook /
by Renford, Beverly. Published 1980

Click on the title to see the full record. Many of the elements of the records can be used as starting points for a new search. Rather than having to enter a person's name to find documents he or she authored, you can just click on the name displayed on the screen to start an author search. The result is no different from doing the same search by entering the person's name in the search box as author (Dimension 2). Try it.

You can start a search for the subject *Library orientation > Handbooks, manuals, etc.*
By clicking on the second element, *Handbooks, manuals, etc*

See what happens if you click on *Library orientation*

Note: When you click on *Library orientation*, the UB catalog does a search

Subject Library orientation <*isSubjectHeadingFor*> Document X

(to be precise: the system finds all documents for which the words library and orientation occur among all assigned subject headings)

That is all a user can do. A user may want to find subjects that are in some way related to Library orientation, which would be the search

Subject Library orientation <*hasST OR hasBT OR hasNT OR hasRT*> Subject X

ST *hasSynonymousTerm*

BT *hasBroaderTerm*

NT *hasNarrowerTerm*

RT *hasRelatedTerm*

There is no provision in the UB catalog to do that (but you will later learn how to use Library of Congress authority databases to do that kind of search). The UB catalog could, upon the user clicking on Library orientation, ask the user

Do you want to find

Books that have the words library and orientation in their subject headings

Books that have the words library and orientation in their subject headings or title

Subject headings related to Library orientation

Courses on Library orientation

Now you can acquire an understanding of the general system proposed by listening or reading or both.

- The recording at www.dsoergel.com/571/UBLIS571Lecture04.1.mp3 takes you through p. 18 - 50. You do not need to read the pages beforehand but you need to have the reading in front of you as the recording refers to the figures.
- You can also just read p. 18 - 50

Lecture 4.2

February 8

Conceptual data schemas and input, storage, and output/presentation formats

(Organizing Information, Sections 9.1, 9.2, 9.4, and 9.5)

www.dsoergel.com/571/UBLIS571Lecture04.2ConceptualDataSchemaExerciseReview.docx

www.dsoergel.com/571/UBLIS571Lecture04.2ConceptualDataSchemaExerciseReview.mp3

<http://dsoergel.com/571/UBLIS571Lecture04.2.mp3> (for Lecture Note pages 72-76)

| | |
|-------------------------------|--|
| Objectives | <p>1 Be able to analyze or design the conceptual data schema of an information system</p> <ul style="list-style-type: none"> • analyze the conceptual data schema underlying an information system; • judge the adequacy of this schema with respect to the queries to be answered; • use the knowledge of the schema to exploit fully the possibilities of obtaining answers from the information system; • design a conceptual data schema for an information system based on user requirements. <p>2 Be able to analyze and design the input formats and output formats used to interact with an information system:</p> <ul style="list-style-type: none"> • input formats that make data entry complete, error-free, and easy • output formats (for reports, such as recurring bibliographies, or the display of search results) that contain all the information needed (and no more) in an easy-to-read form. |
| Practical significance | <p>1 For designing information systems: The success of any information system depends vitally on the completeness of the information included. The conceptual data schema determines what information can be included in the system and what information is elicited from the people that enter data into the system. Input and output formats determine how easy it is to interact with the system.</p> <p>2 For using information systems (including reference tools): To get the most out of an information system in terms of being able to do different types of searches, you need to know its conceptual data schema. To select the appropriate information system, you need to be familiar with the conceptual data schemas of many information systems. To do the kind of power search that draws on multiple information systems simultaneously requires even more knowledge of conceptual data schemas.</p> |

Schema Arrangement of parts in some order, showing interrelationships.

This topic is closely related to document structure design, to be discussed in Lectures 5.2-6.1.

Over

Online class exercise (see schedule on facing page)**Developing the conceptual data schema for the information system of a large computer users' group (such as the Washington Apple Pi, www.wap.org)**

A computer users' group has the purpose of helping members to better use their computers.

Some functions of a computer users group

- a library for members to use
- a newsletter with articles and product reviews
- special interest groups (hold meetings, have a chair)
- a group purchase program
- a list of experts on specific subjects that have agreed to be on call to answer member questions

Sample questions with entity types and relationship types**Who knows about printers?**

Entity types: Person, Subject

Relationship type: Person <*knowsAbout*> Subject

I am looking for a review of Microsoft Word 12

Entity types: Document, SoftwareMakeAndModel

Relationship type: Document <*reviews*> SoftwareMakeAndModel

What is a good word processor for Red Hat Linux

Entity types: New: Function, Quality; already noted: SoftwareMakeAndModel

Relationship types: SoftwareMakeAndModel <*servesFunction*> Function
 SoftwareMakeAndModel <*worksWith*> SoftwareMakeAndModel
 SoftwareMakeAndModel <*hasQuality*> Quality

Your sample questions / reports from the database

In-class exercise. Developing the conceptual data schema. Schedule

| | |
|------------------------------|---|
| W June 6 midnight | Contribute some questions and the resulting entity types and relationship types to the Wiki for week 4 on Ublearns. (See sample questions below) First enter a question and with it the entity types and relationship types needed. Then add new entity types and new relationship types under the appropriate heading. Please do not alter others' contributions, just add to the existing Wiki page or comment on what is there. The object is to contribute to the conceptual data schema |
| Th June 7 6 pm | Instructor will put all contributions together into a sample conceptual data schema |
| | Go through the review files: While looking at the .docx (not for printing) listen to the .mp3 www.dsoergel.com/571/UBLIS571Lecture04.2ConceptualDataSchemaExerciseReview.docx www.dsoergel.com/571/UBLIS571Lecture04.2ConceptualDataSchemaExerciseReview.mp3 |
| | You can now start working on Assignment 6 |

| Entity types | Relationship types |
|---------------------|---------------------------|
| | |

Uses of the different types of information in an information system

A type of information (a fact type, see Lecture 1.2), as defined by a relationship type, may be used for one or more of the following functions.

- **Retrieval, drawing inferences, statistical analysis**

Example: From drug prescription expert system

Disease <*treated with*> Drug

This piece of data is used for

- plain retrieval of medical knowledge;
- inference in conjunction with patient data.

- **Arranging retrieval output**

- Example: Arranging a long list of document records retrieved from an OPAC (Online Public Access Catalog) by call number

- Example: Arranging output from a Web search by URL (Uniform Resource Locator), which would bring pieces of one Web document that consists of several pages together in the output list

- **Providing information to the user**, either the substantive data sought or information about a document that enables the user to judge the relevance of the document.

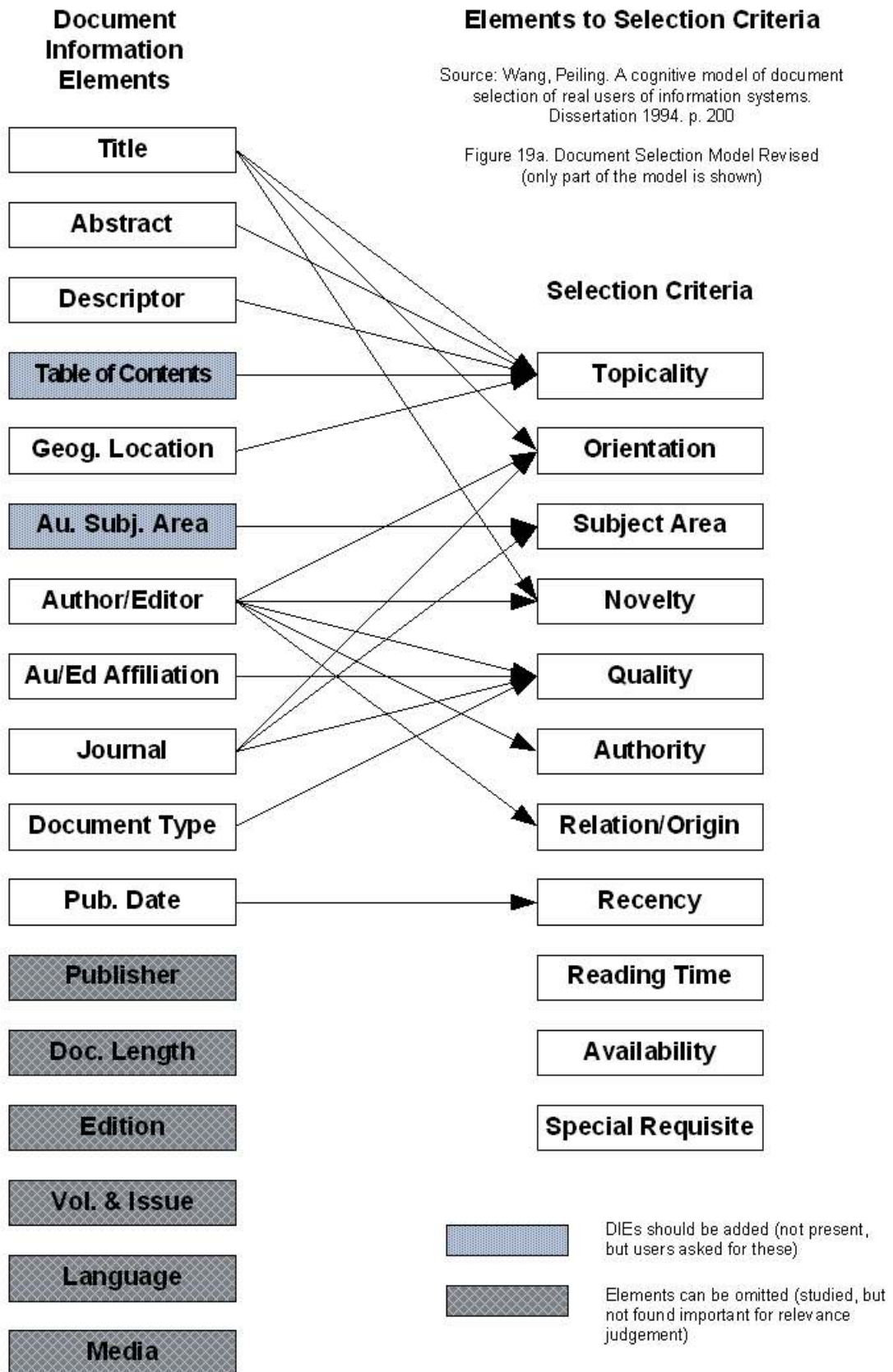
The conceptual data schema designer must weigh the cost for including a type of information against the benefit in terms of these three functions.

As an example, look at Wang's list of "Document Information Elements" users considered in document selection or wished to have available (next page). These results should be used as a guideline in systems design when deciding what information to include in the system and what information to present to users in the output format. Users did consider information elements that are linked to the document indirectly, such as the subject area of the document author. As will be further elaborated below, the information system must assemble all this information about a document, possibly obtaining information from other databases, such as a database about persons.

From Document Information Elements to Selection Criteria

Source: Wang, Peiling. A cognitive model of document selection of real users of information systems. Dissertation 1994. p. 200

Figure 19a. Document Selection Model Revised (only part of the model is shown)



From conceptual data schema to records (Organizing Information, Section 9.2)

| | |
|--|---|
| Record format | <ul style="list-style-type: none"> • A record is an assembly of information about a given entity, such as an event, a person, or a document for input, storage, communication, or display. • The record format determines how the different pieces of information are arranged in the record. • A record is a simple frame. Slot in a frame = data field in a record. • The evaluation criteria for schemes of knowledge representation (Lecture 2.2, Section 3.4) apply to records; see also Organizing Information, Section 9.3. • Many records are extended frames: they incorporate many binary statements that link the focal entity of the record to another entity. Each statement could stand on its own. The presentation of these data in a record is more concise and may be more intuitive and more easily grasped than a series of statements. See the MARC record format (facing page) and the examples in text Chapter 9. |
| Input record. Input format | <p>The best way for eliciting input from the system operator (for example, a cataloger) is often an input record, a form with blanks (slots) to fill in.</p> |
| Storage record. Storage format | <p>Some systems store data internally in tables (relational database), where information about a given entity may be distributed over several tables. Other systems store data internally in records or frames, assembling all information about one entity, e.g., a book, in one place (but making it more difficult to assemble information about entities of other types, e.g., persons).</p> |
| Communication record | <p>A record in a common communication format to transfer data from one system to another. Each system may use its own internal format. Examples: MARC, Z39.50</p> |
| Output/display record. Display format | <p>To present information about an entity in a format easily understood by the user, the information system must assemble the desired types of information into an output record. The information elements in the output record may be linked to the entity directly (for example, the direct link between a document and the person who authored it) or indirectly (for example, the indirect link between a document and the organization with which the author is affiliated).</p> |

| | |
|--------------------|--|
| MARC format | <p>Opposite is a sample record format, the MARC (MACHine-Readable Catalog) format. (MARC was developed by the Library of Congress starting in 1962 for the interchange of bibliographic data and has become a widely used standard). For each data field, the corresponding statement template (relationship type plus entity types related) is given. A more complete list of MARC fields is found with the description of the model catalog in the general readings.</p> |
|--------------------|--|

From relationship types to data fields in a MARC record for documents. (simplified)

| Relationship type | Corresponding MARC field |
|---|---|
| Document <authoredBy> Person (who is main entry) | 100 Main Entry-Personal Name |
| Document <emanatedFrom> Organization | 110 Main Entry-Corporate Name |
| Document <emanatedFrom> Meeting | 111 Main Entry-Meeting Name |
| Document <hasTitle> Text (if Title is main entry) | 130 Main Entry-Uniform Title |
| Document <hasTitle> Text | 245 Title Statement |
| Document <publishedBy> Organization this Organization <locatedIn> Place (chain) Document <publishedIn> Date, ... (distinguished by using subfields) | 260 Publication, Distribution, etc. (Imprint) |
| Document <partOf> Document (which is a Series) | 490 Series Statement |
| Document <dealsWith> Person | 600 Subject Added Entry-Personal Name |
| Document <dealsWith> Organization | 610 Subject Added Entry-Corporate Name |
| Document <dealsWith> Meeting | 611 Subject Added Entry-Meeting Name |
| Document <dealsWith> Document | 630 Subject Added Entry-Uniform Title |
| Document <dealsWith> Topic | 650 Subject Added Entry-Topical Term |
| Document <dealsWith> Place | 651 Subject Added Entry-Geographic Name |
| Document <authoredBy> Person | 700 Added Entry--Personal Name |
| Document <emanatedFrom> Organization | 710 Added Entry-Corporate Name |
| Document <emanatedFrom> Meeting | 711 Added Entry-Meeting Name |
| Document <hasTitle> Text | 730 Added Entry-Uniform Title |
| Document <heldBy> Organization | 850 Holding Institution |

The MARC record is an extended frame with information about a document; it incorporates many statements that link the document to some other entity (a Person, a Date, a Subject, etc.). Each statement could stand on its own; the data in the MARC record could be represented as a series of statements. That is how bibliographic data are represented in the Linked Open Data (LOD) initiative on the Web.

www.oclc.org/multimedia/2011/files/globalcouncil/Buzash_Calhoun_Dunsire_Linked_Data.pdf
http://en.wikipedia.org/wiki/Linked_data XXX URL for LOD bibliographic data set

| MARC Record (simplified) | E-R statements (in this case triples, linked data on the Web) |
|--|---|
| 010 2009038344 | 0838985130 <hasLCCControlNo> 2009038344 |
| 040 DLC \$c DLC \$d YDX \$d MTG \$d CDX \$d BWX | |
| 020 9780838985137 (pbk. : alk. paper) [ISBN 13 digits] | 0838985130 <hasISBN13> 9780838985137 |
| 020 0838985130 (pbk. : alk. paper [ISBN 10digits] | |
| 050 00 ZA3075 \$b .B83 2009 | 0838985130 <dealsWith> LCC#ZA3075, 0838985130 <hasLCCallNo> ZA3075 \$b .B83 2009 |
| 050 4 Z674 \$b .A75 no.61 | 0838985130 <dealsWith> LCC#Z674, 0838985130 <hasLCCallNo> Z674 \$b .A75 no.61 |
| 082 00 028.7071/1 \$2 22 | 0838985130 <dealsWith> DDC#028.7071/1 |
| 092 \$b | |
| 049 SBSM | 0838985130 <> |
| 100 1 Budd, John, \$d 1953- | 0838985130 <authoredBy> Budd, John, \$d 1953- Budd, John <bornIn> 1953 |
| 245 10 Framing library instruction / \$c by John Budd. | 0838985130 <hasTitle> Framing library instruction |
| 260 Chicago : \$b Association of College and Research Libraries, \$c 2009. | 0838985130 <publishedBy> Association of College and Research Libraries Association of College and Research Libraries <locatedIn> Chicago 0838985130 <publishedIn> 2009 |
| 300 v, 197 p. ; \$c 23 cm. | 0838985130 <hasNoOfPages> (v, 197), 0838985130 <hasHeight> 23 cm |
| 490 1 ACRL publications in librarianship ; \$v no. 61 | 0838985130 <isPartOf> (ISSN-0193-1784, 61), ISSN-0193-1784 <hasTitle> ACRL publications in librarianship |
| 504 Includes bibliographical references (p. 186-194) and index. | 0838985130 <includesTypeOfMaterial> BibliographicalReferences 0838985130 <includesTypeOfMaterial> Index |
| 505 0 A framework for instruction -- Beyond information literacy -- In the classroom -- Cognition and clear thinking -- A vision for learning -- Putting it all together. | 0838985130 <hasTOC> A framework for instruction -- Beyond information literacy -- In the classroom -- Cognition and clear thinking -- A vision for learning -- Putting it all together. |
| 650 0 Information literacy \$x Study and teaching (Higher) | 0838985130 <dealsWith> LCSH#Information literacy \$x Study and teaching (Higher) |
| 650 0 Information resources \$x Evaluation \$x Study and teaching (Higher) | 0838985130 <dealsWith> LCSH#Information resources \$x Evaluation \$x Study and teaching (Higher) |
| 650 0 Critical thinking \$x Study and teaching (Higher) | 0838985130 <dealsWith> LCSH#Critical thinking \$x Study and teaching (Higher) |
| 650 0 Library orientation for college students. | 0838985130 <dealsWith> LCSH#Library orientation for college students. |
| 650 0 Academic libraries \$x Relations with faculty and curriculum | 0838985130 <dealsWith> LCSH#Academic libraries \$x Relations with faculty and curriculum |

Lecture 5.1*February 15***Access to information: data structure and search modes**

(Organizing Information, Chapters 10 and 11)

→LIS 518 Reference Sources and Services

www.dsoergel.com/571/UBLIS571Lecture05.1RankingSpreadsheet.xlsx

| | |
|---|---|
| Objectives | <ol style="list-style-type: none"> 1 Understand the purpose of a data structure as answering questions with retrieval as a special case. 2 Understand the basic principle of searching: use all available evidence to predict the degree of relevance of some entity Ranked retrieval and plain Boolean retrieval as special cases. 3 Be able to formulate simple Boolean queries. 4 Be able to analyze the storage structures (tables, record formats) of an information system and design simple storage structures. 5 Be able to analyze data access structures (indexes) in an information system and use the understanding gained for efficient searching. 6 Be able to design simple data structures for access. 7 Be able to apply the principle of hierarchical inheritance to achieving more compact storage. |
| Practical significance, examples | <ul style="list-style-type: none"> • When searching for X, use a reference tool where X can be found in the index. • Store data with minimal redundancy by using hierarchical inheritance. (OCLC and other bibliographic databases are tremendously redundant since they do not use a data structure that exploits hierarchical inheritance.) |

Over

Outline

- 1 Retrieval as prediction.
- 2 Review of Boolean retrieval
- 3 Ranking documents by expected relevance
- 4 Search modes and data structures
 - 4.1 Review of Organizing Information, Chapter 11
 - 4.2 Further elaboration of data structures (Supplement)

1 Retrieval as prediction. *Just read this section*

Query formulation: Find good clues that predict the relevance of an entity

(document, person, computer program, etc.) for a given user with a given purpose

Retrieving a document or other entity is predicting that it is relevant to the problem to be solved at least to a certain degree. The challenge in formulating a query is to find the clues that predict that a document or other entity is relevant:

For documents: What clues can predict that the document will be helpful in solving the problem at hand?

For persons to What clues can predict that a person will do well in the job?
fill a job:

Finding the right clues requires knowledge and may involve some guesswork:

- When searching for documents using free-text retrieval, the searcher must determine what words and expressions the author of a relevant document may have used in the title, the abstract, and the full text; this requires knowledge of how language is being used both in general and by specific scientific schools and even individual authors.
- When using descriptors assigned by an indexer, the searcher must determine what descriptors an indexer would have assigned to relevant documents; this requires knowledge of the index language, the indexing instructions, and the actual indexing practices. (Request-oriented indexing, to be discussed later, seeks to increase the probability that descriptors corresponding to user needs are included in the indexing language and assigned correctly in indexing.)

Of course, free-text terms and assigned subject descriptors are only two kinds of evidence. Many other clues can be considered, such as

- publication date,
- topical focus of the journal in which an article appears,
- reputation of the journal or publisher,
- reputation of the publisher,
- reputation of the author or the author's institution,
- etc.

Again. Think about a user's question where clues other than words expressing the search concept might be useful. To be a good reference librarian, you need to be imaginative in thinking of different clues that could lead to relevant items.

Example:

You want to find documents that list several programs for managing your personal bibliographic database. There are several terms for this concept, such as *bibliographic software*, *bibliography manager*, *reference manager*, *text database management system*. Rather than trying to come up with all possible terms to search, think of two or three programs you know, such as *Endnote* and *Zotero*, and put them in the Google search box. Documents that mention both of these might well mention others as well, and some of these documents will be comparative reviews.

2 Brief review of Boolean retrieval (Organizing Information, Chapter 10)

It is assumed that students know Boolean operators, including the use of parentheses, as in the Google search

ranking AND methods AND (review OR "state of the art" OR overview)
(AND is not needed since it is the default; added for clarity)

In textbook Chapter 10 pay special attention to

Section 10.4 (know when *and* means AND and when *and* means OR

Section 10.6 (in the Figure on text book p. 170, be clear about the relevant documents missed by the AND NOT query, why they are relevant and why they are missed
NOT is tricky, use judiciously.

3 Ranking documents by expected relevance (as in Google)

Boolean retrieval: YES or NO – division of all documents in the system into **two classes**

A document either scores 1 and is retrieved or it scores 0 and is rejected:

class 1: retrieved - expected to be relevant

class 2: rejected - expected not to be relevant

Using three queries to get YES, MAYBE, NO – division of all documents into **three classes**

This method is often useful in practice

Problems with formulation of queries, especially if interactive retrieval not possible. Consider the following query formulation consisting of four descriptors

A Simulation AND B Traffic flow AND C Passengers AND D Rail transport

Perhaps documents that contain any three of the four descriptors might be of some interest; could run the broader query (ABC OR ABD OR ACD OR BCD) in parallel:

Class 1: retrieved in narrow search ABCD – expected to be clearly relevant

Class 2: retrieved in broad search – expected to be somewhat relevant

Class 3: not retrieved even in the broad search expected to be not relevant

Running a focused narrow query and a broader query is often a good idea to give the user a result list that has first the most important items and then more items to explore S a very coarse form of ranking

More refined ranking by expected relevance – continuous scale

Compute for each document a **quantitative measure of expected relevance** to the given search request. Instead of having 3 classes of documents, we then get a **list of documents ranked according to expected relevance**. (In many systems the ranking is poor and does not approximate the user's intuition.) Measure is computed as the nearness or **similarity** between query formulation and document representation, based on the number of descriptors they have in common. Different formulas are possible.

Very simple formula: percentage of query descriptors found in document record. A document would get a score as follows

| | |
|----------------------------|------|
| all four query descriptors | 1.00 |
| three query descriptors | .75 |
| two query descriptors | .50 |
| one query descriptors | .25 |
| no query descriptors | .00 |

For each formula:

- (a) crude form (only the exact term matches, uses no knowledge of concept term relationships) and
- (b) knowledge-based form (uses knowledge of concept and term relationships, for example to match a query term with a synonym or narrower term in the document).

Problems of OR descriptor combinations, as in the following query formulation:

Assume two documents

Document 1 has terms *Traffic congestion, Terminals, Air traffic, and Boston*

Document 2 has terms *Terminals, New York, Boston, and Washington*

The Google query (AND is implied)

Q1 "Traffic congestion" Terminals "Air Traffic" "New York" Boston Washington

treats all terms the same (six components) and would give the same score (simplified: 4/6 or .66) to both documents, even though document 1 is clearly more relevant

Q2 "Traffic congestion" Terminals "Air Traffic" ("New York" OR Boston OR Washington)

considers the conceptual structure and expresses four conceptual components: the three subjects and the place component expressed by ("New York" OR Boston OR Washington). It would give the following scores:

Document 1 gets 1.0 (it matches all four conceptual components of the query)

Document 2 gets .5 (it matches only two conceptual components, *Terminal* and place)

Using OR in Google can give much better results

Start on p.83 **Query formulation Q and document representations D1 - D4**

| | | | | | | | |
|----|------------------------------------|---|--|--------------------------------|---------------|----------------------|----------------------------------|
| Q | B1.2 Rail transport | AND E1.2 Traffic stations | AND J1 Passenger transp. | AND U15 US | | | |
| D1 | B1.2 Rail transport | B3 Air t. | E1.2 Traffic stations | J1 Passenger transp. | U15 US | U20 Europe | Q24 Traffic simulation |
| D2 | B1.2.1 Local rail transport | | E1 Traffic facilities | J1 Passenger transp. | U15 US | | |
| D3 | B1 Ground transp | | E2.1 Vehicles | J1 Passenger transp. | U15 US | | |
| D4 | B1.2 Rail transport | | E1 traffic stations | J1 Passenger transp. | U15 US | | |

The concept numbers or notations (like Dewey numbers) B1, B1.2, B1.2.1, J1, etc. come from an imaginary classification; they do express a hierarchy

Formulas for computing expected relevance

Base formulas **1** and **2**, descriptor matching rules **a** and **b**, gives four formulas: **1a**, **1b**, **2a**, **2b**.

Base formula 1: $R = \# \text{ of matching descriptors } M / \# \text{ of query descriptors } Q$

Base formula 2: $R = \# \text{ of matching descriptors } M / (\# \text{ of query descriptors } Q + \# \text{ of doc. descr. } D)$

Matching rule a: **Exact descriptor match:** A query descriptor produces a match only if the (crude) document representation contains exactly the same descriptor

Matching rule b: **Hierarchically expanded match:** A query descriptor produces matches as (knowledge-based) shown in the following examples:

| Query descriptor | Document descriptor | | Match value |
|----------------------------|------------------------------------|----------|-------------|
| B1.2 Rail transport | B1.2 Rail transport | Same | 1 |
| | B1.2.1 Local rail transport | Narrower | 1 |
| | B1 Ground transport | Broader | 0.5 |

Note: The numbers in the column "Match value" are set arbitrarily for this exercise. One might count a narrower descriptor as 0.75 of a match, for example.

Online: Read Purpose and Task, look carefully at p.90,

In-class exercise: **Ranking of retrieved documents**

| | |
|----------------|--|
| Purpose | <ol style="list-style-type: none"> To give you a better "feel" of how formulas for the computation of expected relevance work and what a rank list of documents looks like. To have you compare the effectiveness of four formulas (two base formulas, each applied with two matching rules). |
| Task | <p>Given are a query formulation and four document representations (descriptors assigned to the documents) and four formulas for the computation of expected relevance. The formulas are deliberately very simple; many more complex formulas are being used.</p> <ol style="list-style-type: none"> Using your own judgement, rank the documents 1-4 by their relevance to the query. Compute for each document the coefficient of expected relevance according to four different formulas and list the documents in rank order by decreasing expected relevance. Compare the four rankings with your intuitive ranking. State which is better. Briefly state why one formula works better than the other. |

Results

Expected relevance score for the query Q

| Docu ment | Formula | | | |
|--------------|---------|----|----|----|
| | 1a | 1b | 2a | 2b |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

Ranking

| | Intui tive | Formula | | | |
|---|---------------|---------|----|----|----|
| | | 1a | 1b | 2a | 2b |
| 1 | | | | | |
| 2 | | | | | |
| 3 | | | | | |
| 4 | | | | | |

Online

Recap: Matching rules and scoring formula

| How "in common" is determined | How the score is computed: Base formula | | | |
|--|---|-----------|----|-----------------|
| | 1 | M / Q | 2 | $M / (Q + D)$ |
| Ma Matching rule a: exact | 1a | M_a / Q | 2a | $M_a / (Q + D)$ |
| Mb Matching rule b: knowledge-based | 1b | M_b / Q | 2b | $M_b / (Q + D)$ |

M Number of descriptors in common

Q Number of descriptors in Query

D Number of descriptors in Document

Instructions for the exercise in

Open the spreadsheet and get familiar with it.

On the top are three tables

The query formulation and document representations for reference

A table to record for each document (row) the expected relevance score. The spreadsheet will record the scores for you

A table to record the rank ordering, first your intuitive ranking, then the ranking by each formula

Then there are four tables to compute the expected relevance score for each document.

If you have trouble with the instructions, take a peek at Sheet 2, it has all the answers filled in.

Step 1. Rank the four documents D1 – D4 intuitively and put the document numbers in the appropriate cell of the ranking table.

Step 2. Do one document at a time. You need to enter for each descriptor the match value.

For **document D1** this easy: For each query descriptor there is a document descriptor, so all match values in row a are 1 and all match values in row b are 1.

For **document D2** this is a bit more difficult:

in Row a (exact match) there is no match for query descriptors B1.2 and E1.2, so put 0
in Row b (knowledge-based match),

For the query descriptor B1.2 there is the narrower document descriptor B1.2.1, which by matching rule b has a match value of 1, so enter that.

For the query descriptor E1.2 there is the broader document descriptor E1, which by matching rule b has a match value of .5, so enter that.

The spreadsheet computed all expected relevance scores and recorded them in the table on the top

Step 3. Arrange the documents by rank according to each formula (highest score is rank 1) in the Ranking Table. For formula 1a, D1 and D4 have the same top score, so they both go into rank 1, rank2 is free, D2 and D3 go into rank 3.

Now compare the ranking with your intuitive ranking.

If you put D1 and D4 both into rank 1 because they have all four query descriptors, followed by D2 and D3 in ranks 3 and 4, respectively, then you will like formula 1b best. Formula 1a does not work as well because it does not consider that D2 has two descriptors that, while not the same, are related to query descriptors while D3 has only one

If you put D4 in rank 1 and D1 in rank 1 because it has extra descriptors while D4 matches the query exactly, then you might like formula 2b best: even though it places D2 above D1, it separates D2 and D3 and formula 2a does not.

In either case, the knowledge-based formula works better than the exact-match (or dumb) formula. Few search engines use knowledge-based formulas because it is hard to assemble the required knowledge for a universal system. A plain search in Medline uses the dumb matching, but an inclusive (hierarchically expanded) search uses knowledge-based match using the knowledge about hierarchy available in MeSH (Medical Subject Headings).

Whether a document should be penalized for having extra descriptors (D1 vs. D4) is a question answered differently by different people. The pro argument is that a document with extra descriptors does not have as much space left to deal with the query descriptors. But what if D1 is 300 pages and D4 10 pages? We would have to run retrieval tests to answer the question, but they might well be inconclusive, or the results might vary from question to question.

There are many ranking methods used by search engines

The following criteria are among those often used in ranking

- 1 **The rarity of a query word:** A document gets more weight for a rare word than for a word occurring in many documents. Extremely common words are ignored for ranking. Rarity is computed as $1/DF$, where DF is the number of documents in the collection that contain the term one or more times; this is also known as IDF (inverse document frequency)
- 2 **The frequency of query words in the document:** Does the word occur once or more than once (2, 3, 4, etc. are all treated equal). This is called TF (Term Frequency)
- 3 **The absolute and relative location of query words:** A document gets more weight for a query word in the HTML title, in an HTML META tag, or in the top portion of the document than for a query word someplace down in the body of the document. Also, a document gets more weight if two query words occur close together.

1 and 2 are often used by ranking algorithms. The contribution of a word in a document to the relevance score in that document is computed as

$$TF * IDF \text{ (Term Frequency in the document * Inverse Document Frequency)}$$

4 Search modes and data structures

To execute a search, a retrieval system must operate on stored data. The problem is to devise a data structure and a search process that makes retrieval fast. We will discuss data structures for Boolean retrieval and data structure in semantic networks.

4.1 Review the data structures described in Chapter 11

Text, Section 11.1. Key messages:

- Need an index for fast searching
- How the index is built. Takes time to build an index
- Index takes up space (not a big problem anymore)
- In database implementation: Build indexes for data fields that are searched

Note: Data structures in semantic networks were discussed in Lecture 2.2.

4.2 Further elaboration of data structures (Advanced, →LIS 506 Information Technology) See Supplement

Part 3*February 15 - February 29***The nature, design, and management of documents and records**

→LIS 506 Information Technology

Lectures 5.2a - 6.1b*February 15 - February 22***Document function, structure, analysis, and design** (No text chapter)

| | |
|--|---|
| Scope | This part of the course requires a clarification of the scope, particularly what is meant by “document,” and how this topic is approached by many disciplines from many angles. |
| Broad definition of “document” | Text has been defined as: "Any passage, spoken or written, of whatever length, that [forms] a unified whole." (Halliday) Document: any presentation of information in any form: <ul style="list-style-type: none"> written or spoken text, still or moving images, or music and sound (a multimedia document combines all of these) in any medium – print, computer screen, TV, radio, etc. |
| Disciplines/ fields dealing with information presentation / document design | Text linguistics, discourse analysis Rhetoric, English composition Document design, including Web design Information architecture User interface design Instructional design Advertising design Graphics design, including, for example, guidelines for transparencies Formatting documents for interpretation by computer programs |

Objectives and practical significance for Lectures 5.2a - 6.2b are given on the following pages.

Discussion question**Design question for multimedia documents:**

What combination is best for given communicative task, for example teaching a concept, persuading people to do something or quit doing something, or instructing in the use of a device? Generalization of text, which refers only to language.

| | |
|---|---|
| Objectives, Lectures 5.2a-6.2b (inherit down to each lecture) | <ol style="list-style-type: none"> 1 (With lectures 1.1-2.2) Understand the principles for developing a good conceptual structure for a body of knowledge and representing that structure for human understanding and for machine processing. 2 Understand the importance of document structure and presentation for the efficient transmission of information. 3 Be able to analyze the structure and design of existing documents as one aspect in assessing the quality and usability of the document for collection development and matching documents found to the user. 4 Appreciate the importance of good document design and what is needed to achieve good design as a basis for further study. |
| Practical significance | Lectures 5.2a - 6.1b especially in the context of the Web. Inherit down to each lecture |
| General | <ul style="list-style-type: none"> • Well-designed information presentation helps people assimilate and understand information and thus to cope with the ever-increasing amounts of information needed to function in a modern society. |
| Document production | <ul style="list-style-type: none"> • Assisting in the manual production of documents is a very important problem particularly in large organizations such as the World Bank. Also important in helping students studying English composition. Note: On a day-to-day level, most information specialists (including librarians) must produce documents all the time (seen the list below). • Automatic or computer-assisted generation of text and documents • Devising guidelines for document design (as in a text on English composition) |
| Document retrieval | <ul style="list-style-type: none"> • Structure for storing information and selecting specific document parts for retrieval and display. Many organizations now organize their documents into large text databases searchable on an intranet. • Devising systems that help users to find just the right documents or portions of documents for a given purpose • Devising computer systems that can assess relevance as a user would, assimilate information from a document, abstract or index a document |

| | |
|--|--|
| <p>Document analysis and assimilation</p> | <p>I. Understanding how people process documents (assess relevance, assimilate information from a document, abstract or index a document)</p> <p>II. Serving as the user's agent in judging the relevance and appropriateness of a document to the user's situation (background and purpose)</p> <p>III. Guidance in the analysis of documents. Reading and evaluating documents, for example, scientific articles or news stories, is much easier if one understands their structure. Document processing by human indexers or machine indexing systems is based on document structure.</p> |
|--|--|

Documents produced by information specialists

- Presentation of search results (bibliography or substantive data)
- New acquisitions list
- Guide to the library, instructional materials
- Guide to information on . . .
- Promotional materials
- Library newsletter
- Meeting notes
- Classification schemes and thesauri

Library Web site

Increasingly, libraries set up Web sites for use by their patrons; these Web sites include (but are not limited to) the kind of information listed above. (Hint: other libraries' Web servers are very useful information sources, for example www.lib.uchicago.edu/LibInfo/)

Outline for Lectures 5.2a - 6.1b

Lecture 5.2a. Knowledge (re)presentation in text and images. Text linguistics (30 min.)

Micro

Lecture 5.2b. Text analysis overview and examples (30 min.)

[5.2c. Natural language processing. Syntactic and semantic parsing, Supplement]

Macro

Lecture 5.2d Document design. Information design (20 min)

Lecture 6.1a. Document macrostructure and inter-document relationships (40 min.)

Document macrostructure. Document templates

[Hypermedia

Inter-document (inter-textual) relationships] [Supplement]

Lecture 6.1b. Formatting documents for interpretation by computer programs.
Document markup languages (20 min, more fully covered in → 506)

Lecture 5.2a (30 minutes)*February 15***Knowledge (re)presentation in text and images. Text linguistics**

www.dsoergel.com/571/UBLIS571Lecture05.2.mp3 (5.2a-d)

| | |
|---|---|
| Objectives (In addition to the objectives inherited from 5.2-6.1) | <ol style="list-style-type: none"> 1 Be aware of the different types of text and the communication purposes they serve. 2 Understand text coherence and cohesion and their role in text understanding by people and by computer programs. 3 In particular, understand the problems of anaphoric reference. |
| Practical significance | <ul style="list-style-type: none"> • Matching text type with user needs is important in answering questions. This has implications for cataloging. → LIS 518 Reference Sources and Services: Judging appropriateness • Understanding text coherence and text cohesion is important for evaluating texts and for good writing. • Knowing about anaphoric references points out limitations in using adjacency commands in free-text searching and the need for linguistic processing to overcome these limitations. → LIS 518 Reference Sources and Services: Query formulation |

| | |
|----------------|--|
| Context | Lecture 2.2 focused on knowledge representation in computer systems and in the brain, so did Lecture 4.2 on data schemas and formats. This sequence of lectures focuses on external knowledge representation with the purpose of helping people to assimilate knowledge. But there is much overlap and interrelationship: <ul style="list-style-type: none"> • Semantic networks are used as external knowledge representation in the form of concept maps. • Record formats can be useful for external representation. • Internal representation often serves the main purpose of letting a computer program create multiple external knowledge representations (Lecture 6.1b on XML). |
|----------------|--|

Outline

- 1 Functions of documents
- 2 Document or text types
- 3 Text structure: cohesion and coherence

Over

1 Document analysis. Document functions

Just read Sections 1 and 2

| Perspectives in document analysis | |
|--|--|
| | <ul style="list-style-type: none"> • Internal document structure • External context or communicative situation of the document: <ul style="list-style-type: none"> • the creator (writer/speaker/designer) and the peruser (reader/listener/viewer) and their relationship • the function the document fulfills |

| Document functions | |
|--------------------------------|--|
| Informing, educating | <p>Enable the reader/viewer to construct, reconstruct, or otherwise update his or her own mental image, to make sense out of the message presented</p> <ul style="list-style-type: none"> • Reporting results of research or scholarly endeavors. Describing objects or events • Educating: helping the reader understand a new field or topic • Providing small pieces of information quickly as needed • Reporting the discourse at a meeting and its results (minutes) • Laying out a plan of action • Giving instructions/prescriptions/orders (transmitting norms) (also below) • Informing documents may be designed for reading in context or to enable the reader/viewer to quickly locate a needed piece of information. |
| Instructing, persuading | <p>Creating or changing beliefs, attitudes, or behaviors. Persuasion.</p> <ul style="list-style-type: none"> • Giving instructions/prescriptions/orders (transmitting norms) (repeated here) • Persuading somebody to do something (vote for a person or issue, buy something) • Assisting in the treatment of mental or behavioral disorders by enabling the reader/viewer/listener to construct or reconstruct emotional/attitudinal structures, including self-image (bibliotherapy) |
| Entertaining | <p>Providing entertainment or enjoyment</p> |
| | <p>A document can serve multiple functions, especially it can entertain while it educates (and educate the better for it, “edutainment”)</p> |

2 Document or text types (summarized from Beaugrande)

The type of a text is determined by its internal structure and the communicative situation, especially the function of text. Major text types are listed and defined briefly below. The viewer/reader/listener can process a document more efficiently if s/he knows the its type
 →LIS 518 Reference Sources and Services: Matching text type with user's purpose.

| Major text types | |
|--|--|
| (There are many, many text types; this list is just the tip of the iceberg.) | |
| Type | Examples |
| <i>Facet 1</i> | |
| Descriptive: | Review article; newspaper article; dictionary definition |
| Argumentative: | Logical proof; Legal argument |
| Didactic: | Textbook |
| Narrative: | Fairy tale; Letter |
| Conversational: | Reference interview |
| <i>Facet 2</i> | |
| Literary: | Prose (e.g., Novel); Poetry (e.g., Limerick) |
| Scientific: | Research study report |

The classification of text types parallels roughly the classification of functions, but there is not a perfect one-to-one correspondence; for example, a poem may educate or persuade or entertain, or all three at once.

Elaboration of text types adapted from Beaugrande *Text, discourse, and process*, VII.1.8
 See supplement.

Question

Why should a reference librarian understand document functions and text types?

How does this help with ascertaining the user's information need?

How does this help with selecting the right documents for the user?

3 Text structure: cohesion and coherence

Just read

Cohesion and coherence as the key devices in determining the internal structure of texts

| | |
|---|---|
| <p>Cohesion (Grammatical)</p> <p>Anaphoric reference</p> | <p>Elements of a text are properly linked grammatically: Properly structured sentences Inter-sentence relationships</p> <p>Use of a pronoun or general noun to refer to an object, action, or thought previously identified in the text.</p> <p>Example:</p> <p>President Bill Clinton gave a speech at Concord High School. He emphasized the need for crime prevention and for the restoration of family values. This made the Republicans angry. They accused him of stealing their issues. Meanwhile, Patrick Buchanan addressed a rally in Manchester. He hammered away at the theme that the jobs of American workers must be protected from low-wage foreign competition. This theme has propelled him to the front in the polls.</p> <p>Importance in the context of information systems:</p> <ul style="list-style-type: none"> - Detecting the relationships in a text. - Proximity searching. |
| <p>Coherence (Lexical- semantic)</p> | <p>Does the document make sense? Does an argument proceed in a logical fashion?</p> <p>IV.If a section requires background knowledge the reader cannot be expected to possess, does the document provide this background knowledge before the reader gets to that section?</p> <p>V.Are there proper transitions to prepare the reader's mind set for new information?</p> <p>VI.Do illustrations fit with the text?</p> <p>VII.In a conversation: Is a question properly answered? Does the contribution of one participant build on previous contributions?</p> <p>Importance in the context of information systems:</p> <p>Design hypermedia systems that support the user in constructing coherent documents</p> <p>Coherence related to document/text macrostructure</p> |

Incohesive text

President Bill Clinton gave a speech at Concord High School. **They** emphasized the need for crime prevention and for the restoration of family values. **This** made the Republicans angry. **She** accused **him** of stealing **her** issues. Meanwhile, Patrick Buchanan addressed a rally in LA. **She** hammered away at the theme that the jobs of American workers must be protected from low-wage foreign competition. **This scandal** has propelled him to the front in the polls.

Cohesive but incoherent text

President Bill Clinton gave a speech at Concord High School. **He** talked about playing the saxophone and mused about Plato. **This** made the Republicans angry. **They** climbed the Mount Everest. Meanwhile, Patrick Buchanan addressed a rally in LA. **He** ran down the street smashing cars. **This courageous action** has propelled him to the front in the polls.

Two related principles from composition:

| | |
|------------------------------|--|
| Frame-style paragraph | Sentences in such a paragraph all have the same grammatical subject or main focus. The paragraph presents a frame focused on one entity; each sentence is a frame slot giving information on that entity, allowing the user to maintain focus rather than jumping back and forth. |
| | <p>Example:</p> <p>Cattle (called cows in vernacular usage) are domesticated ungulates, a member of the subfamily Bovinae of the family Bovidae. They are raised as livestock for meat (called beef and veal), dairy products (milk), leather and as draft animals (pulling carts, plows and the like). In some countries, such as India, they are subject to religious ceremonies and respect. Cattle are estimated to number 1.3 billion in the world.</p> |

| | |
|---------------------------------------|--|
| Spreading activation paragraph | Alternatively, the sentences in a paragraph should be strung together so that the entity mentioned in the previous sentence is taken up at the beginning of the next sentence, like a path through a semantic network. |
| | <p>Example:</p> <p>Cattle are raised for beef and milk. Their milk is an important source of calcium. Calcium is important for growing strong bones and healthy teeth in children and preventing osteoporosis. Calcium is also important for many functions in the body, for example, muscle contraction, which is especially important for athletes.</p> |

Relationship types and their expression in text

The relations given in the Crombie reading and illustrated by the two annotated text examples elucidate further the concept of text cohesion and coherence. To understand a document, a person or a computer program must ascertain the relationships between individual elements. An automated language understanding system needs rules that link the types of relationships with grammatical features. The converse process of text generation needs the same rules. Some of Crombie's relationships have also been proposed for *relational indexing*, which covers not only the concepts treated in a document but also the relationships between these concepts.

Note the relationship between concepts of formal (highly structured) knowledge representation in Lecture 2.2 and the less structured (but of course not completely unstructured) knowledge representation in text:

- frame-style and spreading activation paragraphs,
- the representation of relationships in text (Crombie)

Introduction to the analysis of how relationships are expressed in text

We talk about how relationships are expressed in text to better understand systems that can extract assertions (facts) from text. Such systems exist now commercially (see later readings). They can be very useful to collect a large number of assertions for the knowledge base of an expert system (such as the drug prescription expert system discussed in Week 1) and to answer questions from users by giving the actual answer (substantive data) rather than just pointing to one or more documents from which the user needs to laboriously extract the answer.

Relationship extraction (also called information extraction) from text is one way to cope with information overload. You should consider acquiring such systems for your patrons. We will soon see such systems on the Web.

To extract relationships from text, a computer program needs to be able to figure out what an anaphoric reference refers to. This problem is illustrated in the cohesion part of both Crombie examples through the arrows that point from an anaphoric reference to the person, thing, or action referred to (the referent).

Then we need to analyze the many ways in which relationships can be expressed in text and generate patterns that we could give to a computer program. Look at Crombie example 2 and identify patterns (indicator words, grammatical patterns) that indicate a causative relationship so that a program could extract assertions of the form

A <*causes*> B

Example of an indicator word: Because

You may post your observations to the forum thread

First try to find examples of your own.

Then look at the examples on the page following the Crombie pages

XXX Include one paragraph without the lines so that students can draw their own. Include para with lines also.

Crombie 1

Crombie 2

Examples of patterns for extracting causative assertions

| Text | Pattern | Extracted assertion |
|--|---|---|
| <i>Because</i> the highest electrical conductivity can be achieved only with pure metal, copper is extensively used pure | <i>Because</i> Cause, Effect could also be Effect <i>because</i> cause | highest electrical conductivity can be achieved only with pure metal <causes> copper is extensively used pure |
| Stress <i>due to</i> {the effect of} wind | Effect <i>due to</i> Cause | Wind <causes> Stress |
| The strength [of metal] can be improved <i>by</i> alloying | Effect <i>by</i> Cause | Alloying <causes> Material strength |
| The addition of ten percent aluminum more than <i>doubles</i> the strength of copper | Cause <i>doubles</i> Effect (There are many such verbs that indicate causation such as <i>reduce</i> or <i>increase</i> , see example at the end of the passage) | Addition of ten percent aluminum <causes> Doubled strength of copper |

Lecture 5.2b*February 15***Text analysis overview and examples** (30 minutes)

www.dsoergel.com/571/UBLIS571Lecture05.2.mp3 (5.2a-d)

| | |
|---|--|
| <p>Objectives (In addition to the objectives inherited from 5.2-6.1)</p> | <ol style="list-style-type: none"> 1 Be aware of important text analysis methods. 2 Have an appreciation for and understanding of the potential of automated text analysis for processing vast quantities of information through automated translation, automated indexing, and extraction of data from text |
| <p>Practical significance</p> | <p>Increasing amounts of text need sophisticated linguistic tools for intelligent indexing and data extraction, for example, Convera RetrievalWare, www.convera.com/solutions/retrievalware/default.aspx , Inxight, www.inxight.com/pdfs/linguistics_adding_value.pdf</p> <p>Specifically, linguistic techniques can help with the following functions:</p> <ul style="list-style-type: none"> • Preparing a description of the document <ul style="list-style-type: none"> Descriptive cataloging (e.g. from optically scanned title page) Subject indexing Abstracting, text summarization (e.g., Tools > AutoSummarize in Word, Copernic Summarizer, www.copernic.com/en/products/summarizer/) • Determining the reading level of a document (more generally: the audiences for which the document is appropriate) • Determining the attitudes, beliefs, or emotions underlying the document (content analysis in sociology and political science or in psychoanalytical methods) • Determining authorship or other characteristics of the origin of the document • Preparing a hypertext version of a document, possibly for incorporation into a larger hypertext • Extracting data from a document. Representing the relationships expressed in a document in a more explicit and more easily manipulated way • Machine translation, for example on-the-fly translation of Web documents |

| Scope and limitations of lectures 5.2b-6.1a | |
|--|---|
| Scope | <p>This pair of lectures introduces tools and methods for performing linguistic and statistical analysis of text. This includes</p> <ul style="list-style-type: none">• The use of textual analysis in the building of information retrieval systems and knowledge-based systems.• Frame-based data extraction from text.• Rudiments of parsing sentences (supplement) |
| Limitations | <p>These lectures concentrate on a sub-area of document analysis, namely natural language processing applied to machine-readable text (text available as a stream of individually encoded characters). Text available as sound or graphics can be automatically converted (optical character recognition, speech recognition). Analysis of images (for example, object recognition) and analysis of sounds are other sub-areas of document analysis. Literary and artistic analysis also falls in the broader context of document analysis.</p> |

Approaches to text analysis

Most of these techniques are used by human readers and machine systems alike for the purposes outlined under *Practical significance*.

Human readers may analyze a text for indexing, abstracting, extracting a specific fact or proposition, or for assimilating all the facts or propositions expressed in the text.

- **Statistical**

- A. Word / phrase / concept frequency

- B. Frequency of words that connote an attitudinal/emotional dimension (**content analysis** in psychology/sociology/political science).

- C. Differential frequency.

- D. Looking for the unexpected (such as weighting rare words highly in ranking retrieval results), as in AltaVista's ranking method, Lecture 5.1

The statistical approach is used mostly by computer systems, but perhaps also implicitly by human readers

- **Based on text macrostructure - positional approach**

For example:

- E. Introduction and conclusions useful source for abstract.

- F. Section headings and figure captions useful source for index terms.

- G. First and last paragraphs of sections, first and last sentences of paragraphs

- **Cue words, cue phrases, and cue sentences**

For example, "method", "important result", "new"

- **Syntactic and semantic analysis**

- H. Parsing of sentences (sentence diagraming) or partial parsing to detect noun phrases

- I. Parsing with semantic interpretation

- J. Inter-sentence parsing, resolution of anaphoric references

- **Slot filling in frames using parsing or cues**

In-class exercises and examples illustrating the importance of text analysis through several linguistic techniques

- 1 Importance of **resolving anaphoric references** in free-text searching with proximity operators 122
- 2 Extracting substantive data through **slot-filling in frames**: examples 124
- 3 Extracting data from text, especially importance of **resolving anaphoric references** 128
Lecture 6.1a has an example text with extracted data
- 4 Importance of **recognizing noun phrases** 130
Lecture 6.1a deals with parsing to detect noun phrases
- 5 Importance of **semantic interpretation**, especially **disambiguation of homonyms**, for retrieval and automated translation 132

A further technique, not shown in the examples, is **searching for a word or phrase and its synonyms**

Some of the techniques mentioned here (**in bold**) are applied in Assignment 7

1 Importance of resolving *anaphoric references* in free-text searching with proximity operators

Proximity operators used here (syntax varies from system to system)

WS two words occurring in the same sentence

WP two words occurring in the same paragraph

Texts are from the Columbia University College of Physicians and Surgeons *Complete medical home guide*

Query statement / information need:

What to do about sticky eyelids

Query formulation to search free-text:

eyelid! WS stick!

BLEPHARITIS

Blepharitis is an infection of the edges of the eyelids. They become red, **sticky**,

and crusty, and sometimes the victim has to unstick them to see anything in the morning.

The WS query formulation misses this entry: *eyelid* and *sticky* do not occur in the same sentence.

Query formulation: calcium WS excret!

WS within same sentence

Query formulation: osteoporosis WP vertebr!

WP within same paragraph

OSTEOPOROSIS

BONES NEED CALCIUM to maintain their strength, hardness, and to stay healthy. Milk, the main source of calcium in the diet, is important for the growing skeletons of children and adolescents as well as the bone-forming cells of adults. Regular daily consumption of at least 1 cup of skim or low-fat milk is essential for adults who want to keep their bones strong and to help prevent osteoporosis, a disease in which the body's bone mass decreases and bones become thin and brittle. Bones weakened by osteoporosis, a disease common to postmenopausal women, are prone to fracture if a person falls.

When **calcium** enters the body, it is absorbed into the bloodstream. If there is any excess, it is deposited in the end of the bone shafts where it is stored until the body needs to tap this reserve. (*Some* is also **excreted** via the kidneys.) When the calcium supply is deficient, the blood must take it back from

the bones. If calcium intake remains inadequate over a long period of time, the bones eventually become porous and weak.

It is not known why calcium loss occurs. That postmenopausal women tend to get osteoporosis points in the direction of a hormonal disorder as estrogen in women of this age falls off sharply. Estrogen therapy is one treatment but its ability to decrease calcium loss may last only several years. Increased calcium intake and exercise are other therapies. The links between lack of exercise and osteoporosis are becoming firmer as research into the causes of this disease progresses.

The disease most frequently affects the spinal column, causing backaches and rounded shoulders. In severe cases, the bone becomes as porous as a sponge and can collapse as a result. Collapsing **vertebrae**, which can cause sudden and sharp backaches, is one reason why elderly people tend to get shorter.

2 Extracting data through slot-filling in frames: examples

Understanding and summarizing stories by machine

Based on distinguishing types of stories, such as *corporate merger*, *disaster*, *state visit*. **Each type of story has** a list of items to be included in a summary; these are arranged in **a frame** specific for that type of story.

The summarizing process then proceeds in two steps:

- 1 Detect basic type of story, for example *story about disaster*, and pull up the proper frame
- 2 For filling each slot, fill the slots following the instructions given

Disaster frame – general pattern

| Slot | Instructions: What to look for to find slot fillers |
|-------------------------|---|
| <i>Type of disaster</i> | indicator word such as <i>earthquake</i> , <i>aftershock</i> , <i>hurricane</i> |
| <i>Where</i> | place name (from a large dictionary of place names) |
| <i>When</i> | date line plus words such as <i>today</i> , <i>yesterday</i> , <i>Sunday</i> , <i>recent</i> |
| <i>Number of dead</i> | <i>killed</i> or <i>dead</i> or <i>fatality</i> , and a number close by |
| <i>Amount of damage</i> | (\$ or <i>dollar</i> or and number before or after) or <i>much</i> or <i>heavy</i> , esp. when close to <i>damage</i> or <i>worth</i> or <i>destroyed</i> |

Disaster frame – Event 345

| Slot | Slot filler (for story on facing page) |
|-------------------------|--|
| <i>Type of disaster</i> | earthquake aftershocks |
| <i>Where</i> | central Italy |
| <i>When</i> | October 6, 1997 |
| <i>Number of dead</i> | 10 |
| <i>Amount of damage</i> | \$1 billion |

Disaster frame – Event 406

| Slot | Slot filler (for story on facing page) |
|-------------------------|--|
| <i>Type of disaster</i> | hurricane |
| <i>Where</i> | Mexico's Pacific Coast, Acapulco |
| <i>When</i> | October 9, 1997 |
| <i>Number of dead</i> | 120 |
| <i>Amount of damage</i> | untold millions of dollars |

Frame for extracting data from pesticide reports see Supplement

Aftershocks Jar Central Italy; Repair Cost Put at \$1 Billion

Associated Press

ROME, Oct 6—The ground in **central Italy** rumbled again **today**, and officials said repairing buildings **damaged** by a series of earthquakes could cost more than **\$1 billion**.

The aftershocks in the Umbria and Marches regions have prompted more people to seek temporary shelter, 11 days after a pair of quakes **killed 10** people. The National Geophysics Institute said today's tremors hit about every 30 minutes before dawn, the strongest with a magnitude of 3. No new destruction was reported.

The Sept. 26 quakes damaged the beloved Basilica of St. Francis in Assisi, along with thousands of other buildings.

The less severely damaged buildings will be repaired so that as many people as possible can return to their homes before winter, civil defense chief Franco Barberi said at a news conference. He said it will cost \$875 million to \$1.15 billion to repair damaged buildings.

The government will move about 3,000 units of prefab housing into the region in the next few weeks. Tents and camping vehicles already in place can shelter as many as 50,000 people.

Hurricane Devastates Mexico's Pacific Coast

Floods Kill at Least 120, Most in Acapulco

By Chris Kraul
and Mary Beth Sheridan

Los Angeles Times

ACAPULCO, Mexico, **Oct 9-**

Bearing 115 mph winds and torrential rain, Hurricane Pauline roared out of the Pacific through this coastal resort region before dawn **today**, leaving at least **120** people **dead**, thousands homeless and **untold millions of dollars in damage**.

Most of the dead were counted in and around Acapulco, a sunny port city usually filled with carefree Mexican and foreign tourists. The powerful storm left Acapulco, a city of about 1 million people, "unrecognizable," according to one report—a tangle of uprooted trees, downed power lines, overturned cars and bodies.

Morning light revealed corpses and garbage and the wreckage of countless wood-frame homes floating in oily, four-foot-deep floodwaters that coursed through the streets and washed over La Costera

Miguel Aleman, a fabled promenade skirting Acapulco's ocean-front. City officials said there had been some isolated instances of looting, and army units were called out to patrol the streets. A deluge of rain—20 inches in less than 24 hours—sent floodwater, mud, gravel and boulders rushing down drought-parched hills surrounding Acapulco through several slum neighborhoods, smashing poorly constructed shanties and more substantial houses to flinders and washing away anything not firmly anchored. At least seven mudslides reportedly caused heavy property damage around the city, and local officials fear thick layers of mud coating many neighborhoods may conceal dozens of bodies as well.

While the official death stood at 120 late last night local authorities said it would certainly climb—and perhaps double—as search parties comb through the debris left by the storm. The U.S. Embassy in Mexico City said that no Americans were reported among the dead or missing.

The Red Cross issued a plea for

See HURRICANE, A29, Col. 4

From the Washington Post, October 7, 1997 and October 10, 1997, respectively

3 Extracting data from text, esp. importance of resolving anaphoric references

Consider the following text (only the example marked with | at the left margin is treated in class; explore the other **bold** / *italic* pairs on your own):

VASCULITIS

VASCULITIS, as the name implies, is an inflammation of the blood vessels — both the arteries and the veins. Diseases in this category are relatively rare and comprise some of the most baffling and poorly understood disorders in medicine. Very often, the diagnosis remains unsuspected for long periods because of the variable way in which these disorders behave.

Inflammation of a blood vessel, particularly a small artery, can cause a narrowing of its lumen (internal diameter). If the vessel becomes completely closed, the tissue normally nourished by the diseased artery will die or be severely damaged.

Some forms of vasculitis are believed to result from an allergy or hypersensitivity, such as an adverse reaction to certain drugs. Sulfa drugs were very common causes of vasculitis, particularly in the early days of their use when the preparations were more crude and the dosages given were higher than today.

Patients with vasculitis, particularly when it involves widespread areas in the body, many be extremely ill with a generally poor prognosis. One particular type of vasculitis, which affects older people, involves inflammation of the cranial or temporal arteries, the vessels that serve a portion of the facial, jaw, and tongue muscles, the scalp, and most important, the retina. **Cranial arteritis** is the most common cause of sudden blindness in the elderly. Usually only one eye is involved but sometimes it occurs in both. *This condition* is successfully treated with corticosteroids, provided that treatment is started before there is significant loss of vision. *It* is often associated with a syndrome of severe muscle pain and stiffness called **polymyalgia rheumatica**. *This illness* is also largely confined to the elderly. It is almost always associated with a very high sedimentation rate, which measures the amount of inflammation, and it usually responds dramatically to cortisone-type drugs in low doses. Polymyalgia may occur without cranial arteritis, but because of the association, arteritis should be suspected in patients with polymyalgia.

Another form of vasculitis is called **Wegener's granulomatosis**. *This* is an extremely rare disorder which attacks the respiratory tract, the nasal sinuses, and the kidney in a progressively destructive process. Wegener's granulomatosis was once invariably fatal but now most patients can be treated successfully with cytotoxic or immunosuppressive drugs.

Patients with generalized or systemic vasculitis will often have paralysis of a foot or a wrist as a result of loss of blood supply to the peripheral nerve serving that limb. The blood vessels of the lung may also be affected, resulting in asthmalike symptoms. The development of asthma relatively late in life is very unusual, and may signify vasculitis.

There is another type of vasculitis known as **Takayasu's disease**, which occurs almost exclusively in young women. The inflammation is largely restricted to the branches of the great artery which leaves the heart (the aorta). *It* has also been called "pulseless" disease, for the diseased arteries may be so narrowed that a pulse cannot even be detected at the wrist. Patients with *this disease* will very frequently have symptoms of dizziness, light-headedness, weakness, and difficulty in using the arms, due to muscle pain from even slight physical effort. This is a direct result of lack of oxygen to the muscles, as the narrowed arteries are unable to deliver the increased amount of blood required during muscular effort. Corticosteroid therapy may be effective against Takayasu's disease, but the disease may go into remission without treatment.

These diseases are a few examples of the very broad spectrum of disorders included in the category of vasculitis. They are often difficult to diagnose, for their onset and evolution may be vague and ill-defined. The more classic types are easier to identify, but because of their relative rarity they are often not suspected until late in the course of the illness. Biopsy of an involved organ such as the kidney, muscle, or liver may be required in order to establish that a vasculitic process is indeed present

| | |
|---|--|
| <p>Example of information extraction (entity-relationship statement extraction, relationship extraction) from the previous text</p> <p>Done by hand to illustrate what we want a machine to do</p> | |
| <p>Patients with vasculitis, particularly when it involves widespread areas in the body, may be extremely ill with a generally poor prognosis. One particular type of vasculitis, which affects older people, involves inflammation of the cranial or temporal arteries, the vessels that serve a portion of the facial, jaw, and tongue muscles, the scalp, and most important, the retina. Cranial arteritis is the most common cause of sudden blindness in the elderly. Usually only one eye is involved but sometimes it occurs in both. <i>This condition</i> is successfully treated with corticosteroids, provided that treatment is started before there is significant loss of vision. <i>It</i> is often associated with a syndrome of severe muscle pain and stiffness called polymyalgia rheumatica. <i>This illness</i> is also largely confined to the elderly. It is almost always associated with a very high sedimentation rate, which measures the amount of inflammation, and it usually responds dramatically to cortisone-type drugs in low doses. Polymyalgia may occur without cranial arteritis, but because of the association, arteritis should be suspected in patients with polymyalgia.</p> | <p>Vasculitis <mayCause> Extreme illness Vasculitis <hasPrognosis> Poor</p> <p>Cranial arteritis <definedAs> Vasculitis that involves inflammation of the cranial or temporal arteries, the vessels that serve a portion of the facial, jaw, and tongue muscles, the scalp, and most important, the retina</p> <p>Cranial arteritis <occursIn> Elderly Cranial arteritis <mayCause> (Sudden blindness, Elderly) (Sudden blindness, Elderly) <causedBy> (Cranial arteritis, High percentage of cases)</p> <p>(Blindness <causedBy> Cranial arteritis) <preventedWith> (Corticosteroids, Given early)</p> <p>Cranial arteritis <associatedWith> Polymyalgia rheumatica</p> <p>Polymyalgia rheumatica <definedAs> Syndrome of severe muscle pain and stiffness</p> <p>Polymyalgia rheumatica <occursIn> Elderly Polymyalgia rheumatica <causedBy> (Cranial arteritis, Medium percentage of cases)</p> <p>Sedimentation rate <measures> Degree of inflammation Polymyalgia rheumatica <associatedWith> (Sedimentation rate, High)</p> <p>Cranial arteritis <mayCause> Polymyalgia rheumatica Polymyalgia rheumatica <treatedWith> (Cortisone-type drugs, Dosage: low, Response: very good)</p> |

3 Examples: Importance of recognizing noun phrases for retrieval and translation

Noun phrases and semantic interpretation (word sense disambiguation)

Note: To search for phrases, in most systems use “ ”.

Example 1

information retrieval, retrieval of information, retrieval of legal information
but: information on the retrieval of sunken treasures

Example 2. Noun phrases expressing a unit of meaning

hepatitis A
vitamin A
twelve-step program
route of administration (also known as administration route, medication route,
route of drug entry, method of drug application)
gene pool
breath test
motivational interviewing
blue law
social control
boundary layer flow (aerodynamics)
data link layer (data communication)

peer pressure, pressure by peers
social pressure
vapor pressure
benefits program
safety program
conference program
computer program

Meaning of polysemous words
determined by context in a phrase

Example. Importance of parsing complete sentences for noun phrase identification

- | | | |
|---|--|--|
| 1 | The green vegetables supply calcium. | NP The green vegetables V supply |
| 2 | The green vegetables supply calcium to the body. | |
| 3 | The green vegetables supply digestible calcium. | |
| 4 | The green vegetables supply determines sufficiency of calcium. | NP The green vegetables supply |

4 Importance of semantic interpretation for retrieval and automated translation

| | |
|--|--|
| <p>Example: Importance of semantic interpretation for disambiguating homonyms in searching (sense disambiguation, meaning disambiguation)</p> | |
| <p>Query statement / information need:</p> | <p>Passages referring to white (race/ethnic group)</p> |
| <p>Query formulation to search free-text:</p> | <p>white</p> |
| <p>Passages retrieved:</p> <p>White students were found to hold prejudices against their black and Hispanic peers.</p> <p>White cars are preferred by middle-aged buyers.</p> <p>The white dishwasher laughs</p> <p>The white dishwasher is broken.</p> <p>The black congresswoman won election in a majority white district.</p> <p>Douglas White won the race.</p> <p>A white knight came to the rescue of CSX Corporation in its take-over fight.</p> <p>The family unit is the basis for American society. White units make up 53% of all family units in the state.</p> <p>GE makes microwave ovens. Half the units sold are white.</p> <p>The white drinking fountain</p> <ol style="list-style-type: none"> a. In a story set in the historic segregationist South b. In a travel guide to Italy | |
| <p>A sophisticated free-text retrieval system would analyze the text to determine the meaning of white in each passage and tag the passage accordingly. It would ask the user what meaning of white she was after and find only properly tagged passages. Mistakes in the analysis may cause retrieval of erroneous passages and rejection of relevant passages.</p> <p>This is also known as Word Sense Disambiguation (WSD)</p> | |

| | |
|---|---|
| <p>Example: Importance of semantic interpretation for automated translation</p> | |
| <p>The white dishwasher laughs.</p> <p>German: Der weisse <i>Tellerwäscher</i> lacht.</p> <p>French: Le <i>plongeur</i> blanc rit.</p> | <p>The white dishwasher is broken.</p> <p>German: Die weisse <i>Spülmaschine</i> ist kaputt.</p> <p>French: Le <i>lave-vaisselle</i> blanc est détraqué.</p> |

Semantic interpretation often requires the parsing (diagraming) of complete sentences.

More examples: Semantic interpretation rules

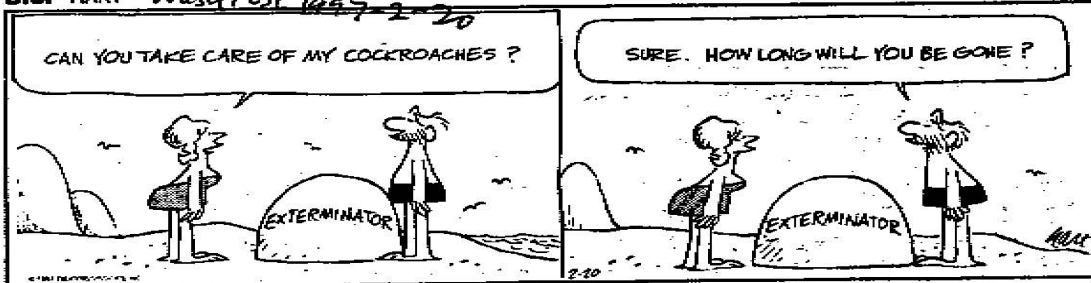
Jokes are often based on deliberately misconstruing the meaning of a word in a given context. Thus, they can focus the attention of the language analyst on words with multiple meanings and on the semantic interpretation rules that distinguish between these meanings. With this in mind, analyze the comic strip episodes below.

BEETLE BAILEY MORT WALKER



Wash Post
1997-3-5

B.C. HART Wash Post 1997-2-20



Washington Post 2000-4-2

by Garry Trudeau



DOONESBURY



5.2c (0 min, see Supplement)**February 15****Natural language processing. Syntactic and semantic parsing**

| | |
|--|---|
| Objective inherited from Lect. 5.2-6.1 | Have a general idea how syntactic parsing and semantic interpretation work. Note: Parsing is the linguists' term for sentence diagraming |
| Practical significance | Natural Language Processing (NLP) is booming both in commercial applications and in academic research. It is used for the following purposes (see also list in Lecture 5.2b and example on next page): <ul style="list-style-type: none"> • Automatic indexing /classification, including document categorization / automatic cataloging / automatic metadata generation • Automatic abstracting, automatic summarization, including creating unified summaries from multiple documents (e.g., multiple news stories on the same topic or event) • Automatic extraction of formatted data from text (information extraction, fact extraction, relationship extraction) (See example on next page) • Question-answering: Within a large document, find the specific sentence or paragraph that answers a question • Automatic translation (e.g., www.google.com/language_tools) • Grammar checking in a word processor (or computer-assisted essay-grading) • Creating textual answers from the data returned by a database query • Automatic speech recognition |
| Note | Working through some detailed examples is necessary to create a good sense of what is going on in natural language processing. However, there is no need for memorizing the details. You will not be required to produce the step-by-step sequence of a parse. <i>Parsing</i> is just a different word for <i>diagraming sentences</i> . <i>Parsing</i> is generally used with the connotation that the diagraming is done by a computer program. |

This outline is just a teaser, not actually a lecture in this course)

From the examples in 5.2b you should understand the enormous importance of natural language processing for providing users with quick access to just the information they need. Huge amounts of information are stored in written text and increasingly in audio text, where it is hard to access and process and correlate with other information. So transforming this information into a more structured form where specific data can be easily retrieved and where data can be processed and correlated is an extremely important function and one companies will pay money for. (Again, text is not unstructured, just not structured enough for easy processing.) We do not have time in this course to even introduce **how** this can be done. Some of you may be interested in learning more, perhaps even taking a course in linguistics. Let me know. The supplement has material on this, but not the guide which would also be needed.

Examples of Natural Language Processing (NLP) Software

Word: <http://office.microsoft.com/en-us/word-help/automatically-summarize-a-document-HA010255206.aspx>

www.copernic.com/en/products/summarizer/

Links to NLP tools, nicely classified:

www-a2k.is.tokushima-u.ac.jp/member/kita/NLP/nlp_tools.html

There is much commercial information retrieval software and content management software that uses NLP techniques

A few examples

Temis - Luxid Semantic content enrichment goes mainstream

www.temis.com/index.php?id=208&selt=6

Autonomy IDOL10

Autonomy's next-generation information platform, Autonomy IDOL 10, is a single processing layer that enables organizations to extract meaning and act on all forms of information, including audio, video, social media, email and web content, as well as structured data such as customer transaction logs and machine-based sensor data. The platform combines Autonomy's infrastructure software for automatically processing and understanding unstructured data with the high-performance, real-time analytics engine for extreme structured data from Vertica, an HP Company.

<http://idol.autonomy.com/>

ConceptSearching

www.conceptsearching.com/Web/home.aspx

Alchemy - Transforming text into knowledge

www.alchemyapi.com/

Lecture 5.2d (20 min)

February 15

Document design (information design)
Formatting documents for understanding by people
External representation of information

www.dsoergel.com/571/UBLIS571Lecture05.2.mp3 (5.2a-d)

| | |
|-------------------------------|---|
| Objectives | Inherited from Lectures 5.2 - 6.1 Gain a feel for good document design where the external form conveys the internal structure well. |
| Practical significance | Inherited from Lectures 5.2 - 6.1 To provide users with documents, Web sites, screens, and other information representations that optimally support understanding, you must be able to select or create such documents or other representations. To summarize: Understanding document design is important for <ul style="list-style-type: none">• creating documents• selecting documents for a collection• matching documents with users |

Some principles for good document design

| | |
|-------------------------------|--|
| <p>Know the reader</p> | <p>Problem to be solved / task to be accomplished Information need Background knowledge</p> |
| <p>Content</p> | <p>Select the information carefully - only what the reader needs to know. Avoid redundancy or use it purposefully</p> |
| <p>Structure</p> | <p>Elaborate in your own mind the intrinsic structure of the topic / the phenomena to be presented - good document design is grounded in a thorough understanding and structuring of the topic.</p> <p>Choose the external representation structure that best facilitates the assimilation of the intrinsic structure — form follows function. Examples of external representation structures: Plain text, typographically structured text (such as a list or a linear arrangement of a hierarchy), table, diagram, picture.</p> <p>A general structure that is often useful</p> <ul style="list-style-type: none"> Make schemas explicit. Provide advance orienters ("Tell them what you are going to tell them"). Give the detail ("Tell them"). Provide opportunity for rehearsal or application to fix the new information in the reader's mind. ("Tell them what you told them.") |
| <p>Layout</p> | <p>Provide guideposts that indicate the overall context . (for example, running heads, navigation chains as in Yahoo).</p> <p>Point out relationships (cross-references, links).</p> <p>In documents intended for looking things up, such as a dictionary: Provide guiding headings at the top of each page. (Counterexamples: Library of Congress Subject Headings, Dewey Decimal Classification index, MeSH.)</p> |

High-level means of expression, media modalities

| | Non-linguistic (depicting) | Linguistic (text, verbal, convention) | | | |
|---|--|---|--------|--|--------------------------------------|
| Auditory Hearing/audio/ sound | Sound Music | Spoken language Speech | | | |
| | | “Audons” | | | |
| Visual Sight/vision/ graphical | Images, pictures (photos, paintings, drawings, charts, diagrams) “Visuals” (including real objects and models) | Written (printed) language “Text” in Information Retrieval | | | |
| | <table border="1"> <tr><td>Still</td></tr> <tr><td>Moving</td></tr> </table> | Still | Moving | <table border="1"> <tr><td>Icons Pictograms Sign language</td></tr> </table> | Icons Pictograms Sign language |
| Still | | | | | |
| Moving | | | | | |
| Icons Pictograms Sign language | | | | | |
| Touch/tactile | Tactile representations, for example a three-dimensional map | Braille | | | |
| Other senses | Smell is of little practical significance here. The kinesthetic senses do not apply (except perhaps in virtual reality applications) | | | | |
| Audiovisual | Sound and images simultaneously (as experienced in real life), often includes speech, may include written text. | | | | |
| Multimedia, hypermedia | Combination of pieces of presentations in different modalities: Display of written text may be followed by still image, or a series of still images with explanation in speech, which in turn may be followed by an audiovisual segment. Multimedia kit. Hypermedia — interlinked segments in several modalities | | | | |

| | | |
|-----------------|--|--|
| Consider | Medium Arsenal of artistic expression, visual vocabulary (icons, symbols in comic strips) Image structure Style | Language (Chinese, English, etc.) Vocabulary Text structure Style |
|-----------------|--|--|

| |
|---|
| Low-level means of expression |
| <p>Typography: Type face, type size</p> <p>Highlight or lead symbol (triangle, bullet, square, pointing hand, etc.)</p> <p>Graphical means for highlighting or de-emphasizing (often used to distinguish between options that are available at the moment and those that are not)</p> <p style="padding-left: 40px;">Bold, blinking (use sparingly), reverse on different background, black vs. gray</p> <p>Color (but 8% of the population are color blind) (also for emphasis)</p> <p>Boxes and other means of grouping</p> |

| | |
|--|--|
| Methods for indicating parts of a document (large or small) | |
| <p>Explicit labels</p> <p>Arrangement</p> <p>Type face</p> | <p>See the examples for different methods of displaying a catalog record in <i>Organizing Information</i>, p. 160 - 161.</p> |

Further elaboration of these principles through a series of

Document design examples

- 1 Two formats for salary data
- 2a Alphabetical vs. meaningful display (Art and Architecture Thesaurus)
- 2b Alphabetical vs. meaningful display (Art and Architecture Thesaurus)
- 3 Examples from the Longman Lexicon of the English Language
- 4 Display of information on buildings on a site in Perseus
- 5 Two displays of the same hierarchy
- 6 Two displays of a catalog record in a public library OPAC (Online Public Access Catalog)

In the Supplement

- 7 Winners and losers in the forecasting game (from Tufte) (supplement)
- 8 Thermal conductivity of tungsten (from Tufte)
- 9 Napoleon's campaign to Russia (from Tufte)
- 10 Classified arrangement of descriptors in a document record for indexing test (Alcohol and Other Drug Thesaurus)
- 11 Contents page from *Alcohol Research*

The syllabus and lecture notes are an example of document design, using boxes, labels, comparative columns, tables showing a concept space that has two dimensions (such as the table in this lecture) and color and striving for consistent format. For example, first pages of lectures follow a common format, so do first pages of assignments. Also running heads as guide posts for orientation in the document.

Example 1. Minimize labeling, enable comparison**Library Jobs by Level, ALA survey 2008***2008 ALA-APA Salary Survey: Librarian – Public and Academic (Librarian Salary Survey)*

| Job title | Average salary |
|---|-----------------------|
| Director/Dean/Chief Officer Public Libraries | 86K |
| Academic Libraries | 95K |
| Deputy/Associative/Assistant Director Public Libraries | 73K |
| Academic Libraries | 80K |
| Dept Head/Branch Mgr/Coordinator/Senior Mgr Public Libraries | 61K |
| Academic Libraries | 61K |
| Manager/Supervisor of Support Staff Public Libraries | 52K |
| Academic Libraries | 54K |
| Librarian Who Does Not Supervise Public Libraries | 48K |
| Academic Libraries | 55K |
| Beginning Librarian Public Libraries | 43K |
| Academic Libraries | 45K |

<http://www.ala-apa.org/salaries/SalarySummary2008.pdf> (Data extracted from Tables 1 and 2)

Same data, different arrangement

| Job title | Public | Academic |
|---|---------------|-----------------|
| Director/Dean/Chief Officer | 86K | 95K |
| Deputy/Associative/Assistant Director | 73K | 80K |
| Dept Head/Branch Mgr/Coordinator/Senior Mgr | 61K | 61K |
| Manager/Supervisor of Support Staff | 52K | 54K |
| Librarian Who Does Not Supervise | 48K | 55K |
| Beginning Librarian | 43K | 45K |

Examples 2a and 2b. Meaningful arrangement**From the Art and Architecture Thesaurus (AAT)**

| | |
|---|---|
| <p><i><size: photograph formats></i></p> <ul style="list-style-type: none"> double whole plate half plate mammoth plate ninth plate quarter plate sixteenth plate sixth plate whole plate <p>Art and Architecture Thesaurus sequence</p> | <p>size: photograph formats</p> <ul style="list-style-type: none"> sixteenth plate ninth plate sixth plate quarter plate half plate whole plate double whole plate mammoth plate <p>Suggested meaningful sequence</p> |
| <p>Figure 1. Alphabetical vs. meaningful sequence on same hierarchical level (Art and Architecture Thesaurus)</p> | |

In the **art genres** example on the next page, notice the advantage of having definitions / scope notes for related terms right next to each other.

| | |
|--|---|
| <p><art genres></p> <ul style="list-style-type: none"> academic art amateur art apocalyptic art art brut children's art commercial art community art SN Includes art undertaken in conjunction with particular communities, often socially deprived, usually with the idea of producing an effect or inspiring response specifically within those communities, with no reference to widely established standards. For art intended to beautify or enrich public places, use public art. computer art court art crafts cybernetic art didactic art dissident art ethnic art fantastic art figurative art folk art funerary art naive art nonrepresentational art primitive art public art SN Use for art whose purpose is to beautify and enrich public places. For art undertaken in conjunction with particular communities, usually to produce an effect or inspire response specifically within those communities, use community art. rock art <ul style="list-style-type: none"> cave art serial art sofa art street art <p style="text-align: center;">a. AAT sequence</p> | <p>art genres</p> <ul style="list-style-type: none"> . art genres by content or other intrinsic characteristics <ul style="list-style-type: none"> . . figurative art <ul style="list-style-type: none"> . . . fantastic art . . . apocalyptic art . . nonrepresentational art . . cybernetic art . . serial art . . crafts . art genres by standard <ul style="list-style-type: none"> . . academic art . . folk art . . dissident art . art genres by type of artist or origin <ul style="list-style-type: none"> . . amateur art . . naive art . . art brut . . children's art . . computer art . . ethnic art . . primitive art . art genres by audience, purpose, or display context <ul style="list-style-type: none"> . . sofa art . . court art . . public art <ul style="list-style-type: none"> . SN Art whose purpose is to beautify and enrich public places. . . . community art <ul style="list-style-type: none"> . SN Public art undertaken in conjunction with particular communities, often socially deprived, usually with the idea of producing an effect or inspiring response specifically within those communities, with no reference to widely established standards. . . . street art . . rock art <ul style="list-style-type: none"> . . . cave art [prehistoric, esp. paleolithic] . . didactic art . . commercial art . . funerary art <p style="text-align: center;">b. Suggested meaningful sequence</p> |
|--|---|

Figure 2. Alphabetical vs. meaningful sequence.

Example from the Art and Architecture Thesaurus (AAT)

Example from Longman

Example 4. Tabular representation of data with same structure**Results of a search for architecture (buildings) whose site is “Amphiaraiion” in the region of Attica (from an old version of Perseus)**

All buildings shown are in the region of Attica

| Name | Summary | Period | Type |
|---|---|------------------|-----------|
| Amphiaraiion, Earlier Temple of Amphiaraios | Small temple; on the western end of the Terrace of Dedications in the Sanctuary of Amphiaraios | Late Clas./Hell. | Temple |
| Amphiaraiion, Klepsydra | Water clock and small annex; southeast of the Sanctuary of Amphiaraios, across the stream and east of the temple of Amphiaraios | Hellenistic | Klepsydra |
| Amphiaraiion, Stoa | Stoa; on the east side of the Sanctuary of Amphiaraios, southeast of Theatre | Late Classical | Stoa |
| Amphiaraiion, Temple of Amphiaraios | Temple; at the Western end of the Sanctuary of Amphiaraios | Hellenistic | Temple |
| Amphiaraiion, Terrace of Dedications | Terrace with retaining wall; on the northwestern side of the Sanctuary of Amphiaraios | Late Classical | Terrace |
| Amphiaraiion, Theater | Hellenistic theater; on the northwestern side of the Sanctuary of Amphiaraios, behind the west half of the Stoa | Hellenistic | Theater |

There is a simple frame for each building; each row shows the frame for a single building, each column is a slot. If all buildings in a display have the same value, we do not need a column for that slot. In the example, the region slot for all buildings has the value Attica. In a display of all theatres, we could omit the column Type but should add a column Region.

Consumer reports contains many tables of test results that are built on the same principle. Features that have the same value for all tested products do not have a column.

Hierarchy from facets, bad example (replaces OCLC Prism)

Hierarchy from facets good example

Example 6. Remove redundancy, have each separate element stand out on its own – easier to scan

Result display in the catalog of the Montgomery County Public Libraries using Sirsi Dynix
www.montgomerycountymd.gov/libtml.asp?url=/content/libraries/find/findbooks.asp

J 599.789 BRE 2006

Giant pandas up close

Bredeson, Carmen.

24 copies available at Aspen Hill Library, Chevy Chase Library, Damascus Library, Davis Library, Fairland Library, Gaithersburg Library, Germantown Library, Kensington Park Library, Noyes Children's Library, Olney Library, Poolesville Library, Potomac Library, Quince Orchard Library, Rockville Library, Silver Spring Library, Twinbrook Library, White Oak Library, and Longbranch Library

Revised design

J 599.789 BRE 2006

Giant pandas up close

Bredeson, Carmen.

24 copies available at

| | | | | | |
|----------------|-------------|---------------|-----------|-------------|--------------|
| Aspen Hill | Chevy Chase | Damascus | Davis | Fairland | Gaithersburg |
| Germantown | Kensington | Park Noyes | Olney | Poolesville | Potomac |
| Quince Orchard | Rockville | Silver Spring | Twinbrook | White Oak | Longbranch |

Each library name should be a link to the page that shows location and opening hours for the library.

Example 6a Arrange citations chronologically to trace the development of an idea in time

As detailed earlier in the section *A Brief Chronicle of Relevance Research*, Cuadra & Katter's ideas were taken up and further developed in the early 1990s when a large group of researchers took a naturalistic approach to studying relevance assessments by real users with a genuine information need in real situations and found a large number of factors beyond topicality which were less studied before, such as novelty, recency, and credibility (Schamber, *et al.*, 1990; Schamber, 1991, 1994; Park, 1992, 1994; Barry, 1993, 1994; Cool, Belkin, & Kantor, 1993; Bruce, 1994; Froehlich, 1994; Hersh, 1994; Janes, 1994; Sutton, 1994; Wang, 1994; Barry & Schamber, 1998; Wang & Soergel, 1998; Wang & White, 1999; Choi & Rasmussen, 2003; Lawley, Soergel, & Huang, 2005; Xu & Chen, 2006). Together these studies identified over 80 "relevance factors" from users, as grouped in an often-cited table by Schamber (1994: p.11). This table lists relevance criteria, types of information used, and other factors affecting relevance judgments.

Review: Notes on the examples

| | |
|--|--|
| <p>1 Importance of resolving anaphoric references in free-text searching with proximity operators p. 122</p> | <p>Searching for the occurrence of two words in the same sentence is an important tool for high-precision searching. But without resolving anaphoric references recall suffers a lot. In the blepharitis example, the second sentence contains <i>eyelids</i> in spirit but not the string "eyelids". Only if the search engine can recognize that <i>they</i> refers to eyelids will the blepharitis entry be found.</p> <p>You should be able to identify the issues in the second example.</p> <p>How does the human reader figure out that in the last paragraph <i>The disease</i> refers to osteoporosis? She applies this heuristic:</p> <p>In a handbook article, if the heading is a disease, then the reference <i>The disease</i> most likely refers to that disease, particularly if it appears at the beginning of a paragraph, so that there is no antecedent disease mentioned in the paragraph. The reader must know that osteoporosis is a disease.</p> |
| <p>2 Extracting substantive data through slot-filling in frames: examples p.124</p> | <p>Read Box 1. The system designer needed to construct a disaster frame intellectually, together with rules or patterns a program could use to fill the slots. Apply the rules given to the story on top of p. 125, and you should get the result as given in the frame instance for Event 345. Such a system is very useful to create a database of disasters gleaned from newspaper stories.</p> <p>This is a very simple system. It makes mistakes, such as: The aftershocks did not kill 10 people, as the system asserts, the main earthquake did</p> |
| <p>3. Extracting data from text, especially importance of resolving anaphoric references. p. 128</p> <p>Lecture 6.1a has an example text with extracted data</p> | <p>The problems are very similar to example 1, and you should be able to figure it out using the bolded and <i>italicized</i> words as your guide</p> |
| <p>4 Importance of recognizing noun phrases, p. 130</p> <p>Lecture 6.1a deals with parsing to detect noun phrases</p> | <p>The examples make clear that noun phrases carry a lot of meaning in English. Take the component words (such as pool, blue, pressure, or program) alone out of context and their meaning thins out into many possibilities. In formulating queries, think about using noun phrases. When a user puts in more than one word, Google looks for the occurrence of one or more noun phrases in what the user entered and boosts the rank of documents that contain the noun phrase rather than just the component words separately</p> |

5 Importance of semantic interpretation, especially disambiguation of homonyms, for retrieval and automated translation¹³²

Word sense disambiguation is a very hard problem; it requires much knowledge. We do it automatically, but computer programs have trouble. How do you know that in "White students" *white* is a race/ethnicity designation, and in "White cars" it is a color. What knowledge do you bring to bear? Answer the same question for the other examples.

The last example is the toughest. It requires knowledge that in the segregationist South drinking fountains were separated by race/ethnicity, so that white refers to the people who were allowed to drink from that fountain, while in Italy there are fountains made of white marble, and so white refers to the color of the object.

Without word sense disambiguation, a free-text search for a word that has many meanings but the user is interested in just one of these meanings will have low precision.

The effect of jokes and comics is often based on word ambiguity (thus they are hard or impossible to translate). Enjoy the examples.

Lecture 6.1a (40 min)*February 22***Document macrostructure, document templates****Inter-document relationships**

→ LIS 506 Information Technology

www.dsoergel.com/571/UBLIS571Lecture06.1.mp3

| | |
|---|---|
| Objectives (in addition to objectives inherited from Lect. 5.2-6.1) | <ol style="list-style-type: none"> 1 Understand the importance of document structure in general and document templates in particular (<i>see practical significance</i>); 2 Understand document type / document template systems with hierarchy and hierarchical inheritance; 3 Be able to design a document template. |
| Practical significance | <ul style="list-style-type: none"> • Document templates make document creation so much easier and thus save a lot of work; • Good document structure makes reading and understanding documents easier; • Good document structure allows for pinpoint retrieval of relevant document sections; • Lesson plan templates facilitate creating and sharing lesson plans. The school library media specialist would be the person to introduce teachers to this idea and find good implementations for the school • Well-structured hypertext / hypermedia allows for reader-directed / learner-directed selection and sequencing of material • If you see that the organization you are working in does not have efficient procedures for creating documents, suggest that they use templates and that they get a system that supports document creation using templates (even MS Word does in a limited way) and perhaps you will be promoted to Chief Information Officer (CIO). |

| | |
|-----------------------------|---|
| Discussion questions | <ol style="list-style-type: none"> 1. How can we design hypermedia systems that support the user in constructing coherent documents? 2. When should sequence be in the writer's hands, and when should it be in the reader's hands? |
|-----------------------------|---|

Document/text macrostructure

Structure of a scientific text - a frame for structuring information (in a full article or in an abstract)

One possible outline

- 1 **Background** (could also be called Problem)
 - 1.1 General problem area (often including a review of the literature)
 - 1.2 Specific problem. Purpose of the study, question to be answered
- 2 **Methods**
 - 2.1 Discussion of the methods used in the study
 - 2.2 Description of the actual conduct of the study
- 3 **Results**
- 4 **Conclusions:** Relationship to existing body of knowledge. Implications for decision making and/or further research

Knowing this structure makes it easier to read a journal article. Having a template with this structure makes it easier to write a journal article. P. 143 gives an example of an abstract using the general outline. All abstracts in the abstracting journal *Alcohol Research* follow this outline making it easy for the reader to quickly peruse each abstract.

Another list of journal article components

(from a study of the human indexing process, indexers look at these components to find index terms)

| | |
|-------------------------------|---|
| Journal title | Introduction |
| Title | Statement of purpose |
| Author | Materials and methods |
| Author's affiliation | Results and discussion |
| Keywords | Conclusions |
| Abstract | Figures, tables, and plates with captions |
| Table of contents (sometimes) | Acknowledgments |
| | Literature cited |

Next page: Structured abstract from *Alcohol Research*, an extremely well designed abstracting journal.

CONIGRAVE KM

abstract 1049

Conigrave KM, Saunders JB, Reznik RB. Predictive capacity of the AUDIT questionnaire for alcohol related harm. *Addiction* 90 (1995) 1479-1485.

'AUDIT can predict a range of harmful consequences of alcohol consumption'

Background

Drinking problems often are not recognized. Most of the people who become alcohol-dependent do not seek help until their problems are obvious. Late diagnosis is of particular concern because effective and low-cost methods of treating problem drinking at an early stage are now available. In 1989, the WHO published a brief 10-item screening questionnaire, the Alcohol Disorders Identification Test (AUDIT) specifically designed to identify problem drinkers before physical dependence or chronic problems have arisen. AUDIT has been reported to have a sensitivity of 92% and a specificity of 94% in detecting hazardous or harmful alcohol use. This study examined the ability of the AUDIT questionnaire to predict which subjects experience medical or social harm from their drinking.

Methods

Subjects were 350 patients who attended a hospital emergency ward in 1984-1985. They underwent a comprehensive assessment of medical history, alcohol use, dependence and related problems in an interview schedule; the AUDIT questions were interspersed among other items. Biochemical variables measured included γ -glutamyltransferase (GGT) and mean corpuscular volume (MCV). Twenty subjects refused to be contacted after 2-3 years or were excluded because of malignant disease. Thus, a cohort of 330 subjects (212 men, 108 women) was left for the longitudinal study; 250 subjects were interviewed again after 2-3 years. Interviewers were blind to the results of the initial assessment. The AUDIT questions were scored from 0 to 4. Subjects who scored 8 or more were classified as potentially hazardous drinkers. AUDIT was examined for its ability to predict a number of end-points including alcohol-related medical disorders, health care utilization, social problems and hazardous drinking at the time of follow-up.

Results

Of those who scored 8 or more on AUDIT at the initial interview, 61% experienced alcohol-related social problems compared with 10% of those with lower scores. They also reported more frequently alcohol-related medical disorders and hospitalization. The AUDIT score was a better predictor of social problems and of hypertension than laboratory markers. Its ability to predict other alcohol-related illnesses was similar to the laboratory tests, but GGT was the only significant marker of mortality.

Conclusions

AUDIT is a brief and convenient questionnaire which can readily be incorporated into the standard medical history. It can predict a range of harmful consequences of alcohol consumption. AUDIT should prove a valuable tool in screening for hazardous and harmful alcohol use so that intervention can be provided to those at particular risk of adverse consequences.

K. M. Conigrave, Centre for Drug and Alcohol Studies, Royal Prince Alfred Hospital, Missenden Road, Sydney, NSW 2050, Australia.

Preview of document templates: A simple mail merge example

The main document: A form letter, a specific case of a template

| | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|
| LITTLE PEOPLE SCHOOL | | | | | | | | | |
| May 15, 2012 | | | | | | | | | |
| «NamePrefix» «FirstName» «LastName» | | | | | | | | | |
| «Street» | | | | | | | | | |
| «City», «State» «ZipCode» | | | | | | | | | |
| Dear «NamePrefix» «LastName», | | | | | | | | | |
| <p>According to our records, «StudentFirstName» does not have a current Emergency Card on file at our school. Because this form is essential to «GenderPossessive» safety while at the Lourie Center, «StudentFirstName» will not be allowed to go on the field trip without it. I have enclosed a copy of this form for you to fill out and return as soon as possible. Please call me if you have any questions or need help with this in any way.</p> | | | | | | | | | |
| Sincerely, | | | | | | | | | |
| Administrative Assistant Little People School | | | | | | | | | |
| Enclosure | | | | | | | | | |

Data source: A MS Access table

| AddressTable | | | | | | | | | |
|--------------|-------------|-----------|----------|-----------------|-------------|-------|----------|-------------------|-------------------|
| ID | Name Prefix | FirstName | LastName | Street | City | State | Zip Code | StudentFirst Name | Gender Possessive |
| 1 | Mr. | Eric | Smith | 504 Flower Ct | Springfield | VA | 22151 | Rebecca | her |
| 2 | Mrs | Elizabeth | Kain | 4801 Thames St. | Springfield | VA | 22151 | Alexander | his |
| 3 | Dr. | Sylvia | Campbell | 3708 Duke St. | Alexandria | VA | 22304 | Mary | her |

Resulting letters: See facing page and next page

The example shows how using a template, in this case a form letter, can save a lot of work. The form letter includes variables or slots whose values are filled in from the database. **Filling in data from a database wherever possible is key to efficient document creation**

LITTLE PEOPLE SCHOOL

May 15, 2012

Mr. Eric Smith
504 Flower Ct
Springfield, VA 22151

Dear Mr. Smith,

According to our records, Rebecca does not have a current Emergency Card on file at our school. Because this form is essential to her safety while at the Lourie Center, Rebecca will not be allowed to go on the field trip without it. I have enclosed a copy of this form for you to fill out and return as soon as possible. Please call me if you have any questions or need help with this in any way.

Sincerely,

Administrative Assistant
Little People School

Enclosure

LITTLE PEOPLE SCHOOL

May 15, 2012

Mrs Elizabeth Kain
4801 Thames St.
Springfield, VA 22151

Dear Mrs Kain,

According to our records, Alexander does not have a current Emergency Card on file at our school. Because this form is essential to his safety while at the Lourie Center, Alexander will not be allowed to go on the field trip without it. I have enclosed a copy of this form for you to fill out and return as soon as possible. Please call me if you have any questions or need help with this in any way.

Sincerely,

Administrative Assistant
Little People School

Enclosure

LITTLE PEOPLE SCHOOL

May 15, 2012

Dr. Sylvia Campbell
3708 Duke St.
Alexandria, VA 22304

Dear Dr. Campbell,

According to our records, Mary does not have a current Emergency Card on file at our school. Because this form is essential to her safety while at the Lourie Center, Mary will not be allowed to go on the field trip without it. I have enclosed a copy of this form for you to fill out and return as soon as possible. Please call me if you have any questions or need help with this in any way.

Sincerely,

Administrative Assistant
Little People School

Enclosure

Note

Mail merge per se is not a topic in 571, just used as an example of document templates. If for some other purpose you are interested in learning about mail merge, here is a useful introduction:
http://extension.oregonstate.edu/esoc/ectu/services/lessons/documents/MailMerge_000.pdf

I am also happy to send you the files I used for this example upon request.

Example. A simple document system

A frame/object hierarchy of document templates and documents

A document template is a frame with a slot (or element) for each part of the document (a part can be a single line or part of a line). Many slots have a procedure attached; the procedure obtains the information from a database, if it is available, or displays a menu of possible values, or asks the user a question. The document templates are arranged in a hierarchy, so that the slots in common to all documents of a class, such as meeting announcements, need to be specified only once; these slots then inherit down to all descendants of the class.

Lecture 6.1b deals with implementing document templates in XML using XML schemas or the older Document Type Definition (DTD)

The simple document system consists of just five document types arranged in a hierarchy:

- Generic memo
 - . Sales report memo
 - . . Content management sales report memo
 - . . Customer relations management sales report memo
 - . Self-assessment memo

For each document type, we give the template and a sample document. Conventions used:

| | |
|----------------|---|
| Bold | A template slot (or element) |
| Arial | An instruction to be carried out when the template is applied to produce a document. Usually these instructions are attached to a slot. |
| <variable> | A variable to filled in with the appropriate value by the system |
| Courier | Text or data filled in by the system or selected by the user from a menu of options displayed by the system |
| Times Roman | Text entered by the user |
| <i>Italics</i> | Comments/explanations (not part of the document) |
| [], [[]] | Inherited, from one level up, two levels up Inheritance is indicated separately for the slot and the content of the slot (the slot may be inherited from the level above, yet the content can be specified at the current level) |
| /* ... */ | Comment |

Top: examples Bottom: explanation

Document template 1: Generic memo

| | |
|---|---|
| Subtype of / child of / inherits from: | Top level |
| Has subtypes / children / inherits to: | Sales report memo, Self-assessment memo |
| Metadata | |
| To: | |
| From: | <name of person signed on to system>, <title of person> |
| Subject: | |
| Date: | <today's date> /* from computer's clock */ |
| Keywords: | |
| URI: | <Universal Resource Identifier> /* to be filled in by system */ |
| MemoBody | |
| PlainText: | |

Picture a company that makes software for business; *content management* and *customer relations management* are two of their product lines. (Just as a matter of interest. customer relations management software supports keeping customers happy, keeping track of their purchases and complaints, suggest new products they might want to buy, and managing email correspondence. One feature of such systems is this: Incoming email is analyzed for its general tone S friendly, laudatory, neutral, angry; an angry email message is moved up the queue for answering.

Each of the product lines has a manager in charge.) The director of sales needs to monitor sales in all areas, and so she must be updated every month on the sales in every product line and problems and opportunities. Having a sales memo template for each product lines greatly reduces the time it takes a product line manager to produce these memos every month.

This lecture introduces such a sales memo template step by step, starting from a plain memo template that should be familiar to you from sending email. Every time you send an email message you save time since the system fills in

- 1 your name and email address under From and
- 2 the date.

Filling in data from a database wherever possible is key to efficient document creation.

Continued on facing page.

Document example 1: Generic memo

To: Sue Feldman, CIO (*Chief Information Officer*)
From: Bob Boiko, content management specialist
Subject: What XML (eXtensible Markup Language) can do for us
Date: February 7, 2001
Keywords: XML; content management; document structure; databases on the Web
URI: www.jasca.com/bboiko/memo20010207-04

XML allows us to define document structures that will make it easier to create documents. Once a document is created, it can be displayed in many different ways (Web page in multiple formats, print, etc.) through applying style sheets (the simple Cascading Style Sheets, CSS2, or the more powerful eXtensible Stylesheet Language for document Transformation, XSLT). A table of contents can be created automatically. Moreover, the document can be displayed selectively using just the parts most appropriate for a given audience. Parts of one document can be reused in another document. In retrieval, specific parts of the document can be targeted; for example, a user could search for just the *results* section of scientific reports.

With XML we can also define documents that hold database records to present databases on the Web. The boundary between text documents and formatted databases becomes blurred.

In the examples I use a hypothetical document management system with templating functionality. There are many real systems that can do most of the things described in the examples. The syntax is different for each system.

< > indicates a value to be filled in by the system, from the computer's clock or a database

/* */ indicates a comment that explains a feature of the template

The memo above is an example of a memo created using the template. Note the different typeface for values filled in by the computer system.

Read the text of the memo; it tells you something useful.

Document template 2: Sales report memo

| | |
|---|--|
| Subtype of / child of / inherits from: | Generic memo |
| Has subtypes / children / inherits to: | Content management sales report memo Customer relations management sales report memo |
| [Metadata] | |
| [To:] | <name of director of sales>, <value = “Director of Sales”> |
| [From:] | [<name of person signed on to system>, <title of person>] |
| [Subject:] | /* to be filled in by memo designer of child template */ <last_month> |
| [Date:] | [<today’s date> /* from computer’s clock */] |
| [Keywords:] | |
| [URI:] | [<Universal Resource Identifier> /* to be filled in by system */] |
| [MemoBody] | |
| [PlainText:] | |
| Sales data table: | header <value = “Sales”> <last_month> <value = “in \$1,000”> Run query /* query to be filled in by designer of child templates */ |
| Data analysis: | |
| Recommendations: | |

Creating document templates for the many kinds of documents in an organization is in itself laborious. Using the generic memo template as a starting point for more specific memos saves work. Slots inherit down from the generic memo to more specific memos, and so on in a hierarchy of document types and their associated templates. A prime example of hierarchical inheritance. This system has the further advantage that all memos have common features making it easier for the reader.

- [] around a slot name indicates that the slot is inherited from the next level up
- [] around a slot filler indicates that the slot filler is inherited from the next level up

Sometimes only a slot inherits, sometimes a slot with its value. For example, in all sales report memos the To: slot is filled with the director of sales as found in the personnel database. The clause <value = “Director of Sales”> means that the string “Director of Sales” is to be displayed here in every memo created using this template

The company in our example has many different products; for each of these a monthly sales report memo is needed. Document template 2 is merely a stepping stone to the more specific sales memo templates that follow. Defining a template that includes everything that is in common to all sales memos makes it easier to define the specific sales memo templates and makes sure that sales memos for all products share a common structure.

The features of this template are explained in the following examples where there is a sample report with text

Again, there is no document example. People just use this template to make more specific templates with values for their specific sales report already filled in, as in the template for Content management sales report memo. Making these specific templates is much easier if one can start from the more general sales memo template .

Document template 3: Content management sales report memo

| | |
|---|--|
| Subtype of / child of / inherits from: | Sales report memo |
| Has subtypes / children / inherits to: | No children |
| [[Metadata]] | |
| [[To:]] | [< <i>name of director of sales</i> >, <value = “Director of Sales”>] |
| [[From:]] | [[< <i>name of person signed on to system</i> >, < <i>title of person</i> >]] |
| [[Subject:]] | <value = “Content management sales report”> [< <i>last_month</i> >] |
| [[Date:]] | [[< <i>today’s date</i> > /* from computer’s clock */] |
| [[Keywords:]] | <value = “content management software”> |
| [[URI:]] | [[< <i>Universal Resource Identifier</i> > /* to be filled in by system */] |
| [[MemoBody]] | |
| [[PlainText:]] | |
| [Sales data table:] | [header <value = “Sales”> < <i>last_month</i> > <value = “in \$1,000”>] [Run query] “ <u>monthly-CM-sales</u> ” |
| [Data analysis:] | |
| [Recommendations:] | |

Underline: Added to the sales report memo template [[]] inherited from two levels up.

This system can run a database query specified in the template and insert the results into the document. **MS Office can include live database query results from MS Access in a MS Word document.** Again, filling in data from a database wherever possible is key to efficient document creation.

The system takes care of assembling and arranging all the data.

The sales manager can focus on **Data analysis** and **Recommendations** and writing the text

Again: **Templates and inheritance**. A slot defined in a broad template, such as the *generic memo* template, occurs in all subordinate templates, such as the *sales report* and *self-assessment memo* templates. The slot may inherit just as a bare shell for content (only the slot name is enclosed in []) or it may inherit with some or all of its content specifications, such as default value, limitations on values, or a procedure to be used to get the content (slot content specification enclosed in []). For example, the From slot always inherits down with the attached procedure: put in the name of the person signed on to the computer. The To slot inherits as an empty shell; the *sales report* template and the *self-assessment memo* template each has its own procedure for filling in a value. However, from *sales report* to *content management sales report* the To slot inherits with the attached procedure.

Document example 3: Content management sales report

To: Joe Bush, Director of Sales
From: Cindy Weaver, Sales Associate
Subject: Content management sales report January 2001
Date: February 5, 2001
Keywords: Content management software
URI: www.jasca.com/rweaver/memo20010210-13

Sales January 2001 in \$1,000

| | | Fed. Gov. | State & local | Fortun e 500 | Small comp. | Total |
|-------------------|------------------|----------------------|------------------------------|-------------------------|------------------------|--------------|
| TeamSite | Dec. 2000 | 500 | 150 | 700 | 200 | 1,550 |
| | Jan. 2001 | 700 | 200 | 900 | 300 | 2,100 |
| Templating | Dec. 2000 | 250 | 30 | 350 | 50 | 680 |
| | Jan. 2001 | 350 | 40 | 450 | 75 | 915 |
| Metatagger | Dec. 2000 | 100 | 20 | 200 | 30 | 350 |
| | Jan. 2001 | 150 | 30 | 250 | 50 | 480 |
| Metafinder | Dec. 2000 | 100 | 10 | 130 | 30 | 270 |
| | Jan. 2001 | 80 | 0 | 90 | 20 | 190 |
| Total | Dec. 2000 | 950 | 210 | 1,380 | 310 | 2,850 |
| | Jan. 2001 | 1,280 | 270 | 1,690 | 445 | 3,685 |

Data analysis:

Smaller organizations make proportionately less use of Templating. Conversations with some customers showed that they do not have the expertise to construct sophisticated templates that would bring great efficiency to their work.

Sales of metafinder are languishing.

Recommendations:

Offer training in the use of Templating and also a consulting service where the consultant would set up the templates for use by the organization's staff.

Promote Metafinder more aggressively through demonstrations of search improvements achieved through its spelling correction and thesaurus lookup features. Also offer a large generic thesaurus with the software so that an organization does not have the expense of constructing its own thesaurus from scratch.

Document template 4: Customer relations management sales report memo

| | |
|---|--|
| Subtype of / child of / inherits from: | Sales report memo |
| Has subtypes / children / inherits to: | No children |
| [[Metadata]] | |
| [[To:]] | [<name of director of sales>, <value = “Director of Sales”>] |
| [[From:]] | [[<name of person signed on to system>, <title of person>]] |
| [[Subject:]] | <value = “ <u>Customer relations management sales report</u> ”> [<last_month>] |
| [[Date:]] | [[<today’s date> /* from computer’s clock */] |
| [[Keywords:]] | <value = “ <u>Customer relations management software</u> ”> |
| [[URI:]] | [[<Universal Resource Identifier> /* to be filled in by system */] |
| [[MemoBody]] | |
| [[PlainText:]] | |
| [Sales data table:] | [header <value = “Sales”> <last_month> <value = “in \$1,000”>] [Run query] “ <u>monthly_CRM_sales</u> ” |
| [Data analysis:] | |
| [Recommendations:] | |

Underline: Added to the sales report memo template

[[]] inherited from two levels up.

Another example just to reinforce how this template system works.

Document example 4: Customer relations management sales report

To: Joe Bush, Director of Sales
From: James Barry, Sales Associate
Subject: Customer relations management sales report January 2001
Date: February 5, 2001
Keywords: Customer relations management software
URI: www.jasca.com/jbarry/memo20010210-13

Sales January 2001 in \$1,000

| | | Fed. Gov. | State& local | Fortun e 500 | Small comp. | Total |
|------------------|------------------|----------------------|-----------------------------|-------------------------|------------------------|--------------|
| Product 1 | Dec. 2000 | 500 | 150 | 700 | 200 | 1,550 |
| | Jan. 2001 | 700 | 200 | 900 | 300 | 2,100 |
| Product 2 | Dec. 2000 | 250 | 30 | 350 | 50 | 680 |
| | Jan. 2001 | 350 | 40 | 450 | 75 | 915 |
| Product 3 | Dec. 2000 | 100 | 20 | 200 | 30 | 350 |
| | Jan. 2001 | 150 | 30 | 250 | 50 | 480 |
| Product 4 | Dec. 2000 | 100 | 10 | 130 | 30 | 270 |
| | Jan. 2001 | 80 | 0 | 90 | 20 | 190 |
| Total | Dec. 2000 | 950 | 210 | 1,380 | 310 | 2,850 |
| | Jan. 2001 | 1,280 | 270 | 1,690 | 445 | 3,685 |

Data analysis:

Smaller organizations make proportionately less use of Product 2. Conversations with some customers showed that they do not have the expertise or training that would allow them to utilize the software.

Sales of Product 4 are very low.

Recommendations:

Offer customer training and a consulting service.

Promote Product 4 through demonstrations, perform user studies and solicit feedback.

Document template 5: Self-assessment memo

| | |
|---|---|
| Subtype of / child of / inherits from: | Generic memo |
| Has subtypes / children / inherits to: | No children |
| [Metadata] | |
| [To:] | <supervisor of person signed on to the system>, <title of supervisor> |
| [From:] | [<name of person signed on to system>, <title of person>] |
| [Subject:] | <value = “Self-assessment for year”> <LastYear> |
| [Date:] | [<today’s date> /* from computer’s clock */] |
| [Keywords:] | <some subject keywords filled in from job description in database> |
| [URI:] | [<Universal Resource Identifier> /* to be filled in by system */] |
| [MemoBody] | |
| [PlainText:] | |
| Accomplishments: | header <value = “Accomplishments in year”> <LastYear> |
| Goals: | header <value = “Goals for year”> <ThisYear r> |
| Training needs: | header <value = “Training needs for year”> <ThisYear> |

Another document type with its template.

Using the self assessment template has many advantages, among them:

1. The supervisor can assimilate each of the memos more quickly
2. The system can produce a report that shows just the goals for each employee s the supervisor can compare the goals
3. The system can compile the **Training need** sections from all self-assessment memos (and even use text analysis to sort these training needs into categories. This report will be very useful for the company's training coordinator

Document example 5: Self-assessment memo

To: Sue Feldman, CIO
From: Bob Boiko, content management specialist
Subject: Self-assessment for year 2000
Date: February 7, 2001
Keywords: Content management; planning; XML; intranet; Web site
URI: www.jasca.com/bboiko/memo20010207-07

Accomplishments in year 2000:

Developed a content management master plan.

Started the development of logical templates for the most important document types and implementation through XML document type definitions.

Developed specifications for the acquisition of content management software and selected a vendor.

Goals for year 2001:

Begin implementation of the content management master plan.

Install software and train staff in intranet-based document creation, deployment, and search.

Redesign the company Web site and use new software to streamline deployment of content on the Web

Training needs:

A course in information architecture

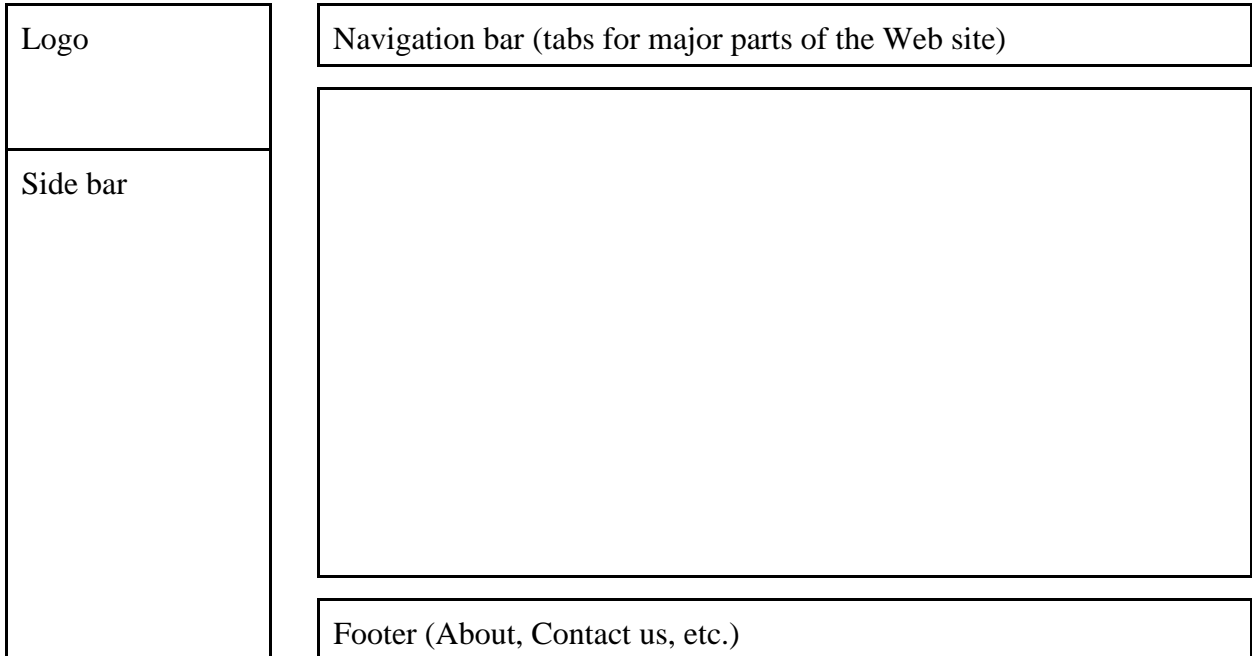
A course in advanced methods in XML, including XLink, XPointer, XPath, and XSLT (eXtensible Stylesheet Language for document Transformation)

Web templates*Just read*

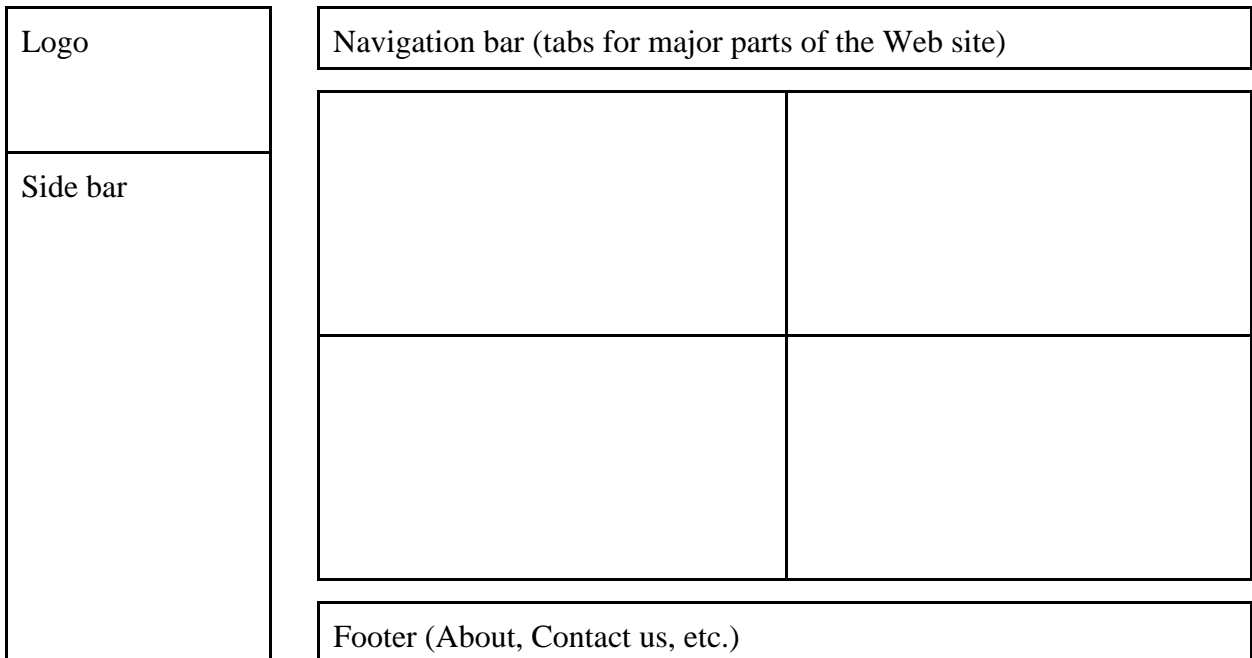
Web templates are very useful for creating and maintaining Web pages. The templates shown have slots and fillers just like the memo templates, but the focus here is on the display area allocated to the data in each slot. The underlying definition is not shown.

When one of these templates is used to create a Web page, each area of the template is linked to a document to be displayed in that area. This may be any multimedia document (text, image, or combination) or it may be a report that is created dynamically from a database. When a document is updated, the update is reflected immediately in all Web pages that link to the document.

Web template 1



Web template 2



Hypermedia/hypertext → LIS 506 Information Technology*Just read*

| | |
|------------------------------------|---|
| Linear text vs. hypertext | <p>Typical text is linear in a sequence set by author: "Begin at the beginning," the King said, very gravely, "and go on till you come to the end: then stop." Lewis Carroll, Alice in Wonderland, Chapter XII</p> <p>Hypertext / hypermedia is a collection of text pieces (and images and sound files) with links; the reader can and often must establish her own order through the text (if indeed the reader goes through the text); this is accomplished by treating the text in blocks (or at least by establishing nodes/locations within the document) and by supplying/permitting links between nodes by which the reader can navigate the text in his or her own order. One could also say that the reader constructs his or her own text. A hypertext can include suggested linear sequences, often indicated by <next> and <previous>.</p> |
| Major features of hypertext | <ul style="list-style-type: none"> • fragmented non-linear text form whose components can be rapidly accessed via machine-supported links/relationships under direction of user • interactive • malleable, modular: it is easy to add or revise small pieces • no strong document boundaries (at least in large hypertexts) |

Hypertext examples

- World Wide Web
- Wikipedia
- A holy book, such as the Koran, the Bible, the Veda, the Vasna, the Tipitaka, or the Book of Mormon (see http://en.wikipedia.org/wiki/Religious_text) (or all of them combined) in hypertext format - links to related verses, to commentary, to dictionary entries
- Fiction examples - choose your own adventure

Inter-document structures

| | | |
|---|---|---|
| Relationships between works often mentioned in cataloging rules | Continuations and sequels Answer key Parodies Critical reviews Concordances | Abstracts Indexes Bibliographies Guides to literature Translation |
|---|---|---|

Lecture 6.1b (20 min)
(Very brief, see →506 for more detail)

February 22

**Formatting documents for interpretation by computer programs.
Document markup languages**

HTML (Hypertext Markup Language) and XML (eXtensible Markup Language)

| | |
|-------------------------------|--|
| Objectives | <ol style="list-style-type: none"> 1 Understand the principles of markup languages and their importance for the implementation of good document design as a basis of further study. 2 Be able to create simple Web pages using HTML markup (for →506) |
| Objectives Elaboration | <p>You will be expected to know the principles of XML, in particular that XML</p> <ul style="list-style-type: none"> • supports the definition of tailor-made templates that divide a document into meaningful sections with all the advantages of templates (Lecture 6.1b); • allows for transforming documents in many different ways for display or export to other systems; • supports a nested structure of document sections <p>You will be expected to know how to use these characteristics of XML strategically in document management in an organization</p> <p>You will not be expected to know the details of how document templates are defined in XML or how an XSLT style sheet for document display or transformation is constructed. (XSLT = eXtensible Stylesheet Language Transformations)</p> |
| Practical significance | <p>Databases of machine-readable text are undergoing an explosion, not only on the Web, but also in intranets and in efforts of creating large text corpora for linguistic and literary studies (the <i>Text Encoding Initiative</i>). Conventions for marking the structure of documents are a prerequisite for creating such databases and for common access and data exchange. Many students will need to access such texts and assist users in the further processing such texts; some students might participate in the setup of text databases.</p> <p>Note: A text corpus is simply a (usually large) body of text in digital form, often with annotations, such as indicating the meaning of each homonym. Examples: The Brown corpus, <i>A Standard Corpus of Present-Day Edited American English</i>, originally created in 1964 at Brown University and updated several times, see http://en.wikipedia.org/wiki/Brown_Corpus Also see www.lancs.ac.uk/fss/courses/ling/corpus/blue/102_1.htm</p> <p>Beyond marking up text, markup languages are now expanded to specify any kind of data structure, blurring the boundary between text and formatted data. There are database management systems that format data with XML-defined tags. The Semantic Web is based on data formatted with XML-defined tags. Under the Linked Open Data initiative (http://en.wikipedia.org/wiki/Linked_data, http://thedatahub.org/) many large data sets have been published on the Web, including library catalogs (for example, http://thedatahub.org/dataset/bluk-bnb) and the LCSUBject Headings (www.oclc.org/research/news/2011-12-14.htm).</p> |

Outline

Brief introduction and basic principles

Definition and general introduction

Principles

HTML, XML, and SGML

Examples

HTML example (simple)

Document with HTML tags (“under the hood”), done directly by the author

Document displayed

XML example

a Template definition

b Document with XML tags

c Style sheet defining appearance

d Document with HTML tags (“under the hood”) produced from XML document

e Document displayed

Note: a-c give XML its power. Not used in the much simpler but less powerful HTML. Style sheets can be used in HTML to display the same HTML document in different ways (Cascading Style Sheet, CSS); not shown in this lecture.

Brief introduction and basic principles*Just read***Definition and general introduction**

| | |
|-----------------------------|---|
| Definition | <p>Markup is the insertion of tags (codes) into a document text or other data stream to specify a structure which can then be used for further processing, in particular for controlling the appearance (or rendering) of a document when it is printed or displayed on a screen.</p> <p>Note: The term <i>markup</i> derives from typesetting. An editor put marks in a manuscript that specified for the typesetter the fonts to be used for a portion of text and other matters of appearance. The meaning of the term has much expanded since then, particularly in the last few years.</p> |
| General introduction | <p>HTML markup tags are designed primarily to direct the display of documents. HTML tags also specify links to other documents to be included automatically at display time (such as images) or available to the user by clicking on the link symbol.</p> <p>Tags defined through XML are much more powerful for expressing document and data structure.</p> |

Principles

| | |
|------------------------|---|
| Physical markup | <p>Tags specify actual appearance properties, such as <i><indent .3"></i>, <i><center></i>, <i><bold></i>, <i></i></p> <p>Problem: What if display device cannot show Times Roman?</p> |
| Logical markup | <p>Tags specify the logical structure of the document, including importance of certain pieces of text. The display is done by a program, possibly in conjunction with style sheets, that renders logical elements in a format determined at output time.</p> <p>Formal (or syntactic) logical elements. Tags specify formal units such as <i><heading level 1></i>, <i><paragraph></i>, <i><numbered list></i>, <i><emphasize></i></p> <p>Content logical elements. Tags specify content units such as <i><from></i>, <i><to></i>, <i><subject></i>, <i><recommendations></i>, <i><warning></i>, <i><methods></i>, <i><conclusion></i></p> <p>defining the content structure of a document. These tags can be used to define record formats even for highly structured data. (XML is used increasingly as a language to define the structure of data in Web-based database applications.)</p> <p>The display program then determines the physical appearance in accordance with the capabilities of the display device and the preferences set by the user. Examples:</p> <p style="padding-left: 40px;">A <i><heading level 1></i> may appear in Times Roman 16 pt bold or in all caps.</p> <p style="padding-left: 40px;">A new <i><paragraph></i> may start with a blank line and no indentation (block style) or without a blank line with the first line indented.</p> <p style="padding-left: 40px;">The document element <i><warning></i> may be displayed in a box with light gray background and a heading Warning.</p> <p>Since logical content markup makes the logical structure of a document explicit, it can be used for information organization and retrieval as well. It can be used to define record formats for straightforward data to be processed by a database management system or to define templates for complex documents (see the examples in Lecture 6.1b). Organizations use markup languages defined in XML to organize large databases of document content, including text and images.</p> |

HTML and XML (and SGML)

| | |
|-------------------------------|--|
| HTML | <p>HTML is a markup language; all tags are predefined. HTML emphasizes logical markup, but the logical elements are primarily formal, and HTML includes an increasing number of physical markup tags (but still not enough to provide tight control over the appearance of a page).</p> <p>An author uses HTML tags to describe the way she wishes the page to display, but parsing and interpretation of the HTML tags is dependent on the Web browser used to display the page. The browser may or may not implement all the features in the same way. For example, look at a complex web page side-by-side with Internet Explorer and FireFox.</p> |
| XML | <p>XML (see the main XML Web site at www.w3.org/XML/) is not a markup language but a language that can be used to define one's own tags, one's own markup language; XML is a markup metalanguage: there are no predefined tags; authors and system administrators define their own tags. Many specific markup languages can be defined using XML. This makes it possible to represent more of a document's semantic structure than HTML does. HTML is one of many markup languages that can be defined in terms of XML.</p> <p>Standards expressed in terms of XML. There are many domains where multiple users have similar kind of documents. They need a format for structuring these documents and for metadata describing them. There are many communities that use XML to define markup languages (domain-specific tagging schemes) for their own domain (with discussion in the whole user community) as a standard to be used by the community; examples are MathML, NewsML, HR-XML for human resource data, etc., financial documents or biological processes (tags for structuring data). This saves thousands of people from having to "reinvent the wheel" for their domain.</p> |
| SGML no longer used | <p>Texts marked up with SGML-defined tags are being migrated to XML</p> <p>SGML (Standard Generalized Markup Language) is a markup metalanguage that was developed primarily by the publishing industry so they could deal more easily with electronic manuscripts. It was created to allow sophisticated and detailed markup for every need of book publishing. The feature richness (or feature excess) made it very difficult for programmers to write practical software for processing SGML documents, so XML started out as a simplified subset of SGML (20% of the complexity, 80% of the functionality). XML has since added features of its own, especially the definition of many data types (such as date and currency) to support databases encoded using XML. Strictly speaking, HTML was defined using SGML as the defining language.</p> <p>See http://en.wikipedia.org/wiki/Standard_Generalized_Markup_Language</p> |

The following gives HTML and XML examples

HTML example (With XML, more steps are required, see XML process diagram)
Should give even a complete novice a sense of the structure of an HTML document.

Document with HTML tags (“under the hood”) (File **d** in the XML process diagram)

Document sections are indicated by a begin tag `< >` and a corresponding end tag `</>`

Here and in the following: **bold** = this is a tag, unless stated otherwise

Read the text; it has useful information

```

<HTML>
  <HEAD>                                /* Note: The <HEAD> section is not displayed. It has metadata*/
    <TITLE>What XML can do for us</TITLE>
    <META NAME="creator" CONTENT="Bob Boiko">
    <META NAME="keywords" CONTENT="XML; content management; document
      structure; databases on the Web">
  </HEAD>
  <BODY>                                /* Note: The <BODY> section is displayed.*/
    <HI><Center> Memorandum </Center></HI>
    To: Sue Feldman, CIO <BR>From: Bob Boiko<BR>
    Date: February 7, 2003<BR><BR>
    Subject: <EM> What XML can do for us</EM>
    <P>XML allows us to define document structures that will make it easier to create
    documents. Once a document is created, it can be displayed in many different ways (Web
    page in multiple formats, print, etc.) through applying style sheets (the simple Cascading
    Style Sheets, CSS2, or the more powerful eXtensible Stylesheet Language for document
    Transformation, XSLT). A table of contents can be created automatically. Moreover, the
    document can be displayed selectively using just the parts most appropriate for a given
    audience. Parts of one document can be reused in another document. . . .</P>
  </BODY>
</HTML>

```

Document displayed by the Web browser under the control of HTML tags (File **e** in the diagram)

Memorandum

To: Sue Feldman, CIO
From: Bob Boiko
Date: February 7, 2003
Subject: **What XML can do for us**

XML allows us to define document structures that will make it easier to create documents. Once a document is created, it can be displayed in many different ways (Web page in multiple formats, print, etc.) through applying style sheets (the simple Cascading Style Sheets, CSS2, or the more powerful eXtensible Stylesheet Language for document Transformation, XSLT). A table of contents can be created automatically. Moreover, the document can be displayed selectively using just the parts most appropriate for a given audience. Parts of one document can be reused in another document. . . .

| | |
|--|--|
| <p>Here: Explanation of XML example</p> | <p>The following pages until the XML example (about four pages) explain the XML example. They are identical to the corresponding audio. So either listen to the audio or take these pages out to look at them with the XML example. You also want to take out the figure XML Documents and Process.</p> |
| <p>This is a short intro to XML</p> | <p>This is a very short introduction to XML and how XML documents are displayed. You should have learned this in LIS 506. Remember from the section HTML and XML (and SGML) above that in XML one can define one's own tags corresponding to meaningful sections of a document; one can define document templates. Often it is possible to copy or modify someone else's template, but here we start from scratch. The figure XML Documents and Process below gives the overall framework; each component is illustrated by an example. You may want to take that page out so you can look at the figure while looking at the examples.</p> |
| <p>Verbal outline of the XML documents and process (below as a diagram)</p> | <p>Each box in this figure is explained in detail with examples on the pages that follow it. So here is just an outline</p> <ul style="list-style-type: none"> a XML on its own has no tags. To create documents (such as the memos in Lecture 6.1a), one must first define appropriate tags; together the tags specify a <i>document template</i>. b Now many <i>document instances</i> (individual documents in which the template slots are filled with text, images, or sound) can be produced using the tags defined. c But how to display these document instances, or documents for short? The browser does not understand the tags we defined. So the document must be transformed into an HTML document that the browser will display in the desired appearance. The instructions for this transformation are given in an XSLT <i>style sheet</i>. An XSLT processing program can read the instructions in the XSLT style sheet and apply them to a document instances. A document instance includes the location of the style sheet d XSLT (= eXtensible Stylesheet Language Transformations) e The result is a document with HTML tags, which the browser then displays on the screen |
| <p>Document a, XML schema definition</p> | <p>On the page following Document a there is a different version (with pictures) of what is explained in this box. Read one or read the other, or read them both, or read this text while looking at the pictures. An XML document consists of <i>elements</i> that are nested into each other; each element is enclosed in begin tag <code><someName></code> end tag <code></someName></code>. See the figure on the page following Document a.</p> |

| | |
|--|---|
| <p>Document a, XML schema definition, continued</p> | <p>The outermost element is the document itself. Elements have a name <i>S</i> that will be the name of the tag. Elements have a <i>type</i> according to their internal structure. The document we are defining has nested in it an element of type <i>metadataContent</i>, which in turn has nested in it elements <i>to</i>, <i>from</i>, <i>subject</i>, <i>date</i>, and <i>keyword</i>. The definitions of all document's elements make up a document template, also called document schema or document type. The schema definition is a special kind of XML document using tags predefined by the XML designers and prefixed by <i>xsd</i>: (XML Schema Definition). An XML processor reads an XML schema document and stores an internal definition of the document template</p> <p>Document a is such a schema definition as indicated in line 2. The third line is a comment <code><! ... ></code>.</p> <p>The following lines name elements, defined from the outside in, starting with the most encompassing element, the document itself. We first give the element a name and assign it a type</p> <pre><xsd:element name="memo" type="memoContent"/></pre> <p>The element has a name (label, <u>tag</u>) that will be used to identify the element in document instances. The element also has an internal structure or content. This internal structure or content is defined through the element's <i>type</i>. The <i>type</i> needs a name; for convenience and ease of understanding, for the <i>type</i> of the element <i>memo</i> we choose the name <i>memoContent</i> because the <i>type</i> specifies the content of the element. (The syntax is not important; the "element tag" is begin and end at the same time, so it encloses no text but has information inside it, called "parameters", in this case "name" and "type")</p> <p>We now need to define for the system what we mean by <i>memoContent</i>. The definition is enclosed in</p> <pre><xsd:complexType name="memoContent"> ... </xsd:complexType></pre> <p>The structure or content of <i>memo</i> consists of a <i>sequence</i> of two elements, named <i>metadata</i> and <i>memoBody</i>. Of course, for each of these elements we need to indicate its content by specifying a type.</p> <p>The definition of <i>metadataContent</i> follows the pattern you just saw: Its structure is a sequence of five elements, each having a <i>type</i>. There is one difference: Defining <i>types</i> has to stop somewhere, otherwise we would end up in an infinite regress. Luckily, the XML designers have pre-defined some basic types as part of the language. We use the basic <i>types</i> <i>string</i> and <i>date</i>; you can guess from the names what they mean.</p> <p>You should be able to follow the rest of the schema definition.</p> |
|--|---|

| | |
|---|---|
| <p>Document b</p> | <p>This is simple, a document instance using the tags we just defined (which together make up a document template). In a real system there would be many such document instances.</p> <p>The first two tags are marked as special by the ?. They tell the version of XML used and what style sheet to use for display.</p> <p>The first regular tag</p> <pre><memo xmlns="www.jasca.com/cm/memo.xs"></pre> <p>tells that this is document instance of memo. <i>xmlns</i> stands for XML Name Space; this is a file that contains the schema definition a; it is the space where all the tags used in the memo instance are defined</p> <p>Digression: Elaboration of the name space idea.</p> <p>Everybody and their brother can define XML tags. Imagine the mess on the World Wide Web where billions of documents are posted containing tags defined by different people meaning different things but given the same name. So tag definitions are divided into name spaces; within a name space there can be only one tag associated with a name. The full name of a tag is namespace:localname; the full names are unique across the Web. For convenience, for a document instance, a name space can be declared at the beginning and all tag names are assumed to be from that namespace, unless explicitly given as full names.</p> <p>Note: The users producing document instances do not need to worry about tags. There are systems that display an input form, the user enters information in the appropriate places, and the system takes care of the XML, much like similar programs for defining Web pages.</p> |
| <p>More on the structure of XML documents</p> | <p>The page facing Documents a and b explains the XML schema definition further, using a diagram</p> <p>The next page shows Document b shown as an arrangement of of nested boxes. This presentation makes it easier to follow the XSLT style sheet (Document c)</p> |
| <p>Document c XSLT style sheets made easy</p> <p>Document c</p> | <p>Now comes the hard part. To display this document in a browser we need to transform it into a HTML document. This is done through the XSLT style sheet. Such style sheets can be exceedingly complex, but this one is actually quite simple once you get the hang of it.</p> |

| | |
|--|--|
| <p>XSLT style sheets made easy, continued</p> | <p>An XSLT style sheet is again a special kind of XML document using tags pre-defined by the XML designers in the name space xsl:. A computer program designed to process XSLT style sheets interprets these tags and applies the appropriate transformations to the input XML document (in our case b).</p> <p>The operative part of the style sheet is between <code><xsl:template match=""/></code> and <code></xsl:template></code></p> <p>Any text that is not preceded by <code><xsl:</code> is transferred to the HTML document as is. This text is highlighted in bold. The text in bold is a template (or skeleton) for all document instances of type memo as transformed into HTML documents (just like the form letter in the introductory example).</p> <p>So now we have to worry about getting data into the template (putting meat on the skeleton). In the form letter example, data come from a database; in the form letter each piece of data is identified by the field name used in the data base. Here, the data come from XML documents. The XSLT style sheets specifies the specific piece of data to be extracted from the input document b into the output document d. For example, what text should go after From: The following is the format of an extraction specification: <code><xsl:value-of select="..." /></code></p> <p>In the specific case: <code><xsl:value-of select="memo/metadata/from" /></code></p> <p>memo/metadata/from is called a path, starting from the root of the document going down the hierarchy of nested elements. To see where that path leads just look at ► in the nested boxes representation of Document b and verify that the correct text was inserted into the target Document d (the HTML document).</p> <p>For another example, what text goes between <code><TITLE></code> <code></TITLE></code>? This time the path is <code><xsl:value-of select="memo/metadata/subject" /></code></p> <p>Again, it is not hard to figure out where that path leads and verify that the correct text was inserted into the target document d (the HTML document)</p> <p>Now that we have an HTML document, a browser can display it</p> <p>You can see that by way of an XSLT style sheet we can rearrange the text (and images, and sound) in an XML document any way we want to. Style sheets can be used to transform an XML documents into any file format imaginable. Data can be sorted and processed in many ways. As an example, consider a database of records on foods displayed with a table of content and an index produced based on definitions in an XSLT style sheet; see www.dsoergel.com/571/SYL2003FaLecturesAppendixNew.pdf, p. 13 -22.</p> |
|--|--|

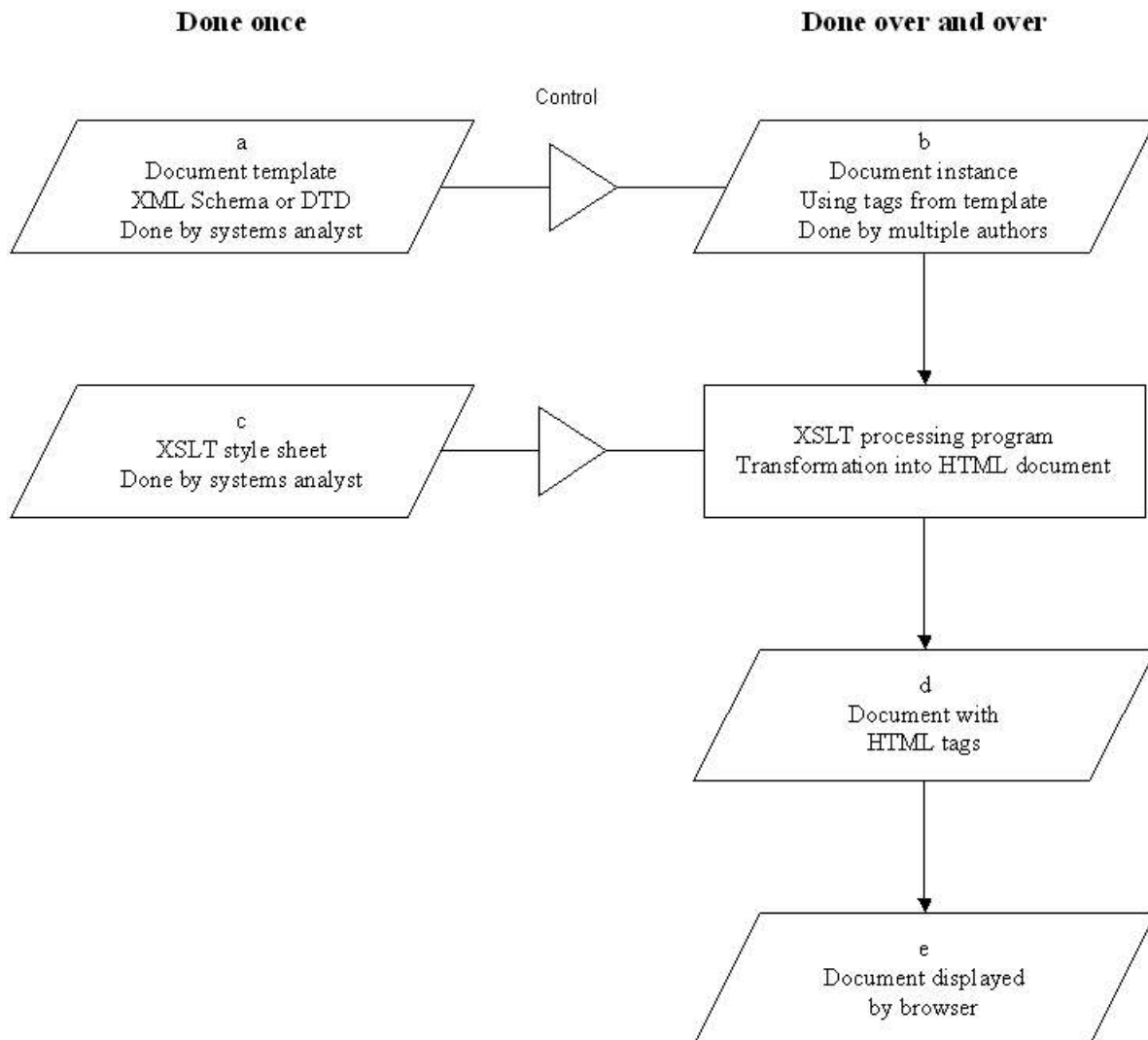
XML example

In XML one must first create an **XML template or schema** which specifies tags for the parts of a document (and thus document structure), in the example for a document type (or class) called

memo. The example assumes that the memo schema is stored at www.jasca.com/cm/memo.xsd. An XML schema is itself a document that follows XML syntax and tags defined by the W3C at the URL www.w3.org/2001/XMLSchema. These tags form a **name space**. To make sure that these tags do not conflict with tags by the same name defined by somebody else, they are prefixed by `xsd:` in the example (`xsd:` is declared as the prefix for the name space defined at the URL).

Most document structure definitions still use a document type definitions (DTD), but XML schemas are more powerful and will replace DTDs. The XML schema syntax is defined in the W3C Recommendation XML Schema (approved May 2, 2001).

XML documents and process



a Definition of template for document *memo* (done once by systems analyst, defines tags)

An XML schema defines a document structure and identifies each element of the structure by a tag. This XML code creates a **memo template or schema**. The documents in the memo class must contain one top-level element, *memo*, which in turn consists of two subordinate elements, *metadata* and *memoBody* (exactly one of each in this order), which in turn contain subordinate elements.

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="www.w3.org/2001/XMLSchema">
  <!-- w3 schema file defines an XML name space; we use prefix xsd. -->
  <xsd:element name="memo" type="memoContent"/>
  <xsd:complexType name="memoContent">
    <xsd:sequence>
      <xsd:element name="metadata" type="metadataContent"/>
      <xsd:element name="memoBody" type="memoBodyContent"/>
    </xsd:sequence>
  </xsd:complexType>
  <xsd:complexType name="metadataContent">
    <xsd:sequence>
      <xsd:element name="to" type="xsd:string"/>
      <xsd:element name="from" type="xsd:string"/>
      <xsd:element name="subject" type="xsd:string"/>
      <xsd:element name="date" type="xsd:date"/>
      <xsd:element name="keywords" type="xsd:string"/>
    </xsd:sequence>
  </xsd:complexType>
  <xsd:complexType name="memoBodyContent">
    <xsd:sequence>
      <xsd:element name="plainText" type="xsd:string"/>
    </xsd:sequence>
  </xsd:complexType>
</xsd:schema>
```

b A document instance of type *memo* (done over and over by authors, uses tags defined in memo template)

```
<?xml version="1.0"?>
<?xml:stylesheet type="text/XSLT"
  xlink:href="www.jasca.com/cm/memo.xslt"?>
<memo xmlns="www.jasca.com/cm/memo.xs">
  <metadata>
    <to>Sue Feldman, CIO</to>
    <from>Bob Boiko</from>
    <subject>What XML can do for us</subject>
    <date>February 7, 2003</date>
    <keywords>XML; content management; document structure; databases on the Web</keywords>
  </metadata>
  <memoBody>
    <plainText>XML allows us to define document structures that will make it easier to create
    documents. Once a document is created, it can be displayed in many different ...</plainText>
  </memoBody>
</memo>
```

Explanation of the structure of a document using XML-defined tags

A document structured using XML-defined tags consists of a hierarchy of nested boxes called **elements**. Each box has

a **tag** that labels the box

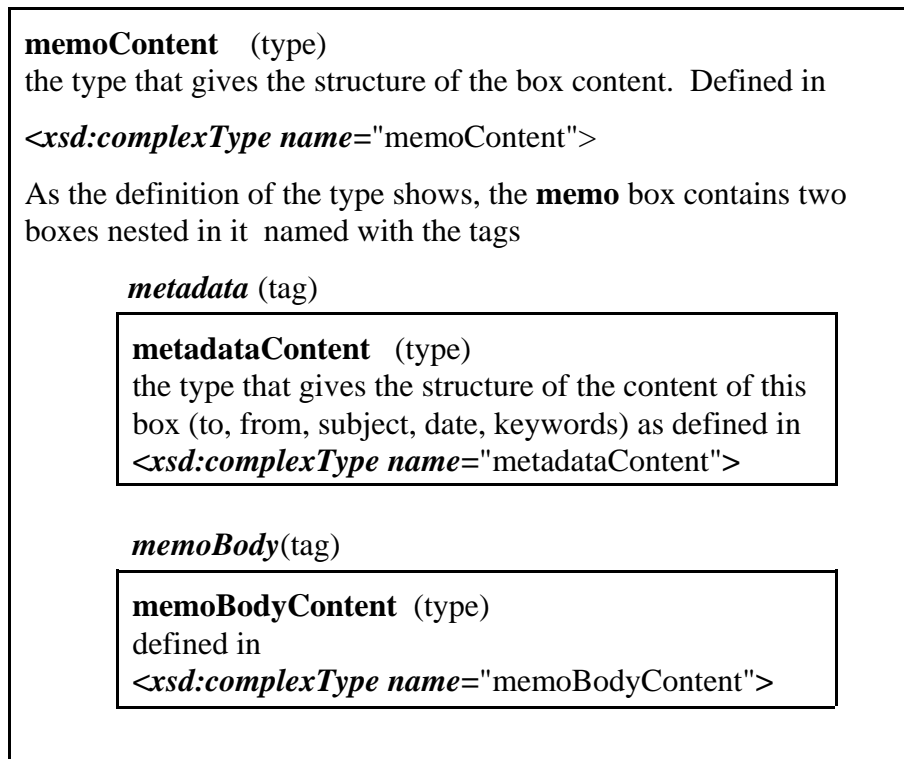
a **type** that defines the content of the box

`<xsd:element name="memo" type="memoContent"/>`

defines a box (element) with the name (tag) "memo"

The content of that box follows the pattern or structure of the type "memoContent"

memo (name or tag of the box)



`<xsd:complexType name="memoContent">`

`<xsd:sequence>`

`<xsd:element name="metadata" type="metadataContent"/>`

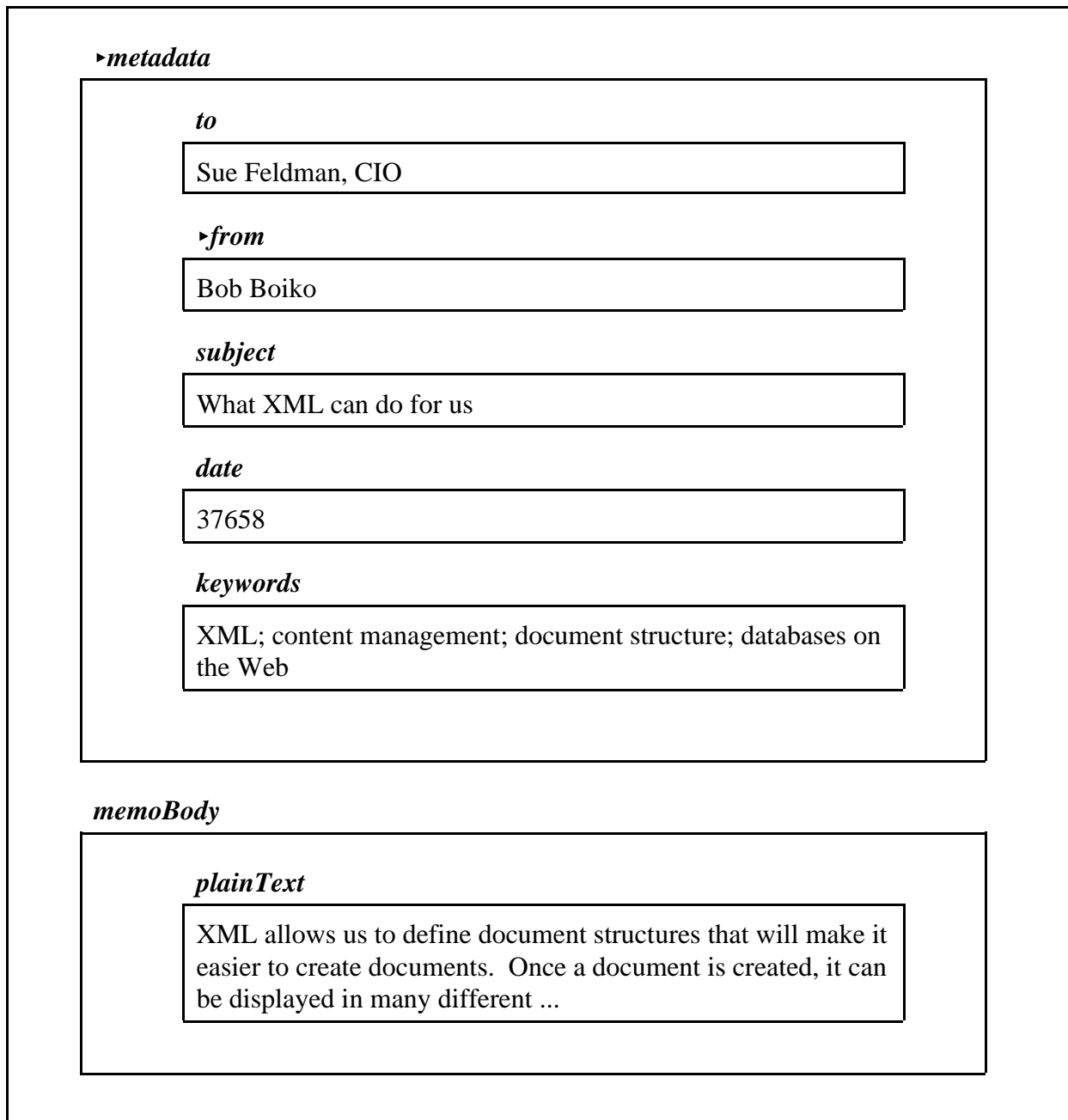
`<xsd:element name="memoBody" type="memoBodyContent"/>`

`</xsd:sequence>`

`</xsd:complexType>`

The sample document shown as nested boxes

▶ A "path" down to a specific element

▶ *memo*▶ **Path:** memo/metadata/from = Bob Boiko

The XSLT style sheet defines a template – the bolded text which goes in all HTML documents produced by this style sheet

To fill each slot in the template, the XSLT stylesheet uses paths to pick out from the whole document the element that goes into the slot.

c XSLT style sheet (Controls the display of the document.) An XSLT processor program uses XML tags to identify pieces of data.. Determines selection of data to be displayed, their arrangement, and the appearance of each element. In the example, the output is an HTML document. But many other types of formatting are possible, e.g., to Wireless Markup Language (WML) for display on a handheld device Here: **Bold** = text that goes as is from the XSLT style sheet into the formatted output document.

```
<xsl:stylesheet
  xmlns:xsl="www.w3.org/TR/WD-XSLT"
  xmlns="www.w3.org/TR/REC-html40">
  <xsl:template match="/">
    <HTML>
      <HEAD>
        <TITLE><xsl:value-of select="memo/metadata/subject"/></TITLE>
        <META NAME="creator" CONTENT="{<xsl:value-of select="memo/metadata/from"/>}" />
        <META NAME="keywords"
          CONTENT="{<xsl:value-of select="memo/metadata/keywords"/>}" />
      </HEAD>
      <BODY>
        <H1><Center> Memorandum </Center></H1>
        To: <xsl:value-of select="memo/metadata/to"/><BR/>
        From: <xsl:value-of select="memo/metadata/from"/><BR>
        Date: <xsl:value-of select="memo/metadata/date"/><BR><BR>
        Subject: <EM><xsl:value-of select="memo/metadata/subject"/></EM>
        <P><xsl:value-of select="memo/memoBody/plainText"/></P>
      </BODY>
    </HTML>
  </xsl:template>
</xsl:stylesheet>
```

d HTML document (same as above) (done over and over, produced by an XSLT processor program)

```
<HTML>
  <HEAD>
    <TITLE>What XML can do for us</TITLE>
    <META NAME="creator" CONTENT="Bob Boiko">
    <META NAME="keywords" CONTENT="XML; content management; document
      structure; databases on the Web">
  </HEAD>
  <BODY>
    <H1><Center> Memorandum </Center></H1>
    To: Sue Feldman, CIO <BR>
    From: Bob Boiko<BR>
    Date: February 7, 2003<BR><BR>
    Subject: <EM>What XML can do for us</EM>
    <P>XML allows us to define document structures that will make it easier to create
      documents. Once a document is created, it can be displayed in many different ways. . . .
    </P>
  </BODY>
</HTML>
```

e Document displayed by the Web browser under the control of HTML tags, see on p. 166

XML schema for a self-assessment memo (see p. 156-157)

Since a *self-assessment memo* is a specific type of memo, we can define its schema by adding to the *memo* schema; the *memo* schema is **reused**.

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="www.w3.org/2001/XMLSchema"
  xmlns="www.jasca.com/cm/memo.xsd">
  <include xsd:schemaLocation="www.jasca.com/cm/memo.xsd"/>
    <!--This schema includes a definition of the type metadataContenr, which is used
    below. -->
  <xsd:element name="selfAssessmentMemo" type="selfAssessmentMemoContenr"/>
  <xsd:complexType name="selfAssessmentMemoType">
    <xsd:sequence>
      <xsd:element name="metadata" type="metadataContet"/>
      <xsd:element name="memoBody" type="memoBodyContent"/>
    <xsd:sequence>
  </xsd:complexType>
  <!-- redefinition of memoBodyContent -->
  <xsd:complexType name="memoBodyContent">
    <xsd:complexType>
      <xsd:extension base="memoBodyContent">
        <xsd:sequence>
          <xsd:element name="accomplishments" type="xsd:string"/>
          <xsd:element name="goals" type="xsd:string"/>
          <xsd:element name="trainingNeeds" type="xsd:string"/>
        </xsd:sequence>
      </xsd:extension>
    </xsd:complexType>
  </xsd:complexType>
</xsd:schema>
```

Reuse is a big theme in the application of XML; reuse can be implemented in several ways.. There are whole collections of defined data types, such as a data type for US states, with the values restricted to a list of two-letter abbreviations of US states, or data types for US address, UK address, France address, Germany address, etc. (all derived from a generic address as a common parent). These type definitions are collected into *vocabularies*, each in its own name space, from which they can be included in any XML document schema, saving the schema creator a lot of work.

Note: The syntax of the XML examples may not be correct in every detail, but it does give the general idea.

There is supplemental material on XML and RDF at www.dsoergel.com/571/SYL2003FaLecturesAppendixNew.pdf In particular a fully worked out example of using RDF to represent the food data from Lecture 2.2 and using style sheets to create several different outputs (a table of contents, a detailed listing, and an alphabetical index) from this database.

Cataloging and metadata. Bibliographic and record control

(no text chapter)

February 22- February 29

| | |
|-------------------------------|---|
| Objectives | <ol style="list-style-type: none"> 4. Understand the use of metadata for finding and interpreting or using any kind of data source (in some interpretations: any kind of object). 5. Understand the fundamental problems of bibliographic control as an application of general principles of Organization of Information. 6. Understand the problems of defining "document", and the problems of defining the relationships between several versions of a document. 7. Be able to apply this understanding to the analysis and design of cataloging codes and to actual cataloging (consulting the appropriate code for details). 8. Understand the complexities in determining the useful entries for a document. 9. Be able to apply some AACR2 rules for entry. 10. Have a general idea of the use of XML and RDF (Resource Description Framework) for expressing and implementing metadata schemas. |
| Practical significance | <ul style="list-style-type: none"> • Good catalogs are more important than ever, just look at the World Wide Web. It is now fashionable to call data about documents (i.e., data about data-carrying objects) <i>metadata</i>. • The question "what is a document" is important for library catalogs but even more important for electronic records, where different versions of the same document proliferate rapidly, especially on the Web. A Uniform Resource Locator or URL does not identify a document but a file storing a document. Many files in different locations can store the same document, creating a burden on the user. There are efforts to define a Uniform Resource Identifier or URI that would identify a document no matter where it is stored (like an ISBN). A URI identifies a <i>document, intellectual work</i>, a URL identifies a <i>document, physical volume</i> (as defined in Chapter 3). However, a system of URIs, while beneficial to users, introduces many difficulties: Who would assign URIs, using what rules? Who would maintain the database(s) with links from each URI to all its physical volumes (URLs)? What happens if one of these files is slightly modified? • Controlling different versions of a document is important for managing document production and access afterwards, as well as for reasons of legal and historical evidence. • Good catalogs that are widely usable and that can share data require standardized cataloging rules. In the World Wide Web domain, there are efforts to agree on a metadata standard. (A minimal standard, the <i>Dublin Core</i> has already found wide acceptance.) In the domain of geographic data there is the <i>Content Standard for Digital Geospatial Metadata</i> issued by the Federal Geographic Data Committee. |

Note: Component lectures do not have pink sheets.

over

Overview of Lectures 6.2 and 7.1

6.2a **General introduction to metadata**

6.2b **Basic concepts in bibliographic control.**

6.2c&7.1a **The conceptual data schema of a library catalog:**

6.2c **Description** (what data describing a document to capture and how for
format them) (AACR2, Part 1) and

7.1a **Access** (what entries to make under which a document can be retrieved and
what form of names and other entries to use) (AACR2, Part 2)

7.1b **Metadata and other structured data on the Web:**
Dublin Core and **RDF** (Resource Description Framework)

Lecture 6.2a (10 min)*February 22***General introduction to metadata**

The audio elaborates on this section

Metadata – Synonymous with cataloging data or pointer data as defined in Organizing Information, Chapter 2.

Data used to **describe other data** and **give context for other data** for the following purposes:

- retrieval
- assessment
- interpretation and use.

Now a hot topic in the context of the Web. Metadata schemas exist or are being developed for many types of “data containers” in the context of many user communities. See examples below.

Note: There is no intrinsic difference between data and metadata. If data are used for the purpose of retrieving, assessing, interpreting, or using other data, they are used as metadata. The data modeling mechanisms are the same no matter what the data modeled are used for.

A slide of Renato Iannella adapted by DS.

| | |
|---|---|
| # | Metadata will improve information discovery across digital repositories |
| # | metadata schema registries (metadata schema = conceptual schema or record format) |
| # | metadata transmission |
| # | metadata repositories |
| # | metadata mappings (for example, from Dublin Core to MARC) |
| # | semantic interoperability (for example, mapping from DDC to LCC) |

Read the next page to get an idea of a range of metadata schemas

Pick one in accordance with your interests to examine in a bit more depth – go to the website indicated.

More on metadata, OAI XXX under construction
Open Archives Initiative
www.openarchives.org/

Review

Lecture 4.2**Conceptual data schemas and input, storage, and output/presentation formats**

Examples of metadata schemas (many expressed in XML)

After the lecture, look at **one** of the examples marked with * ; use URL given in the email I sent.

Bibliographic data: MARC / AACR2. Dublin Core <http://dublincore.org/> (see below)

* **Text: The Text Encoding Initiative (TEI)**

A comprehensive standard for describing (literary) texts, both metadata and the structure of the actual text, www.tei-c.org/

***Archival data: Encoded Archival Description (EAD)**

www.dlib.org/dlib/november99/11pitti.html

www.loc.gov/ead/

For an example repository using EAD see www.cdlib.org/inside/projects/oac/

***Museum data:**

The CIDOC Conceptual Reference Model (CRM), related to FRBR

Introduction: http://cidoc.ics.forth.gr/docs/cidoc_crm_meeting_Prato-1.ppt

Full specification: http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.0.pdf

***Learning objects (instructional materials):**

The Gateway to Educational Materials (GEM)

www.thegateway.org/about/documentation/metadataElements

Learning Technology Standards Committee of the **IEEE**:

http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf

IMS Global: IMS learning resource meta-data information model. (September 2001)

www.imsproject.org/metadata/

Dublin Core Metadata Initiative. DCMI Education Working Group,

<http://dublincore.org/groups/education/> (not much concrete to see there)

CRP Henri Tudor-CITI: Training Exchange Definition: TED.

www.xml.org/xml/schema/8dbca03a/trainingExchangeDefinition.pdf (July 2002)

Geospatial data: will be covered later. Ask me if you are interested

TV programs: TV Anytime Forum. TV Anytime is a set of specifications for the controlled delivery of multimedia content to a user's personal device (Personal Video Recorder (PVR))

www.tv-anytime.org (includes dealing with metadata);

<http://portal.etsi.org/radio/TVAnytime/TVanytime.asp>

BBC: www.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP050.pdf

European Broadcasting Union (EBU): <http://www.ebu.ch/metadata/pmeta/v0102/xml/>

www.ebu.ch/metadata/pmeta/WIP/ESCORT/ESCORT2006.htm (A faceted classification)

Multimedia: MPEG-7,

"Multimedia Content Description Interface", is a standard for describing features of multimedia content: catalog data (e.g., title, creator, rights), semantic data (e.g., the who, what, when, where information about objects and events) and structural data (e.g., the color histogram - measurement of the amount of color associated with an image or the timbre of a recorded instrument). Builds on AV data representation defined by MPEG-1, 2 and 4. www.mpeg.org/MPEG/starting-points.html

PICS (Platform for Internet Content Selection) properties. www.256.com/gray/docs/pics/

Lecture 6.2b (40 min)**February 22****Bibliographic control. General issues**

| | |
|---------------------------|--|
| Introduction | <p>This lecture deals with control of all kinds of documents (all kinds of materials): Regular books and reports, serials, journal and newspaper articles, organizational records, images, sound documents.</p> <p>New dimension of problem: Electronic documents. Ease of copying and modification, cryptic filenames, and online accessibility of electronic documents create special difficulties.</p> <p>A number of general principles of Organization of Information are applicable to the control of any kind of concrete object or "thought object." Each type of material presents its own challenging problems in applying these general principles.</p> <p>Parts of the thinking on descriptive cataloging and the resulting practices are still valid. Other parts have been made obsolete through the greater power of automated systems.</p> <p>Control is mainly access, but also inventory control, including preservation. A distinction is generally made between description and access, but the two are more closely intertwined than many people realize.</p> |
| What is a catalog? | <p>A catalog is a database that contains identifying/descriptive data about objects, such as books (or, more broadly, documents) or data sets (such as geospatial data sets) or merchandise. The coverage of a catalog may be limited to a given physical collection (the books for which physical copies are held in the library, the merchandise items available from a catalog store); that is, the catalog contains only data referring to objects in the given collection. Often the term <i>catalog</i> is defined in this sense of being tied to a physical collection as distinguished from a <i>bibliography</i>, which may include data about documents no matter where physical copies are held. A <i>union catalog</i> refers to objects in multiple collections.</p> <p>If the objects referred to in the catalog are information sources, the catalog data are used primarily as <i>pointer data</i> (see Section 2.5 of Organizing Information,), now commonly called <i>metadata</i>. However, remember, that the distinction between substantive data (data contributing directly to the problem solution) and pointer data lies not in the nature of the data themselves but rather in their use. If data are used to find other data/information sources, they are used as pointer data. If data are directly applied to problem solution, they are used as substantive data.</p> |

Objectives of the library catalog according to Cutter

1. To enable a person to find a book for which either
 - A. the author
 - B. the title
 - C. the subject
 is known
2. To see what a library has
 - D. by a given author
 - E. on a given subject
 - F. in a given kind of literature
3. To assist in the choice of a book
 - G. as to its edition (bibliographically)
 - H. as to its character (literary or topical)

Problems with Cutter's objectives

- (a) Is the user interested in a particular *book* or in the *work* that is embodied in the book? And what is a *book* anyhow?
- (b) In today's world of electronic access, what is a library?

Need to address these before restating objectives.

Fundamental problem in bibliographic control: What are the units we are dealing with?

Look at the examples on the following pages to get a feel for the problem.

Sample documents illustrating problems in defining bibliographic units and in cataloging.

Facing page, some questions to think about here.

Share observations on the examples on the discussion board.

Examine entries (1) – (10) and think how they are related to each other.

How many works?

Who is the author of (4). When cataloging (4), should Halliday be mentioned? If so, how?

How about (8)? What is the status of the marginal notes?

Knowledge update:

Laurence Kerr Olivier, Baron Olivier, OM [Order of Merit] (22 May 1907 – 11 July 1989) was an English actor, director, and producer. One of the most famous and revered actors of the 20th century,[1] he was the youngest actor to be knighted and the first to be elevated to the peerage.[2] Actor Spencer Tracy said that Olivier was 'the greatest actor in the English-speaking world'.[3][
http://en.wikipedia.org/wiki/Laurence_Olivier

Sample documents illustrating problems in defining bibliographic units

- (1) *The man I killed*, by Michael Halliday (i.e. John Creasey). London: Marx Brothers; 1935.
- (2) *The man I killed*, by Michael Halliday (i.e. John Creasey). Large print edition. London: Society for Assistance to the Blind; 1938.
- (3) *The man I killed*, by Michael Halliday (i.e. John Creasey). Audiotape, read by Sir Lawrence Olivier. New York: Books on Tape; 1966.
- (4) *The man I killed*, play by Christopher Wern, based on the novel by Michael Halliday.
- (5) *The man I killed*, a movie version of the play by Christopher Wern, based on the novel by Michael Halliday. On videotape.
- (6) An individual copy of (1) as originally printed.
- (7) An individual copy of (1), produced by making a copy of (6).
- (8) An individual copy of (1), owned by Sir Lawrence Olivier, with many marginal notes in ink.
- (9) A facsimile edition of (8), published by Marx Brothers.
- (10) *The man I killed, completely revised and with a new ending*, by Michael Halliday (i.e. John Creasey). London: Marx Brothers; 1941.
- (11) A legal document with original signatures
- (12) A copy of the same
- (13) A notarized copy of the same

See facing page. Are these all different works? Or is there an identity preserved through all the changes? If so, what does this identity pertain to? A work?

Continuation of text

In all these examples, we are confronted with two fundamental questions:

- 1 What are the units, the types of entities we must deal with in bibliographic control
- 2 How are bibliographic entities related.

The following pages deal with these questions

Knowledge update: Goethe's Faust

Note: This is an example of "Know the reader". To help you appreciate the complexity of the relationship between bibliographic units I need a rich example. Goethe's Faust, one in a family of many works based on a medieval German legend, is such an example, but I was told that even among university students few Americans know about this work (or Goethe, for that matter) and its context. So now that I know my readers I know that I need to provide some cultural background.

Johann Wolfgang von Goethe 1749 – 1832) was a German writer, pictorial artist, biologist, theoretical physicist, and polymath.[2] He is considered the supreme genius of modern German literature.[3] His works span the fields of poetry, drama, prose, philosophy, and science. **His Faust has been called one of the greatest dramatic works of modern European literature.**[3] His other well-known literary works include his numerous poems, the Bildungsroman Wilhelm Meister's Apprenticeship, and the epistolary novel The Sorrows of Young Werther [recently made into a movie]. <http://en.wikipedia.org/wiki/Goethe>

Goethe's Faust is one of the great works of world literature, Goethe's most famous work and considered by many to be one of the greatest works of [German literature](http://en.wikipedia.org/wiki/German_literature)" http://en.wikipedia.org/wiki/Goethe's_Faust

The work is one of many based on a classic German legend:

Faust or **Faustus** (Latin for "auspicious" or "lucky") is the [protagonist](#) of a classic [German legend](#). Though a highly successful [scholar](#), he is dissatisfied, and makes a [deal with the devil](#), exchanging his soul for unlimited knowledge and worldly pleasures. Faust's tale is the basis for [many literary, artistic, cinematic, and musical works](#). The meaning of the word and name has been reinterpreted through the ages. *Faust*, and the adjective *Faustian*, are often used to describe an arrangement in which an ambitious person surrenders [moral integrity](#) in order to achieve power and success: the [proverbial](#) "[deal with the devil](#)". The terms can also refer to an unquenchable thirst for knowledge.^[1]

Plays and comic puppet theatre loosely based on this legend were popular throughout Germany in the 16th century. <http://en.wikipedia.org/wiki/Faust>

Also, many, many works are based on this legend; to give just two examples, [Marlowe's Doctor Faustus](#) and Gounaud's opera Faust ([http://en.wikipedia.org/wiki/Faust_\(opera\)](http://en.wikipedia.org/wiki/Faust_(opera))). Please look at the long list in http://en.wikipedia.org/wiki/Works_based_on_Faust. Goethe's Faust was translated into many languages (it is read in high school in China) and is in turn the basis for derivative works, including [Randy Newman's musical Faust](#) (1993) and [Kamelot's albums Epica](#) (2003) and [The Black Halo](#) (2005). A bit further removed is Thomas Mann's Doctor Faustus ([http://en.wikipedia.org/wiki/Doctor_Faustus_\(Thomas_Mann_novel\)](http://en.wikipedia.org/wiki/Doctor_Faustus_(Thomas_Mann_novel)))

"a re-shaping of the [Faust](#) legend set in the context of the first half of the 20th century and the turmoil of Germany in that period."

If you think cataloging all these works could be more efficient using hierarchical inheritance, you are right.

Edward T. O'Neill and Diane Vizine-Goetz

Table II. Publishing History for *Guide to Reference Books* (adapted and updated by DS)

| Edition | Date | Authors | Publisher | Title |
|---|-------------|---|---|--|
| | 1902 | Kroeger | Houghton, Mifflin & Company | Guide to the study and use of reference books; a manual for librarians, teachers, and students |
| Title edition | 1904 | Kroeger | American Library Association Publishing Board | Guide to the study and use of reference books; a manual for librarians, teachers, and students |
| 2d ed., rev. and enl. | 1908 | Kroeger, Mudge | American Library Association | Guide to the study and use of reference books |
| 3d ed., rev. throughout and much enlarged | 1917 | Kroeger, Mudge | " | Guide to the study and use of reference books |
| [4th ed.] | 1923 | Mudge | " | New guide to reference books |
| 5th ed. | 1929 | Mudge, Kroeger | " | Guide to reference books |
| 6th ed. | 1936 | Mudge, Winchell | " | " |
| 7th ed. | 1951 | Winchell, Mudge | " | " |
| 8th ed. | 1967 | Winchell, Mudge, Sheehy | " | " |
| 9th ed. | 1976 | Sheehy, Keckeissen, McIlvaine, Winchell | " | " |
| 10th ed. | 1986 | Sheehy (ed.) | " | " |
| 11th ed. | 1996 | Balay (ed.), Carrington, Martin | " | " |
| 12th ed. | in prep. | Kieft | " | Guide to Reference Sources, GRS12 [Online; includes Web sources] www.guidetoreference.org/ http://en.wikipedia.org/wiki/Guide_to_Reference |

Notes: This listing does not include supplements issued between editions. Up to edition 9, the person(s) associated with the work are listed thus: *compiled by*, from edition 10 onward *edited by*.

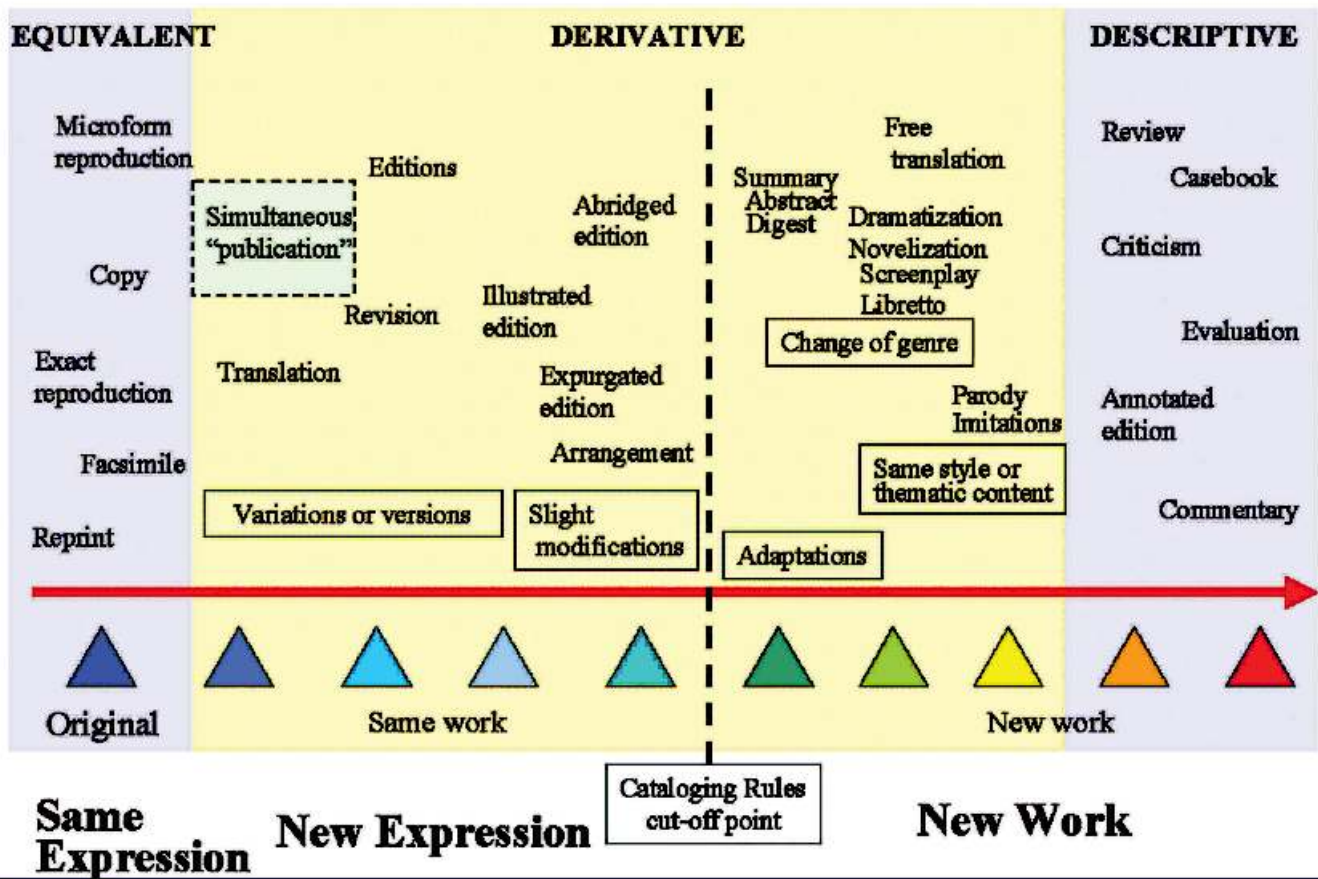
Definition of units in bibliographic and record control

| <p style="text-align: center;">Soergel draft</p> <p>As the most inclusive term that is superordinate to all of the types defined here we will use <i>document</i> (in the broadest sense) or, even broader, <i>resource</i>.</p> | <p style="text-align: center;">FRBR</p> <p style="text-align: center;">Functional Requirements for Bibliographic Records</p> |
|---|--|
| <p>Work</p> <p>Intellectual or artistic entity, as the abstract essence or as a text, image, or piece of music.</p> <p>Range:</p> <p><i>A basic story or theme</i></p> <p style="padding-left: 40px;">the legend of Faust</p> <p style="padding-left: 40px;">the myth of the Great Flood</p> <p><i>A text telling the story, such as</i></p> <p style="padding-left: 40px;">Goethe's Faust</p> <p style="padding-left: 40px;">the account of the Great Flood in the Bible (original Hebrew), a specific version of that text, a Latin version, the account of the same myth in another culture.</p> <p><i>A novel</i> build on the general theme of the legend of Faust but set in a different time with different characters.</p> | <p>Work</p> <p>A distinct intellectual or artistic creation.</p> <hr/> <p>Expression</p> <p>The specific intellectual or artistic form that a work takes each time it is 'realized'</p> |
| <p>Manifestation (also called edition in one meaning of edition)</p> <p>A specific expression or rendering of a work by means of a graphical image or sound, taken in the abstract; the idea of such an expression.</p> <p>Examples:</p> <p style="padding-left: 40px;">The text of Goethe's Faust presented in a particular typeface and layout. (A performance at which the text is recited also renders the text but is a separate, but related, work.)</p> <p style="padding-left: 40px;">A specific score of a given version of Schubert's Fifth. (A performance of that version of Schubert's Fifth also renders the piece of music but is a separate, but related, work)</p> <p>Also the expression or rendering of a work in the form of digital storage that can be transformed to a graphical image or sound, again taken as the abstract pattern of digital signals.</p> | <p>Manifestation</p> <p>The physical embodiment of an expression of a work. As an entity, manifestation represents all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form.</p> <hr/> <p>Printing (not in FRBR, but used)</p> <p>A set of books printed at the same time or printed at different times containing no more than slight variations</p> |
| <p>Item (also individual copy or simply copy or physical copy)</p> <p>The embodiment of a manifestation in a physical object. We can perceive the content of a manifestation only through an individual item (copy) of it (unless we have memorized the contents of a manifestation and can conjure it up from memory). There are works that have only one manifestation of which there is only one item.</p> | <p>Item</p> <p>A single exemplar of a manifestation. The entity defined as item is a concrete entity</p> |

Functional Requirements for Bibliographic Records

| | |
|----------------------|--|
| Work | <p>A work is an abstract entity; there is no single material object one can point to as the work. We recognize the work through individual realizations or expressions of the work, but the work itself exists only in the commonality of content between and among the various expressions of the work. When we speak of Homer's Iliad as a work, our point of reference is not a particular recitation or text of the work, but the intellectual creation that lies behind all the various expressions of the work.</p> |
| Expression | <p>An expression is the specific intellectual or artistic form that a work takes each time it is "realized." Expression encompasses, for example, the specific words, sentences, paragraphs, etc. that result from the realization of a work in the form of a text, or the particular sounds, phrasing, etc. resulting from the realization of a musical work. . . . excludes aspects of physical form, such as typeface and page layout, that are not integral to the intellectual or artistic realization of the work as such. When an expression is accompanied by augmentations, such as illustrations, notes, glosses, etc. that are not integral to the intellectual or artistic realization of the work, such augmentations are considered to be separate expressions of their own separate work(s). Such augmentations may, or may not, be considered significant enough to warrant distinct bibliographic identification.</p> <p>Inasmuch as the form of expression is an inherent characteristic of the expression, any change in form (e.g., from alpha-numeric notation to spoken word) results in a new expression. Similarly, changes in the intellectual conventions or instruments that are employed to express a work (e.g., translation from one language to another) result in the production of a new expression. If a text is revised or modified, the resulting expression is considered to be a new expression.</p> |
| Manifestation | <p>The physical embodiment of an expression of a work. . . . encompasses a wide range of materials, including manuscripts, books, periodicals, maps, posters, sound recordings, films, video recordings, CD-ROMs, multimedia kits, etc. represents all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form.</p> <p>When a work is realized, the resulting expression of the work may be physically embodied on or in a medium such as paper, audio tape, video tape, canvas, plaster, etc. That physical embodiment constitutes a manifestation of the work. . . . Whether the scope of production is broad (e.g., in the case of publication, etc.) or limited (e.g., in the case of copies made for private study, etc.), the set of copies produced in each case constitutes a manifestation. All copies produced that form part of the same set are considered to be copies of the same manifestation.</p> <p>[Changes in physical form result in a new manifestation; examples:] changes affecting display characteristics (typeface, size of font, page layout, etc.), changes in physical medium (e.g., a change from paper to microfilm), changes in the container (e.g., a change from cassette to cartridge as the container for a tape).</p> |
| Item | <p>A single exemplar of a manifestation. . . . a concrete entity. It is in many instances a single physical object (e.g., a copy of a one-volume monograph, a single audio cassette, etc.). There are instances, however, where an item comprises more than one physical object (e.g., a monograph issued as two separately bound volumes, a recording issued on three separate compact discs, etc.). . . . variations may occur from one item to another, even when the items exemplify the same manifestation, where those variations are the result of actions external to the intent of the producer of the manifestation (e.g., damage occurring after the item was produced, binding performed by a library, etc.).</p> |

Family of Works



This diagram from Tillet 2004 illustrates the boundary between work and expression in FRBR

The relationships between the bibliographic entities (Group 1) in FRBR are

| | | | |
|---------------|---------------------|---------------|-------|
| Work | <isRealizedThrough> | Expression | (1:N) |
| Expression | <isEmbodiedIn> | Manifestation | (1:N) |
| Manifestation | <isExemplifiedBy> | Item | (1:N) |

(A work can have many expressions, but an expression is always of one work)

FRBR includes other entity types, namely

Group 2. person (an individual) and **corporate body** (an organization or group of individuals and/or organizations).

| | | | |
|--------------------------|------------|---------------|-------|
| Person or Corporate Body | <creates> | Work | (N:M) |
| Person or Corporate Body | <realizes> | Expression | (N:M) |
| Person or Corporate Body | <produces> | Manifestation | (N:M) |
| Person or Corporate Body | <owns> | Item | (N:M) |

Group 3. Entities that serve as the **subjects of works**. The group includes **concept** (an abstract notion or idea), **object** (a material thing), **event** (an action or occurrence), and **place** (a location).

| | | | |
|------|----------------|---|-------|
| Work | <hasAsSubject> | Concept, Object, Event, Place | (N:M) |
| Work | <hasAsSubject> | Work, Expression, Manifestation, Item, Person, Corporate Body | (N:M) |

You have read a description of FRBR (Functional Requirements for Bibliographic Records) and some more information is provided on the previous pages.

For better or for worse, FRBR is the conceptual basis for the revision of cataloging systems. So you need to be familiar with it.

FRBR takes a small step towards basing cataloging on an entity-relationship model, but is still rooted too much in the past, and its conceptual analysis could be improved.

The FRBR distinction between the entity types *work* and *expression* is poorly conceived. The distinction is so hard to make that it requires a lot of effort from catalogers for no benefit. All the information that is captured in this distinction can be captured better through proper relationships within one entity type *work*. The generic *Story of Faust* is a work, *Goethe's Faust* is a work, and an *English translation of Goethe's Faust* is a work; the relationships between these works are obvious.

Different pieces of information should be pegged at the right level.

Question: In the hierarchy of works, where should we record
<*dealswith*> Pact with the devil

Answer: The *Story of Faust*. This piece of information applies to all works that are based on the *Story of Faust*, so it can be entered into a (properly structured) catalog once and will then inherit down to all the works based on the *Story of Faust*.

Question: Where should <*authoredBy*> Goethe, Johann Wolfgang von appear?

Answer: Once with the record of this work. It will inherit down to the many manifestations of this work.

Question: Where **should** <*publishedby*> Cotta, <*publishedInDate*> [1903], <*printedIn*> [Fraktur, 11 point] (important when large print) appear?

Answer: In the record for a specific manifestation, one of many that renders the text of the work *Goethe's Faust*. Any of these manifestations would do to get the text, provided the prospective reader can read the type face.

An inventory number would be assigned to an individual item

In the cataloging of rare books information on individual items (individual physical copies) is important.

Look at the court case on electronic vs. printed copies of email on the next page as an example of the practical importance of considering different versions of a document.

This ruling is a good example of the importance of discussing the problem of different versions of a document. Emphasis added.

Public Citizen v. John Carlin, Archivist of the United States Oct. 1997
Overturned by Court of Appeals for the District of Columbia Circuit August 1999, but the reasoning is nevertheless important in our context.

Washington Post, Thursday, October 23, 1997, p. A21

Judge Nullifies Rule on Computer Data

Archivist Criticized for Letting Agencies Eliminate Electronic Records

By George Lardner Jr.
Washington Post Staff Writer

A federal judge held yesterday that the head of the National Archives ignored his duties and acted illegally in issuing a regulation that authorizes all government agencies to wipe out their electronic mail and other computerized records regardless of content.

In a 36-page ruling sharply critical of Archivist John W. Carlin, U.S. District Judge Paul L. Friedman declared the controversial rule "null and void" and called the government's defense of it "irrational on its face."

The two-year-old regulation, known in bureaucratic jargon as "GRS [General Records Schedule]-20," permitted all agencies, from the Executive Office of the President on down, to destroy e-mail and wordprocessing records once they have been copied on paper or some other format and deemed "no longer needed for updating and revision."

Historians, researchers and journalists represented by the non-profit advocacy group Public Citizen denounced the provision as an "electronic shredder" and filed suit, accusing Carlin

of abdicating his responsibilities to appraise the value of the records on an agency-by-agency basis.

Friedman agreed. "Simply put," he said, "**electronic communications are rarely identical to their paper counterparts; they are records unique and distinct from printed versions of the same record.**"

Citing an example from the Iran-contra scandal, the judge pointed out that so-called PROF notes-computerized messages between national security adviser John M. Poindexter and White House aide Oliver L North-played an important role in the trials of both men.

"Admiral Poindexter, a computer expert set up a special channel known as "Private Blank Check," which allowed North and Poindexter to relay messages to each other without those messages being accessible to other NSC staff," noted Friedman, who was once an Iran-contra prosecutor. "The communication itself was clearly important to investigators, but the mode of communication and the special channel through which it was sent, which would not have been reflected in paper printouts of the messages, was also important."

In promulgating GRS-20 in 1995, the judge said, Carlin categorically determined that electronic records possess no administrative, legal, research or historical value beyond paper printouts of the same document. In doing this, "the Archivist has absolved both himself and the federal agencies he is supposed to oversee of their statutory duties to evaluate specific electronic records as to their value."

Carlin, the judge said, also exceeded his authority in giving agencies "carte blanche" to destroy electronic versions "whenever agency officials believe they are no longer needed."

The government had argued that GRS-20 was soundly based because such government-wide rules were meant for records of common form, such as "electronic" media.

Lawyers for Carlin had also protested that most federal agencies are not yet equipped to preserve records in electronic format. Friedman said this was "an important concern" but observed that "computers have now become a significant part of the way the federal government conducts its business" and the government must adapt to that reality.

The Archives had no immediate comment.

Structure of a better catalog - next page

Example illustrating the idea of an interlinked catalog

Consider this document:

(4) *The man I killed*, play by Christopher Wern, based on the novel by Michael Halliday

This is Document (4) from the list of sample documents given earlier

The novel referred to is Document (1)

(1) *The man I killed*, by Michael Halliday (i.e. John Creasey). London: Marx Brothers; 1935

How should (4) be cataloged?

Who is the author? Most would say Wern. But should this work also be found in a search for works authored by Halliday? Some users may want to find it, others may not.

In the present system where each book (each manifestation) is cataloged on its own, the cataloging rules must prescribe whether to make the Wern play findable under Halliday or not; the user is then stuck with this decision. The choice is made by the system (the cataloging rules), not the user.

The better solution would be this:

- Catalog (1), with author Halliday (whether Halliday or Creasy is a question addressed in 7.1a).
- Catalog (4) with author Wern and a *<basedOn>* link to (1) but no further information. Of course, (4) *<basedOn>* (1) could also be written as (1) *<basisFor>* (4).
- Have a search system where the user looking for books authored by a person, in the example Halliday, can choose whether to limit the search to books where Halliday is listed directly as the author or whether to follow links *<basisFor>* to more documents such as (4).

Two principles can be derived from this example:

1 **For all information systems: Leave choices to the user.** This means

- 1.1 the systems data need to describe the real situation faithfully and precisely without any distortions due to system rules and limitations;
- 1.2 the system must give the user many options of searching, combining the data in various ways.

2 **For bibliographic catalogs: No bibliographic unit (work, manifestation) is an island.** Bibliographic units are linked in a complex network that should be faithfully represented in the catalog and used in searching. (The hierarchical inheritance examples discussed earlier also fall under this principle.)

How to design a better catalog system

The key to designing an efficient database structure for a catalog lies in analyzing and applying the relationships between bibliographic entities. The root cause for the complexity of many cataloging rules is the attempt to force data with very complex relationships into a simple-minded data structure.

Elements of a conceptual data schema for a database with data about documents.

<isVersionOf>, more specific *<isTranslationOf>*

<isPartOf>

These relationships may hold between works, between manifestations, or between items.

Work *<isRenderedIn>* Manifestation or (inverse) Manifestation *<isRenderingOf>* Work

FRBR chain: Work *<isRealizedThrough>* Expression,
Expression *<isEmbodiedIn>* Manifestation

Manifestation *<generatedFrom>* (Manifestation, RenderingProcess)

(Examples of rendering processes: different screen renderings from same HTML source text through different browsers; facsimile; Optical Character Recognition. The distinction between a rendering process and a reproduction process is fluid.)

Manifestation *<isInstantiatedIn>* Item or (inverse) Item *<isInstantiatedIn>* Manifestation

FRBR: Manifestation *<isExemplifiedBy>* Item

Item *<reproducedFrom>* (Item, reproduction process)

Some problems in a conceptual data schema for bibliographic and record control

Records, originals vs. copies

Permanent copy vs. fleeting copy

Specific printing may use different paper - preservation!

Performance of a work may be more than a mere manifestation since it brings separate creative elements. Perhaps a performance should be considered a work of its own, with the tape (or audio file) on which it is captured being an item of a manifestation of that work (remastering such a tape would create another manifestation).

Definition of "catalog" - elaboration

A work is *covered by a catalog* if the catalog contains data about the work, or any manifestation of the work, or any item (individual copy) of any manifestation of the work. The collection linked with a catalog may be either a collection of items or merely a list of works, manifestations, or copies; a manifestation is said to be *represented in a collection* of items if any item of the manifestation is in the collection. In the electronic age the concept of *collection* becomes more and more fluid. Is the whole World Wide Web one collection, or is a collection confined to the documents (files) stored on one Web server? Likewise, the concept of *library* becomes more and more fluid; there are now digital libraries whose "collections" may be distributed over many sites (whence the term *virtual library*). In fact, there is no sharp distinction between a digital library and any computerized information system. The functional distinctions made in Section 2.6 of the text are useful to clarify some of the issues here.

Objectives of the library catalog - restatement by D. Soergel

The catalog (of a library, a book seller, ...) should be an efficient instrument for ascertaining

- (1) **Criterion search (intellectual access)** [4.1.2]

Which works, manifestations, or items are helpful to a given user for a given purpose, to wit

 - (1a) which works, manifestations, or items covered by the catalog meet a combination of criteria relating to provenance (including authorship), subject, artistic characteristics, and/or other criteria (**retrieval** or **identification**) (in some cases only certain manifestations or individual items may meet the user's search criteria) (**find** a set of resources) [4.1.2];
 - (1b) whether a work, manifestation, or item meets the needs of the user and how several suitable works, manifestations, or items should be ranked (**selection**) [4.3];
 - (1c) how a work, manifestation, or item relates to (another) work, manifestation, or item (**relation**) (for example, *<isRevisionOf>*, *<isReprintOf>*, *<isBasedOn>*);
- (2) **Search for a known work, manifestation, or item (find a single resource)** [4.1.1]

(Confusingly, this is called *known-item search*, a term coined before FRBR)
To ascertain that a resource given in the catalog is the same as the resource in hand (**identity**) [4.2]

 - (2a) whether a known work is covered by the catalog and, if so, which are the manifestation(s) of the work that are covered in the catalog (**coverage**);
 - (2b) whether a known manifestation is covered in the catalog;
 - (2c) whether a known item (specific copy) is covered in the catalog (important for rare books)
- (3) how the user can get **physical access** to some item (copy) of the work (method of access, time, cost) (**acquire** or **obtain**) [4.4].

The objectives are arranged by decreasing complexity and increasing concreteness, not by importance. After achieving objective 1 a user must achieve objective 2 and then objective 3.

The Statement of International Cataloging Principles (a reading), Section 4, presents a somewhat different organization of objectives of the catalog; their numbers given in []

The Statement lists 4.5 to **navigate**, but navigation is a means for achieving any of the objectives, just as query-based search, so it does not belong in this list. Of course, a catalog should support both

Note: Many of these objectives apply to searching for people, organizations, software, or whatever.

Lecture 6.2c (25 min)*February 22*

Bibliographic and record control: Description
Describing texts and documents in a more general context

General principles of description; their application to bibliographic and record control; their implementation in ISBD/AACR2; relationship to the MARC format. User-oriented analysis of elements of description needed.

Description: What needs to be known about an entity?

Relates to catalog objective 1b, ascertaining relevance. Also relates to objective 3, ascertaining whether a given manifestation is indeed the same manifestation that is covered in a catalog record. (The “given manifestation” may be the manifestation requested by the user or the manifestation of which the item in hand is an instance.)

Data about bibliographic entities - conceptual data schema.

Peg each piece of data to the correct bibliographic level (work, manifestation, item).

How to you decide at which bibliographic level (work, manifestation, item) a piece of information should be given? (See also the discussion right after the FRBR picture above)

In used book seller catalog (or union catalog, such as www.abebooks.com) there are descriptions of used books for sale. They often describe the condition of the book, such as

Book Condition: Used - Good.. Shows some signs of wear, and may have some markings on the inside

Hardcover. No dust jacket. Used, good. EX-LIBRARY - has usual library wear/markings/attachments. Small tears/creases on spine/cover. Cover has some edge wear

What bibliographic level does this part of the description pertain to?

Note: For rare books description is much more extensive.

Looking ahead: It is efficient to mark up the description in such a way that entries (access points) can be extracted easily. Cataloging in the MARC format achieves this

Sources of information for cataloging data

Primary: Title page and verso (back of title page),

Secondary: Preface, last page, cover page of a journal issue, etc. (data shown in [])

Which is source is most authoritative?

Sometimes it requires considerable inquiry to find information required by the conceptual data schema (the cataloging rules) but not known from the representation in the item being examined. In the extreme, this involves long research into the authorship of a play or of a painting.

Arrangement of cataloging data

- **in a record** - record format such as MARC
- **in a display** (printed or on a screen)

Many different styles: AACR2, ANSI Standard, Turabian, American Psychological Association (over 4,000 and counting).

Let a computer program, such as Endnote or Library Master, do the work!

MARC is just one record format. There are many others. We will talk later about Dublin core and other Web-based initiatives

Open Archives Initiative www.openarchives.org/

The record format should store data fine-grained so that many different displays can be supported (for example, journal volume, issue number, year, month, day, beginning page, end page should each be a separate piece of data because different display formats arrange these in different ways.

Display format

An OPAC (Online Public Access Catalog) should give the user considerable control over the display format (but they rarely do); systems like Dialog (now part of Proquest) that provide access to bibliographic and other databases are usually much better at this.

Publishers and journals prescribe one of the many available display formats for use with their publications; you should be familiar with some of them and be able to assist users with preparing bibliographies. In many courses you are required to use APA style, a very common style, in formatting bibliographic records in your bibliography.

Unfortunately, these are mostly based on personal preferences of someone in charge or who was in charge 50 years ago rather than on empirically based analysis of what is most functional for users. Thoreau's quote applies to many of these: "Any fool can make a rule, and every fool will mind it". For example, for scanning a bibliography the most important piece of information is the title of the document itself, yet APA style requires highlighting (in italics) the title of the journal in which an article appeared. I always bold the title of the document.

There are many **bibliographic citation managers**, among them

Endnote (on your computer and the Web, www.endnote.com/),

Library Master (more powerful than most but fewer users, www.balboa-software.com),

Zotero (entirely Web-based but may have a way to store data on your computer as well, www.zotero.org)

Mendeley (on your computer and the Web, www.mendeley.com)

These programs have style sheets (many premade, but you can also make your own) for any conceivable display format.

You should use one of these bibliography managers for your work. UB has a site license for Endnote, use version X4 (<http://library.buffalo.edu/libraries/endnote/>)

Lecture 7.1a (40 min)**February 29****Bibliographic and record control 2: Entries**

UBLIS571Lecture07.1aSlides.zip. Slide 2 has audio for p. 191-192

General principles of access; their application to bibliographic and record control; their implementation in AACR2 choice and form of entry. Authority files. User-oriented analysis of access points needed.

| | |
|---|---|
| Definition | An <i>entry</i> is an element, such as an author name, a title, a series title, or a subject descriptor under which a document (or another object) can be found in a catalog or index. (The term comes from book or card catalogs, where an entry for a document is made by writing or filing a card.) Determining entries is a problem of data structure and access. |
| Two issues: | A Which of the data in the description should be made access points for lookup searching? (The answer to that question might have repercussions for description if a data element is important for access but not for ascertaining the relevance of an item.) B What form should each entry take? (Rules for entity values) |
| Main entry Added entry | A document may have many authors / contributors. Most of the time, this does not present a problem: just list them all and provide access from all (“make an entry,” as in a card catalog, for all). But sometimes we want to list a document record only once: In a printed bibliography, in a listing of search results arranged by author, in a card catalog before reproduction equipment when every card had to be written or typed by hand. In that case, we need to select the most significant author / contributor, the one under which the one entry should be made. This is the <u>main entry</u> . (This concept was quite important in the age of card catalogs; it is less important now but still has applications.) |
| In-class exercise | Problems of determining author entry analyzed according to Lubetzky Lubetzky was the foremost thinker on bibliographic cataloging rules. |
| Advanced | Thinking about rules for corporate entry |

The Author approach: Conditions and cases (Lubetzky after Needham)

See next page. Lubetzky's discussion of possible solutions is found in the readings.

| Condition | Issue A: What entries to make | Issue B: What form the entry should take |
|---|--|--|
| 1. Documents having more than one Author | 1.1 Document prepared by an author with the aid of collaborators or contributors. | |
| | 1.2 Document composed by an editor or compiler from the writings of <u>several</u> other people | |
| | 1.3 Document by several authors with no one author more responsible for it than any of the others. | |
| | 1.4 Document in which the writer reports the communication of another person (real or fictitious). | |
| 2. Authors having more than one name | | 2.1 The author has changed his or her name in consequence of marriage, adoption of new citizenship, joining a religious order, or for any other reason. |
| | | 2.2 The author always writes under an assumed name different from his real name, or under his title of nobility, or under part of his name. |
| | | 2.3/4 Author uses more than one name in successive documents |
| | | 2.5 Authors whose names appear in translation in varying forms. |
| 3. Dependent Documents | | |
| 4. Corporate authors | 4.1 The reports and statements of a corporate body are usually prepared by one of its officers or by another person engaged to prepare the statement for it. | 4.2 Many corporate bodies have no proper identifying names of their own but only generic names describing their type and common to most bodies of that type e.g. public library, historical society, dramatic club, etc. |
| | | 4.3 Change of name in corporate bodies. |
| | | 4.4 An organization may act or speak as a whole or through one of its branches, divisions, offices, etc. |

Sample documents for analyzing author entry according to Lubetzky

For the sample documents, think about

- (1) Who should be listed as author for purposes of finding the document?
- (2) For the people selected as authors, in what form should their names be recorded?

How can sample documents (e) and (t) be used to illustrate the idea of an interlinked catalog?

There is an "interactive" PowerPoint presentation for this exercise, UBLIS571Lecture07.1aSlides.zip, starting with slide 3.

- (a) *The record guide* by Edward Sackville-West and Desmond Shaw-Taylor, with Andrew Porter and William Mann.
- (b)* *Studies in the social psychology of adolescence*, by J. E. Richardson, J. F. Forrester, J. K. Shukla, and P. J. Higginbotham; edited with a foreword by C. M. Fleming.
- (c) *The tropics*, by Edgar Aubert de la Rue, Francois Bourliere, Jean-Paul Harroy.
- (d) *Ambit* (a periodical), edited by M. C. O. Bax and Edwin Brock.
- (e)* *Chisholm's handbook of commercial geography*, entirely rewritten by L. Dudley Stamp and S. Carter Gilmour.
- (f)* *Making magical apparatus*, by Jane Reid (i.e. Mrs. David Johnstone).
- (g) *Lord Jim*, by Joseph Conrad (i.e. Josef Theodor Konrad Korzeniowski).
- (h) *The far country*, by Neville Shute (i.e. Neville Shute Norway).
- (i) *The trimmed lamp, and other stories*, by O. Henry (i.e. William Sydney Porter).

* designates items to be analyzed in Assignment 9

- (j) *The scene of the crime*, by John Creasey.
- (k) *The man I killed*, by Michael Halliday (i.e. John Creasey). London: Marx Brothers; 1935
- (l) *A branch for the baron*, by Anthony Morton (i.e. John Creasey).
- (m)* *Schubert: thematic catalogue of all his works in chronological order*, by Otto Erich Deutsch in collaboration with Donald R. Wakeling.
- (n)* *A concordance to the poems of William Wordsworth*, by Lane Cooper.
- (o)* *The poetical works of Wordsworth*, edited by E. de Selincourt.
- (p) *Oxford book of English verse, 1250-1918*, chosen and edited by Sir Arthur Quiller-Couch.
- (q) *Shakespeare's 'Much ado about nothing'*, by N. T. Carrington (Notes on chosen English texts). Text and commentary.
- (r) *The man I killed*, by Michael Halliday (i.e. John Creasey). Audiotape, read by Sir Lawrence Olivier. New York: Books on Tape; 1966.
- (s) *The man I killed*, play by Christopher Wern, based on the novel by Michael Halliday.

- (t) *The Aeneid of Virgil*, retold by N. B. Taylor.
- (u) *Billy Budd* (libretto), adapted from the story by Herman Melville by E. M. Forster and Eric Crozier.
- (v) *Iban agriculture: a report on the shifting cultivation of hill rice*, prepared for the Colonial Office by John Derek Freeman.
- (w) *Essays and studies, 1962: being volume fifteen of the new series of Essays and studies*; collected for The English Association by Beatrice White.
- (x) *The Library Association Record*, edited by Edward Dudley.
- (y) Ministry of Education. 15-18. (The Crowther report).
- (z) *Yearbook of the Institution of Agricultural Engineers* (originally the Institution of British Agricultural Engineers).
- (aa) National Physical Laboratory. *Mathematical tables*. (The National Physical Laboratory is a branch of the Department of Trade and Industry.)
- (bb) *Farm business statistics for south-east England*. Wye College (London University).
- (cc) *Annual report of the Association of Assistant Librarians* (a division of the Library Association).

Observations on form of name

The principle Leave choices to the user and Record data faithfully applies here as well. Consider John Creasey, who wrote under three names: John Creasey, Michael Halliday, and Anthony Morton. Some catalog rules prescribe that in all cases John Creasey should be used as the form of name, the identifier for this one person. However, a user who liked one book by Anthony Morton and wants to find similar books may not be pleased at all by finding books John Creasey wrote under his real name because those books might be quite different (there might be three authorial spirits living in John Creasey's body, and for this user the entity of interest is the authorial spirit, not the actual person). For another example consider someone who has read Alice in Wonderland and wants more books by the same author, listed as Lewis Carroll, and is handed a mathematical treatise the same person wrote under his real name Charles Lutwidge Dodgson. The solution is as follows:

- (1) Catalog books under the name of the author as it appears on the title page.
- (2) Create an authority file of personal names (a database with information about people, including their names) that links together all names that refer to the same person. Such a database exists at the Library of Congress (<http://authorities.loc.gov>) and is used by many libraries. It includes birth and death dates and all names, with time periods during which a name was in effect, if necessary, and other information.

This leaves the choice to the user: She can look for all books authored by the "authorial spirit" Anthony Morton or expand the search to all names listed with the person John Creasey.

Problems of entry for works emanating from corporate bodies

Audio for these pages is on www.dsoergel.com/571/UBLIS571Lecture7.1aSlides.zip, Slide 33

What is a corporate body?

The **definition of a corporate body** according to AACR2:

“21.1B Entry under corporate body

21.1B1. Definition. A corporate body is an organization or a group of persons that is identified by a particular name and that acts, or may act, as an entity. Consider a corporate body to have a name if the words referring to it are a specific appellation rather than a general description. If, in a script and language using capital letters for proper names, the initial letters of the words referring to a corporate body are consistently capitalized, and/or if, in a language using articles, the words are always associated with a definite article, consider the body to have a name. Typical examples of corporate bodies are associations, institutions, business firms, nonprofit enterprises, governments, government agencies, projects and programs, religious bodies, local churches, and conferences.

Note that some corporate bodies are subordinate to other bodies (e.g. the Peabody Museum of Natural History is subordinate to Yale University.)

Consider ad hoc events (such as athletic contests, exhibitions, expeditions, fairs, and festivals) and vessels (e.g., ships and spacecraft) to be corporate bodies.”

Issue A. When to make an entry under a corporate body (corporate entry)?
 (When to establish a relationship between a work and a corporate body?)
 One question to be asked: When does a corporate body have a role similar to the responsibility of an author (corporate authorship)?

AACR2:

“21.1B2. Footnote 2. Consider a work to have emanated from a corporate body if it is issued by that body or has been caused to be issued by that body or if it originated with that body.”

A corporate body can

- be fully responsible, as in a law enacted by a government or the official statement of an organization;
- have commissioned a work;
- have issued / published a work;
- provided the environment in which a work was created, such as a university or Rand providing an environment for a researcher (usually handled, if at all, as author affiliation).

Easiest solution: Make a good list of these relationship types and use the specific relationship when cataloging a work. This does not fit into the prevailing system of cataloging; there are just the MARC fields 110 and 710 for corporate names (corporate body in an author-like role). AACR2 gives rule under what circumstances such entries should be made.

Issue B. What form of name should be used for the corporate body?

See AACR2 Chapter 24

Problems:

- Corporate bodies change their names, for example
Bureau of Foods became *Center for Food Safety and Applied Nutrition*.
- More complex problem: Corporate bodies change, cease to exist, or are merged with other corporate bodies.
- Corporate bodies are part of other corporate bodies, for example
US. Department of Health and Human Services. Public Health Service. Food and Drug Administration. Center for Food Safety and Applied Nutrition. Technical Operations Branch.
Some levels in such a hierarchy are often better known than others, for example,
Food and Drug Administration is better known than
Public Health Service
(possibly a phenomenon similar to basic level concepts)
There are other relationships, such as corporation B being a wholly-owned subsidiary of corporation A (interesting if you want to sue corporation B for damages).
- Corporate bodies often have generic names that are meaningful only in conjunction with a place, such as
Metropolitan Museum of Art
Some rules suggest to put the name of the place first:
New York Metropolitan Museum of Art
Cleveland Metropolitan Museum of Art
- Corporate bodies are often better known under an acronym or short popular name, for example
FDA is better known than *Food and Drug Administration*

These problems occur in any information system that deals in any way with corporate bodies. The best solution is to have a database of corporate bodies including all their names, their life span, their relationships, and information on place. Such a database could be used in conjunction with a bibliographic catalog or other information systems.

Observations on names of corporate bodies

At one point it was thought that users should be able to find all documents emanating from a corporate body under one name, the most recent one, even over time. In a card catalog, that meant when the name of a corporate body changed, all the catalog cards for documents cataloged under the previous name had to be changed to the new name, a process known as superimposition. But then the prevailing opinion changed to prefer the previous system which used the name of the corporate body as found in the document cataloged; so all these catalog cards had to be changed back, a process known as desuperimposition.

Now that computers are available, we can have our cake and eat it too:

Make an authority file for corporate names, a database with information about corporate bodies, including their names over time. Such a database is included in the Library of Congress name authority (<http://authorities.loc.gov>). With such a database, a user could search for all the names under which a corporate body is or was known, or she could restrict her search to one particular name and find only documents associated with that name.

It would be even better if it included hierarchical and other relationships among corporate bodies: subordinate corporate bodies as in the hierarchy of the federal government or the various subordinate organizations within ALA (such as the American Association of School Librarians (AASL)) or the relationship between a holding company and the companies it holds. If the system had that information, the user could do an inclusive search for all documents emanating from a corporate body or any of its subordinate bodies.

Hierarchy of corporate bodies

The federal government is hierarchically organized, as shown in this example:

Department of Health and Human Services

Public Health Service

Food and Drug Administration

Center for Drugs

Center for Food Safety and Applied Nutrition

Technical Operations Branch

The major departments are generally known in the population. But below that there are some levels that stand out, such as FDA, while others above or below these agencies are not often known to the general public. This is a phenomenon similar to basic level concepts.

Advanced exercise: Thinking about rules for corporate entry.

See Supplement

Lecture 7.1b (35 min)

February 29

Metadata, Resource Description Framework (RDF), Dublin Core (DC)

Resource Description Framework (RDF)

www.dsoergel.com/571Lecture7.1b.mp3 (page nos are off)

| | |
|----------------------------|---|
| Definition | <p>A general abstract data modeling method based on the entity-relationship approach, along the lines of what is described in Integrated Information Structure Interface. Developed with metadata in mind, but can be used for any kind of data for any kind of use.</p> <div style="border: 1px solid black; padding: 10px; margin: 10px 0;"> <p>Entities / objects are called <i>resources</i></p> <p>Relationship types are called <i>properties</i></p> <p>A statement is made <i>about</i> a resource (the subject of the statement), using a property and giving another resource as the <i>value</i></p> <p>Example: www.dsoergel.com/571 <creator> “Soergel”</p> </div> |
| Implementation | <p>Major implementation in XML using the</p> <p style="text-align: center;">Resource Description Framework (RDF) Model and Syntax Specification (www.w3.org/TR/REC-rdf-syntax/)</p> <p>The RDF syntax is specified using <i>XML Schema</i>. By convention, RDF syntax elements are identified by the prefix <i>rdf:</i>. RDF is merely a syntax; it does not specify any particular <i>properties</i> (relationship types). These properties can and must be defined by each user community.</p> |
| Difference from XML | <p>Enhancements of RDF syntax over plain XML: The meaning (the semantics) of syntactical constructs are precisely defined, while in XML they are deliberately left to the specific application. In particular, for each property the types of the participating resources can be specified.</p> <p>RDF has many additional features.</p> |
| Example | <p>See next page. More examples in discussion of Dublin Core</p> |
| Name spaces in XML | <p>RDF syntax as a name space</p> <p>Dublin core properties as a name space</p> |

Example of an RDF resource description: simple document description

Here is a simple example of RDF syntax used to describe a resource. This example uses properties defined in the Dublin Core (DC). RDF tags (prefixed with **rdf:** and Dublin Core (DC) tags (prefixed with **dc:**) are defined in files found at the URLs given and then used in the record. Name spaces ensure that there is no collision between tag names in the two syntaxes (imagine what happens when two language syntaxes, each defined in XML, use the same tag name, like "description", to define different data fields. Without the use of distinct name spaces, things would get horribly confusing, both to document creators and the automated systems parsing XML documents).

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.0/">
  <rdf:description rdf:about="http://www.ukoln.ac.uk/metadata/resources/dc/datamodel/WD-dc-rdf/">
    <dc:title> Guidance on expressing the Dublin Core within RDF </dc:title>
    <dc:creator> Eric Miller </dc:creator>
    <dc:creator> Paul Miller </dc:creator>
    <dc:creator> Dan Brickley </dc:creator>
    <dc:subject> Dublin Core; RDF; XML </dc:subject>
    <dc:publisher> Dublin Core Metadata Initiative </dc:publisher>
    <dc:contributor> Dublin Core Data Model Working Group </dc:contributor>
    <dc:date> 1999-07-01 </dc:date>
    <dc:format> text/html </dc:format>
    <dc:language> en </dc:language>
  </rdf:description>
</rdf:RDF>
```

More on RDF

RDF is restricted to expressing binary relationships. This is a serious limitation. Other, more powerful systems will eventually take RDF's place

Many data sets are now put on the Web using RDF as implemented through XML (Linked Open Data, see http://en.wikipedia.org/wiki/Linked_Data and <http://linkeddata.org/guides-and-tutorials>). These data can then be used for many purposes, for example, they could be included in freebase for navigation in a huge information space. A system such as I described in the reading and Lecture 4.1 could now be built using these data.

Logically, RDF structures are visualized as graphs such as the graph shown in Lecture 1.2 for the university database. The structure needs to be represented by series of statements, for example, encoded using tags defined in XML, such as shown in the examples. This is called *serialization*.

When expressed in XML, RDF can be considered a refinement of XML, see below

Dublin Core (DC) (<http://dublincore.org/>)

The Dublin Core is a minimal standard for the description of “document-like objects”

| | | |
|---------------|--------------|---------------------------|
| # Title | # Format | |
| # Creator | # Identifier | |
| # Subject | # Source | |
| # Description | # Language | |
| # Publisher | # Relation | |
| # Contributor | # Coverage | |
| # Date | # Rights | Definitions in Supplement |
| # Type | | |

“Plain” Dublin Core has just 15 properties (relationship types), simplicity both good and bad. The Dublin Core list of properties can be implemented in many ways, among them:

- In the *meta* section of an HTML document (see next page and the model catalog)
- In plain XML
- In XML, using the RDF enhanced syntax

The Dublin Core is a conceptual data schema for bibliographic description that is simpler than MARC and could thus be used widely by people not trained in cataloging. Bibliographic description is understood to cover a broad range of documents and "document-like objects", especially Web pages. The DC set of fields is shown in the box. Expansions and refinements can and have been created for different types of materials, such as learning objects. You must be familiar with the Dublin Core, but do not memorize the DC data fields.

Here are a few lines of an HTML document with Dublin Core metadata

```
<HTML>
  <HEAD>
    <META name="dc.creator" content="Renato Iannella">
    <META name="dc.creator.affiliation" content="DSTC">
    <META name="dc.subject" content="Cats, Fur, Purr">

    Specification of the scheme from which subject descriptors were taken, here LCSH
    <META name="dc.subject" content="(scheme=LCSH) Animals~Felines">

    Language of title or subject descriptors can be specified in like manner
  </HEAD>
  <BODY>
    ...

  </BODY>
</HTML>
```

More complex example illustrating more features of RDF and refinements of the Dublin Core.

```

<?xml version='1.0'?>
<rdf:RDF
  xmlns:rdf="www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc = "http://purl.org/dc/elements/1.0/"
  xmlns:dcq = "http://purl.org/dc/qualifiers/1.0/"
  xmlns:foaf = "http://xmlns.com/foaf/0.1/"> /* foaf = friend of a friend, http://xmlns.com/foaf/spec/ */
  <rdf:description rdf:about="www.ukoln.ac.uk/metadata/resources/dc/datamodel/WD-dc-rdf">
    ...
    <dc:creator>
      <rdf:Description>
        <rdf:type rdf:resource =
          "http://purl.org/dc/terms/1.0/creator/class/Person"/>
        <dcq:creatorType rdf:resource =
          "http://purl.org/dc/terms/1.0/creator/type/Editor"/>
        <rdf:value rdf:resource = "http://411.com/EricMiller"/>
      </rdf:Description>
    </dc:creator>
    ...
  </rdf:description>

  <rdf:description rdf:about = "http://411.com/Eric Miller">
    <foaf:name>Eric Miller </foaf:name>
    <foaf:homepage>rdf:resource="www.oclc.org/~emiller" </foaf:homepage>
    <foaf:img rdf:resource="/images/emiller.jpg" />
  </rdf:Description>
</rdf:RDF>

```


The next few pages show

a worked-out example of representing some bibliographic data in RDF

| | |
|--------------------------------------|---|
| <p>RDF Schema</p> | <p>Before properties can be used in RDF descriptions, they must be defined, just as tags must be defined before they can be used in valid XML documents. For this purpose, RDF provides RDF Schema, which is turn specified using XML Schema.</p> <p>Defining a conceptual data schema involves defining entity types (in RDF: classes) and relationship types (in RDF: properties). This is what is done here through an XML document that uses tags that are predefined in RDF Schema (prefix rdfs:).</p> <p>Please get the gist of this. I will not ask about details of syntax in an exam, but I may present a practical problem were using RDF like that might factor into the solution. And why would I ask a question like that? Because that might happen to you in real life. If it does, you should recognize that RDF representation might be helpful, find out about data sets you could draw on, and find somebody who can implement the idea. Again, some of you might become interested in this and learn enough to be experts at handling data in RDF</p> <p>Below is a very simple piece of an RDF schema definition for bibliographic data.</p> |
| <p>Classes (entity types)</p> | <p>Note the specification of hierarchical relationships among classes, for example the entity type (class) Person is defined as a subclass of LegalEntity using the tag rdfs: subclassOf.</p> <p>Also, there are hierarchical relationships among relationship types (properties)</p> <p>Also note that every class and every property has an internal ID and an external label. They can be the same but do not have to be.</p> <p>XML and RDF schemas for the food database from Lecture 2.2, is given in www.dsoergel.com/571/SYL2003FaLecturesAppendixNew.pdf (see Lecture 6.2b)</p> |

RDF schema definition example: Dublin Core

Definition of an entity-relationship conceptual schema in lengthy syntax

```
<?xml version='1.0'?>
<rdf:RDF
  xmlns:rdf="www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="www.w3.org/2000/01/rdf-schema#"
  xmlns:dc="http://purl.org/dc/elements/1.1"/>
```

Definition of entity types (classes)

```
<rdfs:Class rdf:ID="Book">
  <rdfs:label>Book</rdfs:label>
  <rdfs:comment>The class of books</rdfs:comment>
  <rdfs:subClassOf rdf:resource="www.w3.org/2000/01/rdf-schema#Resource"/>
</rdfs:Class>

<rdfs:Class rdf:ID="LegalEntity">
  <rdfs:label>Legal entity</rdfs:label>
  <rdfs:comment>The class of person or organizations</rdfs:comment>
  <rdfs:subClassOf rdf:resource="www.w3.org/2000/01/rdf-schema#Resource"/>
</rdfs:Class>

<rdfs:Class rdf:ID="Person">
  <rdfs:label>Person</rdfs:label>
  <rdfs:comment>The class of persons</rdfs:comment>
  <rdfs:subClassOf rdf:resource="#LegalEntity"/>
</rdfs:Class>

<rdfs:Class rdf:ID="Organization">
  <rdfs:label>Person</rdfs:label>
  <rdfs:comment>The class of organizations</rdfs:comment>
  <rdfs:subClassOf rdf:resource="LegalEntity"/>
</rdfs:Class>
```

| | |
|---|--|
| RDF Schema Properties (relationship types) | <p>Next we define relationship type (properties). Here is where the real power of RDF comes in. The relationship type (property) <i>creator</i> (as defined in our context or "name space" must have an entity of type book at its left side (domain) and an entity of type LegalEntity (a Person or Organization) at its right side (domain). These restrictions can be checked as data are entered, and incorrect data can be rejected. This is not possible in basic XML.</p> <p>Note the hierarchical structure. What is true for the property <i>creator</i> is also true for the property <i>editor</i>.</p> |
|---|--|

RDF schema definition example: Dublin Core continued**Definition of relationship types (properties)**

```
<rdf:Property rdf:ID="title">
  <rdfs:label>Title</rdfs:label>
  <rdfs:comment>The name given to the resource</rdfs:comment>
  <rdfs:domain rdf:resource="#Book"/>
  <rdfs:range
    rdf:resource="www.w3.org/2000/01/rdf-schema#Literal"/>
</rdf:Property>

<rdfs:Property rdf:ID="creator">
  <rdfs:label>Author</rdfs:label>
  <rdfs:comment>A person or organization responsible for the content of a book</rdfs:comment>
  <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/dcmes.rdf#Creator"/>
  <rdfs:domain rdf:resource="#Book"/>
  <rdfs:range
    rdf:resource="#LegalEntity"/>
</rdfs:Property>

<rdfs:Property rdf:ID="editor">
  <rdfs:label>Editor</rdfs:label>
  <rdfs:subPropertyOf rdf:resource="#creator"/>
</rdfs:Property>
</rdf:RDF>

<rdfs:Property rdf:ID="affiliatedWith">
  <rdfs:label>person affiliation</rdfs:label>
  <rdfs:comment>The organization a person is affiliated with</rdfs:comment>
  <rdfs:domain rdf:resource="#Person"/>
  <rdfs:range
    rdf:resource="#Organization"/>
</rdfs:Property>
</rdf:RDF>
```

Assume this schema is stored in a file with the URL

www.dsoergel.com/DSBibSchema.rdf

We can then refer to that file to use the definitions of entity types (classes) and relationship types (properties).

Data on some books structured according to the rdf schema just given see next page

Data on some books structured according to the rdf schema just given see next page

You may want to take the next page out so you can compare with the definitions of the properties (relationship types)

| | |
|------------------------|---|
| <p>RDF Data</p> | <p>Now we can use the schema just defined to enter data.</p> <p>First we must assign each entity value to an entity type or class. Entity values are identified by URLs, either full URLs or a URL that is internal to the file in which the data are stored (externally referred to as ...#...) The assignment to a class is done using the tag <i>rdfs:type</i>. This should be self-explanatory. (It does seem that the RDF designers are a bit schizophrenic about using the terms <i>type</i> and <i>class</i>)</p> <p>Next: Entering statements using properties (data on books)</p> <p>Look at the first record (about)</p> <p>The whole record is about www.oclc.org/cat#ISBN0126542619, so this is the left side (domain) of all the properties. It is of type/class book (as stated in the class membership definitions).</p> <p>The tag <code><bib:title></code> marks the entity value at the right side (range). It is "some text", which makes it a Literal as required by the definition of <code><bib:title></code>.</p> <p>The tag <code><bib:creator></code> marks the right side (range). According to the definition of <i>creator</i> the right side must be a LegalEntity. Soergel is assigned to the class Person (above); but person is a subclass of LegalEntity, and that makes the person Soergel a LegalEntity</p> <p>You can check this for the other records.</p> <p>The last record does not pass the check. What would the data entry person need to do to make it work?</p> <p>Note: If all the name authority data of the Library of Congress were put on the Web as Linked Open Data in a format that would declare every person as belonging to class Person and every organization to class Organization, the result would be that in almost all cases a correctly spelled name would pass muster.</p> |
|------------------------|---|

```

<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:bib="http://www.dsoergel.com/DSBibSchema">

/* Class membership definitions: Entity values assigned to entity types */

<rdf:Description rdf:about="www.oclc.org/cat#ISBN0126542619">
  <rdf:type rdf:resource="http://www.dsoergel.com/DSBibSchema#Book"/>
</rdf:Description>

<rdf:Description rdf:ID="ISBN081086007">
  <rdf:type rdf:resource="http://www.dsoergel.com/DSBibSchema#Book"/>
</rdf:Description>

<rdf:Description rdf:about="www.oclc.org/cat#ISBN083893529X">
  <rdf:type rdf:resource="http://www.dsoergel.com/DSBibSchema#Book"/>
</rdf:Description>

<rdf:Description rdf:ID="Soergel">
  <rdf:type rdf:resource="http://www.dsoergel.com/DSBibSchema#Person"/>
</rdf:Description>

<rdf:Description rdf:about="www.simmons.edu/faculty#Chan">
  <rdf:type rdf:resource="http://www.dsoergel.com/DSBibSchema#Person"/>
</rdf:Description>

<rdf:Description rdf:ID="ALA">
  <rdf:type rdf:resource="http://www.dsoergel.com/DSBibSchema#Organization"/>
</rdf:Description>

/* Data on books (relating books to persons and organizations) */

<rdf:description rdf:about="www.oclc.org/cat#ISBN0126542619">
  <bib:title> "Organizing information"</bib:title>
  <bib:creator>#Soergel</bib:creator>

<rdf:description rdf:about="#ISBN081086007">
  <bib:title> "Cataloging and classification"</bib:title>
  <bib:creator>www.simmons.edu/faculty#Chan</bib:creator>

<rdf:description rdf:about="www.oclc.org/cat#ISBN083893529X">
  <bib:title> "Anglo-American cataloging rules. 2. rev. ed."</bib:title>
  <bib:creator>#ALA</bib:creator>

<rdf:description rdf:about="www.oclc.org/cat#ISBN083893529X">
  <bib:title> "Information retrieval"</bib:title>
  ** <bib:creator>#Lancaster</bib:creator>
</rdf:Description>
</rdf:RDF>

```

** This line will be rejected since Lancaster has not been defined as a person

Lecture 7.2 is an exercise in descriptive cataloging, starting on the next page

In many instances I have printed out and included what the system puts out to save you time. If you retrace these steps to get the output yourself it will take longer

Midterm exam sample questions

There will be 5 questions. 15-20 minutes = number of points each, for a total of 90 minutes.

- (20) 1. A congregation has a calendar database, which lists for each event (religious service, committee meeting, etc.) the title, time and place, and main participants (celebrant, speaker, etc.). They also have a separate calendar with the dates and names of holy days. They need to produce the following documents.
- 1a A congregational bulletin that lists the week's events.
 - 1b A Web page that is meant to attract prospective members and that gives the week's events.
- 2 For each service a leaflet with the program (date and name of holy day, scripture passages, hymns, main participants).

Sketch a document template for document 1a or 1b and for document 2.

- (15) 2. Describe the process of developing a conceptual data schema for an employment service (an information system on jobs/positions and on people seeking jobs with the purpose of finding matches between jobs and people). Illustrate your discussion with examples, including some of the key entity types and relationship types.
- (15) 3. Assume you have
- a database of recipes that gives ingredients (basic foods) and their amounts for many prepared dishes, and
 - a nutrient database that gives for each basic food the amount of each nutrients (proteins, carbohydrates, fats, vitamins and minerals) it contains.
- You want to find all prepared dishes that contain both vitamin A and vitamin D. Describe a process that could be followed by an automated retrieval system to accomplish such a search.
- (20) 4. There will be a question on restructuring a semantic network with application to bibliographic records.
- (20) 5. This question has to do with possible improvements in retrieval through linguistic techniques. Below is an example consisting of a query and some short passages of text. Assume a straightforward free-text search system that searches for **words**; all passages that contain all query words joined by AND are retrieved. As a refinement, the system provides the proximity operator **WS**, which means **within the same sentence**. Thus, the query formulation *forest WS fire* retrieves all passages in which the two words occur in the same sentence; such passages are considered **retrieved**. See the instructions on the next page.

In the table of passages below, check all passages that are relevant to the user's need as expressed in the query. Then check all passages that are retrieved by the query formulation. Based on these checks, fill in the 3x3 grid and compute recall and discrimination. What refinements could be used to improve retrieval performance? For each refinement, list the passages whose retrieval is affected (give passage numbers). Which of these refinements improve recall, which improve discrimination, and which improve both?

Query topic: Forest fires

Query formulation: forest **WS** fire* (fire* finds fire or fires)

| Passage | R e l e v a n t | R e t r i e v e d |
|--|--------------------------------------|---|
| 30 Forest fires in Indonesia cause serious air pollution in South East Asia. | | |
| 31 The fire in Yellowstone Park destroyed 25% of the forest. | | |
| 32 The fire station is located behind the city forest. | | |
| 33 With fire in her eyes she chased him through the forest. | | |
| 34 The soldiers opened fire into the forest. | | |
| 35 The fire went out of control. It reached the forest and destroyed many acres. | | |
| 36 The animal got scared by the fire burning in the field. It ran into the forest. | | |
| 37 He asked whether he should fire the forest workers. | | |
| 38 Many square miles of forest in the West are burning. | | |
| 39 The dry wooded area went up in flames. | | |

| | Relevant | Not relevant | All |
|---------------|----------|--------------|-----|
| Retrieved | | | |
| Not retrieved | | | |
| All | | | |

Part 4*March 7 - April 18***Part 4. Classification and subject access**

| | |
|-------------------|--|
| Objectives | <ol style="list-style-type: none">1 Understand the problems of vocabulary control. Be able to apply vocabulary control principles to indexing and searching.2 Be able to extend vocabulary control principles to entity types other than Subject, such as names of organizations.3 Understand the pervasive role of classification throughout the human endeavor.4 Understand the functions of classification in information retrieval systems, especially request-oriented indexing and inclusive searching.5 Understand the principles of the structure of subject classification, in particular facet organization and hierarchy, and be able to apply these principles to the analysis of existing schemes and to indexing and query formulation.6 Be able to discern the facet structure of a domain. (There are facets everywhere.)7 Be aware of the variety of classification schemes and other subject access vocabularies and have an acquaintance with the major subject access vocabularies used in American libraries, namely the Dewey Decimal Classification, Library of Congress Classification, and Library of Congress Subject Headings.8 Be able to extend classification principles entity types other than Subject, for example to a hierarchy of organizations and organizational units. |
|-------------------|--|

| | |
|-------------------------------|---|
| Practical significance | <ul style="list-style-type: none">• The practical significance of vocabulary control in indexing and, more importantly, in free-text searching is detailed in Lecture 8.1.• The multiple and pervasive uses of classification have been detailed in the reading for Lecture 8.1. Also remember Lectures 2.1-2.2, The nature of knowledge and knowledge representation.• For IR systems specifically: The index language – the set of subject descriptors used in an IR system and their interrelationships – underlies all activities in subject retrieval. Understanding index language functions and structure - facet structure and hierarchy - is, therefore, at the heart of understanding IR systems. |
| Cross-reference | Lectures 2.1 and 2.2. The nature of knowledge and knowledge representation |

Lecture 8.1 / Small Groups 1

March 7**Explorations in subject access** (based on Assignment 10) (120 min)
www.dsoergel.com/571/UBLIS571Lecture08.1Slides.zip

| | |
|--------------------------------|--|
| Objectives | <p>1 Through your own analysis and discussion, you should arrive at an appreciation of the complexities of subject access and identify the major problems. You do this by working through realistic examples.</p> <p>2 This practical experience and problem awareness form the basis for the treatment of solutions at a more theoretical level in lectures and readings.</p> <p>Note: We have not yet discussed nor have you read about what solutions might exist for these problems. The whole point of this small group exercise is for you to think on your own and figure out solutions to subject access problems yourself. In the remainder of the semester, we will address each problem in turn.</p> |
| Tasks | <p>There are 5 tasks, all exploring problems of the index language in an ISAR system. Tasks a - c deal with a sample collection of transportation documents constructed for this assignment. Tasks d and e deal with the Library of Congress Subject Headings (LCSH) and the Library of Congress Classification (LCC), respectively.</p> <p>Task a Formulate queries to search the rough alphabetical index</p> <p>Task b Design a good alphabetical index</p> <p>Task c Establish an index language for a computerized IR system (no more than 100 descriptors)</p> <p>Task d From the Library of Congress Subject Headings (LCSH) extract a list of headings dealing with Transportation and Traffic.</p> <p>Task e From the Library of Congress Classification (LCC) extract a list of classes dealing with Transportation and Traffic.</p> |
| Materials (attached) | <ul style="list-style-type: none"> • Rough alphabetical subject index to the sample collection • Excerpts from the Library of Congress Subject Headings (LCSH) • Excerpts from the Library of Congress Classification (LCC) <p>Note: LCSH and LCC are two different systems for subject access. The Library of Congress Subject Headings are an alphabetical list of descriptors (subject headings) used in online searching or to arrange a card catalog. The Library of Congress Classification is a systematic arrangement of descriptors for all areas of knowledge; it is used to arrange books on shelves or links on a subject directory web page. At the Library of Congress, one cataloger assigns one or more LC subject headings to a book, another the one LC class that best fits.</p> |

In this small groups exercise emphasis is on identifying the problem and possible solutions. There is not enough time to actually implement a solution. We will do just enough practical work to understand the problem.

Tasks a-c: **Dealing with the sample collection**

Tasks a-d deal with subject access problems in a sample collection created by other students assigning subject descriptors to about 200 documents (six for each student) without any guidance whatsoever. The rough alphabetical index is simply an alphabetized list of the terms assigned in this "instructionless" indexing, with document numbers following each term. (Many terms have been culled to make the index shorter.)

Task a: Identify the terms that should be used in the rough alphabetical index to search for the following queries (no more than 30 terms for each query): (15 min.)

1. Harbors for large tankers.
2. Air cushion craft.
3. The consequences of the development of new types of vehicles for terminal design.
4. Simulation of passenger flow over transportation networks

Task b: Design a good alphabetical index to the sample collection (20 min). Your experience with Task b revealed problems with the rough alphabetical index. How would you design an alphabetical index that would address these problems and make searching easier? Start making some revisions to the index, just enough to get some experience that enables you to address the following points:

- ┆ What should a good alphabetical index look like?
- ┆ How would you go about transforming the rough index into a good index?

Task c: Establish an index language for a computerized ISAR system in the field of transportation (35 min). Assume that the list of terms in the rough alphabetical index is representative of the topics to be searched. So that users can remember all descriptors, the index language is limited to 100 descriptors, yet must allow searching for most of the concepts expressed in the rough alphabetical index without loss of specificity. Remember from Chapter 11 how a computerized ISAR system can be searched (remember its manipulative power.) See whether you can come up with an idea on the nature of such an index language and apply it to a few examples.

Result of the discussion: A sketch of what the index language might look like

Tasks d-e: Dealing with existing systems used in American libraries.

The Library of Congress, and many libraries in the US and around the world (esp. academic and research libraries), use **two separate schemes for subject access**.

- 1 The first is an alphabetical list of subject terms, the **Library of Congress Subject Headings (LCSH)**. These are used to index books, originally for an alphabetical subject catalog on cards and now for search in an **Online Public Access Catalog (OPAC)**. Usually several subject headings are assigned to a book to provide multiple access points. There are about half a million subject headings, listed in four large red volumes (to be found in Baldy 14A, older editions are fine)

- 2 The second is a systematically arranged scheme of subject classes, the **Library of Congress Classification (LCC)**; each class is identified by a class number marking its place in the classified arrangement, for example, BJ 2139 *Etiquette for airplane travel*. LCC is used for the systematic subject arrangements of books on the shelves. Since books are customarily shelved in only one place (even if there are multiple copies), only one LCC class is assigned to a book, providing only one access point. For example, consider a book titled *The history of State Street in Boston, 1870-1930*. This book could be classed under F73.5 *History of Boston > 1865-1950* or under F73.67.P3 *History of Boston > Streets. Bridges. Railroads > Park Street*, but the cataloger has to choose one of these. The 400,000 LCC classes are listed in 30 volumes of *classification schedules* (McKeldin Library Government Documents SU Docs LC 26.9 and McKeldin Reference Z696.U5)

The task is assembling a list of headings from LCSH and a list of class numbers from LCC that deal with transportation and traffic. These lists can serve the following purposes:

- ! A list of subject headings or a classification, respectively, for a transportation library.
- ! A query formulation to regularly search the OCLC WorldCat database for new records in the area of transportation as the basis for book selection in a transportation library. All elements of your list would be connected by OR and the resulting query formulation would retrieve all items on transportation either based on the subject headings or based on the class numbers assigned.
- ! An aid to a user who wants to search a general catalog for transportation topics. Such a user can find the appropriate subject heading(s) or class numbers much more quickly in a selected list than in the full LCSH or full LCC.

The task is the same as that of any user approaching the subject catalog or the shelves with a question, only magnified by the breadth of your topic.

You can get a feel for both schemes from the sample pages in the assignment materials, but you should also look at the actual schemes. Look at the LCC Outline (a thin separate volume).

Task d: Think about what you would do to put together a list of LCSH headings dealing with **Transportation and traffic**. (25 min) (Start with *Transportation*; you might also try *Ships*, *Railroads*, and other broad terms. Explore from there.)

A list of 15 relevant subject headings, at least 5 of which are not simply taken from the cross-references listed under a very broad heading such as *Transportation*, *Ships*, *Railroads*, or *Air transport*. (Cross-references are the Broader Term (BT), Narrower Term (NT), and Related Term (RT) cross-references given in LCSH, as well as the USE and UF cross-references, which are not of interest here.) You may not include subject headings that start with either *transportation* or *traffic*.

Task e: Think about what you would do to put together a list of classes dealing with **Transportation and traffic** (25 min). Can you restrict your efforts to one or two volumes (After all, a classification is supposed to bring all related subjects together)? Why not?

In a given branch of the hierarchy always list the broadest class that falls under transportation, for example, do not list *TF840-851 Technology > Railroads > Elevated railways and subways*, but go up to the broader level still included in transportation, *TF?? Technology > Railroads*.

Rough alpha index

based on “instructionless” indexing of a sample collection on transportation by students in previous classes

For some terms, the document number got lost but the terms are still important

Access criteria

Access criteria 182
 Access study
 Accessibility 92
 Accommodations 100
 Advance acquisition 92
 Aerial camera system 160
 Aerial car transit(act) 195
 Aerial photography 160, 179
 Aero engines 188
 Aerodynamic 79
 Aerodynamics 103, 379
 Aerodynamics improvements 145
 Aesthetics 92
 Air cargo 218, 376, 76
 Air cargo flow 127
 Air cargo traffic 127
 Air cushion craft 109
 Air cushion vehicles 150, 188
 Air flight paths 337
 Air force 155
 Air freight 127, 128
 Air freight directory 385
 Air freight statistics 385
 Air passengers 200
 Air pressure 89
 Air resistance 83
 Air rights 185
 Air shipping statistics 84
 Air terminal 16
 Air traffic 16, 162, 316, 346
 Air traffic control 13, 16, 67, 133, 325, 346, 367, 372
 Air traffic control -SST 194
 Air traffic routes 342
 Air traffic automation 72
 Air traffic -systems and methods 72
 Air transport 65, 204

Airport

Air transportation 29, 45, 91, 93, 110, 372, 373, 374
 Air transportation policy 13
 Airbus 91
 Aircraft 58, 84, 85, 103, 118, 138, 163,211,354,384
 Aircraft accommodation 204
 Aircraft design 199, 358
 Aircraft development 169
 Aircraft developments 85
 Aircraft engineering 193
 Aircraft in Japan 46
 Aircraft maintenance 98
 Aircraft marketing research 169
 Aircraft navigation 358
 Aircraft production 98
 Aircraft tests 54
 Aircraft-wingless 163
 Aircraft--design and construction 47
 Aircraft--future 46, 47
 Airfield master plans 215, 219
 Airline fares 73
 Airline flight paths 37
 Airline networks 37
 Airline operations 113
 Airline parking, new plans 325
 Airline passenger models 337
 Airline passenger transportation 37
 Airline rates 143
 Airline route maps 93
 Airline schedules 73
 Airline schedules and fares 373, 374
 Airline scheduling 37
 Airlines 143
 Airlines-timetables, maps, schedules 74
 Airplane design 128
 Airplanes-passenger information 73
 Airport 162

Airport access

Airport access 31, 38, 42, 52
 Airport accessibility 201
 Airport accessibility 174
 Airport design 128, 180, 201, 360
 Airport facilities 174
 Airport finance 180
 Airport interface 16, 316
 Airport locations 113
 Airport modernization 60
 Airport parking facilities 203
 Airport planning 52, 113, 201,
 Airport redevelopment 203
 Airport roads 154
 Airport satellite terminals 197
 Airport terminal design 200, 203
 Airport terminals 67, 154, 197, 200
 Airport traffic flow 203
 Airport planning 219
 Airports 16, 134, 136, 154, 186, 200, 201
 Airports - access 342
 Airports-England cargo transport 209
 Airports - Baltimore 38
 Airports - parts 197
 Airports --Washington D.C. 38
 Alaska 136
 Alaska railroad 108
 All-cargo airline 76
 All-cargo airport 76
 All-cargo airports 376
 American carriers 53
 American inland waterways 167
 American Lloyds 380
 American Merchant Marine 380
 Aquadromes 121
 Attractiveness criteria 182
 Automated air traffic control 372
 Automated freight ferry 178
 Automatic traffic control 124
 Automatic train control 356
 Automatic train control system 115
 Automatic train operation 356, 56
 Automobile accidents-causes 51
 Automobile design 168

Commercial aircraft survey

Automobiles 153
 Automobile-pleasure driving 75
 Automotive transport 177
 Aviation 65, 100
 Aviation expansion 13
 Baltimore 52
 Baltimore airport 42
 Bay Area rapid transit 99, 115
 Berth occupancy 90
 Berths 90
 Boat facilities 361
 Boating 355, 55
 British Aircraft Indus. 106
 Bus commuting 171
 Bus companies 359, 59
 Bus guide 59
 Bus line reorganization 151
 Bus route design 151
 Bus routes 96
 Bus schedules 59, 96, 151, 359
 Bus service patterns 151
 Bus transportation 110, 151
 Buses 137
 Business jets 40
 Canada
 Canada-freight stations 59
 Canadian carriers 53
 Canadian Inland Waterways 167
 Canadian National Railways 364
 Canadian Railway equipment register 122
 Car ownership 144
 Cargo airlines 376
 Cargo tariffs 376
 Cargo transport 82, 84, 150, 376
 Cargo transport, land 153
 Cargo transportation 36
 Cargo-handling 376
 Coastal Oil Terminal 318
 Coastal Oil Terminals, Italy 18
 Commercial air routes, 17
 Commercial aircraft 54, 186, 204
 Commercial aircraft projections 204
 Commercial aircraft surveys 204

Commercial airliners

Commercial airliners 17
 Commercial jet airplanes 214
 Commercial ports 44
 Commercial shipping- Great Lakes 198
 Commercial train equipment 320
 Commodities shipped 32
 Commodities transport 202
 Common carriers 218
 Communication Center 112
 Communication control 367
 Communications 67, 323
 Commuter aircraft 354
 Commuter parking 183
 Commuter railroads 345
 Commuter transit 86
 Construction 12, 101, 139
 Container docks 125
 Container services 119
 Container shipping
 Containers 125, 127, 178, 180
 Contract motor carriers 366
 Controlled access 92
 Costs 81, 134
 Covered freeways 185
 Criteria 104
 Cross-country 100
 Design 101, 104, 12, 138
 Design concepts 10
 Design of pedestrian tunnels 117
 Design of V-STOL air-craft 17
 Design trends in aircraft 116
 Design variable 182
 Developing countries 45, 81, 381
 Developing country 135
 Development 92
 Diesel 102, 137
 Diesel railway traction 324
 Diesel ships 192
 Diesel-electric ferries 148
 Directories 53
 Directory
 Dock development 119
 Domestic airlines 218
 Domestic transport 82

Freight ferry

Downtown parking areas 183
 Downtown parking systems 207
 Downtown traffic 363
 Driving --requirements for 51
 Economic considerations 175
 Economic development 76
 Economic development of air cargo 376
 Economic effects 382
 Economics 138
 Electric 102
 Electric automobile 168
 Electric motor coaches 88
 Electric vehicle design 370
 Electric vehicles 173
 Engines - marine diesel 210
 Environment 15
 Equipment 102
 Facilities 100
 Feasibility study 58
 Ferries 148, 157
 Ferry design 148
 Ferry engine 148
 Ferry ship Construction 148
 Ferry terminals -design 146
 Ferry terminals - Grimsby 146
 Ferry terminals Immingham 146
 Ferry terminals - Woolwich 146
 Flight operations 98
 Flight safety 98
 Flight schedules 337
 Flight study and training 98
 Flights 337
 Floating airport 121
 Floating platforms 121
 Forecasting 141
 Foreign airlines 218
 Freeway 130
 Freeway design 369
 Freeway ramp traffic 124
 Freeway Systems design. 35
 Freeway systems engr. 35
 Freeway volume 124
 Freeways 69, 92, 159, 335, 369
 Freight ferry

Freight stations

Freight stations 164
 Freight trains 172
 Fuel transport 172
 Future airport planning 316
 Future outlook 127, 128, 129
 Future rail services 184
 Future terminals 170
 Future trip length problems 144
 Garage building 339
 Garages --Boston 39
 Garages --Cost analysis 39
 Gas turbine powered trains 364
 General cargo ports 90
 General Motors Corp. 137
 Glideway system 43
 Glideways 343
 Glideways - economics 343
 Glideways - political implications 343
 Glideways - vehicle design 343
 Government aid 175
 Government legislation 382
 Gravity vacuum transit(GVT) 195
 Great Lakes 212, 217, 322
 Great Lakes guide 61
 Great Lakes Inland Waterways 355
 Great Lakes transportation 319
 Great Lakes Waterways traffic studies 361
 Ground services 100
 Ground transportation 89
 Ground-effects 103
 Guidelines for aircraft selection 354
 Hangar design 180
 Harbor design 107
 Harbor radar design 112
 Harbors 14, 104, 139, 198
 Heavy carriers 218
 Helicopter 118, 377
 Helicopter adaptability 33
 Helicopter future role 118
 Helicopter lines 94
 Helicopter operational history 181
 Helicopter transport 181
 Helicopters 116, 326, 33, 94

International passenger transport

High rail transport 341
 High speed 89
 High speed passenger rains 383, 386, 387
 High speed trains 140, 362
 High speed transportation system 182
 High speeds 88
 High-speed 79
 High-speed ground transport 343
 High-speed trains 156, 379
 High-speed transportation system 43
 Highway capacity 153
 Highway construction 120, 190
 Highway design 158, 190
 Highway engineering 190
 Highway operation 190
 Highway planning 45, 57, 63, 123, 165,
 Highway system 386
 Highway travel 196
 Highway use 153
 Highways 101, 81
 Highways-finance 208
 Highways - India 165
 Highway-transit interchange 177
 History 76
 Hovercraft 150, 152
 Hovercraft apron design 152
 Hovercraft design 150
 Hovercraft terminals 152
 Hovering 103
 Hovering aircraft 150
 Hovering capability-V/STOL 377
 Hovering YS .cruising speed 80
 Hoverports 152
 Inland ports 167
 Inland water terminals 167
 Inland water transportation 382
 Inland waterways 167,55
 Instrument landing systems 325
 Intercity 132
 Intercity jet 316
 Intercity transportation 58,62
 Internal roadway systems 189
 International developments 362
 International passenger transport 373

International railroads

International railroads 166
International survey 112
Interstate commerce 66, 366
Interstate highway system 208
Interstate waterways 167
Interurban transportation systems 195
Inter-state highways 120
Intracity transportation 58
Jet airplane design 214
Jet transport plane 346
Jumbo jets 91, 170, 180, 197
Land approaches to airport 16
Land requirements 92
Land use, transportation 152
Landing strip design 170
Large capacity aircraft 138
Linear capacity tankers 139
Line-haul 78
Line-haul passenger transportation 378
Liverpool dock facility 119
Locomotive designers 324
Locomotive engineers 324
Locomotives 88, 102
Manhattan Island Aquadrome 121
Manufacture-electric vehicles 70
Marine Services 167
Marine supplies 167
Maritime guides 217
Market 131
Mass transit 1 77
Mass Transit 34, 35, 171, 335, 387
Mass transportation 29, 345
Maximum terminal capacity 189
Merchant Marine directory 77
Methods 79
Methods of cargo handling 76
Metropolitan areas 174, 177
Metroport 316
Middle-range aircraft 169
Military Aircraft 155
Modal interchange zones 219
Model experiments 79
Moderate size airliner 193
Modern rail transport 206

Pipelines

Modern transportation 168
Modernization 60
Modifications 79
Motor carriers 202
Motor coach guide 59
Multimode transportation 185
Municipal buses 96
Municipal garages 339
Ocean liners 71, 371
Oil pipelines 28, 318
Oil terminal design 318
Operating problems 327
Operations research analysis 78, 378
Optimal transport aircraft 175
Overseas trade 349
Parking habits 183
Parking innovations 147
Parking lots 363
Parking needs 147
Parking policy 147
Passenger aircraft 138
Passenger airlines 143, 145
Passenger concept 10
Passenger flow 128
Passenger service 96
Passenger ship 95
Passenger ship piers 176
Passenger ship terminals 176
Passenger train 132
Passenger train car equipment, private car lines 20
Passenger train car equipment, railroads 20
Passenger train car officers 20
Passenger train engineering 19
Passenger train equipment 320
Passenger transport 83, 87, 150, 151, 197
Passenger transportation 78, 346
Pedestrian traffic flow 117
Performance 326
Performance model 81
Performance studies 85
Piggyback operations 111
Piggyback traffic 111
Pipelines 28

Planning

Planning 15
 Planning -rural transportation 165
 Port accommodations and facilities 44
 Port authorities 90
 Port dues and charges 44
 Port exports and imports 44
 Port remodeling 131
 Ports 344, 349, 90
 Ports - accommodations and facilities 344
 Ports - directories 344
 Ports - dues and levies 344
 Ports - exports 344
 Ports - imports 344
 Potential 76, 80
 Private train equipment 320
 Prototype-trains 379
 Rail installations 41
 Rail transit 130
 Rail transport 88, 206
 Rail transportation 374, 383, 386, 387
 Railroad car design 64
 Railroad construction 41
 Railroad freight 111
 Railroad freight equipment 122
 Railroad industry 149
 Railroad market guide 206
 Railroad personnel 206
 Railroad signals 41
 Railroad suppliers 149
 Railroad travel 86
 Railroads 64, 149, 156, 166, 184, 206, 341
 Railroads - construction 341
 Railroads - traffic control 341
 Railroads-timetables, maps, schedules 74
 Railway crossings 97
 Railway directories 149
 Railway engineering 34, 62, 362
 Railway finance 34
 Railway finances 334
 Railway locomotives 324
 Railway operating results 334
 Railway schedules and fares 374
 Railway statistics 34, 149
 Railway, underground--design 50

Seadrome design

Railway, underground--Stockholm 50
 Railways 166, 334
 Railways-electric 149
 Railways, electric 99
 Railways, high speed 97
 Ramp entry systems 171
 Rapid rail transit 129
 Rapid tramway-Europe 213
 Rapid transit 56, 86, 87, 130, 171, 327, 356
 Rapid transit construction system 130
 Rapid transit systems 99, 386, 387
 Rapid transit--design and construction 50
 Recent developments 76
 Regional highway planning 123
 Regulations 12
 Research 132, 133
 Rivers 14
 Road capacity 177
 Road design 159
 Road planning 348
 Road planning--condition rating factors 48
 Road planning--rural 48
 Road planning--service rating factors 48
 Road ratings 348
 Road traffic surveys 135
 Road transportation 370
 Road way functions 351
 Roads 158
 Rolling stock 126, 41
 Rotary wing aircraft 118
 Rotary-wing aircraft 116
 Route networks 96
 Runway design 136
 Rural arterial roads 348
 Rural highways-planning 16
 Rural transportation-planning 165
 Safety 91, 112, 326
 Satellite 133
 School transportation 151
 Sea carriers 205
 Sea freight 205
 Sea transport 205
 Sea transportation 32, 371, 374
 Seadrome design 113

Shared facilities

Shared facilities 174
Ship builder 77
Ship building 161,336
Ship data 36
Ship engineering 161
Ship information 380
Ship models 36
Ship names 77
Ship names list 71
Ship owner 77
Ship owners 32
Shipbuilders 71
Shipbuilding 157
Shipping 12,178
Shipping lines in the United States 32
Shipping register 380
Shipping schedules 32
Shipping-Great Lakes 198
Ships 90, 104
Ships-Canada 212
Ships-motor 210
Ships-U.S. 212
Ships--Merchant 36
Short haul air transportation 316
Short haul airliner 193
Short take off and landing 325
Short-haul traffic explosion 216
Short-haul utility aircraft design 216
Short- range transportation 58, 62
Size limits 11
Slow trains 140
Small commercial aircraft 134
Southern Inland Waterway 355
Space vehicles 188
Specialized carriers 218
Specialized trucking 202
Speed 326
Speed problems 79
Speed-volume-density-relationship 179
SST Flight Simulations 194
SST proposals 347
State-of-the-art 62
Statistics 141
Statistics for passenger liners 371

Traffic problems

Steam ships 192
Steamship schedules and fares 374
Steamships - timetables, maps; schedules 74
STOL 121, 186
Streets and highways 335
Subsonic aeronautics 109
Subsonic aircraft developments 216
Subways 129, 327, 356
Super express train 346
Super jet 170
Super sonic transport program 47
Supersonic aircraft 40, 194, 340
Supersonic transport 180, 194
Surface effects ships 12
Swing-way 40
Swing-wing aircraft 340
System analysis 132, 135
Systems analysis 113, 128
S.S.T. 136
Tank truck carriers 53
Tanker 131
Tanker owners 368
Tanker tonnage 368
Tankers 205, 368
Tankers; Sea transport 68
Technical data 188
Technological developments 383, 384
Terminal areas 215, 219
Terminal design 170, 29
Terminal facilities 203
Tilt wing 384
Timetable networks 140
Titanium alloy 347
Tokyo 101
Traffic 105, 141, 369
Traffic aids 179
Traffic aids 151, 153, 177, 186
Traffic control 56, 57, 69, 130, 151, 171
Traffic dynamics 179, 351
Traffic flow theory 369
Traffic lane detectors 1 71
Traffic overload 124
Traffic patterns 33
Traffic problems 153

Traffic projections

Traffic projections 196
Traffic service quality 351
Traffic variables 179
Traffic-intercity 196
Train 79
Train design 362
Train operation 41
Trains 62, 140
Transit systems 177
Transport aircraft 175
Transport capacity 177
Transport safety 83
Transport systems 28
Transport, individual 21
Transport, inter-city 21
Transportation 58, 60, 62, 113, 142, 143, 144, 145, 146, 147, 157, 166, 357
Transportation - innovation 345
Transportation engineering 351
Transportation lines 14, 19, 319, 322
Transportation network 189, 316
Transportation networks 115, 16
Transportation operations 322
Transportation research, government 21
Transportation research, private 21
Transportation Systems - U.S. 343
Transportation-air 211
Transportation-highways 208
Transportation-rapid tramway-Europe 213
Transportation-urban 213
Transportation-water 210, 210
Travel (mode of)--passenger preference 46
Trolley cars 327
Truck freight 111,187
Trucking 66, 202
Trucks 105, 153, 202
Tube vehicles 89
Turbine powered trains 364
Turbine-powered train 10
Turbo trains 142
Turbojet type 80
Turbo-train propulsion 191
Turbo-train structures 191
Turnpikes 141

Wingless aircraft

Underground railway 350
United States
United States-Freight stations 164
Unloading facilities 323
Urban freeway routes 185
Urban freeways 171
Urban highway planning program 123
Urban parking demand 207
Urban parking facilities 207
Urban rail systems 345
Urban renewal 185
Urban traffic engineering 207
Urban traffic flow 207
Urban transit, California 99
Urban transportation 15, 39
Urban transportation systems 195
Urban planning 32
U.S. ports 217
U.S. 13, 59
U.S. Air Service 365
U.S. Air Transport 365
U.S. Rapid Transit 129
Vehicle characteristics 381
Vehicle design 43
Vehicle detectors 179
Vehicle operating costs 381
Vehicle performance model 381
Vehicular parking spaces 321
Vertical-lift aircraft 181
Vessels 198, 322
VSTOL 186
VSTOL jet 316
VSTOL metroport 189
VSTOL 118
V-STOL 80
V-STOL aircraft 17
V/STOL 118, 384, 58
V/STOL aircraft 358, 377
Water transport 82, 205
Water transportation 95
Waterway guide 61
Waterway routes 32
Wind tunnel 379
Wingless aircraft 163

Pages from LCSH

Library of Congress Classification

Broad Outline (Main classes)

- A General works
- B Philosophy. Psychology. Religion
- C Auxiliary sciences of history
- D History: General and outside the Americas
- E-F History of America
 - E History: America General and United States General
 - F History: United States local, Canada, and Latin America
- G Geography
- H Social sciences
- J Political science
- K Law
- L Education
- M Music and books on music
- N Fine arts
- P Language and literature
- Q Science
- R Medicine
- S Agriculture
- T Technology
- U Military science
- V Naval science
- Z Bibliography and library science

The following pages give first a detailed outline and then examples of classes dealing with or relevant to *transportation and traffic*.

Library of Congress Classification. Detailed Outline

A General works

- AC Collections. Series. Collected works
- AE Encyclopedias (General)
- AG Dictionaries and other General reference works
- AI Indexes (General)
- AM Museums (General). Collectors and collecting (General)
- AN Newspapers
- AP Periodicals
- AS Academies and learned societies (General)
- AY Yearbooks. Almanacs. Directories
- AZ History of scholarship and learning. The humanities

B Philosophy. Psychology. Religion

B-BJ Philosophy, incl. BF Psychology

- B Philosophy (General)
 - BC Logic
 - BD Speculative philosophy
 - BF Psychology. Parapsychology. Occult sciences
 - BH Aesthetics
 - BJ Ethics. Social usages. Etiquette
- ### BL-BX Religion
- BL Religions. Mythology. Rationalism
 - BM Judaism
 - BP Islam. Bahaim. Theosophy
 - BQ Buddhism
 - BR Christianity
 - BS The Bible
 - BT Doctrinal theology
 - BV Practical theology
 - BX Christian denominations

C Auxiliary sciences of history

- C Auxiliary sciences of history (General)
- CB History of civilization
- CC Archaeology (General)
- CD Diplomatics. Archives. Seals
- CE Technical chronology. Calendar
- CJ Numismatics
- CN Inscriptions. Epigraphy
- CR Heraldry
- CS Genealogy
- CT Biography [General]

D History: General and Old World

- D History (General). Europe (General)
- DA Great Britain
- DAW Central Europe
- DB Austria, Hungary, Czech Republic, Slovakia
- DC France
- DD Germany
- DE Mediterranean region. Greco-Roman World
- DF Greece
- DG Italy
- DH Netherlands (low Countries). Belgium, Luxemburg
- DJ Netherlands (Holland)
- DJK Eastern Europe
- DK Russia and former Soviet republics. Poland
- DL Northern Europe. Scandinavia
- DP Spain. Portugal
- DQ Switzerland
- DR Balkan peninsula
- DS Asia
- DT Africa
- DU Oceania (South Seas) [Australia. New Zealand]
- DX Roma (Gypsies)

E-F History of America

- E1-143 America (General)
- E151-857 United States (Gen.)
- F1-957 United States: States and Local
- F1001-1140 Canada
- F1201- Other individual countries [mostly Latin America]

G Geography

G - GF Geography

- G Geography (General). Atlases. Maps
- GA Mathematical geography. Cartography
- GB Physical geography
- GC Oceanography
- GE Environmental sciences
- GF Human ecology. Anthropogeography
- GN Anthropology
- GR Folklore
- GT Manners and customs (General)
- GV Recreation. Leisure

H Social sciences

- H Social sciences (General)
 - HA Statistics
- ### HB-HJ Economics
- HB Economic theory. Demography
 - HC- Economic history and conditions
 - HD conditions
 - HE Transportation and communication
 - HF Commerce
 - HG Finance
 - HJ Public finance
- ### HM-HX Sociology
- HM Sociology (General and theoretical)
 - HN Social history. Social problems. Social reform
 - HQ The family. Marriage. Woman
 - HS Societies: secret, benevolent, etc. Clubs
 - HT Communities. Classes. Races
 - HV Social pathology. Social and public welfare. Criminology
 - HX Socialism. Communism. Anarchism

J Political science

- J General legislative and executive papers
- JA-JC Political science
- JA Collections and general works
- JC Political theory
- JF-JQ Political institutions and public administration
- JF General works. Comparative works
- JK United States
- JL Brit. America. Latin America
- JN Europe
- JQ Asia. Africa. Australia. Oceania
- JS Local government
- JV Colonies and colonization. Emigration and Immigration
- JX International law. International relations
- No longer used at LC

K Law

- K Law (General)
- KD United Kingdom and Ireland
- KDZ America. North America
- KE Canada
- KF United States
- KG Central America, Caribbean
- KH South America
- KJ-KK Europe

L Education

- L Education (General)
- LA History of education
- LB Theory and practice of educ.
- LC Special aspects of education
 - LD-LG Individual institutions
- LD United States
- LE America except United States
- LF Europe
- LG Asia, Africa, Oceania
- LH College and school magazines and papers
- LJ Student fraternities and societies in the United States
- LT [Multi-subject] Textbooks

M Music and books on music

- M Music [instrumental and vocal]
- ML Literature of music
- MT Musical instruction and study

N Fine arts

- N Visual arts (General)
- NA Architecture
- NB Sculpture
- NC Drawing. Design. Illustration
- ND Painting
- NE Print media
- NK Decorative arts. Applied arts. Decoration and ornament
- NX Arts in General

P Language and literature

- P Philology and linguistics (Gen.)
- PA Classical languages and lit.
 - PB-PH Modern European lang.
- PB Celtic languages and literature
- PC Romance languages
 - PD-PF Germanic languages
- PD Scandinavian. North Germanic
- PE English
- PF West Germanic
- PG Slavic. Baltic. Albanian languages and literature
- PH Finno-Ugrian. Basque I & I
 - PJ-PL Oriental languages & lit.
- PJ Oriental. Semitic
- PK Indo-Iranian
- PL Languages and literatures of Eastern Asia, African, Oceania
- PM Hyperborean, Indian, and artificial languages
 - PN-PZ Literature**
- PN Literary history and collections
- PQ Romance literature
- PR English literature
- PS American literature
- PT Germanic literature
- PZ Children's literature

Q Science

- Q Science (General)
- QA Mathematics.
 - [Computer science]
- QB Astronomy
- QC Physics
- QD Chemistry
- QE Geology
 - QH-QR Biology
- QH Natural history (General). Biology (General)
- QK Botany
- QL Zoology
- QM Human anatomy
- QP Physiology
- QR Microbiology

R Medicine

- R-RL Medicine**
- R Medicine (General)
- RA Public aspects of medicine
- RB Pathology
- RC Internal medicine. Practice of Medicine
- RD Surgery
- RE Ophthalmology
- RF Otorhinolaryngology
- RG Gynecology and obstetrics
- RJ Pediatrics
- RK Dentistry
- RL Dermatology
- RM-RZ Allied disciplines**
- RM Therapeutics. Pharmacology
- RS Pharmacy and materia medica
- RT Nursing
- RV Botanic, Thomsonian, and eclectic medicine
- RX Homeopathy
- RZ Other systems of medicine [Chiropractic. Osteopathy. Mental healing]

S Agriculture

- S Agriculture (General)
- SB Plant culture
- SD Forestry
- SF Animal culture
- SH Aquaculture. Fisheries. Angling
- SK Hunting

T Technology

- T Technology (General)
- TA-TH General engineering and civil engineering**
- TA General
- TC Hydraulic and ocean eng.
- TD Environmental technology, sanitary engineering
- TE Highway engineering
- TF Railroads
- TG Bridge engineering
- TH Buildings
- TJ-TL Mechanical group**
- TJ Mechanical engineering
- TK Electrical engineering. Nuclear engineering
- TL Motor vehicles. Aeronautics. Astronautics
- TN-TR Chemical group**
- TN Mining, metallurgy
- TP Chemical technology
- TR Photography
- TS-TX Composite group**
- TS Manufactures
- TT Arts and crafts. Handicrafts
- TX Home economics

U Military science

- U Military science (General)
- UA Armies: Organization, description, facilities, etc
- UB Military administration
- UC Maintenance and transportation
- UD Infantry
- UE Cavalry, armor
- UF Artillery
- UG Military engineering. Air forces. Air warfare
- UH Other services

V Naval science

- V Naval science (General)
- VA Navies: Org., descr., fac., etc
- VB Naval administration
- VC Naval maintenance
- VD Naval seamen
- VE Marines
- VF Naval ordnance
- VG Minor services of navies
- VK Navigation. Merchant marine
- VM Naval engineering. Ship-building. Marine Engineering

Z Bibliography and library science.

- Z4-115 Books (General). Writing. Paleography
- Z116-659 Book industry & trade
- Z662-1000 Libraries. [Library science. Information science]
- Z1001-8999 Bibliography

Library of Congress Classification

Sample classes dealing with or relevant to *transportation*

The following list gives a sampling of LC classes dealing with or relevant for *transportation and traffic*. The example classes are in *italics*. For each, the hierarchical chain leading to it is given to provide a sense of context, but neighboring classes are shown only in a few cases for illustration. Some full pages from the classification are also included with examples underlined (unless the entire section is about transportation). The examples have been chosen to make it easy for you to detect patterns on your own.

B **Philosophy. Psychology. Religion**

BJ **Ethics**

- BJ1801-2195 . Social usages. Etiquette
- BJ 2137 . . *Etiquette of travel*
- BJ2139-2156 . . . *Special topics*
- BJ2139 *Airplane travel*
- BJ2140 *Bus travel*

BS **The Bible**

- BS1-680 . General (Whole Bible)
- BS410-680 . . Works about the bible
- BS620-672 . . . Auxiliary topics
- BS647-649 Prophecy
- BS649 Prophecy of special future events, A-Z
- BS649.S8 *Steam engines*

BV **Practical theology**

- BV5-530 . Worship (public and private)
- BV205-287 . . Prayer
- BV229-283 . . . Prayers
- BV283 Other special prayers, A-Z
- BV283.A4 *Air pilots' prayers*
- BV283.T7 *Traveller's prayers*
- BV590-1652 . Ecclesiastical theology
- BV900-1450 . . Religious societies, associations, etc.
- BV950-1280 . . . Religious societies of men, brotherhoods, etc.
- BV955-1280 By period
- BV960-1280 19th-20th centuries
- BV1000-1220 Young Men's Christian Associations
- BV1160-1220 Work with special classes
- BV1175 *Commercial travellers*
- BV1200 *Railroad employees*

- BV2002-3705 . Missions
- BV2610-2695 . . Special types of missions
- BV2660-2695 . . . Work among special classes, by occupation
- BV2695 Other classes, A-Z
- BV2695.R3 *Railroad men*
- BV4000-4470 . Pastoral theology
- BV4200-4317 . . Preaching. Homiletics
- BV4239-4317 . . . Sermons
- BV4309-4316 Sermons and talks to special classes of persons
- BV4316 Other classes, A-Z.
- BV4316.R3 *Railroad men*
- BV4316.S3 *Sailors and seamen*
- BV4400-4470 . . Practical church work. Social work. Work of the layman
- BV4435-4470 . . . Church work with special classes
- BV4457-4459 Soldiers and sailors
- BV4458 *Sailors and seamen*
- BV4485-5099 . Practical religion. The Christian life
- BV4527-4596 . . Religious works for special classes
- BV4588-4591 . . . Soldiers. Soldiers and sailors
- 4590-4591 *Sailors and seamen*
- 4596 . . . Other, A-Z
- 4596.R3 *Railroad men*

C Auxiliary sciences of history

- CB History of civilization**
- CB156 . *Terrestrial evidence of interplanetary voyages*
- CB440-481 . Relation to special topic
- CB440 . . *Astronautics and civilization*

- CJ Numismatics**
- CJ1-4625 . Coins
- CJ161 . . Symbols. devices, etc., A-Z
- CJ161.B2 . . . *Bridges*
- CJ 161.S5 . . . *Ships*
- CJ 161.T73 . . . *Transportation*

3 pages from F

H Social sciences

HD Economic history and conditions [See sample pages following]

HE Transportation and communications [See sample pages following]

HF Commerce

HF5001-6182 . Business

HF5601-5689 . . Accounting. Bookkeeping

HF5686 . . . By business or activity, A-Z

A list of seven pages, including

HF5686.A38 *Air transportation. Airlines*

Railways, see HE2241 [Accounting under Railways]

HF5686.T6 Tobacco

HF5686.T7 *Transportation*

Transportation, air, see HF5686.A38

Transportation, automotive,

see HE5618 Automotive transportation > Finance, accounting, etc.

Transportation, local,

see HE4351 Street railways. Subways. Rapid transit systems > Finance. Accounting. Auditing

HF5686.T73 Travel agents

HF5686.T8 Trustees

HJ Public finance

HJ2240-5957 . Revenue. Taxation. Internal revenue

HJ3231-3696 . . Taxation. Administration and procedure

[Note: Transportation taxes used to be here, but have been moved to

HE: HE196.9 [Taxation under Transportation in general] or

HE384+ [Control, taxation, tolls, etc. under Water transportation]

HQ The family. Marriage. Woman

HQ503-1064 . The family. Marriage. Home

HQ1060-1064 . . Aged, Gerontology (Social aspects). Retirement

HQ1063.5 . . . *Transportation*

HV Social pathology. Social and public welfare. Criminology

HV697-4959 . Protection, assistance, and relief

HV1551-3024 . . Handicapped

HV1568.6 *Transportation and travel*

For transportation of persons with specific handicaps, see the specific class of handicapped persons. [But not all have a subclass transportation.]

HV3011-3024 . . . Physically handicapped

HV3022 *Transportation and travel*

HV3025-3174 . . Special classes. By occupation

HV3025-3164 . . . *Seamen*

Sample pages from HD and HE, an even number, now 10 (2002)

Make this number of pages even

J**Political science****JF-JQ Political institutions and public administration**

- JK . United States
- JK401-1685 . . Government. Public administration
- JK468 . . . Other special, A-Z
- JK468.T7 *Transportation*

K**Law****KF United States — general**

- KF1600-2940 . Regulation of industry, trade and commerce. Occupational law
- KF2161-2849 . . *Transportation and communication*
[10 pages, divided by type of transportation, for example]
- KF2271-2379 . . . *Railroads* [with much detail]

KFC United States — California

- KFC390-547 . Regulation of industry, trade and commerce. Occupational law
- KFC469-543 . . *Transportation and communication*
[3 pages, divided by type of transportation, but less detail.]
- KFC490-499 . . . *Railroads* [only three classes under here]

L**Education****LC Special aspects of education**

- LC65-245 . Social aspects of education
- LC189-214.53 . . Educational sociology
- LC213-214.53 . . . Educational equalization. Right to education
- LC214-214.53 School integration
- LC214.5-53 Special means of integration
- LC214.5-53 *Transportation. Busing*

M**Music and books on music****M Music**

- M1497-5000 . Vocal music
- M1497-1998 . . Secular vocal music
- M1900-1980 . . . Songs (part and solo) of special character
- M1977-1978 By topic, A-Z
[A three-page list, including]
- M1977.R3 *Railroads*
- M1977.T87 *Truck drivers*

Two pages from T

U Military science

- UC** **Maintenance and transportation**
 UC270-360 . *Transportation*
[One-page broad classification of all modes of transportation, e.g.]
 UC310-315 . . *Railroads*

V Naval science

[Almost all of this is relevant, see the detailed LCC outline. Especially]

- VK** *Navigation. Merchant marine and*
VM *Naval engineering. Ship-building. Marine Engineering*
 [Both refer to civilian water transport]

Z Bibliography and library science.

- Z662-1000** **Libraries**
 Z665-718.8 . Library science. Information science
 Z675 . . Classes of Libraries, A-Z [Three-page listing, including]
 Z675.N3 . . . *Naval*
 Z675.T7 . . . *Transportation libraries*
 Z687-718.8 . . The collections. The books
 Z693-Z695.83 . . . Cataloging
 Z695.1 By subject, A-Z
[four-page list, including]
 Z695.1.N3 *Naval art and science*
 Z695.1.R34 *Railroads*
 Z695.1.T73 *Transportation*
 Z696-697 . . . Classification and notation
 Z697 By subject or form, A-Z [Two-page list, including]
 Z697.T7 *Transportation*

- Z1001-8999** **Bibliography**
 Z1001-1121 . General bibliography
 Z1201-4980 . National bibliography
 Z5051-7999 . Subject bibliography
Subjects arranged in alphabetical sequence [sic!]
 Z5811-14 . . Education
 Z5814 . . . Special topics, A-Z [Four-page list, including]
 Z5814.T7 *Transportation of pupils*
 Z7231-7234 . . *Railroads*
 Z8001-8999 . Personal bibliography
Names of individuals arranged in alphabetical sequence

Lecture 8.2a (in Small Groups 1)**March 7****Vocabulary control (terminological control)**

| | |
|-------------------------------|---|
| Objectives | <ol style="list-style-type: none"> 1 Understand the retrieval problems caused by terminological variety–synonymy and homonymy – in language, including any kind of names. 2 Understand and be able to apply vocabulary control to remedy these problems, either through vocabulary control in indexing or through query term expansion in searching. 3 Understand the structure of a thesaurus with its synonym-homonym structure (all terms), classificatory structure (concepts expressed by preferred terms), index language (concepts and corresponding preferred terms selected as subject descriptors), and lead-in vocabulary (all terms that are not subject descriptors). |
| Practical significance | <ul style="list-style-type: none"> • Authority control is applied to terms designating subjects, to names of persons and organizations, to titles of often cited or reprinted works, and in many other cases. It is a major principle underlying many information retrieval systems, especially those used in libraries. • Lack of vocabulary control and authority control more generally is one of the most serious problems impeding the success of end-user searching in free-text searching. The solution lies in the design of systems, including search thesauri, that can assist end users. |

| | |
|---------------------------|---|
| Wider applications | <p>Vocabulary control as a special case of authority control</p> <p>Vocabulary control is the control of subject identifiers. Similar problems arise in the control of the identifiers of other types of entities, such as persons or organizations; thesaurus of organizational names. In the broader sense one speaks of authority control (see Sections 9.1.1 and 9.1.2). The purpose of authority control can also be stated as referential integrity, that is, assuring a one-to-one correspondence between entity values and the character strings or other symbols that refer to them.</p> |
|---------------------------|---|

Review of Organizing Information, Chapter 12 (20 min)

Lexical relationships (10 min.)

Paradigmatic relationships: Synonymy, antonymy, hyponymy

In linguistics: Relationships between terms based on their meanings, that is, on the concepts they designate. If a term has multiple meanings, only one of these meanings participates in the relationships discussed here.

In classification theory/knowledge representation: Relationships between concepts in a classificatory structure.

Paradigmatic relationships are contrasted with **syntagmatic relationships** that bind together words into phrases and sentences or elemental concepts into compound concepts, statements, or larger units of meaning.

| | | | | | | | |
|----------------------------|--|----------------------------|---------------|--------------------|--------------------|-----------------|----------------|
| Synonymy | Two terms designate the same concept. True synonyms can be used interchangeably in sentences without changing the meaning. Core meaning and connotations. Problem of shades of meaning and connotation. | | | | | | |
| Antonymy | Two terms designate opposite concepts. Opposites can be endpoints of a scale, such as <i>light</i> and <i>dark</i> , or exclusive categories, such as <i>male</i> and <i>female</i> | | | | | | |
| Hyponymy | <p>Term A designates concept Concept A', term B designates concept Concept B', and Concept B' is more specific than Concept A'. Examples: <i>flute</i> (in one of its meanings) has as hyponym <i>recorder</i> (in one of its meanings); <i>keyboard instrument</i> has as hyponyms <i>harpsichord</i> and <i>cembalo</i>.</p> <p>Note: In a thesaurus with a controlled vocabulary we would select a preferred term, for example <i>harpsichord</i>, and have the relationships</p> <div style="text-align: center; margin: 10px 0;"> <table style="border: none; margin: auto;"> <tr> <td style="padding-right: 10px;"><i>keyboard instrument</i></td> <td style="padding-right: 10px;">Narrower Term</td> <td><i>harpsichord</i></td> </tr> <tr> <td><i>harpsichord</i></td> <td>Synonymous Term</td> <td><i>cembalo</i></td> </tr> </table> </div> <p>Antonymy and hyponymy are really concept relationships to be dealt with in Chapter 14; hyponymy is the relationship that defines a concept hierarchy. But all three relationships have in common that one term can be exchanged for the other in a sentence and still leave a sentence that has meaning.</p> | <i>keyboard instrument</i> | Narrower Term | <i>harpsichord</i> | <i>harpsichord</i> | Synonymous Term | <i>cembalo</i> |
| <i>keyboard instrument</i> | Narrower Term | <i>harpsichord</i> | | | | | |
| <i>harpsichord</i> | Synonymous Term | <i>cembalo</i> | | | | | |

Homonymy and polysemy

Note: The transition from homonymy to polysemy is gradual

| | | | | | |
|--|--|--|--|--|--|
| <p>Homonymy</p> | <p>Strict definition: two different words or phrases have the same spelling (homography) or the same pronunciation (homonymy in the narrowest sense).</p> <table border="1" data-bbox="548 443 1291 600"> <tr> <td data-bbox="548 443 919 600"> <p>Examples:</p> <p><i>seal (marine mammal)</i> <i>seal (document)</i></p> </td> <td data-bbox="919 443 1291 600"> <p><i>drill (bore a hole)</i> <i>drill (furrow)</i> <i>drill (fabric).</i></p> </td> </tr> </table> <p>Note: While they are spelled the same, the words in each group have different etymological origin.</p> <p>More expansive definition: The same word has two quite different meanings.</p> <table border="1" data-bbox="548 800 1291 957"> <tr> <td data-bbox="548 800 919 957"> <p>Examples:</p> <p><i>drill (bore a hole)</i> <i>drill (training)</i></p> </td> <td data-bbox="919 800 1291 957"> <p><i>seizure (disorder)</i> <i>seizure (law enforcement).</i></p> </td> </tr> </table> <p>Note: In each group, we have the same word (same etymological origin). The word acquired completely different meanings over time.</p> | <p>Examples:</p> <p><i>seal (marine mammal)</i> <i>seal (document)</i></p> | <p><i>drill (bore a hole)</i> <i>drill (furrow)</i> <i>drill (fabric).</i></p> | <p>Examples:</p> <p><i>drill (bore a hole)</i> <i>drill (training)</i></p> | <p><i>seizure (disorder)</i> <i>seizure (law enforcement).</i></p> |
| <p>Examples:</p> <p><i>seal (marine mammal)</i> <i>seal (document)</i></p> | <p><i>drill (bore a hole)</i> <i>drill (furrow)</i> <i>drill (fabric).</i></p> | | | | |
| <p>Examples:</p> <p><i>drill (bore a hole)</i> <i>drill (training)</i></p> | <p><i>seizure (disorder)</i> <i>seizure (law enforcement).</i></p> | | | | |
| <p>Polysemy</p> | <p>The same word has several meanings that can all be traced to a common core of meaning.</p> <table border="1" data-bbox="548 1199 1291 1430"> <tr> <td data-bbox="548 1199 1291 1430"> <p>Example:</p> <p><i>integration (mathematics)</i> <i>integration (psychology)</i> <i>integration (social groups)</i> <i>integration (economic-political)</i> <i>integration (curriculum)</i></p> </td> </tr> </table> <p>All these meanings share a common core meaning: putting together pieces into whole where the pieces are held together in a larger structure.</p> <p>Polysemy is often the result of metaphoric extension of the meaning of a term.</p> <table border="1" data-bbox="548 1604 1291 1797"> <tr> <td data-bbox="548 1604 1291 1797"> <p>Example:</p> <p><i>field (piece of land)</i> <i>field (subject)</i> <i>field (physics)</i> <i>field (mathematics)</i></p> </td> </tr> </table> <p>Even for <i>drill (bore a hole)</i> and <i>drill (train)</i> one can identify the core meaning of <i>repetitive and persistent performance of an operation.</i></p> | <p>Example:</p> <p><i>integration (mathematics)</i> <i>integration (psychology)</i> <i>integration (social groups)</i> <i>integration (economic-political)</i> <i>integration (curriculum)</i></p> | <p>Example:</p> <p><i>field (piece of land)</i> <i>field (subject)</i> <i>field (physics)</i> <i>field (mathematics)</i></p> | | |
| <p>Example:</p> <p><i>integration (mathematics)</i> <i>integration (psychology)</i> <i>integration (social groups)</i> <i>integration (economic-political)</i> <i>integration (curriculum)</i></p> | | | | | |
| <p>Example:</p> <p><i>field (piece of land)</i> <i>field (subject)</i> <i>field (physics)</i> <i>field (mathematics)</i></p> | | | | | |

Lecture 8.2b (in Small Groups 1)**March 7****Index language functions** (Organizing Information, Chapter 13) (60 min)

Subject analysis; abstracting and indexing; types and functions of abstracts

| | |
|-------------------------------|--|
| Objectives | <ol style="list-style-type: none"> 1 Understand the principle of request-oriented (user-centered) indexing and the fundamental role of the index language to communicate users' interests to the indexer. 2 Be able to make intelligent decisions about the type of index language, indexing, and query formulation to be used in a given IR system, considering costs and benefits. 3 Be able to recognize search requests that are difficult to handle in a system that does not use request-oriented indexing and be able to compensate, as far as possible, through creative pursuit of different avenues for the search. |
| Practical significance | Request-oriented indexing (also called problem-oriented indexing or user-centered indexing) is a special case of the maxim that the design and operation of information systems should be based on a thorough understanding of user requirements. Request-oriented indexing is the key to good system performance for the questions that matter to users. Yet in practice it is rarely used. Understanding this will enable students to make the best of existing systems and, more importantly, to go out and change practice. |
| Discussion question | How could request-oriented indexing be implemented in a reference tool addressed to a general audience, such as the <i>Reader's Guide to Periodical Literature</i> ? |

| | |
|--|---|
| <p>Chapter highlights</p> | <ul style="list-style-type: none"> • Derivation of the principle of request-oriented indexing from the problem-oriented approach to information systems introduced in Chapters 1 and 5 (Sections 13.1 and 13.2), • the role of index languages in searching and database organization (Sections 13.3 and 13.4), • design issues (Section 13.5), • review of index language functions (Section 13.6), • culminating in the recognition of an index language as a communication device from users to indexers, so that the indexers understand the users' interests (Section 13.7). <p>Terminology: Filtering technique of indexing (Mooers 1958), Request-oriented indexing (DS 1974), problem-oriented indexing (DS), user-centered indexing (term in vogue now).</p> |
| <p>Questions</p> | <p>Your questions here</p> |
| <p>Discussion question (repeated)</p> | <p>How could request-oriented indexing be implemented in a reference tool addressed to a general audience, such as the <i>Reader's Guide to Periodical Literature</i>?</p> |

Document representation: purpose, structure, process of creation

Abstracts as a different form of document representation

Indicative abstract - merely indicates what the document is about or relevant for, pointer data.

Informative abstract - in addition, includes some of the substantive data given in the document or reports some generalization that can be derived from the document.

Both types of abstract assist the reader in deciding whether to pursue the document further (and incur any costs in doing so). An informative abstract often gives the substantive data needed and thus saves the user the trouble of having to consult the document itself.

Other categorization of abstracts: Reporting vs. analytical-critical. Book reviews

The structure of document representations (abstracts or lists of index terms) discussed in the lecture on document structure. Use of controlled vocabulary

Abstracting and indexing as a cognitive process

Empirical study of document-oriented indexing

Parts of the document considered

Method of information assimilation (reading, interpreting pictures)

Reading/scanning to identify subject matter of interest to users — request-oriented reading

Reading/scanning to fill slots of a frame

Building up mental image

Selecting topics to be included in the abstract or the index terms. Request-orientation comes into play here as well

Choosing a form of expression

Knowledge brought to bear on these operations - from own knowledge or tools (such as thesauri) consulted, for example

- General knowledge of the field

- Knowledge of user needs

- Frames for phenomena in the field

- Knowledge of terminology

- Knowledge of document structure, including knowledge of cue words

Automatic or computer-assisted abstracting and indexing

Conigrave KM, Saunders JB, Reznik RB. **Predictive capacity of the AUDIT questionnaire for alcohol related harm.** Addiction 90 (1995) 1479-1485.

Indicative abstract

This study deals with early identification of alcohol use disorders. It examined the ability of the Alcohol Disorders Identification Test (AUDIT) questionnaire published by the World Health Organization to predict which subjects experience medical or social harm from their drinking. Subjects were 350 emergency room patients who answered the AUDIT questions as part of a comprehensive medical assessment. 250 subjects were interviewed after 2-3 years to determine alcohol-related medical disorders, health care utilization, social problems and hazardous drinking at the time of follow-up. Audit is compared to biochemical indicators for its ability to predict these conditions.

Informative abstract

'AUDIT can predict a range of harmful consequences of alcohol consumption'

Background. Drinking problems often are not recognized. Most of the people who become alcohol-dependent do not seek help until their problems are obvious. Late diagnosis is of particular concern because effective and low-cost methods of treating problem drinking at an early stage are now available. In 1989, the WHO published a brief 10-item screening questionnaire, the Alcohol Disorders Identification Test (AUDIT) specifically designed to identify problem drinkers before physical dependence or chronic problems have arisen. AUDIT has been reported to have a sensitivity of 92% and a specificity of 94% in detecting hazardous or harmful alcohol use. This study examined the ability of the AUDIT questionnaire to predict which subjects experience medical or social harm from their drinking.

Methods. Subjects were 350 patients who attended a hospital emergency ward in 1984-1985. They underwent a comprehensive assessment of medical history, alcohol use, dependence and related problems in an interview schedule; the AUDIT questions were interspersed among other items. Biochemical variables measured included γ -glutamyltransferase (GGT) and mean corpuscular volume (MCV). Twenty subjects refused to be contacted after 2-3 years or were excluded because of malignant disease. Thus, a cohort of 330 subjects (212 men, 108 women) was left for the longitudinal study; 250 subjects were interviewed again after 2-3 years. Interviewers were blind to the results of the initial assessment. The AUDIT questions were scored from 0 to 4. Subjects who scored 8 or more were classified as potentially hazardous drinkers. AUDIT was examined for its ability to predict a number of end-points including alcohol-related medical disorders, health care utilization, social problems and hazardous drinking at the time of follow-up.

Results. Of those who scored 8 or more on AUDIT at the initial interview, 61% experienced alcohol-related social problems compared with 10% of those with lower scores. They also reported more frequently alcohol-related medical disorders and hospitalization. The AUDIT score was a better predictor of social problems and of hypertension than laboratory markers. Its ability to predict other alcohol-related illnesses was similar to the laboratory tests, but GGT was the only significant marker of mortality.

Conclusions. AUDIT is a brief and convenient questionnaire which can readily be incorporated into the standard medical history. It can predict a range of harmful consequences of alcohol consumption. AUDIT should prove a valuable tool in screening for hazardous and harmful alcohol use so that intervention can be provided to those at particular risk of adverse consequences.

Lecture 9.1
Small Groups 2

March 21

Index language structure 1: conceptual (Organizing Information, Chapter 14)

| | |
|--|---|
| Objectives of Lectures 9.1-11.1 | Be able to use the entity-relationship approach, specifically facet analysis, to discern the conceptual structure of a subject. Put differently: Be able to discern the facet structure of a subject. |
| Practical significance | This understanding provides a basis for <ul style="list-style-type: none">• constructing an index language, a task required in setting up specialized information systems and, more importantly, in developing expert systems;• evaluating an index language to determine whether it is suitable for a given application;• indexing, particularly making sure that all applicable facets have been covered;• query formulation, facet analysis of queries. |

In-class exercises: Three steps in the **conceptual analysis and synthesis** in a subject:

- | | |
|---------|---|
| Step 1. | Semantic factoring (results in a list of elemental concepts). |
| Step 2. | Arranging the elemental concepts in a well-structured faceted hierarchy. |
| Step 3. | Fit compound concepts into the framework of the hierarchy (if compound concepts need to be dealt with explicitly) |

In-class exercise: Semantic factoring

Semantic factor the concepts (Dewey classes) from the attached list.
Keep a running list of elemental concepts as they arise.

In-class exercise: Building a hierarchy of elemental concepts

Sort elemental concepts into entity types or facets.
Arrange values within each entity type or facet in a meaningful structure.

In-class exercise: Fitting compound concepts into a hierarchy

In-class exercise, Organizing Information, Chapter 14: Semantic factoring

Factor the following concepts (from Dewey Decimal Classification) **into their semantic components (semantic factors)**. If this is not possible, comment.

Keep a running list of the elemental concepts needed.

Note: A broader class is given in () if necessary to specify the meaning of a term.

| | |
|-----------|---|
| 372.19 | Curriculums of elementary schools |
| 372.35043 | Science in the elementary school curriculum |
| 372.414 | Methods of instruction for reading in elementary schools |
| 372.72043 | Arithmetic in the elementary school curriculum |
| 373.19 | Curriculums in secondary schools |
| 373.243 | Military schools (Secondary Education) |
| 376.63 | Secondary education of women |
| 378.19 | Curriculum of colleges and universities |
| 378.33 | Fellowships (Higher Education) |
| 371.7 | School health and safety |
| 371.855 | Men's social societies and fraternities (Generalities of Education) |
| 371.856 | Women's social societies and sororities |
| 371.911 | Blind and partially sighted students |
| 371.912 | Deaf and hard-of-hearing students |
| 371.95 | Curriculums for gifted students |

In-class exercise: Building a hierarchy of elemental concepts

Sort elemental concepts into entity types or facets.

Arrange values within each entity type or facet in a meaningful structure.

| Elemental concepts Running list | Elemental concepts in a meaningful structure |
|--|---|
| | |

In-class exercise: Fitting compound concepts into a hierarchy

Lecture 9.2
Small Groups 2

March 21

Application of index language structure to searching
(Organizing Information, Section 14.4)

| | |
|--|--|
| Objectives Inherited from Lect. 9.1-11.1 plus these | 1 Understand inclusive searching (hierarchically expanded searching). 2 Be able apply this concept in searching any system. |
| Practical significance Inherited from Lect. 9.1-11.1 plus these | Inclusive searching is an essential technique for searches that emphasize recall. |

In-class exercise: Retrieval of documents in a sample collection

The sample collection consists of about 200 documents on transportation and traffic and is indexed using the index language shown on the following pages (same as the index language used in Assignment 11, *Request-oriented indexing*).

Query statement: I need information on vehicles used in rail transport

Query formulation: E6 Vehicles AND B1.2 Rail transport

Search in a printed index: Look for document numbers listed for both E6 Vehicles and B2 Rail transport (they are marked with = in the entry for B2).

E2.1 Vehicles 10, 12, 13, 24, 25, 26, 30, 36, 40, 46, 47, 50, 53, 54, 58, 59, 62, 64, 70, 76, 77, 79, 80, 81, 85, 91, 92, 94, 95, 100, 101, 102, 103, 104, 105, 106, 108, 109, 110, 116, 118, 121, 122, 126, 127, 132, 133, 134, 138, 148, 150, 151, 153, 155, 168, 169, 170, 171, 173, 174, 176, 178, 180, 181, 186, 187, 188, 191, 192, 193, 194, 199, 202, 204, 205, 207, 210, 211, 212, 213, 214, 216, 218, 219, 322, 330, 332, 333, 336, 337, 340, 346, 347, 353, 354, 355, 356, 357, 358, 362

B1.2 Rail transport =10, =24, 34, 41, 42, 44, =50, 89, 114, =126, 140, 149, 166, 184, =191, 195, =213, 310, 334

B1.2.1 Local rail transit 27, 56, 87, 99, =108, 111, 120, 123, 129, 130, =213, 327, 345, 350, =356

B1.2.2 Intercity railroads =10, 30, =46, =62, =64, =79, 82, 84, 97, =102, =108, 114, =122, =132, 156, 177, =213, 341, 362

Note: The classification shown on p. 292 includes just B1.2 Rail transport. Here the classification is expanded by adding

B1.2.1 Local rail transit (= B1.2 Rail transport : R1 Local systems)

B1.2.2 Intercity railroads (= B1.2 Rail transport : R2.1 Interurban systems)

These are precombined descriptors narrower than B1.2 by combination

Question: Did this search find all relevant documents?

Additional index entry to solve the problem

B1.2 Rail transport, inclusive =10, =24, 27, =30, 34, 41, 42, 44, =46, =50, 56, =62, =64, =79, 82, 84, 87, 89, 97, 99, =102, =108, 111, 114, 120, =122, 123, =126, 129, 130, =132, 140, 149, 156, 166, 177, 184, =191, 195, =213, 310, 327, 334, 341, 345, 350, =356, =362

Search in a peek-a-boo file (some samples distributed)

Like a printed index, but with more manipulative power. Today one uses computers.

Each descriptor has its own card. Each document number has a position on the card. In a printed index, the applicable document numbers are listed after the descriptor. In a peek-a-boo file, the applicable document number positions are punched on the descriptor card. In this particular implementation, document numbers are read off as follows: Find the column number (printed in tiny print), for example **12**). Find the row number (large single digits printed across each row, punched out if the position is punched), for example **6**. The document number is **126**. (One of the first uses of peek-a-boo cards: A bird guide. Retrieval of birds based on their features.)

This peek-a-boo file makes provision for inclusive searching: Each descriptor that has narrower descriptors under it has two cards: An **inclusive card** that includes all the documents from the narrower descriptors as well, and a **general references card** that includes only the documents indexed by the descriptor itself.

To find documents for

E2.1 Vehicles AND B1.2 Rail transport, inclusive

superimpose the two cards and read off the document numbers from the holes that still appear (document numbers in common to both cards).

We will broaden and narrow the search to observe the effects of hierarchy.

File builders and searchers classification display

The descriptors shown in italics with numbers D1.xx are precombined descriptors. Each system using this index language would decide whether to use these precombined descriptors (such as *D1.20 Aircraft*) or whether to index with the corresponding elemental descriptors (in the example D1 Air transport and E6 Vehicles) instead. **In assignment 11 only elemental descriptors are used for indexing.**

Outline: Facets

| | |
|----------|--|
| B | Mode of transportation |
| E | Transportation system elements |
| F | Power supply for vehicles |
| G | Type of propulsion |
| H | Materials to build facilities or vehicles |
| J | Passenger transport vs. freight transport |
| K | Traffic operations |
| L | Transportation providers |
| M | Creation of traffic systems and components |
| N | Organization, administration |
| Q | General and other concepts |
| R | Geographic range |
| S | Geographic location |

The three facets used for arrangement are shown in **bold**.

In the full display

+ signifies descriptors that have Narrower Terms under them

| | | |
|------------|--|-----|
| +A | Transportation and traffic, inclusive | |
| A | Transportation and traffic, general references | |
| +B | Mode of transportation, inclusive | |
| B | Mode of transportation, general references | |
| +B1 | Ground transport, inclusive | |
| +B1 | Ground transport, general references | |
| B1.1 | Road transport | 28* |
| B1.2 | Rail transport, inclusive | |
| B1.2 | Rail transport, gen. ref. | 25 |
| B1.2.1 | Local rail transit BT R1 ¹ | 26 |
| B1.2.2 | Intercity railroads BT R2.1 ¹ | 27 |
| B1.3 | Pipeline transport | |
| B1.4 | Pedestrian mode | |
| B1.5 | Multi-modal ground transport | |
| +B2 | Water transport, inclusive | |
| B2 | Water transport, general references | |
| B2.1 | Inland water transport | 31 |
| B2.2 | Ocean Transport | 32 |
| +B3 | Air transport, inclusive | |
| B3 | Air transport, general references | |
| B3.1 | Supersonic air transport | 35 |
| B4 | Air cushion transport | |
| B5 | Multi-modal transport | |
| B8 | Other specific modes of transportation | |
| B9 | Mode of transportation not applicable | |

D, E Free for expansion

| | | |
|------------|--|----|
| +E | Transportation system components, inclusive | |
| E | Transportation system components, gen. ref. | 15 |
| +E1 | Traffic facilities, inclusive | |
| E1 | Traffic facilities, gen. references | 17 |
| E1.1 | Traffic routes | 18 |
| E1.2 | Traffic stations | 19 |
| E1.3 | Stationary equipment | |
| +E2 | Methods to move persons or freight, incl. | |
| E2 | Methods to move persons or freight, g. r. | |
| E2.1 | Vehicles RT F, G | 20 |
| E3 | Containers | |
| E4 | Self-transport | |
| E8 | Other concepts | |
| E9 | Transp. system components not applicable | |

+ Inclusive (hierarchically expanded, finds documents on all narrower terms as well)

* Edge-notched card hole no.

¹ B1.2.1 (= B1.2 : R1) and B1.2.1 (= B1.2 : R2.1) are

| | | |
|--------------|---|--|
| +F | Power supply f. vehicles, incl. RT E2.1 | |
| F | Power supply for vehicles, gen. ref. | |
| +F1 | Hydrocarbons, inclusive | |
| F1 | Hydrocarbons, general references | |
| F1.1 | Gasoline | |
| F1.2 | Diesel fuel | |
| F1.3 | Hydrocarbons from renewable sources | |
| F5 | Electric power | |
| F6 | Nuclear power | |
| F7 | Animate power | |
| F8 | Other power supply | |
| F9 | Power supply not applicable | |
| +G | Type of propulsion, inclusive RT E2.1 | |
| G | Type of propulsion, gen. references | |
| +G1 | Engine, inclusive | |
| G1 | Engine, general references | |
| +G1.1 | Combustion engine, inclusive | |
| G1.1 | Combustion engine, general ref. | |
| G1.2 | Steam engine | |
| G2 | Turbines | |
| G3 | Walking | |
| G8 | Other type of propulsion | |
| G9 | Type of propulsion not applicable | |
| +H | Materials to build facilities or vehicles, inclusive | |
| H | Materials to build facilities or vehicles, g.r. | |
| H1 | Materials by composition, inclusive | |
| H1 | Materials by composition, gen. ref. | |
| H1.1 | Soils, aggregates | |
| H1.2 | Bitumen | |
| H1.3 | Cement, Concrete | |
| H1.4 | Ceramics, glasses | |
| H1.5 | Wood, paper | |
| H1.6 | Fibers, textiles | |
| H1.7 | Plastics | |
| H1.8 | Rubbers | |
| H2 | Materials by origin, inclusive | |
| H2 | Materials by origin, general references | |
| H2.1 | Petroleum products | |
| H3 | Materials by use, inclusive | |
| H3 | Materials by use, general references | |
| H3.1 | Marking or coating materials | |
| H3.2 | Adhesives, seals | |
| H8 | Other specific materials | |
| H9 | Material not applicable | |

precombined descriptors, narrower than B1.2 by combination

| | | | | | |
|------------|--|----|--------------|--|---|
| +J | Passenger vs. freight transport, incl. | | +Q | General and other concepts inclusive | |
| J | . Passenger vs freight transport, g.r. | 8 | Q | . General and other concepts, gen. references | |
| J1 | . Passenger transport | 9 | Q1 | . Traffic flow | |
| +J2 | . Freight transport, inclusive | | Q2 | . Simulation | 3 |
| J2 | . . Freight transport, general references | 11 | +Q3 | . System characteristic, inclusive | |
| J2.1 | . . Transport of material of heavy weight | 12 | Q3 | . . System characteristics, general references | |
| J2.2 | . . Transport of bulk material | 13 | Q3.1 | . . Noise, vibration | |
| J9 | . Passenger vs. freight transport not applicable | | Q3.2 | . . Pollution | |
| +K | Traffic operations, inclusive | | Q3.3 | . . Quality, performance | |
| K | . Traffic operations, general references | | Q3.4 | . . Durability, life, reliability | |
| +K1 | . Traffic communication, control, safety, I. | | Q3.5 | . . Demand, use | |
| K1 | . . Traffic comm., control, safety, gen. ref. | 4 | Q3.6 | . . Human characteristics | |
| K1.1 | . . Traffic communications | | Q3.7 | . . Community characteristics | |
| K1.2 | . . Traffic control | 6 | Q3.9 | . . Other system characteristics | |
| K1.3 | . . Traffic safety | 7 | +Q4 | . Small vs large capacity, inclusive | |
| +K2 | . Routes and schedules, inclusive | | Q4 | . . Small vs. large capacity, gen. references | |
| K2 | . . Routes and schedules, general references | | Q4.1 | . . Small capacity | |
| K2.1 | . . Routes, route systems, traffic networks | 2 | Q4.2 | . . Large capacity | |
| K2.2 | . . Schedules | | +Q5 | . Civilian vs military, inclusive | |
| K3 | . Handling, loading, unloading | | Q5 | . . Civilian vs military, general references | |
| K8 | . Other specific traffic operations | | Q5.1 | . . Civilian | |
| K9 | . Traffic operations not applicable | | Q5.2 | . . Military | |
| +L | Transportation providers, inclusive | | Q99 | . Other concepts not applicable | |
| L | . Transportation providers, gen. references | | +R | Geographic range, inclusive | |
| L1 | . Organizations, companies | | R | . Geographic range, gen. references | |
| L2 | . Personnel, operators | | +R1 | . Local systems, inclusive | |
| L9 | . Transportation providers not applicable | | R1 | . . Local Systems, general references | |
| +M | Creation of traffic systems& comp. I. | | +R1.1 | . . Urban systems, inclusive | |
| M | . Creation of traffic systems&components, g.r. | | R1.1 | . . Urban systems, general references | |
| +M1 | . Research, design, and evaluation, inclusive | | R1.2 | . . Rural systems | |
| M1 | . . Research, design, and evaluation, g. ref. | | +R2 | . Beyond local systems, inclusive | |
| M1.1 | . . Research and development | | R2 | . . Beyond local systems, general references | |
| M1.2 | . . Planning | | R2.1 | . . Interurban systems | |
| M1.3 | . . Design | | R2.2 | . . State-wide systems | |
| M1.4 | . . Testing, demonstration, evaluation | | R2.3 | . . National systems | |
| M2 | . Manufacturing, construction | | R2.4 | . . International systems | |
| M3 | . Acquisition | | R8 | . Other specific range | |
| M4 | . Training | | R9 | . Geographic range not applicable | |
| M5 | . Maintenance | | +S | Geographic location, inclusive | |
| M8 | . Other specific activities in system creation | | S | . Geographic location, general references | |
| M9 | . System creation not applicable | | +S1 | . North and Central America, inclusive | |
| +N | Organization, administration, incl. | | S1 | . . North and Central America, gen. ref. | |
| N | . Organization, administration, gen. references | | S1.1 | . . Canada | |
| N1 | . Administration, management | | S1.2 | . . U.S. | |
| N2 | . Costs, financing | | S1.3 | . . Central America | |
| N3 | . Marketing | | S2 | . South America | |
| N4 | . Legal aspects | | S3 | . Europe | |
| N8 | . Other specific topics in organization | | S4 | . Asia | |
| N9 | . Organization, administration not applicable | | S5 | . Australia | |
| | | | S6 | . Africa | |
| | | | S8 | . Other geographic locations | |
| | | | S9 | . Geographic location not applicable | |

In-class exercise: Retrieval access and hierarchy

Below are six documents which were indexed in the request-oriented approach you used in Assignment 11. Each descriptor is on a separate line. Using the hierarchy of the index language in the *File builder's and searcher's display* (see preceding pages), do the following:

- 1 For each descriptor (index term), list the descriptor(s) under which the document should be found on the basis of this index term.
- 2 Give some query formulations retrieving the document. The query formulations should illustrate how a search for a combination of two broad concepts finds documents indexed by more specific concepts.

Document 1 is a filled-in example

Document 1

Automatic control of freeway ramp traffic, P.J.ATHOL. SAE—Analysis & Control of Traffic Flow Symposium—Conf Proc. Jan 9-10 1968 paper 680172 p 61-5.

Major problem in operating transportation system is traffic overloading demands at peak periods; expressway Surveillance Project was formed to improve efficiency of highway system through application of electronic automation and traffic engineering to problem of traffic congestion.; by providing means for quick response in case of accidents and fast removal of hindrances. Volume capacity of freeways was effectively increased during peak periods; use of ramp metering controls achieved reduction in delay, safer merging characteristics, and reduced freeway accidents.

Descriptors assigned

B1.1 Road transport
 E1.1 Traffic routes
 K1 Traffic communication, control, safety
 M1.2 Planning
 M1.3 Design
 M1.4 Testing, demonstration, evaluation
 Q1 Traffic flow

Descriptors under which the document should be found

B1.1; B1 inclusive
 E1.1; E1 inclusive
 K1 **gen. ref.**; K1 inclusive
 M1.2; M1 inclusive
 M1.3; M1 inclusive,
 M1.4; M1 inclusive,
 Q1

Query formulations

B1 Ground tr., incl. AND E1 Tr. fac., incl.
 B1.1 Road tr. AND K1 Tr. comm., control, safety incl.

Document 2

Antwerp's new container dock, K.W.Flitcroft for the Antwerp Harper Committee. Dock & Harbor Authority v 49 n 571 May 1968 p 28-30.

Dock described is protected by locks from rise and fall of tides; spreader is employed in lifting of containers and is adaptable in spread to handle both long and short types; containers can be stored on quay and special connections for powering of plants of refrigerated containers are set in concrete paving every 10 ft.; set of rail tracks runs along quay between high legs of container cranes to bring rail-hauled containers directly for lifting off.

Descriptors assigned

B2.2 Ocean transport
E1.2 Traffic stations
J2 Transport of freight, material, cargo
K3 Handling, loading, unloading
R2.1 International system
S3 Europe

Descriptors under which the document should be found**Query formulations**

Document 3. **Rolling Stock for London Transport's Victoria Line**

Descriptors assigned

B1.2.1 Local rail transit
 E2.1 Vehicles
 F5 Electric power
 G1 Engine
 M1.3 Design
 M3 Acquisition
 Q3.9 Other characteristics (automation)
 R1.1 Urban systems
 S3 Europe

Descriptors under which the document should be found

Query formulations

Document 4. **Air Transp. 1975 and Beyond - Systems Approach**

Descriptors assigned

B3 Air transport
 E Transportation system components
 J Passenger transp. vs. freight transp.
 K Traffic operations
 M1.1 Research and development
 N Organization, administration
 Q5 Civilian vs. military
 R2.3 National systems
 S1.2 U.S.

Descriptors under which the document should be found

Query formulations

| Document 5. Technical and Economic Prospects of Air Cargo Traffic | |
|--|---|
| Descriptors assigned | Descriptors under which the document should be found |
| B3 | Air transport |
| E2.1 | Vehicles |
| F1.3 | Hydrocarbons from renewable sources |
| G2 | Turbines |
| J2 | Transport of freight, material, cargo |
| K3 | Handling, loading, unloading |
| M1 | Research, design, and evaluation |
| Q1 | Traffic flow |
| Q3.5 | Demand, use |
| Q3.9 | Other characteristics (automation) |
| R2.4 | International system |
| S | Geographic location |
| | Query formulations |

| Document 6. United States Subway Requirements 1968-1990 | |
|--|---|
| Descriptors assigned | Descriptors under which the document should be found |
| B1.2.1 | Local rail transit |
| E1 | Traffic facilities |
| M2 | Manufacturing, construction |
| N2 | Costs, financing |
| N3 | Marketing |
| Q2 | Simulation |
| Q3 | System characteristic |
| R1.1 | Urban systems |
| S1.2 | U.S. |
| | Query formulations |

Small Groups 3. Lecture 10.1

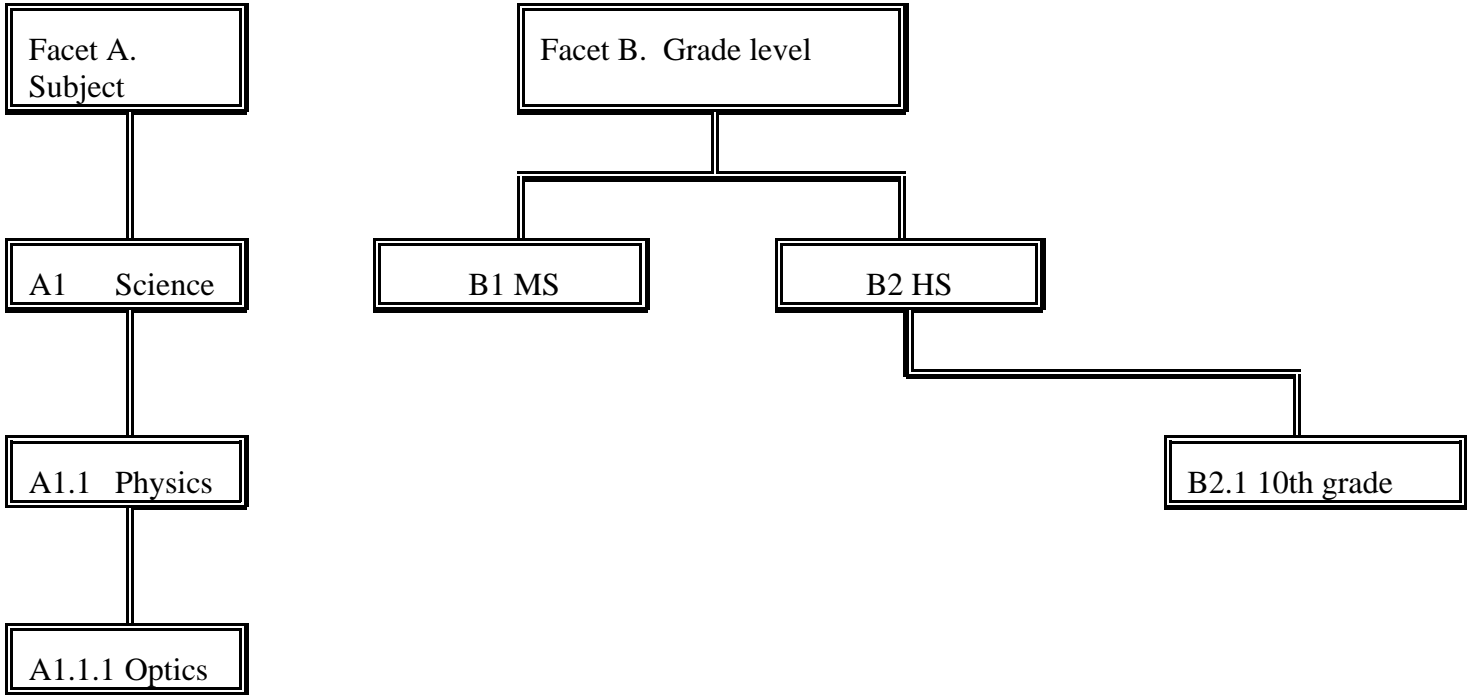
Hierarchy from Facets

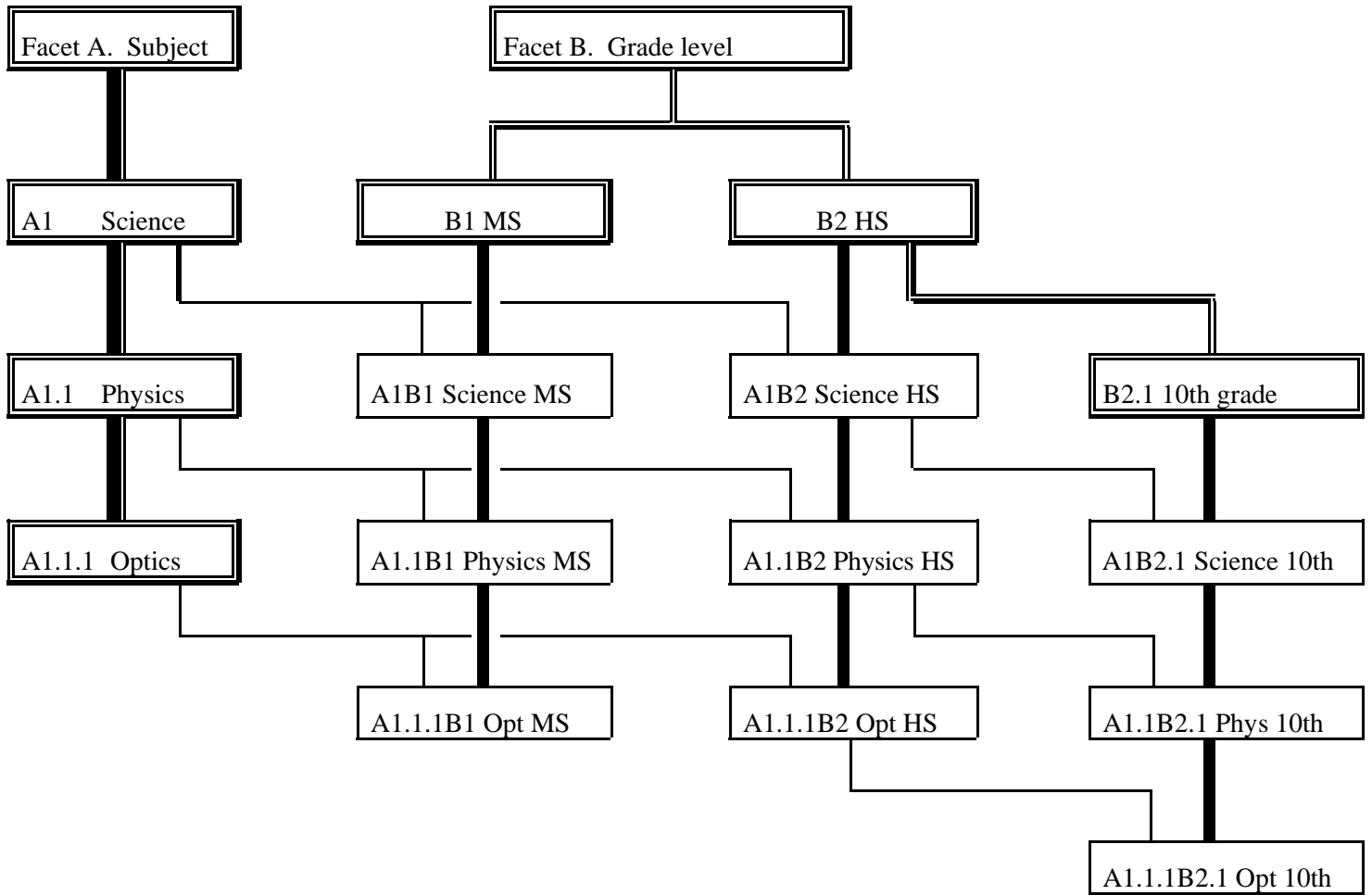
March 28

Information system of instructional materials. Two facets, only between-facet combinations

| | |
|--------------------------------------|--|
| Objectives Inherited, plus | Understand complex hierarchies that result from combining hierarchically structured facets. |
| Practical significance | Basis for understanding search Basis for understanding structure of DDC, LCC, LCSH, and similar systems |

| | |
|---------------------------------|---|
| Process | Step 1: Form all possible between-facet combinations (do not combine with facet heads). Step 2: Find all hierarchical relationships. (Specifying all BT one level up defines the hierarchy completely.) Step 3: Represent the hierarchy Step 3a: – as a two-dimensional graph Step 3b: – as a linear arrangement with indention plus cross-references. |
| Application to retrieval | In a system using only elemental descriptors In a system using precombined descriptors with multiple entry (such as LC Subject Headings) In a system using precombined descriptors with single entry (such as Library of Congress Classification) |





A Facet A. Subject

- . A1 Science
- . . A1B1 Science MS NT A1.1B1; BT B1
- . . A1B2 Science HS NT A1.1B2; BT B2
- . . . A1B2.1 Science 10th grade NT A1.1B2.1; BT B2.1
- . . A1.1 Physics
- . . . A1.1B1 Physics MS NT A1.1.1B1; BT A1B1
- . . . A1.1B2 Physics HS NT A1.1.1B2; BT A1B2
- A1.1B2.1 Phys 10th gr NT A1.1.1B2.1; BT A1B2.1
- . . . A1.1.1 Optics
- A1.1.1B1 Optics MS BT A1.1B1
- A1.1.1B2 Optics HS BT A1.1B2
- A1.1.1B2.1 Optics 10th grade BT A1.1B2.1

B Facet B. Grade level

- . B1 MS NT A1B1
- . B2 HS NT A1B2
- . . B2.1 10th grade NT A1B2.1

A Facet A. Subject

- . A1 Science NT B1A1, B2A1
- . . A1.1 Physics NT B1A1.1, B2A1.1
- . . . A1.1.1 Optics NT B1A1.1.1, B2A1.1.1

B Facet B. Grade level

- . B1 MS
- . . B1 A1 MS Science BT A1
- . . . B1 A1.1 MS Physics BT A1.1
- B1 A1.1.1 MS Optics BT A1.1.1
- . B2 HS
- . . B2 A1 HS Science NT B2.1A1; BT A1
- . . . B2 A1.1 HS Physics NT B2.1A1.1; BT A1.1
- B2 A1.1.1 HS Optics NT B2.1A1.1.1; BT A1.1.1
- . . B2.1 10th grade
- . . . B2.1 A1 10th grade Science BT B2A1
- B2.1 A1.1 10th grade Physics BT B2A1.1
- B2.1 A1.1.1 10th grade Optics BT B2A1.1.1

Lecture 10.2a**March 28****Brief Introduction to Assignments 13.1-4. Examination of KOS**

Analysis of Knowledge Organization Systems (ontologies, classification schemes, thesauri, etc.) based on their hierarchical structure, facet structure, and citation order. See the calendar or the assignment page for schedule of assignment activities and due dates.

| | |
|-------------------------------|--|
| Objectives | <ol style="list-style-type: none"> 1 Solidify general principles underlying all KOS and develop a more complete understanding of the general concepts of classification structure by applying them to several concrete schemes. 2 Get an overview of different types of KOS (Knowledge Organization Systems) and the wide variety of systems in which KOS are used and the wide variety of purposes for which they are used. In particular: Gain practical experience with a semi-faceted classification used on the Web (Assignment 13.2 Yahoo), create familiarity with specific schemes that are widely used in libraries in the US (Assignments 13.1 DDC, 13.2 LCC, and 13.3 LCSH), and become acquainted with a wide range of schemes used for a wide variety of purposes (Assignment 13.4 ERIC and Lecture 13.1, Exploration of classification schemes and thesauri). You can grasp the structure of these schemes better by applying a general conceptual framework to their analysis. Assignments 13.1 - 13.4 help you understand specific schemes by application to specific problems in cataloging (indexing) and query formulation for searching. Where available, these assignments introduce the electronic form of a scheme as well. |
| Practical significance | <ul style="list-style-type: none"> • A good working knowledge of faceted classification principles is important for the conceptual analysis of queries as a basis for developing good query formulations in any system. • Knowledge of specific schemes is important for searching catalogs and indexes based on those schemes (including catalogs of Web documents). • Knowledge of the variety of schemes that exist for different purposes is important for being able to work in many different applications and for recognizing where classification could be useful. • Knowing the general principles that underlie all KOS will enable you to evaluate KOS, to improve existing KOS, and to build new KOS (after taking LIS 514 Indexing and Surrogation). • You will be able to “sell your skills” to a wider variety of organizations, increasing opportunities for work |

over

Turn to Assignments 13.1-4, Assignments p. 115 (or thereabouts)
Read p. 115-116 (information for all assignments 13.1-13.4).

Listen to presentation

Lecture 10.2b, going through the Dewey lecture/PowerPoint will make this clearer.

Lecture 10.2b**March 28****Introduction and in-class exercise:
Assignment 13.1. Dewey Decimal Classification (DDC)**

| | |
|--|--|
| Objectives Inherited from 10.2a | Specifically, gain a first understanding of the Dewey Decimal Classification with emphasis on the general structure. Knowledge where specific subjects are placed in DDC will come with practice |
| Practical significance Inherited from 10.2a | Specifically, DDC is used in most public and school libraries in the US. It is also used in many other countries OCLC, the owner of DDC, is pushing its use for organizing subject directories on the Web |

Read the golden page for Assignment 13.1, DDC

With a study group, start on the DDC worksheet and do as much as you can in no more than two hours (perhaps less). The idea is to get at least a start on figuring things out for yourself without banging your head against the wall.

Then go to the presentation, look at the pdf while listening to the mp3 audio

Lecture 11.1

Introductory discussion and in-class exercise on Assignment 13.2 Yahoo

We will start going through the worksheet, index a document, and formulate a query.

Lecture 11.2

Introductory discussion and in-class exercise on Assignment 13.2 LCC

We will start going through the worksheet, index a document, and formulate a query.

Lecture 12.1

April 4

Index language structure 2: database organization (Organizing Info., Chapter 15)

| | |
|---|---|
| <p>Objectives Inherited from Lect. 9.1-11.1 plus these</p> | <ol style="list-style-type: none"> 1 Understand postcombination and precombination - more generally, the degree of precombination — and how they relate to the retrieval mechanism used. 2 Be able to match the index language structure to the database organization and search mechanism available. 3 Understand the effect of precombination on index language structure and searching and be able to apply this understanding to the analysis of classification schemes such as DDC and LCC and improved searching with such schemes. 4 Understand the access mechanisms that help a user find the proper descriptors in a large classification scheme with many precombined descriptors, in particular cross-references and a descriptor-find index. 5 Understand principles of meaningful arrangement of search results. |
| <p>Practical significance Inherited from Lect. 9.1-11.1 plus these</p> | <p>In conjunction with Chapter 14, this chapter establishes the foundation for understanding</p> <ul style="list-style-type: none"> • the structure of systems used in libraries — and increasingly for the arrangement and display of electronic information — such as the Dewey Decimal Classification (DDC), the Library of Congress Classification (LCC), the Yahoo classification, and the Library of Congress Subject Headings (LCSH); • the structure of Web directories designed for browsing; • ad-hoc arrangement of retrieval results based on the analysis of noun phrases as compound concepts, as in the next-generation Web search engines. |
| <p>Discussion question</p> | <p>Consider the design of an interface to a public-access online catalog in an academic library that would assist users in finding the appropriate LC class number and the appropriate LC subject headings.</p> |

Part 1. Introductory example**Part 2. Discussion of Organizing Information, Chapter 15**

| | |
|------------------------------|--|
| Section 15.1 and 15.2 | <p>Further examination and explication of postcombination vs. precombination of the concepts chosen as descriptors and their relationship to database organization and search mechanism.</p> <p>Interpretation of postcombination and precombination in terms of the entity-relationship approach and semantic networks (see figures on the following two pages).</p> <p>Examples of applying these concepts to a better understanding of index languages such as the Library of Congress Classification and the Library of Congress Subject Headings.</p> |
| 15.3 and 15.4 | <p>Emphasis on looking at precombination as a matter of degree.</p> <p>Introducing precombined descriptors as an example of restructuring semantic networks using hierarchical inheritance.</p> |
| 15.5 | <p>Methods for organizing an index language for access.</p> <p>Emphasis on understanding the idea of a descriptor-find index.</p> <p>Section 15.5.2 on how then arrangement of precombined concepts is important for understanding classification schemes and for arranging search output or any type of information, in print or online.</p> |
| 15.6 | <p>A look into the future: the idea of a conceptually unified index language for different search mechanisms.</p> |

Part 3. In-class exercises concluding Lectures 8.1-10.1 and 12.1

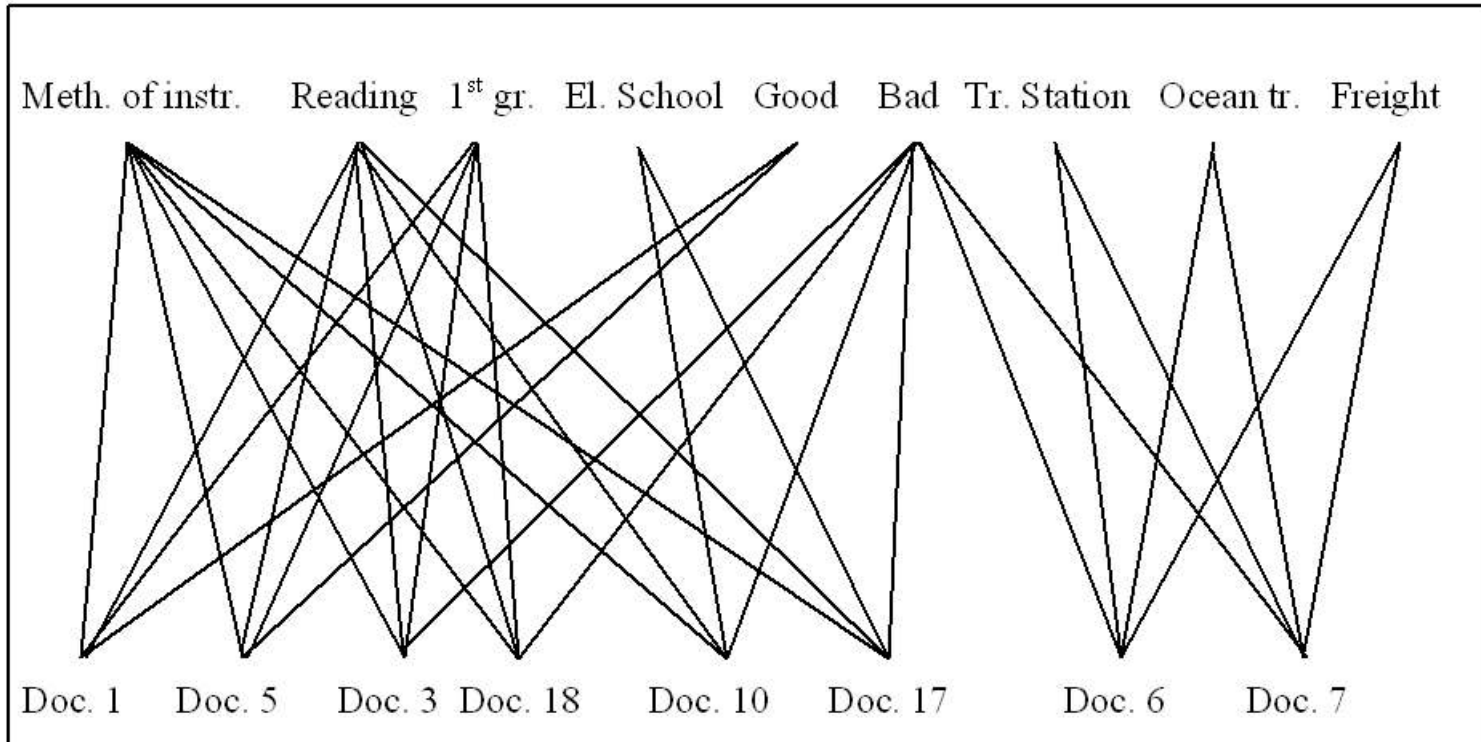
Semantic networks and precombined descriptors

Vocabulary control and hierarchical structure

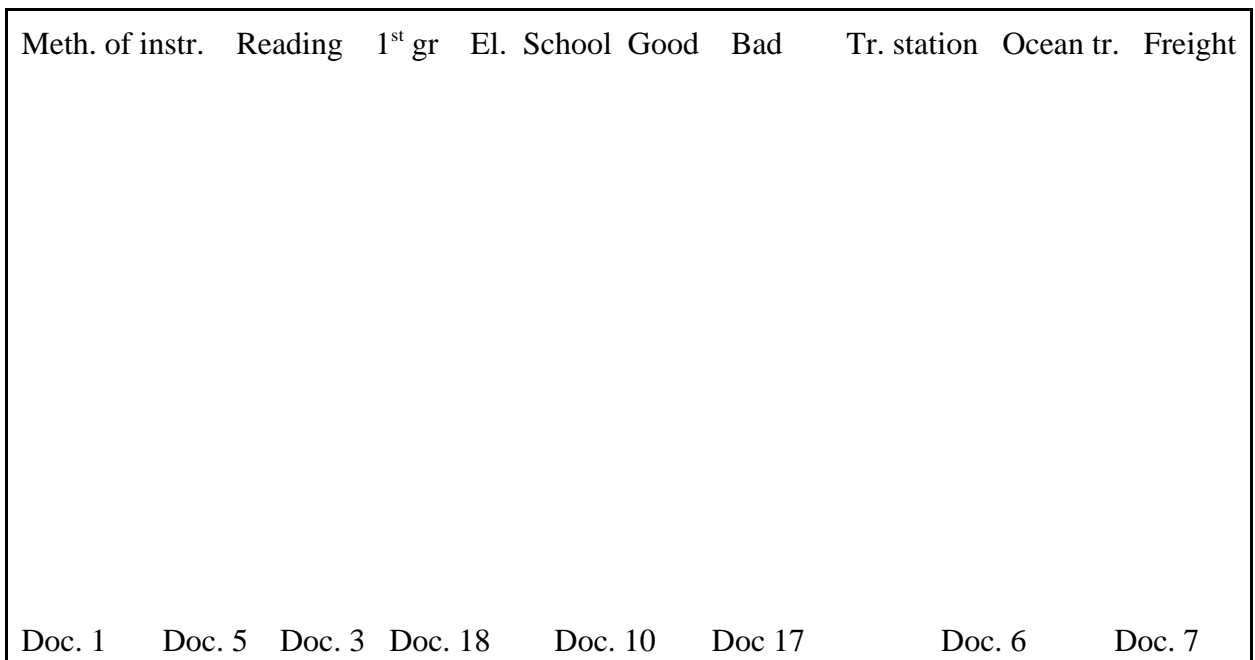
Conceptual analysis and synthesis

Discussion question on OPAC (Online Public Access Catalog) interface

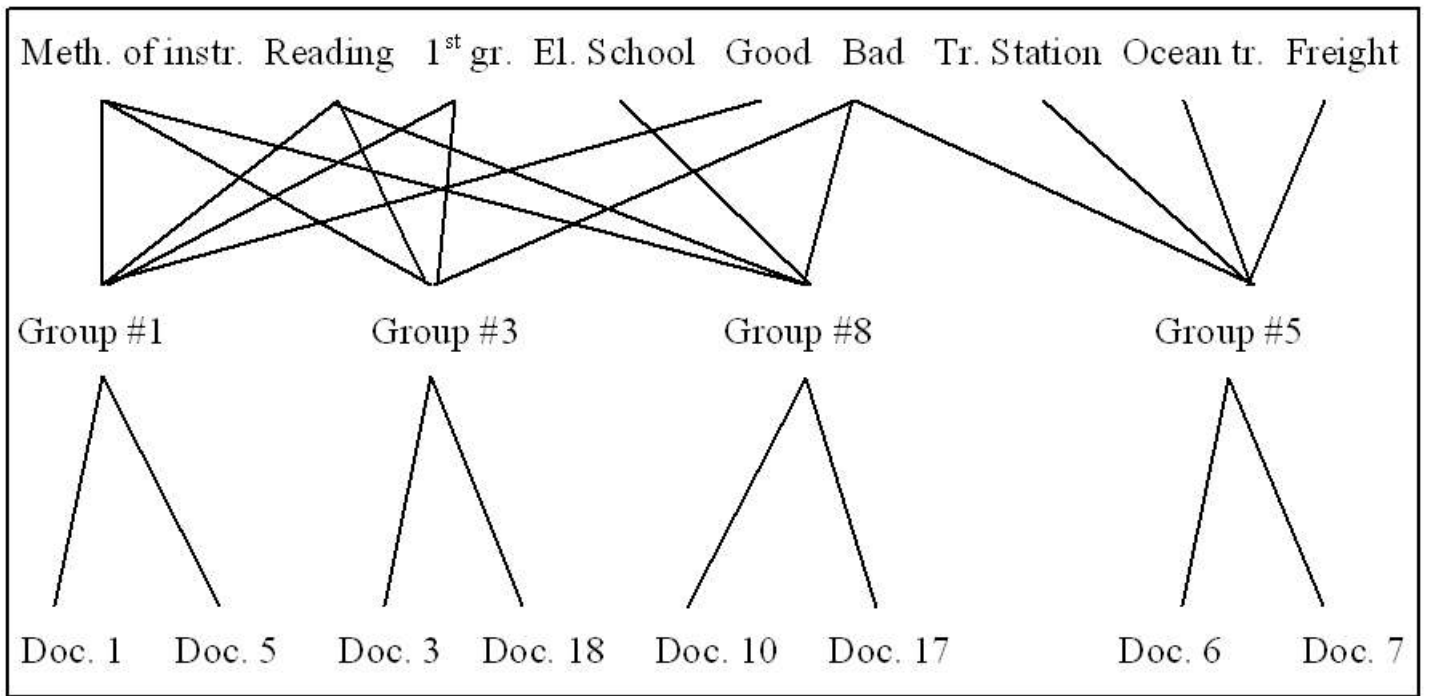
Semantic network representation of Database 15.1 (Organizing Info., p. 296)



In-class exercise: Semantic networks and precombined descriptors. Simplify the network.



Semantic network representation of Database 15.2 (Organizing Info., p. 296)



In-class exercise: Vocabulary control and hierarchical structure

The following is a list of terms that have occurred in query statements and in document titles. Organize it for purposes of information retrieval.

| | |
|----------------------|--|
| Book | |
| Campaign | |
| Candidate | |
| Department of State | |
| Elections | |
| Foreign Office | |
| Issue | |
| Journal | |
| Movement | |
| Periodicals | |
| Roll-call vote | |
| Running for Governor | |
| Running for Office | |
| State Department | |
| Vote | |

This task calls on you to apply your knowledge from Soergel, Chapters 12-15. Therefore, no further guidelines are provided. (You may have to do this on your own on a much larger scale, in real life.) Since the list of terms is so small, facet analysis and synthesis is not required in this task.

In-class exercise: Conceptual analysis and synthesis

Organize the following list of terms for purposes of information retrieval.

| Terms to work on | Additional terms (just to think about) |
|-------------------------------|---|
| U.S. Congress | Foreign Office |
| State Court | British Parliament |
| County administration | United Nations |
| State legislature | Prime minister |
| Federal court | House of Commons |
| U.S. Senate | House of Lords |
| U.S. House of Representatives | UN Secretary-General |
| State administration | UN Security Council |
| State senate | UN General Assembly |
| State assembly | World Court |

Procedure: Facet analysis and synthesis

| | |
|---------|---|
| Step 1: | Factor concepts into semantic components, resulting in elemental concepts |
| Step 2: | Organize the resulting elemental concepts in facets |
| Step 3: | Combine the facets (form all combinations) |

The resulting hierarchical structure is to be shown graphically as well as in a linear sequence with cross-reference.

Note: The combinations produced in step 3 show gaps in the original list of terms.

Discussion question

Consider the design of an interface to a public-access online catalog in an academic library that would assist users in finding the appropriate LC class number and the appropriate LC subject headings.

Lecture 12.2 (brief)*March 28***Short Media Streams Classification demo**

www.dsoergel.com/571/UBLIS571Lecture12.2Video.avi

XXX Video to be posted

An interesting classification for the purpose of indexing movie segments and video clips. In keeping with indexing visual materials it uses icons as descriptors

The video presentation will walk through the major sections of the classification and show examples of descriptors.

As you watch, reflect on the structure of the classification. Post comments on the discussion board.

Lecture 13.1*December 7***Exploration of Knowledge Organization Systems (KOS)
(ontologies, classification schemes, thesauri)**

Read all pages for this lecture before listening to the audio

| | |
|--|---|
| <p>Objectives</p> <p>Inherit from Lecture 110.2a In addition</p> | <ol style="list-style-type: none"> 1. The student should have an appreciation for the wide range of classifications and subject access vocabularies used for a variety of purposes 2. The student should have an improved understanding of general principles of KOS structure and of methods for presenting this structure and be able to apply these principles in analyzing KOS. 3. The student should gain some acquaintance with a number of important schemes. |
| <p>Practical significance</p> <p>Inherit from Lecture 10.2a In addition</p> | <ul style="list-style-type: none"> • You will be able to “sell your skills” to a wider variety of organizations, increasing opportunities for work. • Knowing the general principles that underlie all KOS will enable you to evaluate KOS, to improve existing KOS, and to build new KOS (after taking LIS 514 Indexing and Surrogation). |

| | |
|------------------------------|---|
| Materials and methods | <p>The schemes to be explored are listed below</p> <p>Each scheme is represented by a judiciously chosen selection of pages. These sample pages are not meant to be read word for word but rather to be examined with the goal of forming an overall image of the scheme. In particular, look at any sample entries marked by underline or *; they are usually part of a coordinated example illustrating the structure and relationships between parts. The schemes are separated by a blue sheet.</p> <p>These pages are meant for exploration, not for word-to-word to reading. Get an overview and pick out some examples to understand the structure of a scheme.</p> <p>Two elements to consider when examining a system</p> <ol style="list-style-type: none">1 Function: What is this system used for? What could it be used for2 Structure: Facets, hierarchy <p>The lecture will go through the schemes, discussing examples from each. You can examine the schemes before listening to the lecture (lecture will make more sense) and/or after listening to the lecture.</p> <p>Please explore all schemes and select two schemes for closer examination and post an observation on at least one scheme on the free write for this week. Select the two schemes according to your interests, For example, if you are in School Library Media Program, examine the <i>Taxonomy for learning, teaching, and assessing</i> (the revision of <i>Blooms Taxonomy of educational objectives</i>, if you are interested in art and/or museums, examine the <i>Art and Architecture Thesaurus</i>.</p> <p>The last page for this lecture has a listing of schemes illustrating more functions. Just read it.</p> <p>For more in-depth information, see www.dsoergel.com/SoergelKOSOverview.pdf</p> |
|------------------------------|---|

Schemes to be examined

- Bloom** Anderson, L. W., & Krathwohl, D. R.
A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.
New York: Addison Wesley Longman; 2001.
- SOC** **Standard Occupational Classification 2000**
Bureau of Labor Statistics (BLS) + other agencies
<http://stats.bls.gov/soc/home.htm>
The SOC is augmented by the **Occupational Information Network (O*NET)**, a database with additional occupational titles, definitions, and features of occupations.
<http://www.doleta.gov/programs/onet>
- CSDGM** **Content Standard for Digital Geospatial Metadata 1998**
Federal Geographic Data Committee (FGDC)
<http://www.fgdc.gov/metadata/constan.html>
- AOD** **The Alcohol and Other Drug Thesaurus.** 3rd ed., 2000.
National Institute on Alcohol Abuse and Alcoholism (NIAAA)
<http://etoh.niaaa.nih.gov/AODVol1/Aodthome.htm>
- MeSH** **Medical Subject Headings.**
National Library of Medicine (NLM)
No longer published in print
www.nlm.nih.gov/mesh/meshhome.html
- AAT** **Art and Architecture Thesaurus.** 2nd ed 1994
Getty Art History Information Program
www.getty.edu/research/tools/vocabulary/aat/index.html
- WordNet** **WordNet Lexical Database.** Version 3.0 Aug. 2003
Princeton University, Cognitive Science Laboratory
(www.notredame.ac.jp/cgi-bin/wn.cgi)
- CYC Ontology** (CYC Corporation) 1997
To get an idea of what is in CYC:
http://cyc.com/cyc/technology/whatis_cyc_dir/whatdoes_cyc_know
to search: <http://sw.opencyc.org/>

Additional KOS examples illustrating different functions

| | |
|---------------|---|
| HS | Harmonized Commodity Description and Coding System. World Customs Organization, Brussels. Info: http://pacific.commerce.ubc.ca/trade/HS.html |
| NAICS | North American Industrial Classification System "common industry definitions for Canada, Mexico, and the US. Developed in cooperation with the US Economic Classification Policy Committee, Statistics Canada, and Mexico's Instituto Nacional de Estadística, Geografía e Informática to better compare economic and financial statistics and ensure that such statistics keep pace with the changing economy. NAICS will replace the countries' separate classification systems (in the US: Standard Industrial Classification, SIC) with one uniform system for classifying industries." Info: www.census.gov/epcd/www/naics.html , www.naics.com www.cdc.gov/nchs/about/major/dvs/icd10des.htm |
| ICD-10 | The International Statistical Classification of Diseases and Related Health Problems, tenth revision. Produced by the World Health Organization. Published in many languages. Info: www.who.int/whosis/icd10/index.html , www.cdc.gov/nchs/about/major/dvs/icd10des.htm |
| CPT | Physicians' Current Procedural Terminology. www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt.page (Info: www.ama-assn.org/ama/pub/category/3113.html , listing of codes https://webstore.ama-assn.org/index.jhtml) |
| | Health Care Finance Administration (HCFA) Common Procedure Coding System (HCPCS) for Medicare reimbursement for hospital outpatient services. It has three levels - CPT (level 1), HCPCS or National (level 2), and Local (level 3). In its data collection the Agency for Health Care Policy and Research (AHCPR) uses data standards that are based on those used by the Census Bureau, the American Hospital Association, the Health Resources and Services Administration (Area Resource File), the National Center for Health Statistics, and codes for clinical diagnosis and procedures such as ICD-10 and CPT 1998. These standards facilitate data analysis and use by ensuring comparability, quality and interoperability. Further, uniform health care data advance medical and health care services research, the efficiency of the private sector health care delivery system, and quality improvement measurement. |
| | Further type of classification: biological taxonomies. Used in biology, agriculture, food science, and medicine. Several rivaling schemes for major areas (kingdoms) and many publications on specific areas. www.itis.gov/ www.ucmp.berkeley.edu/help/taxaform.html |
| | See www.obofoundry.org/ for many biomedical ontologies |

Lecture 13.2a

Questions on Assignment 13.3 LCSH / Sears and 13.4 ERIC
or anything else (15 min)

Lecture 13.2b *April 4*
Indexing and system performance (Organizing Information, Chapter 16)

| | |
|-------------------------------|---|
| Objectives | <ol style="list-style-type: none"> 1 Understand the concepts of exhaustivity and specificity of indexing and their effect on searching. 2 Understand the concept of weights in indexing. 3 Be able to ascertain the exhaustivity and specificity of indexing in a given system and apply this knowledge to appropriate query formulation. 4 Be able to apply indexing weights in query formulation (including analogous techniques in free-text searching). 5 Be able to determine the proper levels of exhaustivity and specificity of indexing for a new IR system based on user requirements. |
| Practical significance | <p>An examination of indexing parameters, especially exhaustivity and specificity and term weighting, their measurements, their effect on retrieval performance (which is often oversimplified), their dependence on various factors in the indexing process, and their costs (Chapter 16). A correct understanding of these relationships is important for optimal query formulation in online systems, including the more sophisticated Web search engines, as well as for system design.</p> |

| | |
|-----------------------------|--|
| Discussion questions | <p>How could one gauge the exhaustivity of indexing in a database if indexers' instructions are not available? How could one tell if within one and the same database exhaustivity varies from subject to subject?</p> <p>Give examples of exhaustivity and specificity of indexing for other types of entities and relationships.</p> |
|-----------------------------|--|

XXX Combine Ch. 16 and logical evidence paper into new Ch. 16

Discussion of Text Organizing Information, Chapter 16 and the reading

| | |
|---------------------------------------|--|
| <p>Text Section 16.2.1</p> | <p>Definitions of exhaustivity and specificity. Indexing weights.</p> <p>Put in the context of the conceptual data schema of a system.</p> <p>Indexing specificity has to do with the entity values for the entity type subject (or of other entity types, for example Date, to which the concept of specificity can be applied).</p> <p>The rules for exhaustivity in indexing are a special case of rules for establishing relationships, such as relationships between a document and subjects. Analogous rules can be defined for many types of relationships.</p> <p>Indexing with weights requires three-place relationships, such as</p> <p style="text-align: center;">Document <i>deals with or is relevant for</i> (Subject, Weight)</p> |
| <p>16.3.1</p> | <p>Effects of indexing exhaustivity on retrieval performance</p> <p>Important conclusion: The query formulation must be adapted to the exhaustivity of indexing for best retrieval results.</p> |
| <p>Other questions</p> | <p>Questions on the remainder of the chapter and the reading.</p> |

Discussion questions (repeated)

How could one gauge the exhaustivity of indexing in a database if indexers' instructions are not available? How could one tell if within one and the same database exhaustivity varies from subject to subject?

Give examples of exhaustivity and specificity of indexing for other types of entities and relationships.

Conclusion

April 25

Lectures 14.1-14.2

Final review

Numbers at left margin indicate number of minutes = number of points.

- 15 1. There are a wide variety of "documents" on the World Wide Web ("Web pages" and "Web sites"). In a catalog of Web documents, it might be useful to include an indication of the type of document in the catalog record. **Develop a typology of Web documents for this purpose.** (A typology is a list or classification of types).

- 30 2. **Home page design**

A naturalist organization keeps an inventory of (rare) plants and animals in Western New York.

They provide access to several general databases on plants and animals.

They maintain a database of locations where (rare) plants are found and of sightings of (rare) animals. They collect these data from individual members (or anybody, for that matter, with some check of credentials). Each occurrence/ sighting report is recorded by place, date, and time. Some of this information is kept confidential so as to not make poaching easy.

They also have events (lectures, excursions) they want to announce and make registrations for

Design a home page for such an organization.

15 3. **Reorganize thesaurus information to take less reading and less storage space.**

The ERIC Thesaurus has the following entries:

Autoinstructional aids

- RT Audiovisual aids
- RT Computer assisted instruction
- RT Courseware
- RT Individualized instruction
- RT Learner controlled instruction

Programmed instructional materials

- RT Audiovisual aids
- RT Computer assisted instruction
- RT Courseware
- RT Learner controlled instruction
- RT Workbooks

Teaching machines

- RT Computer assisted instruction
- RT Courseware
- RT Learner controlled instruction
- RT Pacing

Another example

free participation

- RT health care delivery and administration
- RT health care economics

payment-based participation

- RT health care delivery and administration
- RT health care economics

subsidized payment

- RT health care delivery and administration
- RT health care economics

full cost-recovery payment

- RT health care delivery and administration
- RT health care economics

- 20 4. You have to **design a controlled-vocabulary IR system** (with human indexing) that gives the searcher the option of emphasizing either discrimination (one factor determining precision) or recall. List the features that are important for achieving this flexibility.
5. **Query formulation for free text.** A user needs information on the following topic:
Validity of the evaluation of instructors through undergraduate students in social science courses.
- A free-text search for this topic is to be made in a bibliographic database
- (1) in database 1, searching is by terms occurring in the **title** of the document,
 - (2) in database 2, searching is based on terms that occur in the **title and/or the abstract of the document.**
- 20 a. For each database give the **conceptual query formulation** that you would use (do not worry about terminology at this point). Give your rationale.
- 10 b. **Give the free-text query formulation for database 2.** Assume that the search from the previous question is to be made in the system searching on **titles and abstracts** (system 2). Any word or phrase (multi-word term) occurring in the title or abstract can be used as descriptor for searching. Briefly describe how you would go about developing the query formulations in terms of descriptors (3 min.) Start doing it (7 min.)
- 20 6. This question deals with **retrieval in archives**; sufficient background is provided so that you can answer it even if you are not familiar with archives. Archives are a collection of documents (letters, memoranda, reports, etc.) produced by an organization, its various units, and the persons working in the units. (Assume an organization of the complexity of the Federal Government with many organizational units interrelated hierarchically and otherwise.) The organization of archives usually allows for easy retrieval of all documents produced by an organizational unit or a person; a document is linked to its producer at its creation so that the archivist need not do additional indexing to provide this type of access. Date when created, receiving organizational unit or person, and often related documents are also known for each document. It is usually too expensive to assign subject descriptors to individual documents, yet subject searches are frequent. The archivist doing a subject search uses her - more or less - complete knowledge of organizational units and persons and the subjects they have been dealing with at certain times to find relevant documents to look under appropriate units and persons.
- Sketch a conceptual data schema for a computerized retrieval system for archives that implements in a formal way the approach described. Describe how this system does searches for subject.

- 40 7. You are appointed as head of a medium-sized IR-system (about 200,000 documents) that uses three different systems for subject access:

- (1) an alphabetical subject catalog of books;
- (2) shelving books by subject;
- (3) an independent classification scheme for filing newspaper clippings

Your analysis shows that the subject heading list and the shelving classification are both far from satisfactory. The subject headings have grown without control and no listing is available. But a cost-benefit analysis rules out major changes or revision, like introducing new schemes, especially in view of the large costs for re-indexing the old collection. On the other hand, the cost-benefit analysis also shows that some costs would be justified to improve the usability of the IR-system. What do you suggest should be done? How would you implement your suggestions?

- 40 8. You are given the task to design an IR system. One problem is to determine **how much money should be spent for indexing**. Discuss the data you need/the considerations on which you would base your decision.

- 40 9. You are given the task of **developing an index language and thesaurus** for

- (1) a newly set up information center in a company, or
- (2) a public information center in the inner city (choose **one**).

What are the main points you have to take into consideration in performing this task?

- 20 10. **Assist users in coping with large Web search results**. A search in a Web directory, such as Yahoo or the Open Directory Project (<http://dmoz.org/about.html>), or a search engine, such as Google, AltaVista, or Lycos, often returns hundreds of documents. What could the system do to help the user to cope with these large numbers?

- 15 11. Discuss **exhaustivity** in the context of **hypertext links** made in a system.

- 15 12. A large subject index is to be put on microfiche. The system has two parts:

- (1) The actual index on microfiche. This is an ordinary index: Under each descriptor the entries for the documents (or other retrieval objects) indexed by that descriptor are listed.
- (2) To help the user find the appropriate microfiche, there is a hard copy "index to the index." This is simply a list of all descriptors, giving for each the microfiche number and the frame number on the microfiche.

Question: Should the subject index on microfiche be arranged in classified or in alphabetical order? How should the hard-copy "index to the index" be arranged?

Assume a microfiche reader where the user must manually insert the fiche and find the frame.

- 40 13. You are charged with the design and development of an online information retrieval system for courses at the University of Maryland. The system should serve
- (1) students in course selection and
 - (2) curriculum committees who want to know what courses exist in a given area (such as *statistics* or *communication in organizations* before approving a new course.
- Discuss your approach (describe the workings of the system you propose to the extent feasible in 40 minutes; bulleted lists for some pieces are fine)
- 12 14. Compare a system using shelf arrangement based on an index language like LCC or DDC with a system based on postcombination (such as a computerized IR system) with respect to the exhaustivity and specificity of indexing that can be achieved. What can you say about retrieval performance in both cases?
- 30 15. Assume you have to **design a large lexical and classification database** that has the ambitious objective of serving as a tool for both natural language processing and indexing and retrieval. What information should be included for each term or concept?

Final review. Natural language processing (NLP). See Supplement

Final review. Precombination vs postcombination See Supplement

Both important

Dagobert Soergel

LIS 571
Organization and Control
of Recorded Information

Summer 2012

Assignments

Assignments 1 - 3

Searching three information systems

| | |
|--|--|
| Perseus (plus Freebase) | A hypermedia system on classics |
| Medline | A database of journal articles in medicine |
| OCLC WorldCat and ArticleFirst on FirstSearch | A database of books on all subjects and a database of journal articles |

| | |
|--------------------------------------|---|
| Objectives / concepts learned | <p>These assignments give you experience in how information retrieval (IR) systems work and provide many examples for concepts discussed throughout the course, esp. in the reading for Lecture 4.1, <i>Design of an integrated information structure interface</i>.</p> <p>The goal is to create an intuitive understanding now as the basis for more formal treatment later</p> |
|--------------------------------------|---|

These assignments relate to →LIS 518 Reference Sources and Services; they will make you a better searcher.

If you want to examine still another information system where the entity-relationship model is hard at work, look at

[/www.ontopia.net/omnigator/models/index.jsp](http://www.ontopia.net/omnigator/models/index.jsp)

(The opera example is best.)

Also <http://wiki.dbpedia.org/About> and www.freebase.com (Freebase is now part of Ass. 1)

Note on controlled vocabulary

All three assignments touch on this concept in a preliminary way. This concept will be discussed in depth later. Briefly, a controlled vocabulary is a vocabulary in which one term is chosen from a group of synonyms to represent a concept. From the terms *popular culture* and *mass culture* choose *popular culture*; *popular culture* is used in indexing and should be used in searching. This way, no matter which term the author uses, a user searching for the concept will find it.

Assignment 1
Lecture 1.2

Assigned: May 21
Due: May 28

Hypermedia exploration: Perseus and Freebase

| | |
|---------------------------|--|
| <p>Objectives</p> | <ol style="list-style-type: none"> 1. Experience an integrated IR system that covers many entity types (art works, buildings, documents, and words) that are related in many ways. Get a good understanding how the entity-relationship approach can be used in searching through a rich network of relationships 2. Improve understanding of information structure concepts: entity-relationship modeling, conceptual data schema. 3. Understand searching through navigation and through explicit queries. Optional: A first look at controlled vocabulary vs. full-text. <p>Perseus assignment</p> <p>The purpose of this assignment is to get a good understanding of how the entity-relationship approach can be used in searching through a rich network of relationships.</p> |
| <p>Deliverable</p> | <p>Individual essay (one for each student)</p> <p>3/4 (three-fourths) page. The Essay should be a reflection on what you experienced / learned, what you got out of this assignment, and/or what you found neat or deficient about this system. You should analyze the system as to its strengths and weaknesses and/or think about how you might use this system (or a system like it in a different domain) in the kind of job your aspire to, or how this or a similar system could be used for patrons of the kind of library/information system you are interested in.</p> <p>Do not write about the quirks and unreliability of the system.</p> <p>Include a comparison of Perseus and Freebase</p> |
| <p>Tasks</p> | <ol style="list-style-type: none"> 1. Do a guided exploration of Perseus. You can get a basic view or you can explore the system more fully by following the optional parts. 2. Do your own searches on Perseus. (optional) |
| <p>Materials</p> | <ol style="list-style-type: none"> 1 Description of Perseus 2 Step-by-step guide for exploring the features of Perseus <p>You need a computer with a high-speed Internet connection (your computer at home or any of the computer labs in the Lockwood Library, the Cybrary, or Silverman Library).</p> <p style="text-align: right;">Over</p> |

| | |
|---------------|--|
| Caveat | If Perseus does not work for you exactly as it did when the assignment was last checked, “go with the flow” and figure out what to do. |
| Time | 2 ½ hours |

Assignment materials for Perseus

1 Description of Perseus and Freebase

1.1 What is Perseus? What is Freebase?

Perseus is a Web-based hypermedia and information retrieval system that contains a vast amount of information about classics (and recently other things), including descriptions and images of archaeological sites, buildings, sculpture, vases, and coins; Greek literary works and their English translations; a Greek-English dictionary; a time line of events; a classics encyclopedia.

Section 1.2 gives a more complete description of the contents in form of a *conceptual data schema*.

In the exploration, you will, among other things

- ! visit a Greek archaeological site, look at a site plan, get information about the buildings, and look at images of the buildings, with an excursion to an encyclopedia article about Greek theaters in general;
- ! consult the Perseus Encyclopedia;
- ! locate a passage of Greek text with links to dictionary entries;
- ! do some searches in the English-Greek dictionary;
- ! [locate literature relevant to Thucydides research and follow some hypertext links;]
- ! do a free-text search of the entire Perseus database.

Notes 1 Perseus is under continuous construction; as a consequence, everything work does not always work as advertised.

2 An older version of Perseus had several cool features the assignment can no longer demonstrate. These are summarized at the end of the assignment.

"**Freebase** is an open, Creative Commons licensed repository of structured data of more than 12 million entities. An entity is a single person, place, or thing." Entities are connected through binary relationships, forming statements. This can be visualized as an entity-relationship graph that users can follow starting from one entity, such the person *Verdi*, to find connected entities, such as operas he composed. There is a large community of people wh contribute data to Freebase. You will explore Freebase on your own. www.freebase.com

1.2 Conceptual data schema for Perseus (draft, suggestions for improvements welcome)

A conceptual data schema defines the types of data that are included in an information system. In the *entity-relationship approach* (which is quite prevalent in the database world, in data representation for the semantic Web, and, since FRBR and RDA, in bibliographic cataloging), types of data are modeled by giving the types of entities covered in the information system and the types of relationships that can be used to combine entities into statements, such as:

| | | |
|---------------------|--|---|
| Statement templates | Entity type <i>Relationship</i> Entity type Text object < <i>authoredBy</i> > Person | Entity type <i>Relationship</i> Entity type Building < <i>containedIn</i> > Site. |
| Statements | Entity value <i>Relationship</i> Entity value Case of Plataia < <i>authoredBy</i> > Crane | Entity value <i>Relationship</i> Entity Parthenon < <i>containedIn</i> > Acropolis |

Why is the conceptual data schema (which is not easily visible in Perseus) elaborated here?

- 1 Understanding the conceptual data schema makes it easier to use Perseus to full advantage.
- 2 Applying a conceptual data schema in navigation / search makes it easier to understand.

Entity types

| | |
|---|--|
| <p>Object</p> <p>Note: For many objects, a distinction between intellectual/artistic work and item (physical copy) can be made; for example, <i>The Republic</i> by Plato as an intellectual work and a particular item (physical copy) with the text of that work. (More on that in Week 7)</p> <ul style="list-style-type: none"> . MediaObject (document) <ul style="list-style-type: none"> . . TextObject <ul style="list-style-type: none"> . . . LiteraryWork . . . Text describing an object . . Image <ul style="list-style-type: none"> . . . Photograph . . . DrawingOfRealObject . . . SchematicDrawing . Art object <ul style="list-style-type: none"> . . Building . . Sculpture . . Vase . . Coin <p>Person</p> | <p>Place</p> <ul style="list-style-type: none"> . Region . Site . MapLocation (in coordinates) <ul style="list-style-type: none"> . . Spot (a specific spot on a site) <p>Direction (as in northeast)</p> <p>ObjectType</p> <p>Values are all the object types listed above and more specific object types, such as theater under building or fortified city under site; used in Perseus to find all objects of a given type.</p> <p>TimeEntity</p> <ul style="list-style-type: none"> . Period . Date <p>Event</p> <p>Language</p> <p>WordEntity</p> <ul style="list-style-type: none"> . Word (a word underlies a group of forms) . WordForm (a string of characters) <p>Concept</p> <p>identified/named by keywords in Perseus</p> |
|---|--|

Relationship types

When you click on a hyperlink or a *Search* button in Perseus, you follow a relationship. For example, Perseus includes the statement Troad (a Region) *contains* Assos (a Site). When Troad is highlighted, clicking the *Search* button follows the *contains* link to find Assos (and other sites). A list of Perseus relationship types, along with the entity types they connect, is given on the next page.

| Relationships: Statement templates and sample statements | | |
|---|---------------------------|--|
| Entity | <isa> | Entity (for example, concept hierarchy) (isa = is a) |
| Object | <isa> | Object type (building, vase, coin, document, etc) |
| Assos | <isa> | Site |
| Object | <depicts> | Entity |
| Schematic drawing (site plan) | <depicts> | Site |
| Tampa 82.14.1 | <depicts> | Pillar (where Pillar is an Object type, "some pillar") |
| Image Smith 1989 | <depicts> | Assos (where Assos is a specific object) |
| Media object | <dealsWith> <mentions> | Object |
| Perseus Building Catalog, p. 535 | <deals with> | Assos Theater |
| Entity | <contains> | Entity (whole-part), e.g. for example |
| Text object | <contains> | Text object |
| Text object | <contains> | Word |
| Region | <contains> | Site |
| Troad | <contains> | Assos |
| Site | <locatedAt> | Map location |
| Assos | <locatedAt> | (39.5845 N, 26.2082 E) |
| Site | <contains> | Building |
| Assos | <contains> | Assos Theater |
| Word | <belongsTo> | Language |
| polis | <belongsTo> | Greek |
| Text object | <writtenIn> | Language |
| Herodotus 9.28.6 | <writtenIn> | Greek |
| Text object | <authoredBy> | Person, |
| Text object | <refersTo> | Text object |
| Crane 1996 | <refersTo> | Herodotus 9.28.6 |
| Text object | <translationOf> | Text object |
| Word | <translationOf> | Word |
| city | <translationOf> | polis |
| Word | <hasForm> | Word form |
| polis | <hasForm> | poleô̄s (poleos is the genitive singular) |
| Image | <takenAt> | (Spot, Direction) |

| Inverse relationships | | |
|----------------------------------|-----------------------|----------------------------------|
| Entity | <includesSpecific> | Entity |
| Object type | <includesSpecific> | Object |
| Site | <includesSpecific> | Assos |
| Entity | <depictedIn> | Object |
| Site | <depictedIn> | Schematic drawing (site plan) |
| Pillar | <depictedIn> | Tampa 82.14.1 |
| Assos | <depictedIn> | Image Smith 1989 |
| Object | <dealtWithIn> | Media object |
| Assos Theater | <dealtWithIn> | Perseus Building Catalog, p. 535 |
| Entity | <containedIn> | Entity |
| Text object | <containedIn> | Text object |
| Word | <containedIn> | Text object |
| Site | <containedIn> | Region |
| Assos | <contained in> | Troad |
| Map location | <isLocationOf> | Site |
| (39.5845 N, 26.2082 E) | <isLocationOf> | Assos |
| Building | <contained in> | Site |
| Assos Theater | <contained in> | Assos |
| Language | <isLanguageOf> | Word |
| Greek | <isLanguageOf> | polis |
| Language | <isLanguageOf> | Text object |
| Greek | <isLanguageOf> | Herodotus 9.28.6 |
| Person | <authorOf> | Text object |
| Text object | <referredToIn> | Text object |
| Herodotus 9.28.6 | <referredToIn> | Crane 1996 |
| Text object | <translationSourceOf> | Text object |
| Word | <translationSourceOf> | Word |
| polis | <translationSourceOf> | city |
| Word form | <isFormOf> | Word |
| pole $\acute{\alpha}$ s (= city) | <isFormOf> | polis |
| (Spot, direction) | <isReferencePointFor> | Image |

2 Step-by-step guide for exploring the features of Perseus

General Hints:

Use a computer with a fast Internet connection. Close other applications.

Perseus works better in Firefox than in Explorer, where pages often look a bit weird.

The WWW version of Perseus often shows only the top of a screen while the real results you want are further down. **Always scroll down a bit before assessing the result.**

Sometimes, Perseus opens a new browser window when you follow a link. You might have to click on the icon for that new window in the Window's task bar to actually see the window.

The Perseus system is not entirely reliable. The things in the exercise generally work, at least on the second try. Watch for your browser's indicator that it is downloading a lengthy document (esp. an image). The system did respond; it just takes a while. **Do not click the mouse while the browser is downloading the page.** The browser stores your clicks and acts on them as soon as the result of the previous click is available. So when you finally could see the results you have been waiting for, the browser immediately moves on to the next thing.

Hyperlinks, that is, elements that can be used as starting points in a next search step, are usually blue (and often underlined). Purple means you have been there.

In this assignment, an underlined word or phrase means: select this item by clicking on it.

In Perseus there are **two types of hyperlinks** (even though all hyperlinks appear the same)

- **Hardcoded hyperlinks** that lead to a specific place. These hyperlinks follow a specific relationship introduced between the two linked entities (see the conceptual data schema).
- **Hyperlinks that** simply call the *Look-up Tool* to **search** for the underlined word or phrase.

Search

Search works in two different modes.

(1) Search in document metadata. By default, when you press the search button (or just [Enter], Perseus searches only the author and the title of documents. (This is called *low exhaustivity of indexing*.)

(2) Search in full document text and for other types of entities This is a second step, invoked by clicking on [here](#) on the first result screen. Example: Search for *vase*.

Notes: (a) If none of the initial search terms occurs in the metadata, the system skips (1) and goes directly to (2). (b) Perseus looks for exactly the string you input even if "Search for all possible forms" is checked. (c) Several search terms are connected with OR as the default.

A note of caution. Perseus is a system very rich in information, linkages, and navigational possibilities. But its interface, while improved, may be hard to follow. If you are confused it might well be the system's fault. If you think there should be a better way to do some of the interactions, there probably is!

| Start of exploration | |
|--|---|
| Instruction on what to do <u>Underline</u> means: Click on this | Explanation |
| Go the Web address: <u>www.perseus.tufts.edu</u> | Note: Access to Perseus is sometimes unstable and slow. If that happens, use at off hours or try again. |
| First, look at some help files | |
| <u>Help</u> | Inside the logo band at the top of the page. Explore help topics if you wish, for example <u>Classics Collection Overview</u> You can follow links from there |
| <u>Perseus 4 Quick Start Guide</u> <u>Searching the collection</u> | You can read this if you like; the assignment always gives you instructions what to do. |
| Now look for sites, buildings, and a lot of information about them. Search using navigation | |
| [Collections/Texts] | Inside the logo band at the top of the page |
| <u>Art & Archaeology Artifact Browser</u> | |
| Browse one of the following artifact types: • <u>Site</u> | This defines the target of your search , the type of entities (namely sites) you want to find. Perseus displays two ways to search for sites, allowing you to select an entity type as the starting point in the next search step . |
| Choose a property of the Site artifact: • <u>Region</u> | Perseus displays a list of regions (starting entity values) to choose from. |
| Scroll down the list. Click on <u>Troad</u> | Perseus displays a list of all sites X for which Region Troad <contains> Site X is true. |
| <u>Assos</u> | Leads to a full description of the site. |

| | |
|---|--|
| | We will now do a search for images of Assos |
| <u>View Thumbnails (89)</u> | You can look at some of the images by clicking on the thumbnail |
| In the search box (upper right hand corner), type Assos and press [Enter] or click on <u>Search</u> | The 89 thumbnails do not include all images of Assos, so we are going to search another way which searches the entire collection. Look over the result screen and observe how it is divided by type of entity found |
| In the Matching Artifacts section on the right, click on <u>Assos, Theater</u> [Building] | See a description of the building, found following the link Object <dealt with in> Media object Note the uniform template used for building descriptions |
| <u>View Thumbnails</u> [10] | Lower SW corner and Upper NE section are particularly good images. You might also look at the Plan These images are not among the 89 shown under Assos |
| | Another way to find all buildings on the site Assos |
| [Collections/Texts] <u>Art & Archaeology Artifact Browser</u> | |
| <u>Building</u> | Defines the target of your search , the type of entities (in this case buildings) you look for. Now Perseus displays five ways to search for buildings, allowing you to specify the entity type to be used as starting point in the next search step . |
| Choose a property of the Building artifact: • <u>Context</u> | Context is the same as Site. Perseus displays a list of sites (starting entity values) to choose from. |
| • <u>Assos [9 Buildings]</u> | Displays a list of the 9 buildings. Site <contains> Building |
| <u>Assos, Theater</u> | Brings up the description you have seen before Object <dealt with in> Media object |

Second example of navigation search: *vases showing a pillar*

| | |
|--|---|
| <p>[Collections/Texts] Art & Archaeology Artifact Browser</p> | |
| <p>Browse one of the following artifact types:</p> <ul style="list-style-type: none"> • Vase | <p>You want to find objects of type vase (the search target)</p> |
| <p>Choose a property of the Vase artifact:</p> <ul style="list-style-type: none"> • Keyword | <p>You want to select a keyword as starting point for the search.</p> |
| <p>View Vases whose Keyword category is...</p> <ul style="list-style-type: none"> • Buildings [20 keywords] | <p>Browse the keyword list by category</p> |
| <p>View Vases with keyword category: Buildings...</p> <ul style="list-style-type: none"> • pillar [8 Vases] | <p>Now you select the specific keyword "pillar" as the starting point of your search Shows thumbnail images of 8 vases whose Keyword is "pillar" Object (Vase) <depicts> Object (Pillar)</p> |
| <p>Scroll down to Vase 7, Tampa 82.14.1</p> | <p>This vase has a distinct image of a pillar.</p> |
| <p>Click on the thumbnail image titled Woman at right</p> | <p>See a larger image.</p> |
| <p>Back, then click on Tampa 82.14.1: Read the Decoration Description, note "pillar"</p> | <p>Shows record with information about the vase. Some vases have a lot more information.</p> |

| Search through the search box. Topic: vases showing a pillar | |
|---|--|
| Type "pillar" in the search box top right and click [Search] | Finds all documents, images, artifacts, and dictionary entries in some way associated with "pillar" ("pillar" occurs in the text itself or in the description of an object) |
| Under Matching Artifacts (33) Uncheck all object types (entity types) except vases and click [View] | Narrow the scope of the search targets All the vases from before are found, plus three more. |
| Explore the Perseus Encyclopedia. More hypertext examples | |
| In the search box top right, enter <i>encyclopedia</i> and click on <u>Search</u> Under 3 from Greek and Roman Materials <u>Perseus Encyclopedia</u> : | Leads to the start of the alphabetically arranged Perseus Encyclopedia. Note: There used to be a very useful table of contents by broad subject categories, each leading to a listing of encyclopedia articles in that category. The alpha table of contents is not useful. |
| In the search results <u>Perseus Encyclopedia</u> (English) | |
| In the search box for the encyclopedia , type <i>Assos</i> and press [Enter]. | |
| Click on <u>More (2)</u> , then click on <u>entry assos</u> | The text fo this entry has not yet been entered, but there is a cross-reference |
| Optional: Click on [XML] Go back with browser Back button | Examine the entry with XML tags. You can do this for other pages as well |
| <u>Paus. 6.4.9</u> | This will open in a new tab or new window in your browser. You may have to click on that tab to see the text. Note: The first half sentence of the text is missing; it is "Sodamas from Assos in the Troad, a city ..." |
| Close the Pausanias tab or window | |
| Optional: Click on Pausanias, <i>Description of Greece</i> , <u>6.4.9</u> | In a new tab or Window, the Greek text of that section appears (this time complete) |

| | |
|---|--|
| <p>In the Perseus Encyclopedia search box, type <i>theater</i> and press [Enter].</p> | <p>Read the text. Note: <i>theater</i> and <i>theatre</i> are different strings and are treated separately. Here <i>theatre</i> does not find anything. Question: What would the system need to make it easier for the user?</p> |
| <p>Click on More (43)</p> | <p>See all results (encyclopedia entries, all in the same document, namely the encyclopedia). entry theater is toward the bottom of the list.</p> |
| <p>Scroll toward the bottom of the list, click on <u>entry theater</u></p> | <p>Examine the entry. Lot's of linked information.</p> |
| <p>Scroll down to a list of related encyclopedia articles. Click on any one of them</p> | <p>See related encyclopedia articles. These in turn have cross-references</p> |

| Locate a passage of Greek text and follow links from there More examples of hypertext. | |
|---|--|
| In the Perseus Encyclopedia, Theater, find at Athens: Paus. 1.21.1, Paus. 5.12.4, completed by Lycurgus: Paus. . . . Click on 1.29.16 | Leads to Greek Text, English translation. (This may open a new browser tab or window.) |
| In the Paus, box, delete .16 (so that only 1.29 remains) and press [Enter] | |
| Scroll down to see the entire section Click on footnote symbol 3, then click on the footnote symbol there to get back into the text. Follow a path to another cross-reference and back | Looking at the XML is instructive |
| Move around the text. Use the browser's Find (Ctrl-F) to quickly locate theater in the text. | |
| Ctrl-Home (to get to the top) Use the blue right arrow just above 29 to go to the next Section. | |
| Use the blue left arrow to get back to 1.29 | |
| Under Places (automatically extracted) View a map | |
| Zoom so the map focuses on the Mediterranean. Click on some of the red push-pins to see the name of the place. (Close one balloon before clicking on the next red push-pin.) | |
| Back In the bar Greek (1903) to the right click on load | Loads the Greek text on the side. (Clicking on focus puts the Greek text in the middle.) |
| Ctrl-End to get to the end of the page Under Display Preferences In the drop-down box Greek Display : select Latin transliteration [Update Preferences] | You can read the Greek text transliterated On transliteration, see See http://en.wikipedia.org/wiki/Transliteration |

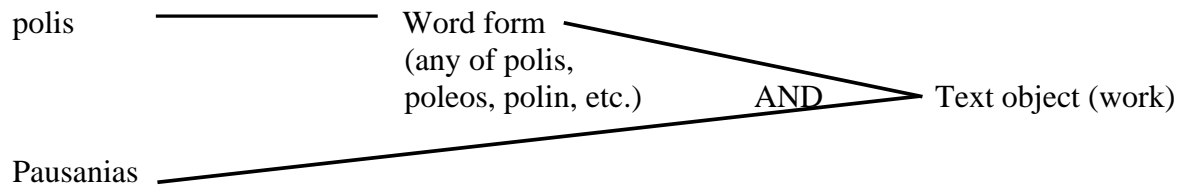
| Final section of the Perseus exercise. OPTIONAL, ADVANCED | |
|--|---|
| Words as entity types of interest. | |
| In the second paragraph of the transliterated text, a few words in, click on <u>poleôs</u> | In its own browser tab or window (you may need to click to open this tab), displays morphological analysis for this inflected Greek word. Word form (poleôs) <is form of> Word (polis) polis = city, poleos is the genitive singular |
| <u>Middle Liddell</u> | Briefly study this dictionary entry. (Users are asked to contribute word sense disambiguation data, which can be used as a training set by a computer program that learns automatic word sense disambiguation.) |
| Under Max, click on <u>1,223</u> The click on <u>More</u> to see all results (The cursor does not change as you mouse over more, but clicking on it works anyway.) | This uses an enhanced version of Perseus Search tool to search is for <i>polis</i> in Pausanias. Results in a list of Pausanias passages containing <i>polis</i> in various inflected forms (nominative, genitive <i>poleôs</i> etc., accusative <i>polin</i> , lemmatized search). Displays (after a while) in a new browser tab or window. The list includes Book 1, Chapter 29. |
| Click on <u>book 1, chapter 29</u> | The chapter is displayed with all inflected forms of <i>polis</i> highlighted |
| Close the tab or window with the search results | |
| <u>LSJ</u> | Have a look at the entry. Each of the references to an author is a hyperlink to the passage in question. |
| Follow on of the links | the text displays in a different tab; you may need to click on it. (If you follow an Od[yssey] link, the tab will be labeled Homer ...) |
| Close the tab/window Greek Word Study Tool | |
| You should be back in the text tab. If you do not see the Pausanias text, use the browser back button until you are back there. | |

More explanation of the Greek Word Search

You put in two **search criteria or starting points for the search:**

The Word *polis* and the Person *Pausanias*

Starting from *polis*, Perseus uses the relationship **Word** <has-form> **Word form** to find different forms of *polis*, such as *poleos* (genitive) and *polin* (accusative). In a second step, Perseus follows the relationship **Word form** <contained-in> **Text object** to find text objects (works). In parallel, Perseus starts from *Pausanias* and follows the relationship **Person** <author-of> **Text object** to find text objects (works). If a work is found both ways it meets both search criteria and is retrieved



| Search in the dictionary (continuation of Optional Advanced) | |
|--|--|
| <p>Under the general search box, click All Search Options</p> <p>In the line English-to-[Language] lookup, click Show</p> <p>In the drop-down box Search for, select the exact word</p> <p>then type <i>city</i> and press Enter</p> | <p>You will do a search for all lexicon entries for which the English definition contains the word <i>city</i>. More specifically, all dictionary entries in which <i>city</i> is part of the definition of the Greek head word</p> <p>This gives a (sort of) alphabetical list. It would be very nice if one could sort the list by frequency. As it is,</p> |
| <p>go to page 2 and find polis (the most frequent)</p> <p>Middle Liddell</p> | <p>The same entry you saw before when you started a word search from a word in a text</p> |
| <p>Back</p> <p>On page 1, find philopolis (Max. Inst. 51).</p> <p>Ctrl-F philo works best</p> | |
| | <p>In Perseus 3 one could get a list of Greek authors indicating how often each uses the word <i>philopolis</i>. One could then select an author and get to his passages in which <i>philopolis</i> appears.</p> <p>In Perseus 4, one can only find a list of all passages in which the word is used, sorted by author; the nice summary page is not available</p> |
| <p>In the philopolis line, click on 51.</p> | <p>A list of passages from all Greek authors that include the word philopolis, giving the sentence in which the word occurs.</p> |
| <p>You can click on the ID of the passage (blue) to get to the full passage.</p> <p>From there, you can click on any word to get to its dictionary entry</p> | |
| <p>In the same way, do a dictionary search for <i>theatre</i></p> <p>Watch the spelling, <i>theater</i> finds no dictionary entries</p> | |
| <p>End of Advanced</p> | |

After this guided tour, **do your own exploration**

You should note, that while Perseus is a wonderful system, it is quirky (especially Perseus 3), and its idiosyncracies get sometimes in the way of seeing principles. For example, in Perseus 3, full text search for a word does not find images that have been manually indexed by that word, only images that have the word in their caption. Sometimes, it just seems that the system is buggy. So if it does not do what it logically should do, that may be the reason

To log out, just close all browser windows.

Perseus 3 features lost in Perseus 4

Perseus 3 does not have as nice an interface and it is less stable, but it has a number of features that got lost in Perseus 4. Perseus 3 is still available at <http://perseus.mpiwg-berlin.mpg.de/>. An older version of the Perseus exercise using Perseus 3 will be posted on UBlerns.

The site descriptions were more complete and better structured

The sources and bibliography entries in the site descriptions were hyperlinks, and one could get a nicely formatted bibliographic record.

There was a button to plot all the sites mentioned in a site description (just as is now possible for texts)

On the site page there was a link [Buildings on this site](#) that would lead directly to a list of all the buildings on the site without having to do a search.

There was a site map with the many little circles with embedded arrows; clicking on a circle would bring up an image taken from that spot in the indicated direction. So one could take a walk around the site. Very cool, but does not work any more even in Perseus 3 on the Berlin site.

The dictionary search was more powerful, but the interface was not intuitive.

Assignment 2

Lecture 1.2

Assigned: May 21**Due: May 28****Bibliographic retrieval system exploration: Medline**

| | |
|---|--|
| Objectives | <ol style="list-style-type: none"> 1 You should understand the role of hierarchy in searching. 2 Optional: You should understand some problems of free-text searching. <p>This assignment is used a lot as an example in later lectures</p> |
| Deliverables to submit | <ol style="list-style-type: none"> 1 Answer Set 1 shown in Figure 4 with descriptors responsible for retrieval highlighted (for Documents 1 - 8) as explained in the assignment . 2 The filled-in work sheet (the work sheet asks some questions for analysis; it is found at the end of the assignment) (Part 1 required, Part 2 optional) 3 A brief individual essay (three-quarters of a page) summing up your experience with this assignment. (What have you learned? What is the significance of it?) |
| Tasks (details in assignment) | <p>Analyze results of a given search</p> <ol style="list-style-type: none"> 1 to explore hierarchy and facets in searching (1a and 1b) 2 (optional) to explore free-text search: Uncontrolled vocabulary |
| Materials | <p>This assignment is based on a search in Medline, an online database of medical journal articles produced by the National Library of Medicine and searchable, among other places, with DIALOG (an online search system). There is no need to search Medline; the search statements and the resulting lists of references (called <i>answer sets</i>) are presented as part of the assignment, slightly reformatted for ease of analysis. In this assignment, you will first read the explanations (Section 0) and then do your own analysis (Sections 1 and 2).</p> <p>The assignment assumes that you have done some online searching so you can imagine the actual search. Even having done the Perseus assignment gives some idea, and surely everyone has searched an online library catalog. You have learned or will learn the Dialog system in LIS 518 Reference Sources and Services.</p> <p>I will try to arrange Dialog access to search Medline for those who would like to do so. Ask the TA.</p> <p>Note: You will need highlighters or colored pencils in two contrasting colors.</p> |
| Time | 3 hrs |

Note: Focus your attention on the conceptual aspects of the search. The syntax of the commands and the presentation format of the answer sets are incidental; you are not required to do your own search so you do not need to concern yourself with these details.

Outline

0 Introduction

- 0.1 General introduction: Search topic and hierarchy
- 0.2 Hierarchically expanded searching or the mighty exclamation point.
- 0.3 Searching with combination. General explanation

1 Explore hierarchy and facets in searching

- 2.1 Combination search with hierarchically expanded searching: Answer Set 1
- 2.2 Focusing the search: Using more specific descriptors or adding query components.

2 Explore free-text search: Uncontrolled vocabulary. Two searches, AND vs. (With)

0 Introduction

0.1 General introduction: Search topic and hierarchy

Medline is a database of documents, more precisely, of document records (rather than the full text). Its main entity type is Document. The relationship type of primary interest in this assignment is Document *deals with* Concept. Concepts are expressed as *descriptors* or subject headings drawn from a controlled list (*controlled vocabulary*), namely **MeSH (Medical Subject Headings)**.

The search is about the general topic of **Hearing tests** assisted by computers (**Computing methodologies**); thus the topic is a combination of **Hearing tests** and **Computing methodologies**. An important concept illustrated in this search is the use of hierarchy in searching. To this end we introduce various subtopics in which one or both components are made more specific, for example, **Audiometry** (under Hearing tests) using **Signal processing, computer-assisted** (under Computing methodologies). Relevant excerpts from the **MeSH Hierarchy ("Tree Structures")** are shown in Figure 1. Look it over so that you can recognize some of the terms as you analyze retrieval results.

| | | | |
|----------------------------|---|----------------------------|---|
| C9 | Otorhinolaryngologic diseases | L1 | Information science |
| C9.218 | . Ear diseases | L1.700 | . Medical informatics |
| C9.218.458 | . . Hearing disorders | L1.700.508 | . . Medical informatics applications |
| C9.218.458.288 | . . . Deafness | L1.700.568 | . . . Medical informatics computing |
| C9.218.458.500 | Hearing loss, partial | L1.700.568.80 | Computer systems (<i>with more NT</i>) |
| C9.218.458.500.481 | Hearing loss, sensorineural | L1.700.568.110 | Computing methodologies |
| C9.218.458.500.481.432 | Hearing loss, central | L1.700.568.110.50 | Algorithms |
| C9.218.458.500.481.460 | Hearing loss, noise-induced | L1.700.568.110.65 | Artificial intelligence |
| C9.218.458.500.481.772 | Presbycusis | L1.700.568.110.65.190 | Expert systems |
| | | L1.700.568.110.65.250 | Fuzzy Logic |
| | | L1.700.568.110.65.580 | Natural language processing |
| | | L1.700.568.110.65.605 | Neural Networks (computer) |
| | | L1.700.568.110.65.630 | Robotics |
| E1 | Diagnosis | L1.700.568.110.85 | Automatic data processing |
| E1.276 | . Diagnosis, otorhinolaryngologic | L1.700.568.110.85.800 | Punched-card systems |
| E1.276.299 | . . Diagnosis, ear | L1.700.568.110.108 | Computer graphics |
| E1.276.299.375 | . . . Hearing tests | L1.700.568.110.108.150 | Computer-aided design |
| E1.276.299.375.100 | Acoustic impedance tests | L1.700.568.110.160 | Computer simulation |
| E1.276.299.375.297 | Audiometry | L1.700.568.110.308 | Image processing, computer-assisted |
| E1.276.299.375.297.45 | Audiometry, evoked response | L1.700.568.110.308.380 | Image-enhancement (<i>with more NT</i>) |
| E1.276.299.375.297.92 | Audiometry, pure-tone | L1.700.568.110.680 | Mathematical computing |
| E1.276.299.375.297.105 | Audiometry, speech | L1.700.568.110.680.310 | Decision support technics |
| E1.276.299.375.297.105.890 | Speech discrimination tests | L1.700.568.110.680.310.380 | Data interpretation, statistical |
| E1.276.299.375.297.105.902 | Speech reception threshold test | L1.700.568.110.680.310.428 | Decision theory |
| E1.276.299.375.330 | Dichotic listening tests | L1.700.568.110.680.310.480 | Decision trees |
| E1.276.299.375.570 | Recruitment detection (audiology) | L1.700.568.110.680.310.600 | Neural networks (computer) |
| E1.276.299.816 | Vestibular function tests | L1.700.568.110.680.700 | Numerical analysis, computer-assisted |
| E1.276.299.816.250 | Caloric tests | | |
| E1.276.299.816.435 | Electronystagmography | L1.700.568.110.800 | Signal processing, computer-assisted |
| E1.276.591 | . . Laryngoscopy | L1.700.568.180 | Computing milieu (<i>with more NT</i>) |
| E1.276.660 | . . Nasal provocation tests | L1.700.568.810 | Software (<i>with more NT</i>) |
| | | L1.725 | . . Pattern recognition |
| | | L1.725.500 | . . . Neural networks (computer) |
| | | L1.737 | . . Publishing (<i>with more NT</i>) |
| | | L1.906 | . . Systems analysis (<i>with more NT</i>) |

Note: The term numbers (also called codes or notations) make the connection between an alphabetical index and the hierarchy listing.

Figure 1. **Medical Subject Headings (MeSH). Hierarchy ("Tree structures"). Excerpts**

0.2 Hierarchically expanded searching or the mighty exclamation point.

A subject descriptor (a term or concept used to index a document), such as **Hearing tests**, that has narrower descriptors, for instance, **Audiometry**, **Acoustic impedance tests**, **Speech Discrimination Tests**, **Dichotic listening tests**, can be used for searching in two ways:

| | | |
|---|--------------------------------|---|
| <p>Query Intro 1 general references search</p> <p>998 documents Answer Set Intro 1 see Figure 2</p> | <p>S Hearing tests</p> | <p>Simple search, narrower descriptors not included (S is DIALOG's command for entering a search statement and having the system retrieve an answer set.) Answer Set Intro #1 includes only documents indexed by the broad descriptor Hearing tests itself. This is called a <u>general references</u> search because it finds only documents that deal with Hearing tests in general.</p> |
| <p>Query Intro 2 inclusive search</p> <p>7667 documents Answer Set Intro 2 see Figure 3</p> | <p>S Hearing tests!</p> | <p>Inclusive search, narrower descriptors included More often than not, a user asking for a search on Hearing tests expects to find not only documents on Hearing tests in general but also documents on all the specific types of Hearing tests as shown in the MeSH hierarchy (Figure 1); this is referred to as an inclusive or hierarchically expanded search. To do a hierarchically expanded search in DIALOG, one puts an ! after the descriptor, for example.</p> |

Check out the descriptors responsible for retrieval (shown in bold) in the general references search shown in Figure 2 and in the hierarchically expanded search shown in Figure 3. Of course, the hierarchically expanded search answer set includes all documents found in the general references search.

Note 1. In the document records some descriptors are marked with *. These descriptors indicate major topics in the document.

Note 2. Within a search, answer sets are numbered and within an answer set, documents are numbered.

Figure 2. **Answer Set Intro 1. General references search** (998 documents)
(Descriptors responsible for retrieval are shown in **bold**)

| | | |
|--|---|--|
| Query 1 | <p>?B 154 File 154:MEDLINE®) 1985-1998/Feb W4</p> <p>?S hearing tests Result: S1 998 documents found</p> <p>?T 1/8/1-5</p> | <p>Begins a new search in database (file) 154, Medline ? is the Dialog prompt, followed by the Dialog command</p> <p>General references search</p> <p>Print (Type) from set 1 in format 8 (title and descriptors) records 1-5.</p> |
| <p>1/8/1 ! Speech perception performance of children with a cochlear implant compared to that of children with conventional hearing aids. II. Results of prelingually deaf children. <i>Tags:</i> Human <i>Descriptors:</i> *Cochlear Implants; *Deafness--Rehabilitation; *Hearing Aids; Age of Onset; Deafness--congenital; Deafness--Etiology; Hearing Tests; Meningitis--Complications; Speech Discrimination Tests</p> <p>1/8/2 ! Speech perception performance of children with a cochlear implant compared to that of children with conventional hearing aids. I. The "equivalent hearing loss" concept. <i>Tags:</i> Comparative Study; Human <i>Descriptors:</i> *Cochlear Implants; *Deafness--Rehabilitation; *Hearing Aids; *Hearing Loss, Partial--Rehabilitation; *Speech Perception; Child ; Hearing Tests; Speech Discrimination Tests</p> <p>1/8/3 ! Audiometric evaluation of hearing loss in children. <i>Tags:</i> Human <i>Descriptors:</i> *Deafness--Diagnosis--DI; Child, Preschool; Cochlear Implantation; Deafness--Rehabilitation; Hearing Aids; Hearing Tests; Infant; Infant, Newborn</p> <p>1/8/4 ! Desferrioxamine in chronic progressive multiple sclerosis: a pilot study. <i>Tags:</i> Female; Human; Male <i>Descriptors:</i> *Antidotes--Pharmacology; *Deferoxamine--Pharmacology; *Multiple Sclerosis--Drug Therapy; Adult; Chronic Disease; Hearing Tests; Middle Age; Neurologic Examination; Pilot Projects</p> <p>1/8/5 ! Functional hearing results in revision stapes surgery. <i>Tags:</i> Comparative Study; Female; Human; Male <i>Descriptors:</i> *Hearing Loss, Conductive--Diagnosis; *Hearing Loss, Conductiveiology; *Ossicular Prosthesis; *Prosthesis Failure; *Reoperation; *Stapes Surgery--Adverse Effects; Adult; Aged; Auditory Threshold; Ear Ossicles--Surgery; Follow-Up Studies; Hearing Tests; Middle Age; Retrospective Studies</p> | | |

Figure 3. **Answer Set Intro 2. Hierarchically expanded search** (7667 documents found)

| | | |
|--|---|--|
| Query 2 | <p>?S hearing tests! Result: S2 7667 documents found. ?T 2/8/1-30</p> | <p>! means hierarchically expanded or inclusive search Print from set 2 in format 8 records 1-30</p> |
| <p>2/8/1 ! Comparison of TEOAE with Play audiometry for screening hearing problems in children. <i>Tags:</i> Comparative Study; Female; Human; Male <i>Descriptors:</i> *Audiometry, Pure-Tone; *Auditory Threshold--Physiology; *Mass Screening; *Otoacoustic Emissions, Spontaneous--Physiology; *Play and Playthings; Acoustic Impedance Tests; Child, Preschool; Hearing Loss, Conductive--Diagnosis; Hearing Loss, Conductive--Prevention and Control; Hearing Loss, Conductive--Physiopathology; Hearing Loss, Sensorineural--Diagnosis; Hearing Loss, Sensorineural--Prevention and Control; Hearing Loss, Sensorineural--Physiopathology; Reference Values; Sensitivity and Specificity</p> <p>2/8/5 ! Electrocochleographic evaluation of the guinea pig model of endolymphatic hydrops. <i>Tags:</i> Animal; Comparative Study; Female; Male <i>Descriptors:</i> *Audiometry, Evoked Response--Methods; *Disease Models, Animal; *Endolymphatic Hydrops--Diagnosis; Action Potentials; Auditory Threshold; Dehydration--Complications; Endolymphatic Hydrops--Chemically Induced; Glycerol; Guinea Pigs; Meniere's Disease--Physiopathology; Osmolar Concentration; Recruitment (Neurology); Reproducibility of Results; Urea</p> <p>2/8/8 ! Otitis-prone children and controls: a study of possible predisposing factors. 2. Physical findings, frequency of illness, allergy, day care and parental smoking. <i>Tags:</i> Female; Human; Male; Support, Non-U.S. Gov't <i>Descriptors:</i> *Otitis Media--Etiology; Acoustic Impedance Tests; Acute Disease; Audiometry; Case-Control Studies; Causality; Child Care; Child, Preschool; Hypersensitivity--Complications; Otitis Media--Microbiology; Otorhinolaryngologic Diseases--Complications; Parents; Physical Examination; Questionnaires; Recurrence; Retrospective Studies; Smoking--Adverse Effects</p> <p>2/8/13 = 1/8/3 ! Audiometric evaluation of hearing loss in children. <i>Tags:</i> Human <i>Descriptors:</i> *Deafness--Diagnosis; Child, Preschool; Cochlear Implantation; Deafness--Rehabilitation; Hearing Aids; Hearing Tests; Infant; Infant, Newborn</p> <p>2/8/34 ! Frequency discrimination of stylized synthetic vowels with a single formant. <i>Tags:</i> Female; Human; Male; Support, Non-U.S. Gov't <i>Descriptors:</i> *Phonetics; *Speech Perception; *Speech, Alaryngeal; Adult; Middle Age; Speech Discrimination Tests; Time Factors</p> <p>2/8/44 ! Developmental instability and cerebral lateralization. <i>Tags:</i> Female; Human; Male <i>Descriptors:</i> *Developmental Disabilities--Physiopathology; *Laterality--Physiology; Adolescence; Adult; Child; Cognition--Physiology; Developmental Disabilities--Psychology; Dichotic Listening Tests; Face; Middle Age; Neuropsychological Tests; Psychomotor Performance--Physiology</p> | | |

0.3 Searching with combination. General explanation

The topic of this search has two components, **Hearing tests** and **Computing methodologies**. Each component expresses a different aspect or **facet** of the search topic, namely **Diagnosis** and **General methodology** (with the more specific concept **Computing methodology** under it). **Both** component descriptors must be present to make a document relevant; the descriptors are combined with Boolean **AND**.

Hearing tests AND Computing methodologies.

The hierarchy under each descriptor is shown in Figure 1.

The user may be interested in all **Computing methodologies** (hierarchically expanded search) or only in a specific type, e.g. **Signal processing, computer-assisted**, or the user's interests may lie in between: She may need all documents that deal with, **Signal processing, computer-assisted** or **Computer simulation**, or both; the descriptors are combined by **OR**.

Signal processing, computer-assisted OR Computer simulation

Continuing the search, here is an example of the difference between general reference and hierarchically expanded searching when combining two descriptors (just look at the numbers):

| | | |
|--|------------|------------------------------|
| ?S computing methodologies | Result: S3 | 31 documents found |
| ?S computing methodologies! | Result: S4 | 73447 documents found |
| ?S hearing tests AND computing methodologies | Result: S5 | 0 documents found |
| ?S hearing tests! AND computing methodologies! | Result: S6 | 214 documents found |

1 Explore hierarchy and facets in searching

We now start a new search to explore the effects of using more and more specific descriptors in the query formulation and of adding more components joined with AND. Answer sets are numbered from 1 in the new search.

?B 154

Begin a new search

File 154:MEDLINE®) 1985-1998/Feb

?limitall/maj, eng

Limits the search to major descriptors (documents in which the topic is major), and to documents in English in all subsequent searches

1.1 Combination search with hierarchically expanded searching: Answer Set 1

| | |
|----------------|---|
| Query 1 | <p>?S hearing tests! AND computing methodologies! (now limited to major descriptors, English)</p> <p>Result: S1 29 documents found Answer Set 1, shown in Figure 4</p> <p>?T 1/8/1-29 Print Answer Set 1 in format 8</p> |
|----------------|---|

Figure 4. **Answer Set 1. Combination search with hierarchic expansion** (For Tasks 1a and 1b)

Task 1a: Visualize the combination of descriptors from two facets in retrieval, with hierarchy in each facet. For Documents **1 - 8** (enough to get the idea), highlight the descriptors responsible for retrieval, using red for the descriptors from the **Diagnosis facet** (e.g., **Hearing tests** or **Audiometry**) and green for the descriptor from the **Methodology facet** (e.g., **Algorithms**) (see hierarchy in Figure 1 or Figure 5). The descriptors of interest are bolded to facilitate your task. In this search only major descriptors (marked with *) are used for retrieval. **To be handed in.**

| Answer Set 1/Format 8/Document no | For the analysis required in 1.2, you need the answer sets in which document is included as given here. |
|--|---|
| <p>1/8/1 ! Speech recognition in noise. Development of a computerized test and preparation of test material. <i>Tags:</i> Human <i>Descriptors:</i> *Computer-Aided Design; *Noise--Adverse Effects; *Speech Perception; *Speech Reception Threshold Test; Audiometry, Speech ; Hearing Loss, Sensorineural</p> | <p>Answer Set 1</p> |
| <p>1/8/2 ! Effect of sampling frequencies and averaging resolution on medical parameters of auditory brainstem responses. <i>Tags:</i> Human <i>Descriptors:</i> *Audiometry, Evoked Response--Methods; *Evoked Potentials, Auditory, Brain Stem--Physiology; *Signal Processing, Computer-Assisted; Acoustic Stimulation; Adolescence; Adult; Aged; Artifacts; Child; Child, Preschool; Middle Age; Pattern Recognition; Reaction Time</p> | <p>Answer Sets 1, 2, 3, 4, 5</p> |
| <p>1/8/3 ! Composite probability modeling of increasing resolution where diagnostic covariates are unmeasurable for some subjects. <i>Tags:</i> Human <i>Descriptors:</i> *Algorithms; *Audiometry, Evoked Response--Statistical and Numerical Data; *Evoked Potentials, Auditory, Brain Stem--Physiology; *Models, Statistical; *Neuroma, Acoustic--Diagnosis; *Reaction Time--Physiology; Adult; Brain Stem--Physiopathology; Cochlear Diseases--Diagnosis; Cochlear Diseases--Physiopathology; Diagnosis, Differential; Neuroma, Acoustic--Physiopathology; Prognosis; ROC Curve; Sensitivity and Specificity</p> | <p>Answer Set 1</p> |
| <p>1/8/4 ! A neural network approach to the prediction of pure tone thresholds with distortion product emissions. <i>Tags:</i> Comparative Study; Female; Human; Male; Support, Non-U.S. Gov't <i>Descriptors:</i> *Audiometry, Pure-Tone; *Auditory Threshold; *Hearing Disorders--Diagnosis; *Neural Networks (Computer); *Otoacoustic Emissions, Spontaneous; Adolescence; Adult; Aged; Aged, 80 and over; Discriminant Analysis; Middle Age; Predictive Value of Tests; Reproducibility of Results</p> | <p>Answer Set 1</p> |
| <p>1/8/5 ! Automated electrophysiologic hearing testing using a threshold-seeking algorithm. <i>Tags:</i> Female; Human; Male; Support, U.S. Gov't, P.H.S. <i>Descriptors:</i> *Algorithms; *Audiometry, Evoked Response--Methods; *Auditory Threshold; *Hearing Disorders--Diagnosis; Adolescence; Adult; Aged; Aged, 80 and over; Audiometry, Pure-Tone; Child, Preschool; Diagnosis, Computer-Assisted--Methods; Evoked Potentials, Auditory, Brain Stem; Infant; Middle Age; Regression Analysis; Reproducibility of Results</p> | <p>Answer Set 1</p> |

1/8/6

Answer Sets **1, 2**

! Measurement of acoustic impedance and reflectance in the human ear canal.

Tags: Female; Human; Male*Descriptors:* ***Acoustic Impedance Tests**; ***Computer Simulation**; *Ear Canal--Physiology; *Hearing--Physiology; Adolescence; Adult; Auditory Threshold--Physiology; Calibration; Neural Networks (Computer) ; Pitch Perception--Physiology; Reference Values

1/8/7

Answer Sets **1, 2, 3, 4, 5**

! Electrocochleography in syphilitic hearing loss.

Tags: Female; Human; Male*Descriptors:* ***Audiometry, Evoked Response--Instrumentation**; *Labyrinthitis--Diagnosis; *Neurosyphilis--Diagnosis; ***Signal Processing, Computer-Assisted--Instrumentation**; Adult; Aged; Deafness, Sudden--Diagnosis; Deafness, Sudden--Physiopathology; Diagnosis, Differential; Edema--Diagnosis; Edema--Physiopathology; Endolymph--Physiology; Hearing Loss, Sensorineural--Diagnosis; Hearing Loss, Sensorineural--Physiopathology; Labyrinthitis--Physiopathology; Meniere's Disease--Diagnosis; Meniere's Disease--Physiopathology; Microcomputers; Middle Age; Neurosyphilis--Physiopathology

1/8/8

Answer Sets **1, 2, 3, 4, 5**

! Intraoperative electrocochleography of endolymphatic hydrops surgery using clicks and tone bursts.

Tags: Female; Human; Male*Descriptors:* ***Audiometry, Evoked Response--Instrumentation**; *Edema--Surgery; *Endolymph--Physiology; *Endolymphatic Shunt; *Intraoperative Monitoring--Instrumentation; *Meniere's Disease--Surgery; ***Signal Processing, Computer-Assisted--Instrumentation**; Acoustic Stimulation--Methods; Adolescence; Adult; Aged; Child; Cochlear Nerve--Physiopathology; Edema--Physiopathology; Meniere's Disease--Physiopathology; Middle Age

1/8/9

Answer Sets **1, 2, 3**

! An inexpensive alternative for recording middle ear muscle activity (MEMA) during sleep.

Tags: Comparative Study; Female; Human; Male*Descriptors:* ***Acoustic Impedance Tests--Instrumentation**; *Ear, Middle--Physiology; *Muscle Contraction--Physiology; *Polysomnography--Instrumentation; ***Signal Processing, Computer-Assisted--Instrumentation**; *Sleep Stages--Physiology; *Transducers, Pressure; Adult; Arousal--Physiology; Equipment Design; Middle Age; Reference Values

1/8/10

Answer Set **1**

! Komputerowa analiza, identyfikacja I graficzna prezentacja badan ABR--system audiometrii klinicznej. [Computer analysis, identification and graphic representation of ABR research--clinical audiometric system]

Tags: Human*Descriptors:* ***Audiometry, Pure-Tone**; ***Computer Graphics**; *Diagnosis, Computer-Assisted; *Hearing Disorders--Diagnosis; *Microcomputers; Programming Languages; Software Design

1/8/11

Answer Sets **1, 2, 3, 4**

! Analysis of the analog circuit's SNR in the selection of ADC bit resolution [letter; comment]

Tags: Human*Descriptors:* ***Audiometry--Instrumentation**; *Equipment Design--Standards; ***Signal Processing, Computer-Assisted--Instrumentation**; *Speech Intelligibility

1/8/12

Answer Sets 1, 6

! New draft ANSI standard enhances efforts in hearing conservation.

Tags: Human*Descriptors:* ***Audiometry--Methods**; ***Data Interpretation, Statistical**; *Ear Protective Devices--Standards; *Hearing Loss, Noise-Induced--Prevention and Control; *Mass Screening--Methods; *Occupational Diseases--Prevention and Control; Databases, Factual; Equipment Failure; United States

1/8/13

Answer Sets 1, 2, 3

! Simulating reflex induced changes in the acoustic impedance of the ear.

Tags: Human; Support, U.S. Gov't, Non-P.H.S.*Descriptors:* ***Acoustic Impedance Tests--Instrumentation**; *Reflex, Acoustic; ***Signal Processing, Computer-Assisted--Instrumentation**; Auditory Threshold; Equipment Design; Reference Values

1/8/14

Answer Sets 1, 2, 3, 4, 6

! An evaluation of two signal-processing hearing aids.

Tags: Comparative Study; Human; Support, U.S. Gov't, Non-P.H.S.*Descriptors:* *Hearing Aids; *Hearing Loss, Sensorineural--Rehabilitation; ***Signal Processing, Computer-Assisted--Instrumentation**; ***Speech Reception Threshold Test**; Aged; Equipment Design; Middle Age; Perceptual Masking

1/8/15

Answer Sets 1, 2

! Optimization of automated hearing test algorithms: a comparison of data from simulations & young children.

Tags: Comparative Study; Human; Support, U.S. Gov't, P.H.S.*Descriptors:* ***Algorithms**; *Auditory Threshold; ***Computer Simulation**; ***Hearing Tests--Methods**; *Models, Biological; Analysis of Variance; Child, Preschool; Infant

1/8/16

Answer Sets 1, 2

! Optimization of automated hearing test algorithms: simulations using an infant response model.

Tags: Human; Support, U.S. Gov't, P.H.S.*Descriptors:* ***Algorithms**; *Auditory Threshold; ***Computer Simulation**; ***Hearing Tests--Methods**; *Models, Biological; Analysis of Variance; Infant

1/8/17

Answer Sets 1, 2, 3, 4

! New hearing threshold measurements for pure tones under free-field listening conditions.

Tags: Female; Human; Male; Support, Non-U.S. Gov't*Descriptors:* ***Audiometry, Pure-Tone--Instrumentation**; *Auditory Threshold; ***Signal Processing, Computer-Assisted--Instrumentation**; Loudness Perception; Psychoacoustics; Reference Values

1/8/18

Answer Sets 1, 2, 3, 4, 6, 7

! On enhancement of spectral contrast in speech for hearing-impaired listeners.

Tags: Female; Human; Male; Support, U.S. Gov't, P.H.S.*Descriptors:* *Hearing Aids; *Hearing Loss, Sensorineural--Rehabilitation; *Microcomputers; ***Signal Processing, Computer-Assisted--Instrumentation**; *Sound Spectrography--Instrumentation; ***Speech Discrimination Tests**; Adult; Phonetics; Sensory Thresholds

1/8/19

Answer Sets 1, 6

! Classification of audiograms by sequential testing using a dynamic Bayesian procedure.

Tags: Human; Support, U.S. Gov't, P.H.S.*Descriptors:* ***Algorithms**; ***Audiometry, Pure-Tone--Classification--CL**; *Bayes Theorem; *Hearing Loss, Sensorineural--Diagnosis; Attention; Auditory Threshold; Computer Simulation; Reproducibility of Results

1/8/20

Answer Sets **1, 2, 3, 4**

! A comparison of the variability in thresholds measured with insert and conventional supra-aural earphones.

Tags: Comparative Study; Human*Descriptors:* ***Audiometry--Instrumentation**; ***Audiometry, Pure-Tone--Instrumentation**; *Auditory Threshold; ***Signal Processing, Computer-Assisted**; Adult; Equipment Design; Middle Age; Reference Values

1/8/21

Answer Sets **1, 2**

! Computer simulation of the patient for training in audiometry.

Tags: Human*Descriptors:* ***Audiometry**; ***Audiometry, Pure-Tone**; ***Computer Simulation**; Allied Health Personnel--Education

1/8/22

Answer Set **1**

! Comparison of manual and computer-controlled self-recorded audiometric methods for serial monitoring of hearing.

Tags: Comparative Study; Human*Descriptors:* ***Algorithms**; ***Audiometry--Methods**; ***Audiometry, Pure-Tone--Methods**; ***Automatic Data Processing**; Auditory Threshold--Physiology; Evaluation Studies; Follow-Up Studies

1/8/23

Answer Sets **1, 2, 3, 4, 5, 6, 7**

! Auditory brainstem evoked potentials peak identification by finite impulse response digital filters.

Tags: Female; Human; Male*Descriptors:* ***Audiometry, Evoked Response--Instrumentation**; *Brain Stem--Physiopathology; *Hearing Loss, Central--Diagnosis; *Hearing Loss, Sensorineural--Diagnosis; ***Signal Processing, Computer-Assisted**; Adult; Algorithms; Evoked Potentials, Auditory; Hearing Loss, Central--Physiopathology; Hearing Loss, Sensorineural--Physiopathology; Reaction Time--Physiology

1/8/24

Answer Sets **1, 2, 3, 4**

! Developmental changes in high-frequency sensitivity.

Tags: Female; Human; Male; Support, Non-U.S. Gov't*Descriptors:* ***Audiometry--Instrumentation**; ***Audiometry, Pure-Tone--Instrumentation**; *Child Development; *Hearing Loss, High-Frequency--Diagnosis; *Microcomputers; *Pitch Discrimination; ***Signal Processing, Computer-Assisted**; Adolescence; Adult; Auditory Threshold; Child; Child, Preschool; Hearing Loss, Noise-Induced--Diagnosis; Hearing Loss, Partial; Infant; Reference Values; Sound Localization

1/8/25

Answer Sets **1, 2, 3, 4**

! A comparison between coupler gain and insertion gain of hearing aids as used in partially hearing units in five schools in Belfast.

Tags: Comparative Study; Human*Descriptors:* ***Audiometry--Instrumentation**; ***Audiometry, Pure-Tone--Instrumentation**; *Deafness--Rehabilitation; *Education, Special; *Hearing Aids; *Microcomputers; ***Signal Processing, Computer-Assisted**; Auditory Threshold; Child; Psychoacoustics

1/8/26

Answer Sets **1, 2**

! Evaluation of stopping rules for audiological ascending test procedures using computer simulations.

Tags: Human; Support, U.S. Gov't, Non-P.H.S.*Descriptors:* ***Audiometry--Standards**; ***Computer Simulation**; *Models, Statistical; Auditory Threshold; Psychometrics; Random Allocation; Sampling Studies

Documents 27 - 29 skipped to save space

1.2 Focusing the search: Using more specific descriptors or adding query components.

(Continuation of the search started in 1.1, answer set numbering continues)

This page presents the remainder of the search statements (query formulations) for the main search. The answer sets are not printed; Figure 4 records for each document the answer sets in which it is included. The hierarchy of the search statements is shown in Figure 6 and the search results are summarized in Figure 7.

Task 1b: Analyze the results and answer questions 1.1 - 1.4 in the worksheet on p. 45.

| | | |
|-----------------------|---|--|
| <p>Query 2</p> | <p>?S hearing tests! AND (signal processing, computer-assisted OR computer simulation)</p> <p>Result: S2 20 documents found. Answer Set 2</p> <p>?T 2/6/1-20 (only two sample documents shown)</p> <p>2/6/1 ! Effect of sampling frequencies and averaging resolution on medical parameters of auditory brainstem responses.</p> <p>2/6/2 ! Measurement of acoustic impedance and reflectance in the human ear canal.</p> | <p>Second component narrowed</p> |
| <p>Query 3</p> | <p>?S hearing tests! AND signal processing, computer-assisted</p> <p>Result: S3 15 documents found. Answer Set 3</p> <p>Note: The print statements are not shown in the remainder</p> | <p>Second component narrowed further</p> |
| <p>Query 4</p> | <p>?S audiometry! AND signal processing, computer-assisted</p> <p>Result: S4 11 documents found. Answer Set 4</p> | <p>First component narrowed</p> |
| <p>Query 5</p> | <p>?S audiometry, evoked response AND signal processing, computer-assisted</p> <p>Result: S5 4 documents found. Answer Set 5</p> | <p>First component narrowed further</p> |

| | | |
|----------------|---|---|
| Query 6 | ?S s1 AND hearing loss, sensorineural! Result: S6 6 documents found. Answer Set 6 | Third component added, further restricting results |
| Query 7 | ?S s6 AND adult Result: S7 3 documents found. Answer Set 7 | Fourth component added, still further restricting results |

| | | | |
|----------------------------|--|----------------------------|---|
| C9 | Otorhinolaryngologic diseases | L1 | Information science |
| C9.218 | . Ear diseases | L1.700 | . Medical informatics |
| C9.218.458 | . . Hearing disorders | L1.700.508 | . . Medical informatics applications |
| C9.218.458.288 | . . . Deafness | L1.700.568 | . . . Medical informatics computing |
| C9.218.458.500 | . . . Hearing loss, partial | L1.700.568.80 | . . . Computer systems (<i>with more NT</i>) |
| C9.218.458.500.481 | Hearing loss, sensorineural | L1.700.568.110 | Computing methodologies |
| C9.218.458.500.481.432 | Hearing loss, central | L1.700.568.110.50 | Algorithms |
| C9.218.458.500.481.460 | Hearing loss, noise-induced | L1.700.568.110.65 | Artificial intelligence |
| C9.218.458.500.481.772 | Presbycusis | L1.700.568.110.65.190 | Expert systems |
| | | L1.700.568.110.65.250 | Fuzzy Logic |
| | | L1.700.568.110.65.580 | Natural language processing |
| | | L1.700.568.110.65.605 | Neural Networks (computer) |
| | | L1.700.568.110.65.630 | Robotics |
| E1 | Diagnosis | L1.700.568.110.85 | Automatic data processing |
| E1.276 | . Diagnosis, otorhinolaryngologic | L1.700.568.110.85.800 | Punched-card systems |
| E1.276.299 | . . Diagnosis, ear | L1.700.568.110.108 | Computer graphics |
| E1.276.299.375 | . . . Hearing tests | L1.700.568.110.108.150 | Computer-aided design |
| E1.276.299.375.100 | Acoustic impedance tests | L1.700.568.110.160 | Computer simulation |
| E1.276.299.375.297 | Audiometry | L1.700.568.110.308 | Image processing, computer-assisted |
| E1.276.299.375.297.45 | Audiometry, evoked response | L1.700.568.110.308.380 | Image-enhancement (<i>with more NT</i>) |
| E1.276.299.375.297.92 | Audiometry, pure-tone | L1.700.568.110.680 | Mathematical computing |
| E1.276.299.375.297.105 | Audiometry, speech | L1.700.568.110.680.310 | Decision support technics |
| E1.276.299.375.297.105.890 | Speech discrimination tests | L1.700.568.110.680.310.380 | Data interpretation, statistical |
| E1.276.299.375.297.105.902 | Speech reception threshold test | L1.700.568.110.680.310.428 | Decision theory |
| E1.276.299.375.330 | Dichotic listening tests | L1.700.568.110.680.310.480 | Decision trees |
| E1.276.299.375.570 | Recruitment detection (audiology) | L1.700.568.110.680.310.600 | Neural networks (computer) |
| E1.276.299.816 | . . . Vestibular function tests | L1.700.568.110.680.700 | Numerical analysis, computer-assisted |
| E1.276.299.816.250 | Caloric tests | | Signal processing, computer-assisted |
| E1.276.299.816.435 | Electronystagmography | L1.700.568.110.800 | Computing milieu (<i>with more NT</i>) |
| E1.276.591 | . . Laryngoscopy | L1.700.568.180 | Software (<i>with more NT</i>) |
| E1.276.660 | . . Nasal provocation tests | L1.725 | . . Pattern recognition |
| | | L1.725.500 | . . Neural networks (computer) |
| | | L1.737 | . . Publishing (<i>with more NT</i>) |
| | | L1.906 | . . Systems analysis (<i>with more NT</i>) |

Note: The term numbers (also called codes or notations) make the connection between an alphabetical index and the hierarchy listing.

Figure 5. Medical Subject Headings (MeSH). Hierarchy ("Tree structures"). Excerpts (Figure 1 repeated)

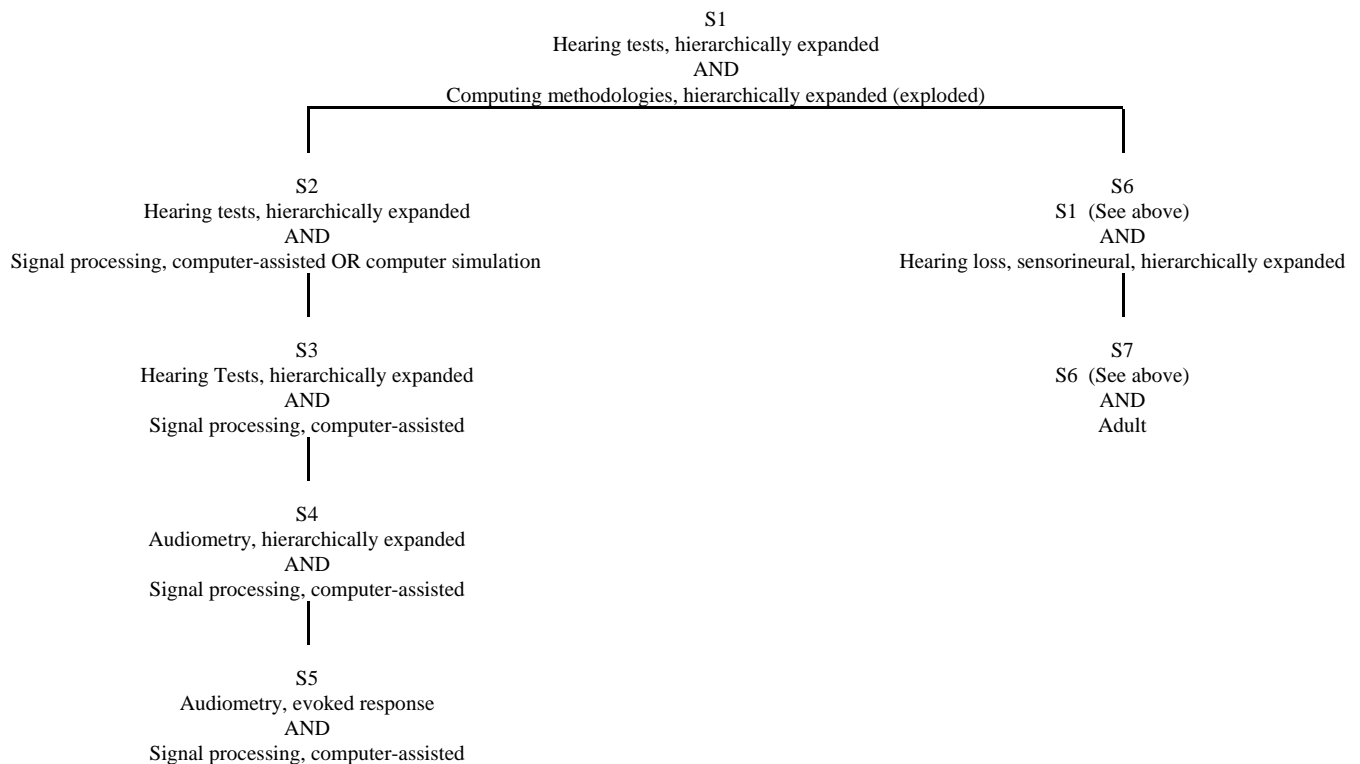


Figure 6: Hierarchy of query formulations

| | First title words | S1 | S2 | S3 | S4 | S5 | S6 | S7 |
|----|--------------------------------|----|----|----|----|----|----|----|
| 1 | Speech recognition in noise | ! | | | | | | |
| 2 | Effect of sampling frequency | ! | ! | ! | ! | ! | | |
| 3 | Composite probability model | ! | | | | | | |
| 4 | A neural network approach | ! | | | | | | |
| 5 | Automated electrophysiologic | ! | | | | | | |
| 6 | Measurement of acoustic | ! | ! | | | | | |
| 7 | Electrocochleography in syph | ! | ! | ! | ! | ! | | |
| 8 | Intraoperative electrocochleog | ! | ! | ! | ! | ! | | |
| 9 | An inexpensive alternative | ! | ! | ! | | | | |
| 10 | Komputerwa analiza, identyf | ! | | | | | | |
| 11 | Analysis of the analog circ | ! | ! | ! | ! | | | |
| 12 | New draft ANSI standard | ! | | | | | ! | |
| 13 | Simulating reflex induced | ! | ! | ! | | | | |
| 14 | An evaluation of two signal- | ! | ! | ! | ! | | ! | |
| 15 | Optimization...children | ! | ! | | | | | |
| 16 | Optimization...infant | ! | ! | | | | | |
| 17 | New hearing threshold meas | ! | ! | ! | ! | | | |
| 18 | On enhancement of spectral | ! | ! | ! | ! | | ! | ! |
| 19 | Classification of audiograms | ! | | | | | ! | |
| 20 | A comparison of the variabil | ! | ! | ! | ! | | | |
| 21 | Computer simulation of the | ! | ! | | | | | |
| 22 | Comparison of manual and | ! | | | | | | |
| 23 | Auditory brainstem evoked | ! | ! | ! | ! | ! | ! | ! |
| 24 | Developmental changes in | ! | ! | ! | ! | | | |
| 25 | A comparison between coupl | ! | ! | ! | ! | | | |
| 26 | Evaluation of stopping rules | ! | ! | | | | | |
| 27 | Validation of addition and | ! | | | | | | |
| 28 | Temporal characteristics of | ! | ! | ! | | | | |
| 29 | Click-evoked oto-acoustic | ! | ! | ! | | | ! | ! |

Figure 7: Search result summary. Documents found with search statements S1 - S7

2 Exploring free-text search: Uncontrolled vocabulary. AND vs. (With)

(Task 2, Worksheet questions 2.1 - 2.3 on p. 46) This section is optional

| | | |
|-----------------------|---|---|
| <p>Query 8</p> | <p>?S hearing/ti AND test?/ti</p> <p>Result: S8 183 documents found Answer Set 8, Figure 8.</p> | <p>Search words in the title (not assigned descriptors) The title must contain both words somewhere ? is the truncation symbol (test, tests, testing are all found)</p> <p>75% of the documents found are relevant (the restriction to title words helps); the selection here over-represents false retrieval so that the reasons for false retrieval can be explored.</p> |
| <p>Query 9</p> | <p>?S hearing/ti (W) test?/ti</p> <p>Result: S9 36 documents found Answer Set 9, Figure 9.</p> | <p>Search in the title, word 1 followed immediately by word 2.</p> <p>97% of all documents found are relevant</p> |

Figure 8. Answer Set 8: hearing/ti AND test?/ti

8/8/1

! **Hearing tests** in extended high frequency range in pre-school age children. Initial results

Tags: Female; Human; Male

Descriptors: *Audiometry, Pure-Tone--Instrumentation; *Hearing Loss, High-Frequency--Prevention and Control; *Mass Screening; Adult; Auditory Threshold; Berlin; Child; Child, Preschool; Hearing Loss, High-Frequencyiology; Reference Values

8/8/2

! IHAF loudness contour **test:** reliability and effects of approach mode in normal-**hearing** subjects.

Tags: Comparative Study; Human

Descriptors: *Hearing--Physiology; *Loudness Perception; Audiometry, Pure-Tone; Hearing Aids; Random Allocation; Reproducibility of Results; Speech Perception

8/8/3

! Effects of otitis on **hearing** in dogs characterised by brainstem auditory evoked response **testing**.

Tags: Animal; Support, Non-U.S. Gov't

Descriptors: *Dog Diseases--Physiopathology; *Evoked Potentials, Auditory, Brain Stem--Physiology; *Hearing Loss, Functional--Veterinary; *Otitis--Veterinary; Dogs; Hearing Loss, Functional--Complications; Hearing Tests--Methods; Hearing Tests--Veterinary; Otitis--Complications; Otitis--Physiopathology

8/8/4

! New computerized Finnish speech in noise **test** and binaural **hearing**.*Tags:* Human*Descriptors:* *Hearing--Physiology; *Speech Perception--Physiology--PH ; *Speech Reception Threshold Test--Methods; Reference Values

8/8/6

! Otoacoustic emission and auditory efferent function **testing** in normal subjects and patients with sensori-neural **hearing** loss*Tags:* Human*Descriptors:* *Auditory Pathways--Physiopathology; *Cochlea--Physiopathology; *Hearing Loss, Sensorineural--Physiopathology; *Otoacoustic Emissions, Spontaneous; Acoustic Impedance Tests; Adult; Evoked Potentials, Auditory, Brain Stem; Neuroma, Acoustic--Physiopathology

8/8/7

! **Test** of the health promotion model as a causal model of construction workers' use of **hearing** protection.*Tags:* Female; Human; Male; Support, U.S. Gov't, P.H.S.*Descriptors:* *Ear Protective Devices; *Health Promotion; *Hearing Loss, Noise-Induced--Prevention and Control; *Models, Nursing; *Models, Psychological; *Occupational Diseases--Prevention and Control; Adult; Analysis of Variance; Factor Analysis, Statistical; Health Behavior; Knowledge, Attitudes, Practice; Predictive Value of Tests; Questionnaires; Reproducibility of Results

8/8/8

= 9/8/1

! Preauricular tags and pits in the newborn: the role of **hearing tests**.*Tags:* Human*Descriptors:* *Audiometry; *Ear, External--Abnormalities; Evoked Potentials, Auditory, Brain Stem; Hearing Disorders--Congenital; Hearing Disorders--Complications; Infant; Infant, Newborn; Prospective Studies

8/8/11

= 9/8/2

! Use of the distraction **hearing test** in children with congenital ocular motor apraxia.*Tags:* Case Report; Female; Human*Descriptors:* *Apraxia--Complications; *Hearing Loss, Sensorineural--Complications; *Hearing Loss, Sensorineural--Diagnosis; *Hearing Tests--Methods; Child; Eye Movements

8/8/21

= 9/8/5

! Early **hearing testing** of still critically ill neonates.*Tags:* Female; Human; Male; Support, Non-U.S. Gov't*Descriptors:* *Audiometry, Evoked Response--Methods; *Hearing Loss, Sensorineural--Diagnosis--DI; *Intensive Care Units, Neonatal; Audiometry, Evoked Response--Instrumentation; Chi-Square Distribution; Critical Illness; Evoked Potentials, Auditory, Brain Stem

8/8/22

! Computer-stimulated **test** fitting of an implantable **hearing** aid using implantable **hearing** aid using three-dimensional CT scans of the temporal bone: preliminary study.

Tags: Human

Descriptors: *Cochlear Implant; *Image Processing, Computer-Assisted; *Prosthesis Fitting; *Temporal Bone--Radiography; *Tomography, X-Ray Computed; Deafness--Rehabilitation

8/8/29

! Efficiency of screening procedures for assigning levels of the Stanford Achievement **Test** (eighth edition) to students who are deaf or hard of **hearing**.

Tags: Human

Descriptors: *Deafness; *Educational Measurement; *Mathematics; *Reading

8/8/36

! Classroom attributes and achievement **test** scores for deaf and hard of **hearing** students.

Tags: Human; Support, U.S. Gov't, Non-P.H.S.

Descriptors: *Achievement; *Deafness; *Hearing Disorders; Adolescence; Child; Cohort Studies; Mathematics; Reading; Schools; Students

8/8/41

= 9/8/7

! Correlation between the American Academy of Otolaryngology-Head and Neck Surgery five-minute **hearing test** and standard audiologic data.

Tags: Female; Human; Male

Descriptors: *Hearing Disorders--Diagnosis; *Hearing Tests; Aged; Middle Age; Questionnaires; Sensitivity and Specificity

8/8/47

= 9/8/8

! Automated electrophysiologic **hearing testing** using a threshold-seeking algorithm.

Tags: Female; Human; Male; Support, U.S. Gov't, P.H.S.

Descriptors: *Algorithms; *Audiometry, Evoked Response--Methods; *Auditory Threshold; *Hearing Disorders--Diagnosis; Adolescence; Adult; Aged; Aged, 80 and over; Audiometry, Pure-Tone; Child, Preschool; Diagnosis, Computer-Assisted--Methods; Evoked Potentials, Auditory, Brain Stem; Infant; Middle Age; Regression Analysis; Reproducibility of Results

8/8/54

= 9/8/11

! Fair **hearing testimony**.

Tags: Female; Human

Descriptors: *Medical Staff Privileges--Standards; *Obstetrics --Standards; *Physicians, Family; Cesarean Section; Expert Testimony; Missouri; Pregnancy; Societies, Medical; United States

8/8/65

! RNs **testify** at DHS **hearing**: 'our patients need protection!'

Tags: Human

Descriptors: *Health Services--Legislation and Jurisprudence; *Nurses ; *Patient Advocacy--Legislation and Jurisprudence; California; Expert Testimony; Societies, Nursing

8/8/68

! New immunobiological **tests** in the investigation of Meniere's disease and sensorineural **hearing** loss.

Tags: Female; Human; Male

Descriptors: *Autoantibodies--Analysis; *Hearing Loss, Sensorineural --Immunology; *Labyrinth--Immunology; *Meniere's Disease --Immunology; Adult; Autoantigens--Immunology; Blotting, Western; Hearing Loss, Sensorineural--Diagnosis; IgG--Analysis; Lymphocyte Transformation--Immunology; Meniere's Disease--Diagnosis--DI CAS Registry No.: 0 (Autoantibodies); 0 (Autoantigens); 0 (IgG)

8/8/98

! A **test** suite for **hearing** aid evaluation.

Tags: Human; Support, U.S. Gov't, Non-P.H.S.

Descriptors: *Acoustics; *Hearing Aids--Standards; *Signal Processing, Computer-Assisted; Analog-Digital Conversion; Mathematical Computing; Sound

8/8/115

! Carolina Picture Vocabulary **Test**: validation with **hearing**-impaired students.

Tags: Comparative Study; Female; Human; Male

Descriptors: *Deafness--Psychology; *Intelligence Tests; *Manual Communication; *Sign Language; *Vocabulary; Child; Wechsler Scales

8/8/129

! New **tests** for identifying **hearing**-impaired students with visual perceptual deficits: relationship between deficits and ability to comprehend sign language.

Tags: Human

Descriptors: *Hearing Loss, Partial--Diagnosis; *Manual Communication ; *Perceptual Disorders--Diagnosis; *Psychological Tests; *Sign Language; *Visual Perception; Cognition; Psychometrics; Students --Psychology

8/8/155

! Classification of **hearing** loss in industrial workers for the purpose of expert **testimony**

Tags: Human

Descriptors: *Expert Testimony; *Hearing Disorders--Classification; *Mining; *Occupational Diseases--Classification; Disability Evaluation; Hearing Loss, Noise-Induced--Classification; Noise, Occupational --Adverse Effects; Poland

8/8/159

! The use of acoustical **test** fixtures for the measurement of **hearing** protector attenuation. Part II: Modeling the external ear, simulating bone conduction, and comparing **test** fixture and real-ear data.

Tags: Human

Descriptors: *Bone Conduction; *Ear Protective Devices--Standards; *Ear, External--Physiology; *Models, Anatomic; *Protective Devices --Standards; Auditory Threshold; Ear Canal--Physiology--PH; Perceptual Masking; Pitch Discrimination; Psychoacoustics

Figure 9. Answer Set 9: hearing/ti (W) test?/ti

| | |
|---|----------|
| 9/8/1 | = 8/8/8 |
| ! Preauricular tags and pits in the newborn: the role of hearing tests . | |
| <i>Tags:</i> Human | |
| <i>Descriptors:</i> *Audiometry; *Ear, External--Abnormalities; Evoked Potentials, Auditory, Brain Stem; Hearing Disorders--Congenital; Hearing Disorders--Complications; Infant; Infant, Newborn; Prospective Studies | |
| 9/8/2 | = 8/8/11 |
| ! Use of the distraction hearing test in children with congenital ocular motor apraxia. | |
| <i>Tags:</i> Case Report; Female; Human | |
| <i>Descriptors:</i> *Apraxia--Complications; *Hearing Loss, Sensorineural --Complications; *Hearing Loss, Sensorineural--Diagnosis; *Hearing Tests--Methods; Child; Eye Movements | |
| 9/8/5 | = 8/8/21 |
| ! Early hearing testing of still critically ill neonates. | |
| <i>Tags:</i> Female; Human; Male; Support, Non-U.S. Gov't | |
| <i>Descriptors:</i> *Audiometry, Evoked Response--Methods; *Hearing Loss, Sensorineural--Diagnosis; *Intensive Care Units, Neonatal; Audiometry, Evoked Response--Instrumentation; Chi-Square Distribution; Critical Illness; Evoked Potentials, Auditory, Brain Stem | |
| 9/8/7 | = 8/8/41 |
| ! Correlation between the American Academy of Otolaryngology-Head and Neck Surgery five-minute hearing test and standard audiologic data. | |
| <i>Tags:</i> Female; Human; Male | |
| <i>Descriptors:</i> *Hearing Disorders--Diagnosis; *Hearing Tests; Aged; Middle Age; Questionnaires; Sensitivity and Specificity | |
| 9/8/8 | = 8/8/47 |
| ! Automated electrophysiologic hearing testing using a threshold-seeking algorithm. | |
| <i>Tags:</i> Female; Human; Male; Support, U.S. Gov't, P.H.S. | |
| <i>Descriptors:</i> *Algorithms; *Audiometry, Evoked Response--Methods; *Auditory Threshold; *Hearing Disorders--Diagnosis; Adolescence; Adult; Aged; Aged, 80 and over; Audiometry, Pure-Tone; Child, Preschool; Diagnosis, Computer-Assisted--Methods; Evoked Potentials, Auditory, Brain Stem; Infant; Middle Age; Regression Analysis; Reproducibility of Results | |
| 9/8/11 | = 8/8/54 |
| ! Fair hearing testimony . | |
| <i>Tags:</i> Female; Human | |
| <i>Descriptors:</i> *Medical Staff Privileges--Standards; *Obstetrics --Standards; *Physicians, Family; Cesarean Section; Expert Testimony; Missouri; Pregnancy; Societies, Medical; United States | |

Worksheet for Assignment 2. Medline

Questions on the main search queries using assigned subject descriptors (Section 1.2, Task 1.2)

- 1.1 What effect does focusing the query formulation have on the number of documents found?

Note for questions 1.2 - 1.4: The listing for Answer Set 1 (Figure 4) also gives for each document the other answer sets in which it was found.

- 1.2 Which descriptors are responsible for retrieving Document 6 into Answer Set 2? How about Document 7? Why is Document 7 retrieved into Answer Set 3 but not Document 6?

- 1.3 Why is Document 18 retrieved into Answer Set 4 but not into Answer Set 5?

- 1.4 In Answer Set 6 (Query: **s1 AND hearing loss, sensorineural!**), Documents 14, 18, 19, 23, and 27 all have the descriptor **hearing loss, sensorineural**, but why was Document 12 retrieved?

Questions on the free-text queries (Section 2, Task 2) (optional)

- 2.1 In this question you compare the controlled vocabulary search (based on descriptors / subject headings assigned from a controlled list) with free-text search (based on words occurring in the document; in our search restricted to title, but one could also look in the abstract or the full text). Look at Answer Set 1 (Figure 4); among the first 10 documents, how many would be retrieved by free-text query 8, **hearing/ti AND test?/ti**?

- 2.2 Comparing Answer Sets 8 and 9, would you advise a user who needs complete retrieval to use query 9? Why does query 9 miss so many relevant documents?

- 2.3 Answer Set 8 includes a number of irrelevant documents. List the reasons for false retrieval that you can observe. Which of these reasons persist into query 9?

Assignment 3
Lecture 1.2

Assigned: May 21
Due: June 4

Online Catalog Search

| | |
|-------------------------------|---|
| Objective | Students should understand the significance of catalog structure, especially the record format used, for searching. |
| Deliverables to submit | No deliverable. (This assignment is like a reading: You are expected to know what can be learned from it.) |
| Task | Explore OCLC's Connexion system for searching. This is OCLC's system for information professionals; later you will use this system for cataloging. OCLC's system for the public is FirstSearch, which can be used for searching several databases, including WorldCat and ArticleFirst. |
| Time | 1.5 hrs |

Many of the examples are based on the following record from the model catalog.

| | | |
|---|-------------|---------------|
| Deese-Roberts, Susan | | |
| Library instruction : a peer tutoring mode. / [by] Susan Deese-Roberts [and] Kathleen Keating | | |
| Englewood, Colo. : Libraries Unlimited, 2000. | | |
| xiv, 212 p. : ill. ; 26 cm | | |
| Bibliography: p. 359-356 | | |
| 1563086522 (pbk.) | | |
| Z711.2 | 371.2 | 00-55850 |
| [LC Call no.] | [Dewey no.] | [LC Card no.] |
| 1. Library orientation for college students -- United States. 2. Peer-group tutoring of students -- United States. I. Deese-Roberts, Susan, 1953- , joint author. II. Keating, Kathleen,; 1958- joint author. III. Title (Series) | | |

You will search for Seymour Lubetzky, the preeminent figure in cataloging in the 20th century

To save time, in the searches described below, you can copy the queries from the emailed Word document and paste them into the *Search for* box. (Windows tip: Use Alt-Tab (hold down the Alt key while pressing Tab) to switch between the browser and Word.)

Search on OCLC/Connexion

Log in`

<http://connexion.oclc.org/>

Authorization (Name): 100062747

Password: SILS2

Click on Logon

In the bar on the top,

click on ⇒Search

Note: You can also explore other functions in Connexion, especially the Resource Catalog (Cataloging > Search > select Resource Catalog in drop-down box). It lets you view bibliographic records for all kinds of documents, including many Web sites, in a variety of formats, including MARC and Dublin Core.

Note: The prompt numbers you see may be different from the ones given in the following instructions. The prompt numbers may change from session to session.

Searching in OCLC/Connexion

This will all become obvious as you do it.

In the drop-down box under *Search*, select WorldCat

In the top box labeled **Search for:** enter search criteria, then click on [\[Search\]](#).

Searching in the top box is easiest. Use Boolean AND to combine search criteria. Prefixes that specify the role of the search criterion, such as **au:** for Author or **su:** for Subject can be found in the drop down box associated with the second Search for box.

In the result list, click on an item number to see the detailed record. If you want to step through the entire list looking at detailed records, click on arrow next to the little box that shows the result number.

To go back to the result list, click on [\[Search Results\]](#) (the browser's Back button does not work)

. You can edit the **Search for** box and click on the [\[Search\]](#) button for a new search. Of course, you can copy and paste to and from the **Search for** box.

There are more things you can explore following the instructions on the screen

To learn more, consult Help (not needed for this assignment)

Introduction

To search for books authored by Seymour Lubetzky, that is

Seymour Lubetzky (known starting point) $\xrightarrow{\langle authorOf \rangle}$ Book X (unknown, to find)

enter **au: Seymour Lubetzky** and click search (Lubetzky as author is in field 100 or 700)

or to search for books about Seymour Lubetzky, that is

Seymour Lubetzky (known starting point) $\xrightarrow{\langle treatedIn \rangle}$ Book X (unknown, to find)

enter **su: Seymour Lubetzky** and click search (Lubetzky as subject is in MARC field 600)

The box **Indexed in:** has a drop-down list of field labels (corresponding to relationship types).

Start of searching (instructions are first detailed, then you just follow the same pattern)

in the Search for box, enter **au: Seymour Lubetzky** and click [Search]

In Group 2, click on Books

In the Search results, click on no. 96

You see a MARC (MACHINE Readable Cataloging) record. Ignore the top gray box. You can also ignore the field tags and just read the data in the record (scroll down to the end of the record). Or you can go to Model Catalog document (the first reading) to see what the tags mean)

To see a more user-friendly record, In the drop-down box labeled **View** select **Display Find in a Library**. A new browser tab opens. Look at the record

Note: You can click on anything blue to start a new search. For example, you could start an author search for Robert L Maxwell by just clicking on the name. Or you could search for either of the subject headings displayed.

You can find libraries based on you zip code, but that is not the focus here

Next search: **su: Seymour Lubetzky**

Select result 10 (author Hayes) and display in the same way

In the library display screen, be sure to examine the entire record (scroll all the way down)

For the rest of the exercise, just search queries are given with very few further instructions or comments. Select one or two items to display for each search; generally the MARC record will do.

au: Seymour Lubetzky AND su: Seymour Lubetzky

su: Henry Kissinger AND au: Henry Kissinger Autobiographies

au: Susan-Deese-Roberts

au: Kathleen Keating (there are several people by that name who authored books)

au: Kathleen Keating AND ti: instruction

Next search for all documents whose title (or subtitle) starts with *library instruction*. This is called a **phrase search** and indicated by = after the field label.

First try

ti: library instruction (searched as ti: library AND ti: instruction)

Finds all documents for which both words *library* and *instruction* occur anywhere in the title, the subtitle, or an alternate title, or field 5xx. **Finds too many documents.** If we could restrict the search to the title proper, it might be ok. This is a question of exhaustivity of indexing, as we shall see.

Now do

ti= library instruction

Finds only documents for which the phrase *library* and *instruction* occurs at the beginning of the title, the subtitle, or an alternate title, or field 5xx (*the* and other articles at the beginning are ignored). Instead of over 2000 records, this finds 59

Examine the results. Note the duplication of the same title, either referring to exactly the same work or to different editions. (Note: If there was a book titled *The library instruction manual*, it would also be found)

bn: 1563086522 International Standard Book Number

As you surely have noticed in earlier searches: One book, many records in OCLC

The remainder of this exercise is concerned with **subject searching**. The sample topic is *library instruction*. There are two dimensions of the query formulation that can be varied independently:

1 Format of the search argument, which can be

- **words** (which can optionally be truncated) connected with AND
- **words connected with a proximity operator** (several choices specifying distance and whether or not sequence should be preserved). We just do w1 (words with at one word in between, sequence does not matter)
- a **phrase** (phrase must be at the beginning of an element in the record, such as title, subtitle, subject heading, or subheading)

2 Choice of data field(s) to search

All available fields are listed in the drop-down box *Indexed in*: Some important fields:

- kw title, subject headings, and other fields that contain words (not se or co)
- ti only title-like fields
- su subject headings and other MARC 6XX fields (also de for descriptor)
- lc Library of Congress Classification class number
- dd Dewey Decimal Classification class number
- other fields, such as **series**, **corporate name**, **conference name**

Try the searches in the table and compare the results. You may need to look at a full record to understand why a document was found. (CTRL-F in the browser will help you find the word you are looking for on the page displayed.)

To get search results of manageable size, append to each query **and yr: 2000-2009**

| | Boolean | Proximity | Phrase |
|----|--|---------------------------------------|--------------------------------|
| kw | kw: library AND kw: instruction | kw: library w1 kw: instruction | |
| ti | ti: library AND ti: instruction | ti: library w1 ti: instruction | ti= library instruction |
| su | su: library AND su: instruction | su: library w1 su: instruction | su= library instruction |

Notes:

From left to right, the searches go from broad to more focused.

In Boolean searches, the words could be searched in different fields, for example one word in ti, the other in kw.

It is also possible not to specify a field at all. For example, searching for just *amaranth* finds documents by or about people or organizations whose name includes Amaranth, or whose title includes amaranth (referring to the grain, a person, or a place, such as Amaranth, ON), or that are indexed by the subject heading amaranth.

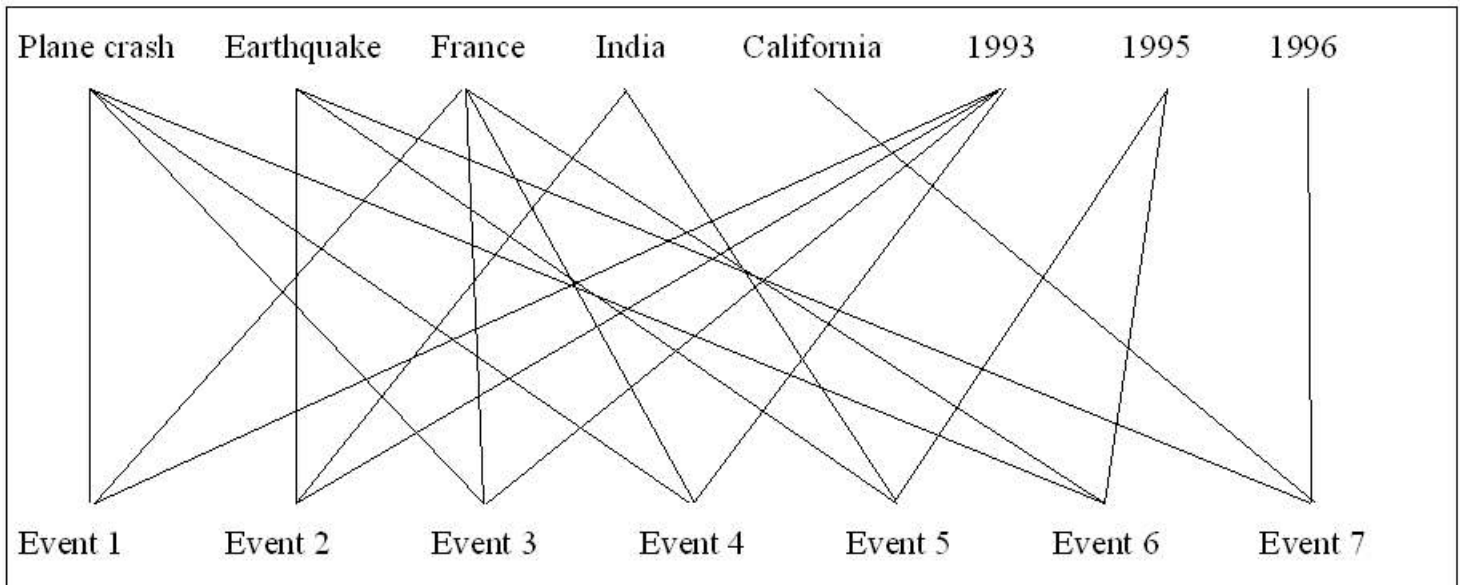
Assignment 4.
Lecture 2.2

Assigned: May 28
Due: June 4

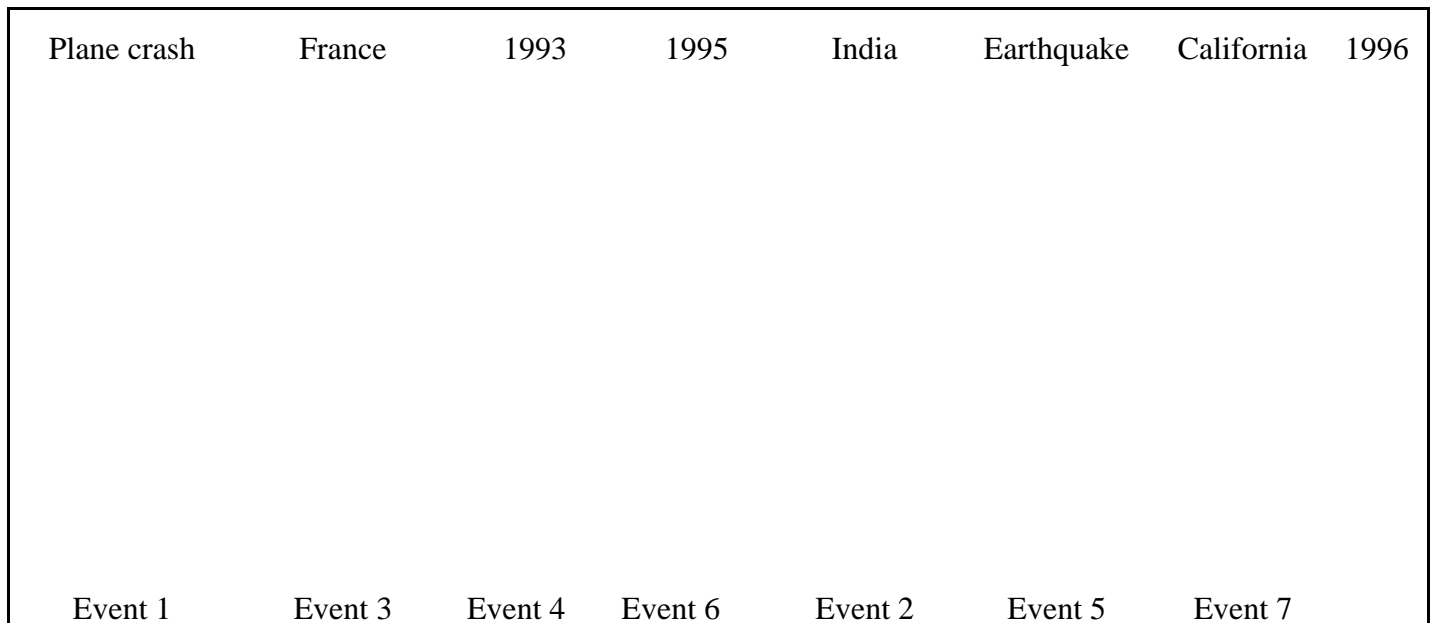
Restructuring a semantic network

| | |
|-------------------------------|---|
| Objectives | Solidify understanding of hierarchical inheritance and its role in efficient information storage. |
| Deliverables to submit | <ol style="list-style-type: none"> 1 The graphically restructured network (p. 72) 2 The restructured set of records (p. 73) |
| Tasks | <ol style="list-style-type: none"> 1 Restructure a semantic network given in a graphic representation. 2 Restructure a set of records for more efficient storage. <p>Note: In essence, these two tasks are the same. The same information is represented in two different ways that are completely equivalent; one representation can be transformed into the other without difficulty</p> |
| Further Explanation | <p>In the semantic network detect commonalities among the bottom nodes and introduce new nodes that capture the commonalities. (These nodes would be above the bottom nodes.)</p> <p>In the database detect commonalities among the records and create new, generic records that capture the commonalities. Indicate the hierarchical inheritance relationships among the records in the restructured database.</p> |
| Time | 1.5 hours |

Original semantic network



Restructured semantic network



Note 1: The original semantic network is drawn from the original database by drawing from each event a line to its *Type*, *Place*, and *Time*. So the database shown on the next page and the semantic network represent exactly the same data.

Note 2: In the outline for the restructured semantic network the events have been reordered to make drawing the restructured network easier.

Original database**Event no. 1**

Type: Plane crash
Place: France
Time: 1993
Number killed: 10

Event no. 2

Type: Earthquake
Place: India
Time: 1993
Number killed: 20

Event no. 3

Type: Plane crash
Place: France
Time: 1993
Number killed: 100

Event no. 4

Type: Plane crash
Place: France
Time: 1993
Number killed: 30

Event no. 5

Type: Earthquake
Place: India
Time: 1995
Number killed: 400

Event no. 6

Type: Plane crash
Place: France
Time: 1995
Number killed: 20

Event no. 7

Type: Earthquake
Place: California
Time: 1996
Number killed: 5

Restructured database

Assignment 5

Lecture 3.1

Assigned: May 28**Due: June 4****Analytical description of an information system**

| | |
|-------------------------------|--|
| Objectives | <ol style="list-style-type: none"> 1 You should enhance your understanding of the framework for the analysis of an information system, esp. the ISAR system component, as developed in the text, Chapter 5 by applying it to an actual library or other information system. 2 You should develop analytical skills in general. |
| Deliverables to submit | Your paper can be in the form of a large diagram with explanations typed in the boxes. Or it can be in text form (2 - 3 pages), with a heading and brief text (telegraphic style is ok) for each function and file. Functions and files can be discussed separately or in one integrated sequence. |
| Tasks | <p>Analyze and describe the organizational and technical set-up of the library or information center or library where you work or some other system to which you have access. If you work in a branch library, your analysis needs to cover central functions where these are essential to the operations of your branch. Your analysis must include the following information (you may use your own arrangement):</p> <ol style="list-style-type: none"> 1 Identify the organizational units performing the functions (processes) shown in the combined information system diagram (see Materials below) (If there is no unit for a given function, please say so.) Briefly indicate how each function is performed. 2 Identify the files shown in the diagram. Briefly indicate for each file purpose and form (computer file, card file, book, etc.). 3 Name the document(s) that describe the conceptual data schema and other rules used in the system; characterize the rules briefly. <p>The closer you follow the combined system diagram (Notes for Lecture 3.1) the better.</p> |
| Materials | <p>Text Chapter 5, especially Sections 5.1 and 5.3. The combined information systems diagram given in the Lecture Notes for Lecture 3.1 or in the text Figure 5.1c, <i>Information systems. Complete structure</i>, and Figure 5.6, <i>The structure of an ISAR system</i></p> <p>Three sample analyses from students in previous semesters. [Supplement]</p> |
| Time | 3 hours |

Put the total information system diagram (Lecture notes around p. 62) here and say it is repeated from Lecture 3.1

Assignment 6

Lecture 4.2

*Assigned: June 4**Due: June 11***Developing a conceptual data schema**

| | |
|-------------------------------|--|
| Objectives | <p>Gain a deeper understanding of the concepts discussed in Text Chapter 3 and Section 9.1 by applying them to a concrete example.</p> <p>Be able to construct a conceptual data schema for a given information system</p> <p>In particular, you should be able to</p> <ol style="list-style-type: none"> 1 identify entity types that reflect only intrinsic properties and not relationships to other entities (Person not Author, Text not Title); 2 identify relationship types; 3 find easy-to-understand terms for entity types and relationship types. |
| Deliverables to submit | <p>A list of sample questions</p> <p>An entity-relationship conceptual data schema XXX example</p> |
| Tasks | <p>Picture the information system for one of the following types of business:</p> <ul style="list-style-type: none"> ! a hospital ! a movie production company ! an airline <p>Design a conceptual data schema for the information system you selected. Start from 5 - 7 sample questions that the information system must be able to answer. (See below for examples of good and bad questions.)</p> <p>List the entity types and relationship types required.</p> <p>For each entity type: Indicate how the permissible values are specified.</p> <p>For each relationship type (statement template): Give one example of an actual statement that would be included in the database</p> <p>Focus on some of the central functions of the business that are supported by data in a database. For example "In which bed is Fred Smith?" is a central question in a hospital information system; the hospital cannot function if it is not known where the patients are. "Who is the nurse employee of the month?" is at the fringe; the hospital can function without this information.</p> <p>Your schema need not be complete; order of magnitude: 10 entity types, 7 - 10 relationship types.</p> |
| Time | 1.5 hours |

Definition of *conceptual data schema* on next page

Definition of *conceptual data schema* (as a reminder)

The conceptual data schema of an information system is a specification of the types of information to be stored. In this course, this specification is given by listing the entity types that are covered in the information system and the relationship types that can be used to combine entities into statements. Put differently, the conceptual data schema defines statement templates:

Relationship type,

Statement template,

connects entity types, part of
the conceptual data schema

entity type

Book, intellectual work <*authoredBy*>

entity type

Person

Each piece of information in the system can be seen as a statement that follows one of these statement templates, replacing each entity type, such as Person, by an entity value belonging to that type, such as Eye, Glen G. For example:

Statement,

connects entity values, not part
of the conceptual data schema

entity value

ISBN-0060419512

<*authoredBy*>

entity value

Eye, Glen G.

Put differently, each entity type in a statement template is a place holder for an entity value. By filling in the empty spaces with entity values we get a statement. For a given statement template, there may be thousands or even millions of statements that follow the template. In this assignment, you need to give for each relationship type / statements template on actual statement; this is to make sure that the statement template makes sense.

Examples of poor and good questions

Since this assignment is about developing the conceptual data schema for a database, good questions are questions that would be answered from a database rather than questions that would be answered from a policies and procedures manual, as illustrated below,

| Policies and Procedures Manual question | Database question |
|--|---|
| Database for the operation of an airline | |
| What are the maximum dimensions for carry-on? | On which flight is the bag with tag UA32651? |
| How many pieces of luggage can I bring on a flight? | |
| Does the airline serve food? Note: For a single airline, this is a policy question. For a database of all the world's airlines, this would be a database question | What meals are served on Flight 763? (if the airline serves food) |
| How can I upgrade to first class? | Are upgrades available on Flight 763 on February 22, 2010? |
| How do I book a flight online? | Is there an aisle seat available on Flight 212 on March 15, 2010 |
| | |
| Database for the operation of a university | |
| What is the procedure for making up an Incomplete? | Give me a list of all students who have an Incomplete |
| How do I change my advisor? | Who is the advisor of Joseph Smith? |
| | I need a list of advisees of Prof. Jeng |
| What is the first day of registration? | How many spaces are open in LIS 581 (as of today) |
| | What are the prerequisites for LIS 569? |
| | Am I eligible to take LIS 569? (as of today) (inferred from prerequisite information and the courses the student asking the question has completed) |

Term paper assigned: May 21

Description (this form) due: June 11

Term paper due: Aug. 15 (After Aug 13 by email)

Name:

Term paper description

Title:

One-sentence statement of purpose (absolutely required):

Description (½ page):

See back for further instructions for the term paper
It is recommended that you discuss your topic with the instructor

Term paper — end-of-term essay.

A reflective essay on a theme of your choice, guideline 4,000 - 5,000 words. Make connections between ideas gleaned from the readings, assignments, and class discussions. Reflect on what you have learned and what insights you gained. Consider implications and applications, particularly to your present or expected future work or other courses. You can identify issues, questions, problems for further study. This can be a personal essay that reflects your personal experience and views.

As one option, you may analyze an information system – existing or to be built – with focus on the information retrieval (IR) system, applying all the concepts from the course. This option may help you to better understand course concept by seeing how they work together in one system. For this option:

- ! Focus on the IR subsystem; do not do the entire information system (as you did in Ass 4).
- ! Begin with a section on user needs.
- ! Include an entity-relationship (E-R) conceptual data schema.
- ! Analyze the index language(s) used.
- ! Analyze the indexing process and the parameters of indexing.
- ! Discuss how searches can be done (through queries or through navigation).
- ! Be critical, considering how well the system, and its different components, serve user needs; for example, does the conceptual data schema provide for all the types of information needed to answer users' questions. You can intersperse critique throughout or have a section at the end.

Throughout the course, as ideas and connections occur to you, jot them down.

This essay can be written based on the required readings and lectures without the need for consulting further sources. This is a task of integration, not of expansion. In the past some students have chosen topics that required further reading due to their own interests; not prohibited, but not encouraged.

Format The paper must have a title

Begin with a one-sentence statement of the general purpose of your essay. The remainder of the first paragraph should elaborate on the purpose, possibly stating several specific purposes. Enlighten the reader as to why s/he should read your paper.

Divide the paper into **numbered sections with informative headings.**

See “Soergel’s supplemental style rules” in the readings for Lecture 6.2

To make application of course concepts explicit, include at least five specific references to readings or lectures. Only if your essay does draw on other sources are further citations required in line with general practice; citation format is up to you.

Length. Expected 4,000 - 5,000 words, single spaced, 1" margins, Times Roman 12 points (8 - 10 pages), no upper limit. Ultimately it is the content that matters.

Term paper grading criteria. The following criteria contribute to a holistic assessment

- ! Level of understanding of course concepts shown in the paper. Insights discussed
- ! How well are ideas integrated and related
- ! How well are course concepts applied in discussing the term paper topic or in analyzing or designing an information system.
- ! Connections drawn to other courses
- ! Implications for present and/or future work
- ! Innovative and interesting ideas

Guidelines for term papers on existing IR systems

If you do for your term paper a **description, analysis, and critique of an existing IR (Information Retrieval) system** such as database available through EBSCO, Dialog, or on a website or a library catalog **or the design of a new system, keep the following in mind:**

Note: I will not repeat these comments for each individual term paper proposal that deals with the description and analysis of a specific IR system

- (1) **Focus on the information retrieval subsystem; do not do the entire information system (as you did in Assignment 4)**
- (2) **Begin with a section on user needs**
- (3) **Include an entity-relationship (IR) conceptual data schema.** This may take some effort to construct. Often you can work from a record format that is given, but sometimes you need to look at sample records or at the search criteria that can be used.
- (4) **Analyze the index language(s) used and how suitable it is for the needs of the intended audience.** The index language could be natural language (just English or any language); then you need to discuss how useful free-text search is for the intended audience
- (5) **Discuss how the conceptual data schema and the index language are used to enter data** into the database (indexing very broadly defined). Who does the indexing? (Could be the users as in social tagging). Using what approach / process. (Text chapters 13 and 16 are particularly important here)
- (6) **Discuss how searches can be done** (through queries, through navigation, through a combination). Doing some sample searches and see how they work is a good idea; include sample searches and results in the paper)
- (7) **Be analytical and critical**, considering how well the system, and its different components, serve user needs; for example, does the conceptual data schema provide for all the types of information needed to answer users' questions. You can intersperse critique throughout or have a section at the end.
- (8) **You must apply course concepts.** It will make your analysis better and it will demonstrate that you understand

Assignment 7*Assigned: June 11*

Lecture 5.2b

*Due: June 18***Applying linguistic techniques to retrieval problems**

| | |
|----------------------------------|---|
| Objectives | Understand, through exploration, the possible improvements in free-text retrieval that can be achieved through linguistic techniques such as <ul style="list-style-type: none"> ! word sense disambiguation by syntactic analysis to determine part of speech and by semantic interpretation (from the multiple meanings of a homonym or polyseme, pick out the one that applies in the context); ! using all terms that designate a query concept (all synonyms of the query term); and ! resolution of anaphoric references (what do <i>it, she, they, the machine,</i> ... refer to). |
| Deliverables to submit | The filled-out table. A list of linguistic retrieval techniques indicating the passages affected. |
| Tasks | Explore possible improvements in free-text retrieval through linguistic techniques , using the example on the next page: some short passages of text and a query to be applied to this “collection”. <ol style="list-style-type: none"> (1) analyze retrieval performance of a query using the WS operator and (2) (main task) to suggest linguistic techniques that could be added to the retrieval system to improve retrieval. See the next page for more detailed instructions. You should still adhere to the requirement that the two concepts must be mentioned in the same sentence. |
| Explanation of the query. | The proximity operator WS requires that the two words occur within the same sentence . Thus the query formulation <i>forest WS fire</i> retrieves all passages in which the two words occur in the same sentence. This is the operator used in the baseline query formulation in the assignment. Most IR systems will take this query quite literally and look for the <u>words</u> (and that is how you need to analyze retrieval performance in Task 1. But the user is interested in the close mention of two <u>concepts</u> . That is where linguistic techniques come in. |
| Time | 2 hours |

over

- (1) **Prelude:** Check all passages that are relevant to the user's need as expressed in the query. Then check all passages that are retrieved by the query formulation. Fill in the 3x3 grid and compute performance measures: recall, discrimination, precision.
- (2) **Main point:** What **linguistic techniques** (see objectives) could be used to improve free-text retrieval performance? (Adding index terms to the passages is not an option.) You can analyze each passage in turn, giving for each the applicable linguistic technique(s). Which of these techniques improve recall, discrimination, or both?

Query statement (description of information need / topic): **Forest fires**

Query formulation: forest WS fire* (fire* finds fire or fires)

| Passage | R e l e v a n t | R e t r i e v e d |
|---|--------------------------------------|---|
| ! Forest fires in Indonesia cause serious air pollution in South East Asia. | | |
| ! The fire in Yellowstone Park destroyed 25% of the forest. | | |
| ! The fire station is located behind the city forest. | | |
| ! With fire in her eyes she chased him through the forest. | | |
| ! The soldiers opened fire into the forest. | | |
| ! The fire went out of control. It reached the forest and destroyed many acres. | | |
| ! The animal got scared by the fire burning in the field. It ran into the forest. | | |
| ! He asked whether he should fire the forest workers. | | |
| ! Many square miles of forest in the West are burning. | | |
| ! The dry wooded area went up in flames. | | |

| | Relevant | Not relevant | All |
|---------------|----------|--------------|-----|
| Retrieved | | | |
| Not retrieved | | | |
| All | | | |

Recall: _____

Discrimination: _____

Precision: _____

relevant correctly retrieved
all relevant

irrelevant correctly rejected
all irrelevant

relevant correctly retrieved
all retrieved

Assignment 8**Assigned: June 18**

Lecture 7.1A

Due: June 25+**Descriptive Cataloging Practice**

| | |
|---|--|
| Objectives | <ol style="list-style-type: none"> 1 To gain an understanding of what it takes to describe a work and to determine the entries according to AACR2, consulting the rules in all but the simplest cases, using both the MARC format and the Dublin Core to structure the bibliographic records. 2 Gain minimal experience with another cataloging code (e.g., APA style). 3 Do not spend too much time on getting every detail right; concentrate on the general idea. |
| Deliverables to submit | <p>Three MARC records (typed or printed), one Dublin Core record.</p> <p>You can work on this assignment in teams but you must submit the deliverable individually.</p> |
| Tasks | See detailed description on the next page |
| Materials | <p>Excerpts from AACR2, Part 1 in the readings</p> <p>Note: We did not talk about this in class yet; the readings should do it</p> |
| Time | 4 hours |
| Note on limitations | <p>The objective of this assignment is just to give you a first taste of descriptive cataloging. You need not be concerned with the intricacies of the MARC format. Emulating the Model Catalog records (especially for the Web page) is ok, even encouraged. Do not be obsessed with details; as long as you get the basic record correct and are aware that there are many detailed rules, you are ok even if you do not get all the details exactly right.</p> <p>More information about the MARC format can be found at lcweb.loc.gov/marc/marcdocz.html</p> <p>and, for cataloging Web documents, at oclc.org/support/documentation/worldcat/cataloging/electronicresources/default.htm</p> |
| Descriptive cataloging is not hard | Do not over-think this; it is very simple, just follow the instructions and look at the example. Many students come to this with preconceived but erroneous notion that cataloging must be hard. There are documents that are hard to catalog, but the assigned documents are not. So just get rid of the notion that all cataloging is hard, and Assignment 8 becomes a breeze. |

over

| | |
|---------------------------------------|--|
| <p>Task 1 MARC records</p> | <p>Prepare MARC records for documents 1 - 3, using AACR2 rules (deliverable).</p> <ol style="list-style-type: none"> 1 Lewis, Differentiating the teaching staff 2 Silberman, The open classroom reader 3 MathMagic (Web site) <p>Title page and verso (back of title page) for each item are attached. The number of pages is indicated on the verso if necessary; the physical dimensions of the book can be seen from the title page copy. The books themselves can be examined in Baldy 14A.</p> <p>Use the second level of description (as defined in AACR2) unless indicated otherwise. Use only MARC fields 100 - 599, 700-759 (added author/title entries), and 800-840 (series added entries), as they apply to the document being cataloged. Optional: use subfield indicators \$a, \$b, or #a, #b etc. to divide the information in a MARC field</p> <p>Two simplified records as required in this assignment are shown below. For a listing of the MARC fields and many sample records see the model catalog; Record 26 illustrates the cataloging of a website.</p> <p>For one of these documents prepare a Dublin Core record (deliverable).</p> <p>Use the <i>Dublin Core Metadata Template</i> at worthingtonmemory.org/DC_Form.cfm to make the Dublin Core record. Simply copy and paste the information from your MARC record into the appropriate slots.</p> |
| <p>Task 2</p> | <p>Examine documents A1- A4 (title page and verso attached) for descriptive cataloging problems (no deliverable)</p> <ol style="list-style-type: none"> A1 Women's organizations and leaders directory A2 Social science data file directory A3 Inventory of information resources in the social sciences A4 Anglo-American cataloging rules. 2. ed. 1988 rev |
| <p>[APA style record]</p> | <p>You prepare APA style records for works cited all the time for other courses, starting with LIS 505. This is a form of descriptive cataloging, just using a different set of rules.</p> |

Sample MARC records. The records you create should look like this

100 Conant, James Bryant, #d 1893-1978
 245 The comprehensive high school; #b a second report to interested citizens #c by James B. Conant.
 250 [1st ed.]
 260 New York, #b McGraw-Hill #c [1967]
 300 vi, 95 p. #c 21 cm.

100 Richards, James McDowell, #d 1931-
 245 Assessing student performance in college / #c James M. Richards, Jr.
 260 Washington: #b ERIC Clearinghouse on Higher Education, George Washington University #c [1970]
 300 12 p. #c 28 cm.
 410 ERIC Clearinghouse on Higher Education #t Report; #f 1970, v no. 2
 505 Bibliography: p. 10-12.
 710 ERIC Clearinghouse on Higher Education

Materials for assignment 8 inserted here

must be even number of pages, each title page on a front.

Assignment 9
Lectures 6.2a - 7.2
Problems of entry

Assigned: June 18
Due: June 25+

| | |
|---|---|
| Objectives | <ol style="list-style-type: none"> 1 Improve understanding of the problems in determining entries. 2 Gain enough familiarity with AACR2 rules for entry to find the rule that applies to a given situation. |
| Deliverables to submit | The pages <i>Sample documents . . .</i> (or a copy) with your answers written in. |
| Tasks (If you have trouble, see the TA) | <p>For the starred items in the attached <i>Sample documents for analyzing author entry according to Lubetzky</i> (copied from Lecture 7.2) indicate</p> <ul style="list-style-type: none"> • The Lubetzky condition that applies to this case (see reading and the overview in the notes for Lecture 7.2A) • The AACR2 rule that applies (see reading) • The correct author entry formatted according to AACR2 (no other entries are required; for example, do not do title or subject entries) |
| Materials / reading | This assignment is based on Part 2 of AACR2, Chapter 21 - 26 (not in your package, multiple copies of AACR2 are available in Baldy 14A). Skim the pages listed in the AACR2 reading; get the gist of the rules, no need to read all examples. Use the entire book for reference when doing the assignment. Alternatively you can use RDA, also in Baldy 14A |
| Time | 1.5 hours |

| | |
|-------------------------------------|--|
| Note on the concept of entry | <p>(As defined in AACR2 reading) An author entry is simply the name of a person or corporate body under which the document should be found in a catalog. For actual cataloging in the US, and for this assignment, the name should be given in the form prescribed by AACR2. In a MARC record, these names go into fields 100 or 110 (if the main entry) or in fields 700 or 710 if an added entry (any entry other than the main entry).</p> <p>So for the Assignment 8 deliverable you need to give for each of the documents marked with * the personal or corporate name(s) which should be used as entries (you need not be concerned with specifying which is the main entry). You do need to give for each entry the Lubetzky condition that applies and the applicable AACR2 rule(s) that you used</p> <p>(A) in the decision to make an entry for this person or organization and (B) in determining the form of the name to be used in the catalog record.</p> |
|-------------------------------------|--|

Sample documents for

- **analyzing author entry according to Lubetzky and**
- **determining author entry according to AACR2**

(selected from the longer list used in Lecture 7.2A)

Use a separate line for each entry

- (b)* ***Studies in the social psychology of adolescence*, by J. E. Richardson, J. F. Forrester, J. K. Shukla, and P. J. Higginbotham; edited with a foreword by C. M. Fleming.**

Lubetzky condition:

AACR2 rule:

Main Entry:

Added Entry (if applicable):

- (e)* ***Chisholm's handbook of commercial geography*, entirely rewritten by L. Dudley Stamp and S. Carter Gilmour.**

Lubetzky condition:

AACR2 rule:

Main Entry:

Added Entry (if applicable):

- (f)* ***Making magical apparatus*, by Jane Reid (i.e. Mrs. David Johnstone).**

Lubetzky condition:

AACR2 rule:

Main Entry:

Added Entry (if applicable):

(m)* ***Schubert: thematic catalogue of all his works in chronological order, by Otto Erich Deutsch in collaboration with Donald R. Wakeling.***

Lubetzky condition:

AACR2 rule:

Main Entry:

Added Entry (if applicable):

(n)* ***A concordance to the poems of William Wordsworth, by Lane Cooper.***

Lubetzky condition:

AACR2 rule:

Main Entry:

Added Entry (if applicable):

(o)* ***The poetical works of Wordsworth, edited by E. de Selincourt.***

Lubetzky condition:

AACR2 rule:

Main Entry:

Added Entry (if applicable):

Assignment 10
Assigned in Week 7 for Lecture 8.1

Assigned: June 18
Due: June 25

Free indexing of three documents and preparation for Small Groups 1

| | |
|-------------------------------|--|
| Objectives | <ol style="list-style-type: none"> 1 To get a sense of how free indexing works in preparation for Small Groups 1. 2 To prepare for seeing the contrast to request-oriented indexing as done in Assignment 11. |
| Deliverables to submit | The sheet Document abstracts for free indexing filled in. |
| Tasks | <ol style="list-style-type: none"> 1 Free indexing. Index the three documents given on the page Document abstracts, using whatever terms you can think of as good access points for finding this document. 2 Look over the tasks for Small Groups 1 and the accompanying materials in the notes for Lecture 8.1 (Small Groups 1) <p>You need to complete both tasks before the Small Groups 1</p> |
| Time | 2 hours |
| Materials | A page with three document abstracts (the back of this page) |
| Grading | This assignment will be graded as complete/incomplete |

Document abstracts for Assignment 10

Document 1

N69-17257 *# National Aeronautics and Space Administration. Langley Research Center. Langley Station, VA

Supersonic transport operating practices during simulated operations in future air traffic control systems environments

Milton D. McLaughlin and Richard H. Sawyer Washington Feb. 1969 65 p refs
(NASA-TN-D-5018) Avail: CFSTI CSCL 01B

Operating practices of the supersonic transport (SST) during simulated operations in air traffic control (ATC) system environments conceived for the time period of introduction of the SST into service are presented. An SST flight simulator and the Federal Aviation Administration ATC simulation facilities were used to create the real-time simulations. The SST flight simulator was operated by airline crews and the ATC simulation facilities by experienced air traffic controllers. The test program included departure and arrival operations under instrument flight rule conditions in the New York and Los Angeles terminal areas with two design study configurations of the SST. The design study configurations were representative of variable-sweep and fixed-wing design. Both designs had a variable-incidence forebody.

Free index terms

Document 2

A Plan for a New Consolidated Passenger Ship Terminal in the Port of New York.

Distributed free by Port of New York Authority, 111 Eighth Avenue, New York, NY 10011. 1967. pp. iv+102.

Passenger ship activity in New York, existing passenger ship piers, design criteria, alternate solutions, proposed custom examination system, proposed terminal layout.

Free index terms

Document 3 **A system for bus rapid transit on urban freeways.**

Traffic Quarterly Oct. 1969

Describes the design of a system of buses sharing the general freeway lanes with regular traffic but having separate entrance and exit ramps. A sophisticated traffic control system would give preference to buses to guarantee predictable travel times. Compares cost and effectiveness (as measured by passengers carried in peak travel times) with other rapid transit systems.

Free index terms

Assignment 11
Lecture 8.2B, Text Chapter 13

Assigned: June 25
Due: July 2

Request-oriented indexing

| | |
|-------------------------------|--|
| Objectives | <ol style="list-style-type: none"> 1 Gain experience with the checklist technique of indexing 2 Practice building precombined descriptors with a faceted classification |
| Deliverables to submit | Three filled-in document indexing forms |
| Tasks | <ol style="list-style-type: none"> 1 Checklist indexing. Index three documents with the attached index language (a faceted classification of transportation), consciously applying the checklist technique of indexing. Optional 2 Building precombined descriptor for arrangement. After you have assigned the applicable elemental descriptors, make a precombined descriptor for the document. |
| Materials | <p>Document indexing forms for three documents, the same documents you indexed with free terms in Assignment 10</p> <p>An example of a filled-in document indexing form is included in the materials for this assignment</p> |
| Time | 2 hours |

The pages in this assignment are organized as follows:

First the **index language** to be used, a faceted classification of transportation
Have a look at it so that the explanation of the tasks makes more sense.

Then **detailed explanations of Task 1 and Task 2** with examples.

And last the Document Indexing Forms

Index language / classification for request-oriented indexing

This is a faceted classification of transportation to be used for Assignment 11.

For indexing use the **Indexer's classification display** included here.

For the in-class searching exercise building on this assignment you will use the **File builder's and searcher's classification display** included in the Lecture Notes.

Outline: Facets

- B** **Mode of transportation**
- E** **Transportation system components**
- F Power supply for vehicles
- G Type of propulsion
- H Materials to build facilities or vehicles
- J** **Passenger transport vs. freight transport**
- K Traffic operations
- L Transportation providers
- M Creation of traffic systems and components
- N Organization, administration
- Q General and other concepts
- R Geographic range
- S Geographic location

The three facets used for arrangement (optional Task 2) are shown in **bold**.

| | |
|---|---|
| <p>A Transportation and traffic</p> <p>B Mode of transportation</p> <p>B1 . Ground transport</p> <p>B1.1 . . Road transport</p> <p>B1.2 . . Rail transport</p> <p>B1.2.1 . . . Local rail transit BT R1</p> <p>B1.2.2 . . . Intercity railroads BT R2</p> <p>B1.3 . . Pipeline transport</p> <p>B1.4 . . Pedestrian mode</p> <p>B1.5 . . Multi-modal ground transport</p> <p>B2 . Water transport</p> <p>B2.1 . . Inland water transport</p> <p>B2.2 . . Ocean Transport</p> <p>B3 . Air transport</p> <p>B3.1 . . Supersonic air transport</p> <p>B4 . Air cushion transport</p> <p>B5 . Multi-modal transport</p> <p>B8 . Other specific modes of transportation</p> <p>B9 . Mode of transportation not applicable</p> <p>D, E Free for expansion</p> <p>E Transportation system components</p> <p>E1 . Traffic facilities</p> <p>E1.1 . . Traffic routes</p> <p>E1.2 . . Traffic stations</p> <p>E1.3 . . Stationary equipment</p> <p>E2 . Methods to move persons or freight</p> <p>E2.1 . . Vehicles</p> <p>E3 . . Containers</p> <p>E4 . . Self-transport</p> <p>E8 . Other concepts</p> <p>E9 . Transp. system elements not applicable</p> | <p>F Power supply for vehicles</p> <p>F1 . Hydrocarbons</p> <p>F1.1 . . Gasoline</p> <p>F1.2 . . Diesel fuel</p> <p>F1.3 . . Hydrocarbons from renewable sources</p> <p>F5 . Electric power</p> <p>F6 . Nuclear power</p> <p>F7 . Animate power</p> <p>F8 . Other power supply</p> <p>F9 . Power supply not applicable</p> <p>G Type of propulsion</p> <p>G1 . Engine</p> <p>G1.1 . . Combustion engine</p> <p>G1.2 . . Steam engine</p> <p>G2 . Turbines</p> <p>G3 . Walking</p> <p>G8 . Other type of propulsion</p> <p>G9 . Type of propulsion not applicable</p> <p>H Materials to build facilities or vehicles</p> <p>H1 . Materials by composition</p> <p>H1.1 . . Soils, aggregates</p> <p>H1.2 . . Bitumen</p> <p>H1.3 . . Cement, Concrete</p> <p>H1.4 . . Ceramics, glasses</p> <p>H1.5 . . Wood, paper</p> <p>H1.6 . . Fibers, textiles</p> <p>H1.7 . . Plastics</p> <p>H1.8 . . Rubbers</p> <p>H19 . . Metal</p> <p>H2 . Materials by origin</p> <p>H2.1 . . Petroleum products</p> <p>H3 . Materials by use</p> <p>H3.1 . . Marking or coating materials</p> <p>H3.2 . . Adhesives, seals</p> <p>H8 . Other specific materials</p> <p>H9 . Material not applicable</p> |
|---|---|

J Passenger vs. freight transport

- J1 . Passenger transport
- J2 . Freight transport
- J2.1 . . Transport of material of heavy weight
- J2.2 . . Transport of bulk material
- J9 . Passenger vs. freight transp. not applicable

K Traffic operations**K1 . Traffic communication, control, safety**

- K1.1 . . Traffic communications
- K1.2 . . Traffic control
- K1.3 . . Traffic safety

K2 . Routes and schedules

- K2.1 . . Routes, route systems, traffic networks
- K2.2 . . Schedules

K3 . Handling, loading, unloading

- K8 . Other specific traffic operations
- K9 . Traffic operations not applicable

L Transportation providers

- L1 . Organizations, companies
- L2 . Personnel, operators
- L9 . Transportation providers not applicable

M Creation of traffic systems and components**M1 . Research, design, and evaluation**

- M1.1 . . Research and development
- M1.2 . . Planning
- M1.3 . . Design
- M1.4 . . Testing, demonstration, evaluation

M2 . Manufacturing, construction

- M3 . Acquisition
- M4 . Training
- M5 . Maintenance

M8 . Other specific activities in system creation

- M9 . System creation not applicable

N Organization, administration

- N1 . Administration, management
- N2 . Costs, financing
- N3 . Marketing
- N4 . Legal aspects
- N8 . Other specific topics in organization
- N9 . Organization, administration not applicable

Q General and other concepts**Q1 . Traffic flow**

- Q2 . Simulation

Q3 . System characteristic

- Q3.1 . . Noise, vibration
- Q3.2 . . Pollution
- Q3.3 . . Quality, performance
- Q3.4 . . Durability, life, reliability
- Q3.5 . . Demand, use
- Q3.6 . . Human characteristics
- Q3.7 . . Community characteristics
- Q3.9 . . Other system characteristics

Q4 . Small vs large capacity

- Q4.1 . . Small capacity
- Q4.2 . . Large capacity

Q5 . Civilian vs military

- Q5.1 . . Civilian
- Q5.2 . . Military

Q8 . Other general concepts

- Q9 . None of these concepts applicable

R Geographic range**R1 . Local systems**

- R1.1 . . Urban systems
- R1.2 . . Rural systems

R2 . Beyond local systems

- R2.1 . . Interurban systems
- R2.2 . . State-wide systems
- R2.3 . . National systems
- R2.4 . . International systems

R8 . Other specific range

- R9 . Geographic range not applicable

S Geographic location**S1 . North and Central America**

- S1.1 . . Canada
- S1.2 . . U.S.
- S1.3 . . Central America

S2 . South America**S3 . Europe****S4 . Asia****S5 . Australia****S6 . Africa****S8 . Other geographic locations**

- S9 . Geographic location not applicable

Detailed explanations and instructions for Tasks 1 and 2

Have a look at the sample document indexing form given below and examine the index language given above so you have concrete examples in mind as you read the explanations.

Task/step 1. Index the documents with elemental descriptors

The elemental descriptors are found in the faceted classification *schedule*. Use the indexing forms provided. In the space for each facet, enter the applicable descriptor(s) from that facet. When a facet does not apply to a document, you must circle *Not applicable*; you cannot skip over a facet without making an explicit decision.

Use the most specific descriptors that cover the subject. Use B2.1 *Inland water transport* if the document addresses just *inland water transport* (transport on rivers and lakes); use B2 *Water transport* only if the document addresses all kinds of *water transport* (inland and ocean). Hierarchically expanded searching (searching with hierarchic expansion as practiced in the Medline exercise) will make the specific documents accessible through the broader term, so there is no need to index with the broader term. There may be more than one descriptor from any given facet.

Optional. Task/step 2. Building precombined descriptors

Elemental descriptors support computer searching using Boolean query formulation, but they do not support meaningful arrangement of documents as in a display of a large number of search results, in a Web subject directory, or on library shelves. For meaningful arrangement we need to build precombined descriptors.

Thus when a faceted classification is to be used for arrangement, we need an additional step, building precombined descriptors. The notation (descriptor number) for a precombined descriptor can be formed by gluing the notations for the elemental components together. In the example, only three facets are used: B, E, and J, and they are combined in that order. In real life, selection and combination order of facets are determined in accordance with user requirements. See the supplement for more explanation and an arrangement example.

Even without consulting the supplement, you can arrange the precombined descriptors for Documents 1 - 3 to form a mini subject directory.

Indexing. Document 0 (sample document)

Amsterdam's new container dock, K.W.Flitcroft for the Amsterdam Harbor Committee. Dock & Harbor Authority v 49 n 571 May 1968 p 28-30.

Dock described is protected by locks from rise and fall of tides; spreader is employed in lifting of containers and is adaptable in spread to handle both long and short types; containers can be stored on quay and special connections for powering of plants of refrigerated containers are set in concrete paving every 10 ft.; set of rail tracks runs along quay between high legs of container cranes to bring rail-hauled containers directly for lifting off.

| Write in appropriate concept(s) | | merely circle |
|---------------------------------|--|-----------------------|
| B Mode of transportation | B2.2 Ocean transport | not applicable |
| E Transp. system elements | E1.2 Traffic stations E3 Containers | not applicable |
| F Power | | <u>not applicable</u> |
| G Propulsion | | <u>not applicable</u> |
| H Materials to build | | <u>not applicable</u> |
| J Passenger vs. freight | J2 Freight transport | not applicable |
| K Operations | K3 Handling, loading, unloading | not applicable |
| L Provider | | <u>not applicable</u> |
| M Creation | | <u>not applicable</u> |
| N Administration | | <u>not applicable</u> |
| Q General and others | | <u>not applicable</u> |
| R Geographical range | R2.4 International system | not applicable |
| S Geographical location | S3 Europe | not applicable |

Optional. Task/step 2: Build a precombined descriptor for arranging documents

B2.2E1.2J2 Document 0 (the descriptors from facets B, E, and J)

Document 1. **Supersonic transport operating practices during simulated operations in future air traffic control systems environments**

Operating practices of the supersonic transport (SST) during simulated operations in air traffic control (ATC) system environments conceived for the time period of introduction of the SST into service are presented. An SST flight simulator and the Federal Aviation Administration ATC simulation facilities were used to create the real-time simulations. The SST flight simulator was operated by airline crews and the ATC simulation facilities by experienced air traffic controllers. The test program included departure and arrival operations under instrument flight rule conditions in the New York and Los Angeles terminal areas with two design study configurations of the SST. The design study configurations were representative of variable-sweep and fixed-wing design. Both designs had a variable-incidence forebody.

| Write in appropriate concept(s) | merely circle |
|--|----------------------|
| B Mode of transportation | not applicable |
| E Transp. system elements | not applicable |
| F Power | not applicable |
| G Propulsion | not applicable |
| H Materials to build | not applicable |
| J Passenger vs. freight | not applicable |
| K Operations | not applicable |
| L Provider | not applicable |
| M Creation | not applicable |
| N Administration | not applicable |
| Q General and others | not applicable |
| R Geographical range | not applicable |
| S Geographical location | not applicable |

Optional. Precombined descriptor:

Document 2 **A Plan for a New Consolidated Passenger Ship Terminal in the Port of New York.**

Passenger ship activity in New York, existing passenger ship piers, design criteria, alternate solutions, proposed custom examination system, proposed terminal layout.

| Write in appropriate concept(s) | merely circle |
|--|----------------------|
| B Mode of transportation | not applicable |
| E Transp. system elements | not applicable |
| F Power | not applicable |
| G Propulsion | not applicable |
| H Materials to build | not applicable |
| J Passenger vs. freight | not applicable |
| K Operations | not applicable |
| L Provider | not applicable |
| M Creation | not applicable |
| N Administration | not applicable |
| Q General and others | not applicable |
| R Geographical range | not applicable |
| S Geographical location | not applicable |

Optional. Precombined descriptor:

Document 3 **A system for bus rapid transit on urban freeways.** Traffic Quarterly Oct. 1969

Describes the design of a system of buses sharing the general freeway lanes with regular traffic but having separate entrance and exit ramps. A sophisticated traffic control system would give preference to buses to guarantee predictable travel times. Compares cost and effectiveness (as measured by passengers carried in peak travel times) with other rapid transit

| Write in appropriate concept(s) | merely circle |
|--|----------------------|
| B Mode of transportation | not applicable |
| E Transp. system elements | not applicable |
| F Power | not applicable |
| G Propulsion | not applicable |
| H Materials to build | not applicable |
| J Passenger vs. freight | not applicable |
| K Operations | not applicable |
| L Provider | not applicable |
| M Creation | not applicable |
| N Administration | not applicable |
| Q General and others | not applicable |
| R Geographical range | not applicable |
| S Geographical location | not applicable |

Optional. Precombined descriptor:

Assignments 12.1 - 12.3 **Assigned: July 2/9**
 Small Groups 2 and 3, Text Chapter 14 **Due: July 9/16**
Conceptual analysis and synthesis

12.1 and 12.2 assigned after **Small Groups 2** and due at **Small Groups 3, July 9.**
12.3a and 12.3b assigned after **Small Groups 3** and due on **July 16.**

| | |
|-------------------|---|
| Objectives | <p>Solidify understanding of classificatory structure through practicing the process of conceptual analysis and synthesis as discussed in Chapter 14 and illustrated through the in-class exercise. Specifically:</p> <ol style="list-style-type: none"> 1. Understand semantic factoring through practicing it (the only way). 2. Understand how to build a hierarchy through applying the pragmatic definition of <i>A is broader than B</i> in a limited set of elemental concepts. 3. Understand the interaction between concept combination and hierarchy. |
| Tasks | See individual assignments in this group |

Each part of the assignment is concerned with one step in the process:

| | |
|-----------------|--|
| Assignment 12.1 | Semantic factoring (results in a list of elemental concepts). |
| Assignment 12.2 | Arranging the elemental concepts in a well-structured faceted hierarchy. |
| Assignment 12.3 | <p>Fit compound concepts into the framework of the hierarchy (if compound concepts need to be dealt with explicitly)</p> <p>Assignment 12.3a is an exercise in facet combination unconnected to the set of concepts from Assignment 12.1. The point is to practice building a hierarchy from facet combination</p> <p>Assignment 12.3b returns to the set of concepts from Assignment 12.1, applying the principles learned from Assignment 12.3a</p> |

Note: The list of concepts given for Assignment 12.1 deliberately covers two domains, *Medicine* and *Transportation*. There might be elemental concepts that apply to both!

Over for more explanation of the flow of the sub-assignments

Flow of Assignments 12.1, 12.2, 12.3a, 12.3b

Some students in the past have had difficulty with this, so here is another explanation. Sometimes hearing or reading the same thing twice helps.

There are three steps in the **conceptual analysis and synthesis** in a subject (repeated from Lecture 9.1)

| | |
|---------|---|
| Step 1. | Semantic factoring (results in a list of elemental concepts). |
| Step 2. | Arranging the elemental concepts in a well-structured faceted hierarchy. |
| Step 3. | Fit compound concepts into the framework of the hierarchy (if compound concepts need to be dealt with explicitly) |

We practiced these steps in the small group sessions; in the sub-assignments you practice what you have learned in the small groups. So the overall sequence is as follows:

Small groups 2, Lecture 9.1. Semantic factoring and facet organization

Assignment 12.1 Step 1. Semantic factoring (results in a list of elemental concepts).
You learned how to do this in Small Groups 2, Lecture 9.1, first half

Assignment 12.2 Arranging the **elemental concepts** in a well-structured faceted hierarchy.
You learned how to do this in Small Groups 2, Lecture 9.1, second half

Submit your results quickly so they can be returned with corrections before you use them for the next step

Small groups 3, Lecture 10.1. Building a hierarchy of compound concepts from facets.

Assignment 12.3 Fit compound concepts into the framework of the hierarchy (if compound concepts need to be dealt with explicitly)

This is divided into two pieces.

Assignment 12.3a is an exercise in facet combination to practice what you learned in Small Groups 3 with a different example. In this example the facets are given. The subject domain is language/verbal ability, quite **unconnected** to the set of concepts from Assignment 12.1. Again, the point is to practice building a hierarchy from facet combination

Now you are ready to apply what you learned in Assignment 12.3a about building a hierarchy from facets to the set of concepts from Assignment 12.1. This is done in Assignment 12.3b.

Assignment 12.1**Assigned: July 2****Due: July 9****Semantic factoring***Do before you come to Small Groups 3.*

| | |
|-------------------------------|--|
| Objectives | Inherited from Assignment 12 (page 109) |
| Deliverables to submit | <ol style="list-style-type: none"> 1 A list of the compound concepts with their semantic factors. (You can write the semantic factors on the assignment sheet and hand that in or you can use the template.) 2 A list of the elemental concepts used, each concept on its own line. (This will be the basis for Assignment 12.2.) |
| Tasks | <p>Determine the semantic factors of each of the concepts designated by the following terms, that is, express each concept by a combination of elemental concepts (or what you consider elemental concepts). Some definitions to help you with this task are given on the next page. If you cannot find semantic factors, or if there are different sets of semantic factors because of term ambiguity, write a comment.</p> <p>To determine the semantic factors (or elements of meaning) of a concept, you need to understand the concept. See the definitions on the back of this page.</p> <p>Since you have no list of elemental concepts to choose from, you must make up your own elemental concepts and choose the terms to express them. (This happens often in the construction of index languages.) But be consistent: if the same elemental concept occurs more than once, use the same term each time.</p> <p>Note: Use the most specific elemental concept for each aspect of the concept to be expressed. For example, if the disease is a type of cancer, use the specific elemental concept <i>cancer</i> as the semantic factor, not the more general elemental concept <i>disease</i>.</p> <p>When you are finished with semantic factoring, prepare a list of the elemental concepts you used. For example, your list should have one entry for <i>cancer</i>, which occurs several times as a semantic factor. This list will be the basis for Assignment 12.2.</p> |
| Time | 1.5 hours |

Over for definitions

Definitions you may need

The terms in this assignment (and the concepts designated by these terms) come from the domains of medicine and transportation. You may not be familiar with these terms, but that will happen to you often in practice. You are information specialists, so if you do not know something you should know where to find out. In this case you need definitional knowledge, so you consult a dictionary. To save you time, I compiled the definitions you need here.

| | |
|----------------|---|
| Leukemia | Cancer of white blood cells (also called leukocytes) |
| Mononucleosis | An infectious disease of white blood cells caused by a virus |
| Pneumonia | An inflammation of the lungs |
| Conjunctiva | The mucous membrane covering the anterior surface of the eyeball and lining the eyelids Note: Consider body parts to be elemental concepts |
| Conjunctivitis | An inflammation of the conjunctiva. Some forms of conjunctivitis are infectious, others are not. |
| Wind tunnel | A tubular chamber or structure in which a steady current of air can be maintained at controlled velocity, equipped with devices for measuring [the aerodynamic] forces and moments on scale models of complete aircraft [or cars] or of their parts or of full-scale aircraft [or cars] or their parts. (Random House Dictionary) |

Assignment 12.1. Concept list for semantic factoring

| | |
|-----|---|
| 1. | Leukemia |
| 2. | Leukemia - diagnosis |
| 3. | Leukemia - drug therapy |
| 4. | Leukemia - radiation therapy |
| 5. | Pneumonia |
| 6. | Pneumonia treatment |
| 7. | Lung cancer |
| 8. | Mononucleosis |
| 9. | Conjunctival cancer |
| 10. | Conjunctival cancer - radiation treatment |
| 11. | Conjunctivitis |
| 12. | Conjunctivitis - drug therapy |
| 13. | Highway repair |
| 14. | Diagnosis of car problems |
| 15. | Car repair |
| 16. | Wind tunnel |

Assignment 12.2*Assigned**July 2**Due:**July 9***Building a faceted classification of elemental concepts***Do before you come to Small Groups 3.*

| | |
|-------------------------------|---|
| Objectives | Inherited from Assignment 12 (page 109) |
| Deliverables to submit | A faceted classification of <u>elemental</u> concepts shown as a linear arrangement with indentation (with cross-references as needed) Use the list of elemental concepts you prepared in Assignment 12.1 |
| Tasks | See next page |
| Time | 1.5 hours |

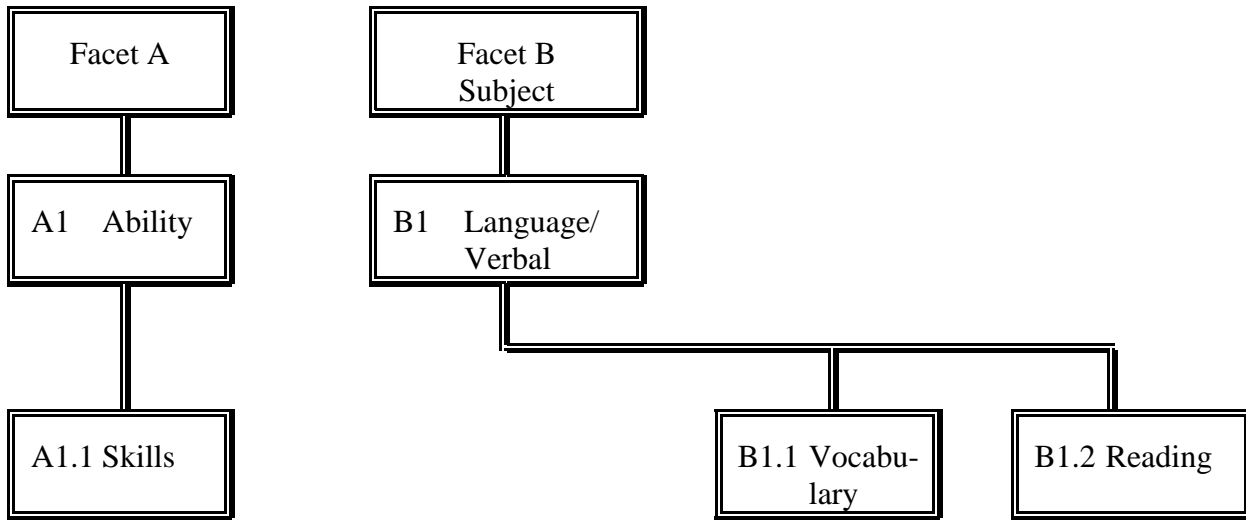
| Task: | Build a faceted classification of elemental concepts |
|--|---|
| Include only elemental concepts | <p>In Assignment 12.1 you started from a list of compound concepts and derived elemental concepts through semantic factoring. You prepared a list of elemental concepts.</p> <p>Here you start from this list (do not include any other concepts) and establish the hierarchical relationships among them. For example, in the hierarchy</p> <p style="padding-left: 40px;">do include <i>Cancer</i> (elemental concept)</p> <p style="padding-left: 40px;">do not include <i>Leukemia (Cancer of Leukocytes)</i> (compound concept)</p> <p>You will have a chance to fit compound concepts into a hierarchy in Assignment 12.3b, where you will build a hierarchy that includes all the concepts used in Assignment 12.1 in a well-structured arrangement..</p> <p>Avoid a frequent mistake: In a hierarchy such as</p> <p style="padding-left: 40px;">Disease</p> <p style="padding-left: 80px;">. Cancer</p> <p style="padding-left: 80px;">. . Diagnosis</p> <p>“Diagnosis” does not refer to the elemental concept <i>Diagnosis</i> but to the combination</p> <p style="padding-left: 40px;"><i>Cancer > Diagnosis</i></p> <p>The elemental concept <i>Diagnosis</i> belongs to different facet.</p> |
| All hierarchical relationships | <p>Be sure to introduce all hierarchical relationships that are useful for searching and/or the checklist technique of indexing.</p> <p>Before you stipulate that concept A has a Narrower Term B, ask: Does a user searching for A want to find all entities dealing with or relevant for B?</p> |
| May need additional broad concepts | <p>You may need to introduce additional broad concepts to make for a more logical, more easily understood hierarchy. In many places additional <u>specific</u> concepts will suggest themselves from the logic of your hierarchy; you need not add those, but you may add a few examples.</p> |
| Overall arrangement | <p>Your hierarchy may include one part for <i>Medicine</i>, one part for <i>Transportation</i>, and one part for concepts needed in both domains (if any) or not fitting in any of the two domains.</p> |
| Represent the hierarchy in outline format | <p>The hierarchical relationship A has Narrower Term B should be shown by arrangement and indention (outline format) where possible and through a cross-reference otherwise. A hierarchy shown in a linear arrangement is often easiest to construct and always easiest to read. Since the hierarchy contains only elemental concepts there will be few if any cross-references needed.</p> |

Assignment 12.3a**Assigned: July 9****Due: July 16****Hierarchy from facet-combination. General practice**

| | |
|-------------------------------|---|
| Objectives | <p>Inherited from Assignment 12 (page 109)</p> <p>Assignment 12.3a is an exercise in facet combination unconnected to the set of concepts from Assignment 12.1. The point is to practice building a hierarchy from facet combination</p> |
| Deliverables to submit | <ol style="list-style-type: none"> 1 Hierarchy in graphical arrangement (can be on the assignment sheet). 2 Hierarchy in linear arrangement with indention, with cross-references. 3 Hierarchy in another linear arrangement with indention, with X-ref. |
| Tasks | <p>Construct the hierarchy of elemental and compound concepts generated by the two facets given on the following sheet. Use only between-facet combinations, that is, only combinations of an A-concept with a B-concept. Do not combine the facet headings (Facet A, Facet B) with anything. Represent this hierarchy graphically. (You can draw on the assignment sheet.)</p> <p>Then represent the hierarchy in outline form (a linear arrangement with indention) with cross-references. Repeat this choosing a different possibility for the linear sequence.</p> <p>Note 1: When you do a linear arrangement with cross-references, many hierarchical relationships are shown by the arrangement. These relationships do not need to be shown again through cross-references. Only relationships not shown through the arrangement require cross-references.</p> <p>Note 2. This assignment uses classification concepts and procedures already discussed but you may still find it difficult, particularly since this is the first case with hierarchy in both generating facets. This assignment shows you (and the instructor) to see just what you do or do not know and forces you to thoroughly think about the problem. The assignment prepares you for Small Groups 3.</p> <p>Note 3: This assignment deals with the problem of constructing a hierarchy from facet combination in the context of a simple example unconnected to the set of concepts from Assignment 12.1. You deal only with the concepts given on the following page and their combinations; do not introduce any other concepts.</p> |
| Time | 2 hours |

Name:

Assignment 12.3a



Assignment 12.3b**Assigned: July 9****Due: July 16**

**Hierarchy from facet-combination.
Application to the set of concepts from Assignment 12.1**

Do after Small Groups 3.

| | |
|-------------------------------|--|
| Objectives | Inherited from Assignment 12 (page 109) Assignment 12.3b returns to the set of concepts from Assignment 12.1, applying the principles learned from Assignment 12.3a |
| Deliverables to submit | A hierarchy in linear arrangement with indentation and cross-references for the concepts presented in Assignments 12.1 and 12.2. |
| Tasks | See next page |
| Time | 2 hours |

Task. Construct hierarchy from facets for the Assignment 12 subject domain

You will now apply the principles you have learned from Assignment 12.3a to the set of concepts from the Assignment 12 domain. Arrange all the concepts from Assignments 12.1 and 12.2 (both the compound concepts which are given and the elemental concepts that you derived) in a hierarchy. Represent the hierarchy in a linear arrangement with cross references; **graphical representation is not required and would be too complex** (there are too many concepts).

Start from the hierarchy of elemental concepts you developed in Assignment 12.2 and fit the compound concepts given in Assignment 12.1 into the structure. A compound concept has more than one broader concept, one for each semantic factor; choose one place for the arrangement and make a cross-reference from the other. **Do not create all possible combinations of elemental concepts (as you did in Assignment 12.3a); just use the compound concepts actually listed in Assignment 12.1**, possibly adding a few new concepts to fill in hierarchical steps or otherwise make the hierarchy more logical .

Hints

You will get a clearer view of your hierarchy if you use a word processor rather than hand-writing. Copy the hierarchical arrangement of elemental concepts you created in 12.2 (or an improved version you created after you got the answer sheet). Then copy and paste each of the terms for the compound concepts you were given for Assignment 12.1 into one of its possible places in the hierarchical arraignment, using the semantic factors as your guide; make a cross-reference to and from the other possible place(s). Remember that you already dealt with the meaning of all these terms when you did semantic factoring in Assignment 12.1; now you just use the semantic factors as a guide for fitting the compound concepts into the hierarchical arrangement with cross-references.

Write out each term so it can stand on its own outside the hierarchy:

Not

Leukemia
 . Diagnosis

But

Leukemia
 . Leukemia - Diagnosis [LCSH/SSH form. Yahoo form: Leukemia > Diagnosis]

Important note: Do not introduce cross-references for hierarchical relationships that can be seen from the arrangement. Cross-references are used to show **additional** hierarchical relationships.

Assignments 13.1 - 13.4 (Lecture 10.2) (There is no Assignment 13 per se)**Analysis of and practice with Knowledge Organization Systems (KOS)****P. 121-122 apply to ALL Assignments 13.1-13.4** (Example of **hierarchical inheritance**)

| | |
|--|---|
| Objectives | <p>Be able to use index languages/classifications to index and to search.</p> <ol style="list-style-type: none"> 1 General: Understand the principles governing the structure of Knowledge Organization Systems (KOS) and be able to apply these principles to an analysis of existing schemes 2 Specific: Gain some familiarity with the index languages covered and grasp their basic structure and be able to use them to index documents and to formulate queries in simple cases. |
| Schemes covered | <p>13.1 Dewey Decimal Classification (DDC): Discussed July 9, due July 16, .</p> <p>13.2 Yahoo: Yahoo classification (a semi-faceted scheme) Dis July 16, due July 23 OR LCC: Library of Congress Classification Discussed July 16, due July 23 OR DDC 2 Not discussed , due July 23</p> <p>13.3 LC/Sears Subject Headings (LCSH): Not discussed, due July 30</p> <p>13.4 ERIC Thesaurus: Not discussed, due July 30,</p> |
| Deliverables to submit | <ul style="list-style-type: none"> • For DDC, Yahoo or LCC, and ERIC: The worksheet answers, in the worksheet or on a separate page, labeled with the acronym for the scheme. • For every scheme you do: Document forms and query forms filled in • Only for LCSH/SSH: The filled-in Outline for the Analysis of KOS (purple) |
| Tasks | For every scheme |
| 1. Guided exploration (worksheet) | <ul style="list-style-type: none"> • Read introductory materials. Attend lecture / follow presentation. Includes familiarizing yourself with the scheme and how it is presented based on the printed volumes (from some library), sample pages for each scheme in the Assignment Packet, and exercise introducing the electronic version • Finish the worksheet as applicable. |
| 2. Indexing and query formulation | <p>See front page of Indexing and Query Forms (p. 305) for general instructions. Specific instructions are given with each scheme</p> |
| 3. Summary analysis | <p>Read and contemplate, the Outline for the analysis of Knowledge Organization Systems (purple), Only for LCSH/SSH: fill in the purple sheet using the filled-in sheets for Yahoo and DDC as examples. Telegraphic style ok. If an answer would require extensive study, give your best guess and mark it "(guess)".</p> |

You can work on these assignments in study groups.
You may want to schedule a lab session with instructor or TA.
You must submit the deliverables individually.

Pages in the Assignment Packet

| | |
|--|--|
| Separate for each scheme | <p>Overview, list of readings and other materials (yellow)</p> <p>Worksheet and introduction to electronic version, if any (white)</p> <p>Outline for the analysis of Knowledge Organization Systems (KOS) (purple)</p> <p>Instructions for indexing and query formulation (white)</p> <p>Sample pages from print version of the scheme (white)</p> |
| Use for all schemes: Indexing and query forms | <p>Document indexing forms A -C (at the end of the Assignment Packet, p. 305)</p> <p>Query forms D - F (after the Document indexing forms)</p> <p>The same documents and queries are used for all schemes</p> <p>Title page and other relevant pages from the documents to be cataloged are found at the very end of the Assignment Packet, p. 319.</p> |

Readings and materials

| | |
|----------------------|--|
| Readings | <p>Needham, Christopher D. Organizing knowledge in libraries: An introduction to information retrieval. 2nd ed. New York: Seminar Press, 1971.</p> <p style="padding-left: 40px;">Ch. 7 Review of classification principles, p. 109-132 (review of classification principles, esp. facet analysis and synthesis)</p> <p style="padding-left: 40px;">Ch. 8 Schemes of classification, esp. p. 133, DDC, p.140-152 LCC, p. 163-168</p> <p style="padding-left: 40px;">Still the clearest exposition of timeless classification principles In Packet, Required</p> <p>Other readings given with the schemes and in the weekly schedule</p> |
| Model catalog | <p>The Model catalog is helpful as a source of examples (Paper and electronic) Covers 25 education documents indexed and arranged by the various index languages. Can be used to study the application of the various index languages and to compare them.</p> |

Assignment 13.1**Assigned: July 9****Due: July 16****Dewey Decimal Classification**

| | |
|--|--|
| Objectives | Inherited from Assignment 13 (page 121) |
| Deliverables to submit | The supplemental worksheet (to come) The filled-in document forms and query forms (Examine the <i>Outline for the analysis of Knowledge Organization Systems</i> (purple), which is already filled-in) |
| Tasks | Inherited from Assignments 13, p. 121 The main worksheet will be done in a presentation. Work on it first in a study group to see what you can figure out for yourself, then go to the presentation. After the presentation do the supplemental worksheet. |
| Materials | <ol style="list-style-type: none"> The newest edition is edition 23 (2012), but you can also work with edition 22 (2003) or 21(1996) available in Baldy 14A or many libraries. Most of you will use WebDewey, which is updated to edition 23. The structural principles and the basic layout of the scheme have remained the same over time. General summary of DDC and more detailed summaries of the areas of Education and some sample pages (Assignment Packet, p. 149-). Resources on the Web (optional). http://www.oclc.org/dewey/ |
| Readings (repeated from calendar page for Unit 10) | <ol style="list-style-type: none"> Needham, Christopher D. Organizing knowledge in libraries: An introduction to information retrieval. 2nd ed. New York: Seminar Press, 1971. Ch. 7 Review of classification principles, p. 109-131 Ch. 8 DDC, p. 133, 140-152 Still the clearest exposition of timeless classification principles. In Packet, Requ. For further study Chan, Lois Mai; Comaromi; Mitchell, Joan Dewey Decimal Classification: Principles and Application. 3. ed. Dublin, Ohio : OCLC, 2003. xi, 216 p. This version not in UB Libr, Optional On the new ed. 23: www.oclc.org/dewey/versions/print/default.htm Optional |
| Time | 4 hours |

DDC Main Worksheet Name:

| | |
|---|---|
| 65 min. 1. General layout and principles of use of DDC | |
| General layout | <p>Look at the volumes (if you have access to a printed version of DDC) or look at the sample pages starting on p.145 and familiarize yourself with the general layout of the classification by reading through the first and second "summary" (including the summary of the "Tables"). Then skim the third summary, reading thoroughly only the section 370, 380, and 620 (all attached). Then skim 370 and 380 in the schedules (vol. 2), skim 629.2 (vol.3).</p> <p>Do the WebDewey exercise (below), parts A and B now, before going through the rest of the worksheet, part C later.</p> |
| Fundamental rules of use | <p>(1) Never assign a class number simply taken from the alphabetical (relative) index; always check the number in the schedules before using it. Finding the appropriate class number is often quicker in the schedules (scan the right section) than in the alphabetical (relative) index with subsequent check in the schedules.</p> <p>(2) Always read the instructions given in the schedules and in the manual very carefully. There are many exceptions to general rules!</p> |
| Focus of document | <p>For determining the proper DDC class always ask yourself:</p> <p>What is the focus of this document? Where does this document belong basically?</p> <p>See examples on the next page</p> |

Document focus examples

Example 1. Sheep

| The anatomy of sheep | Keeping sheep. A farmer's manual | Sheep hunting for profit and pleasure |
|---|---|--|
| 500 Natural sciences and mathematics 590 . Animals [natural history and descriptive biology] 599 . . Mammalia (Mammals) 599.6 . . . Ungulates 599.64 Bovidae 599.649 Sheep | 600 Technology (Applied Sciences) 630 . Agriculture and related 636 . . Animal husbandry 636.3 . . . Sheep | 700 The arts. Fine & decorative 790 . Recreational and performing 799 . . Fishing, hunting, shooting 799.2 . . . Hunting 799.27 Specific big game 799.27 6 Ungulates 799.27 64 Bovidae 799.27 649 Sheep |

- **Always interpret the meaning of a class caption in context**
 599.649 means the *natural history and descriptive biology of sheep*
 799.27|649 means the *hunting of sheep*
- The classes 799.27|6, 799.27|64, and 799.27|649 are built from the pattern under 599; more on building classes in Section 3. (the symbol | separates components of a built class number.)
- All these numbers are found under *sheep* in the relative index where you can also see that *sheep's milk* and *sheep's wool* are in different places still.

Example 2. Taxicabs. Document: **Urban transportation services using taxicabs**

The document is clearly focused on *Commerce, communication, transportation*, specifically local transportation. It would therefore be wrong to use

629.222 32 Taxicabs

which is part of 600 *Technology* and means *Technical aspects of taxicabs*. The correct class is

| | |
|--------------------|--|
| 300 | The social sciences (the main class) |
| 380 | . Commerce, communication, transportation |
| 388 | . . Ground transportation |
| 388.4 | . . . Local transportation [<i>Urban transportation</i> is local] |
| 388.41 | Vehicular and pedestrian traffic |
| 388.413 | Activities and services |
| 388.413 2 | Vehicular services |
| 388.413 21 | Services by private passenger automobiles |
| 388.413 214 | Taxicabs and limousines |

Now give the class number for

The economics of trucking services (Hint: check out 629.22 and 388.3 and note that trucking services could be local or long-distance, thus do not fit under 388.4):

►A

| | |
|---|---|
| 40 min. 2. Building new precombined classes May need to ask in class about this | |
| Introduction | <p>The DDC schedules (vols 2 and 3) enumerate many precombined descriptors, but packing all the concepts of a document theme into one class often requires still more precombination. DDC allows for building new precombined descriptors (classes). Since each built class is represented by a built class number, this is often called number building.</p> <p>Any Dewey class number (enumerated or built) has a decimal point after the third digit and a blank space after the sixth digit. We show the division of a built number into its components through . (The symbol setting off components of a built number is not part of the “official” Dewey number. Nevertheless, in this assignment include the for your own understanding.)</p> <p>To build the number for a new precombined class, one starts with a class number given in the schedule and appends digits from another place, often from a table, as explained in this part of the worksheet.</p> <p>Note: In Dewey number building instructions, “add” it means “append”. In LCC number building instructions, “add” means “use a calculator”.</p> |

| | |
|---|--|
| 2.1 Building new precombined classes 1: General tables | |
| General tables | <p>The most prominent number-building feature in DDC is the use of the general <i>Tables</i> (v. 1), which apply to all Dewey classes. Numbers from these tables are never used alone but combined with base numbers from the schedules.</p> <p>The following instructions introduce Table 1. Standard subdivisions and Table 2. Geographical areas.</p> |
| Table 1. Standard Subdivisions | <p>Numbers from Table 1 may be combined freely with any number from the schedules, e.g.,</p> <p><i>modern Indic languages</i> (491.4), <i>classification of</i> (Table 1 / -012): 491.4 012.</p> <p>(Note: The '-' in front of 012 just shows that this number must not be used alone; the hyphen does not become part of the number: 491.4-012 is wrong!)</p> <p>Example 2:</p> <p>530 .0712 <i>12th grade physics</i> (530 Physics, Table 1 / -0712 Secondary ed.)</p> <p>Now build the class number for</p> <p>Railroads (385) directory (of persons and organizations) (T1 / -025):</p> <p>►B</p> |

2.1 Building new precombined classes 1: General tables, continued

Table 2.
Areas.

Areas from Table 2 can be added according to the following rules:

(1) **By direct addition when so noted in an instruction in the schedules**

Examples:

378.73 Higher education in the US (as instructed at 378.4-.9)

372.9 *Historical, geographical, persons treatment of elementary education* has a note to add directly from Table 2, T2 / -52 is *Japan*, thus

372.9|52 *Elementary education in Japan*

(2) **Through the interposition of "Standard Subdivisions" notation -09 from Table 1** whenever there is no special instruction.

Example:

385 *Railroad transportation* has no special instruction about adding areas directly, thus

385|.09|52 *Railroad transportation in Japan*

The "linkage" function of the standard subdivision -09 is fully explained under

Table 1, -093-099 Treatment by specific continents, countries, localities; extraterrestrial worlds

History and description by place, by spec. instance of the subject

Add "Areas" notation 3-9 from Table 2 to base number - 09. e.g., the subject in the *United States* - 09|73, in *Brazil* -09|81, in *North America* -09|7.

Class persons associated with the subject regardless of area, region, place in -092; treatment by areas, regions, place not limited by continent, country, locality in -091; history and description of the subject among groups of specific kinds of persons in -088, among specific racial, ethnic, national groups nondominant in their continents, countries, localities in -089.

Now build class numbers for the following::

River transportation (386.3) in the U.S. (T2 / -73) (no instruction with 386.3)

►C

The role of trucks in transportation in Germany (area -43)

►D

Again: Add area numbers without intervening 09 only if instructed to do so.

| 2.1 Building new precombined classes 1: General tables, continued | |
|--|---|
| Other general tables | Never add numbers from Tables 3-7 unless an instruction tells you that they can be used. |
| Combination from several tables | When so instructed (and only then), you can build a very specific class by adding from two tables. As an example, go to Table 1 / -025 and read the instruction. You see that one can build the class 385 .025 52 <i>Railroad directories for Japan</i> . |

| 15 min. | 2.2 Building new precombined classes 2: Local tables |
|---------|---|
| | <p>Local tables are found in the schedules and apply only to a specified range of classes. For example, at the beginning of 372.3 - 372.8 <i>Elementary education in specific subjects</i> (DDC21 vol.2, p. 779), there is a table that applies only to this range of classes. This table “hijacks” -04 of Table 1. Read this table and the instructions, and then build the class number for</p> <p>Curriculums for elementary school science:</p> <p>►E</p> <p>For a more extensive local table see 362-363 <i>Specific social problems and services</i> (vol.2, p.682), which gives such concepts as 1 <i>Social causes</i> and 62 <i>Standards</i>, or 616.1- 616.9 <i>Specific diseases</i> (v 3, p.68).</p> |

| | | | | | | | | | | | |
|-----------------------|--|-----|--------------------------------------|-----|---|-----|---------------------------|-------|--------------------------------|--------|------------------|
| 15 min. | 2.3 Building new precombined classes 3: Applying the pattern of subdivision found in a different place in the schedules as a model | | | | | | | | | | |
| Pattern models | <p>Consider a variation on the taxicab example from Section 1 (Example 2). In Section 1 emphasis was on transportation services using a type of vehicles, here the focus is on the vehicles themselves, but in a transportation context (rather than their technical construction), such as a document that is a register of taxicabs in Chicago. This is an example of choosing the proper class based on focus on a detailed level, but the main focus here is on building classes. There is no enumerated class for this topic, the classification only goes as far as shown here:</p> <table data-bbox="435 604 1258 787"> <tr> <td>300</td> <td>The social sciences (the main class)</td> </tr> <tr> <td>380</td> <td>. Commerce, communication, transportation</td> </tr> <tr> <td>388</td> <td>. . Ground transportation</td> </tr> <tr> <td>388.3</td> <td>. . . Vehicular transportation</td> </tr> <tr> <td>388.34</td> <td>. . . . Vehicles</td> </tr> </table> <p>To class more even more specifically, follow the instruction given for 388.342-.348: Add to 388.34 the numbers following 629.22 in 629.222-228, e.g., <i>taxicabs</i> 388.34 232. In other words, the subdivision of vehicles given in the technical class is reused here, saving space in the schedule.</p> <p>Now build the class number for</p> <p>Trucks (transportation focus)</p> <p>►F</p> | 300 | The social sciences (the main class) | 380 | . Commerce, communication, transportation | 388 | . . Ground transportation | 388.3 | . . . Vehicular transportation | 388.34 | Vehicles |
| 300 | The social sciences (the main class) | | | | | | | | | | |
| 380 | . Commerce, communication, transportation | | | | | | | | | | |
| 388 | . . Ground transportation | | | | | | | | | | |
| 388.3 | . . . Vehicular transportation | | | | | | | | | | |
| 388.34 | Vehicles | | | | | | | | | | |

2.4 More examples for building classes

| | |
|----------------------|---|
| More examples | <p>For further examples see the model catalog. But note that in the model catalog, numbers for area and historical period have been added far beyond what is usual, resulting in long numbers. The common practice often omits the area 73 U.S. and the modern time periods. DDC 18 allowed for even more specific (and longer) numbers: 09 followed by area notation followed by time notation taken from 930-990; the official rules of DDC21 (started in DDC19) allow only subdivision by place and not further subdivision by time.</p> <p>Example:</p> <p style="padding-left: 40px;">DDC18: 386.4 09 73 0917 Canal transportation in the U.S. in the Administration of Franklin D. Roosevelt</p> <p style="padding-left: 40px;">DDC21: 386.409 73 Canal transportation in the U.S.</p> <p>Any individual system can, of course, apply DDC any way its designers see fit if they are not worried about compatibility.</p> |
|----------------------|---|

| | |
|----------------------------------|---|
| 45 min. | 3. Probing the conceptual cross-reference structure (Broader Term, Narrower Term, Related Term) |
| Cross-reference structure | <p>For each example, give the class number and list 5 - 7 cross-references (class number and full caption) that <i>should</i> be there, followed by (Sched) if they are given in the schedules, (Rel) if they can easily and obviously be located through the Relative Index, and (No) if they are not given either place. Remember that semantic factoring can be used to detect cross-references needed. Use the Relative Index to find the class numbers for the concepts to which cross-references should be made, but do not spend more than 45 min. on this section.</p> <p>Note: A cross-reference "should" be there if it would help the user to know about the classes referred to. Of the cross-references that should be there by this definition, Dewey includes very few. Your task is to supply the missing cross-references, or at least some of them.</p> <p>Example XXX</p> |

► **G School nursing services**

Class number and full caption:

Cross-references (class number and full caption)

1

2

3

4

5

6

7

► **H Blind students**

Class number and full caption:

Cross-references (class number and full caption)

1

2

3

4

5

6

7

►I Elementary education in music composition

Class number and full caption:

Cross-references (class number and full caption)

1

2

3

4

5

6

7

4. Comparison of exhaustivity and specificity of indexing in two schemes

Introduction and definition of task

Exhaustivity and specificity of indexing are important parameters that affect retrieval performance. (See Organizing Information, Chapter 16, for definitions.) This part of the assignment will help you gain a better understanding of these parameters. For three documents, you will compare DDC with the faceted London Education Classification (LEC).

Examine the entries for the three documents below (each has a grid to be completed by you).

Your task is to compare the exhaustivity and specificity achieved with the LEC with that achieved with DDC (the one Dewey class used for the call number) as shown in the example.

The descriptors from LEC and; the Dewey class are given. As discussed in Chapter 16, in order to compare exhaustivity and specificity of two sets of descriptors assigned to the same document, one must first semantic factor the descriptors to arrive at two lists of elemental concepts, which can then be compared. The LEC concepts are already elemental. **Determine the semantic factors of the Dewey class; all you need for this is the class caption, such as *Organization and activities in secondary education > curricula*. Enter the resulting elemental concepts in column 2, matching them up with the LEC concepts where possible.** Use the blank row if there is a concept from DDC without a corresponding concept from LEC. The elemental concept from the Dewey class may be broader or narrower than the LEC concept. For some LEC concepts there may be no elemental concept from the Dewey class; conversely, an elemental concept from the DDC class may have no corresponding LEC concept. In column 3 circle or underline the scheme that has the more specific concept. Can you detect a pattern? Discuss below at ►J.

| LEC outline | | Some LEC examples | |
|-------------|---|-------------------|---|
| B | Education: foundations, principles, policy, etc | Bap | Educational opportunities; Bept Statistics of educ. Betm Financial resources; Bid Government Center |
| D | School buildings, bldg. services, equipment | Def | Music Room; Dvo Computer |
| F | The teaching profession, personnel management | Fal | Responsibility; Fas Recruitment |
| G | Type of personnel in education | Gan | Dean; Get Librarian; Gon Nurse |
| H | Management of educational institutions | Hab | Management of education; Heb Admission |
| Hs-z | Human biology. Health and hygiene | Htw | Motor ability; Huv Neurosis |
| J | Psychology of education. Educ. measurement | Jze | Emotion; Jed Student discipline; Jud Counseling |
| K | Educand's work (study method, interests, voluntary vs. compulsory work) | Kad | Study methods; Kib Children as writers |
| L | Teaching method | Lah | Team teaching; Lep Group work; Lus Library |
| M-P | Subject taught (Curriculum) | Mab | Curriculum, syllabus, in general; Mok Biology |
| Rab-Sus | Grade level and type of institution | Rib | Comprehensive school; Rek Secondary school - upper |
| Svb-Tvp | Educands (by age, by exceptional requirements, and other characteristics) | Svg | School child, pupil; Teb Blind and partially sighted |

Example

373.19 Organization and activities in secondary education > curricula

Note: This class caption is all you need to semantic factor the DDC class to arrive at the elemental concepts to put in column 2.

Conant, James Bryant, 1893-1978

The comprehensive high school; a second report to interested citizens

| LEC concepts | Elemental concepts from DDC class | Which scheme is more specific? | | |
|--|-----------------------------------|--------------------------------|---|---|
| Rib Comprehensive (type of school) | | SAME LEC DDC | | |
| Rek Secondary school-upper | Secondary education | SAME <u>LEC</u> DDC | | |
| Mab Curriculum, syllabus, in general | Curricula | <u>SAME</u> LEC DDC | | |
| Bept Statistics of education | | SAME LEC DDC | | |
| Bap Educational opportunities, access to education | | SAME LEC DDC | | |
| No. of concepts | | 1 | 1 | 0 |

| Exhaustivity | | Specificity | |
|-----------------------------|---|-----------------------------------|---|
| LEC elemental concepts | 5 | LEC same as DDC | 1 |
| DDC elemental concepts | 2 | LEC more specific | 1 |
| | | DDC more specific | 0 |
| More exhaustive: LEC | | Overall more specific: LEC | |

In this example: LEC indexing has 5 concepts, DDC indexing 2, thus LEC is more exhaustive
 Of the two common concepts, one is more specific in LEC and one is the same specificity, thus,
 for this document, indexing with LEC is more specific overall.

Example, repeated**373.19 Organization and activities in secondary education > curricula**

Conant, James Bryant, 1893-1978

The comprehensive high school; a second report to interested citizens

| LEC concepts | Elemental concepts from DDC class | Which scheme is more specific? | | |
|--|--|---------------------------------------|---|---|
| Rib Comprehensive (type of school) | | SAME LEC DDC | | |
| Rek Secondary school-upper | Secondary education | SAME <u>LEC</u> DDC | | |
| Mab Curriculum, syllabus, in general | Curricula | <u>SAME</u> LEC DDC | | |
| Bept Statistics of education | | SAME LEC DDC | | |
| Bap Educational opportunities, access to education | | SAME LEC DDC | | |
| | No. of concepts | 1 | 1 | 0 |

| Exhaustivity | | Specificity | |
|-----------------------------|---|-----------------------------------|---|
| LEC elemental concepts | 5 | LEC same as DDC | 1 |
| DDC elemental concepts | 2 | LEC more specific | 1 |
| | | DDC more specific | 0 |
| More exhaustive: LEC | | Overall more specific: LEC | |

In this example: LEC indexing has 5 concepts, DDC indexing 2, thus LEC is more exhaustive. Of the two common concepts, one is more specific in LEC and one is the same specificity, thus, for this document, indexing with LEC is more specific overall.

Document 1

378.111 **Higher Education > Organization and activities in higher education > Personnel management > Administrators**

Dibden, Arthur James, 1919-

The academic deanship in American colleges and universities

| LEC concepts | Elemental concepts from DDC class | Which scheme is more specific? |
|--------------------------------------|-----------------------------------|--------------------------------|
| Sab Institutions of Higher Education | | SAME LEC DDC |
| Hab Management of Education | | SAME LEC DDC |
| Gan Dean | | SAME LEC DDC |
| Ban Sociology of Education | | SAME LEC DDC |
| | | SAME LEC DDC |
| | No. of concepts | |

| Exhaustivity | | Specificity | |
|-------------------------|--|-------------------------------|--|
| LEC elemental concepts | | LEC same as DDC | |
| DDC elemental concepts | | LEC more specific | |
| | | DDC more specific | |
| More exhaustive: | | Overall more specific: | |

Document 2

379.1214 **Public policy issues in education > Specific elements of support and control of public ed. > Support by specific level of government > Support by national governments > National support of higher ed.**

Wakefield, Rowan Albert, 1919-

Sources of Federal support for higher education. Experimental systems for a national information network.

| LEC concepts | Elemental concepts from DDC class | Which scheme is more specific? | | |
|------------------------------|--|---------------------------------------|--|--|
| Sab Higher education | | SAME LEC DDC | | |
| Dvo Computer | | SAME LEC DDC | | |
| Buxt Information services | | SAME LEC DDC | | |
| Bid Government: central | | SAME LEC DDC | | |
| Betm Financial resources | | SAME LEC DDC | | |
| Bepd Economics and education | | SAME LEC DDC | | |
| | | SAME LEC DDC | | |
| | No. of concepts | | | |

| Exhaustivity | | Specificity | |
|-------------------------|--|-------------------------------|--|
| LEC elemental concepts | | LEC same as DDC | |
| DDC elemental concepts | | LEC more specific | |
| | | DDC more specific | |
| More exhaustive: | | Overall more specific: | |

Document 3

027.8 General libraries, archives, information centers > School libraries

Delaney, Jack J. *The school librarian. Human relations problems.*

| LEC concepts | Elemental concepts from DDC class | LEC or DDC more specific? |
|--|-----------------------------------|---------------------------|
| Svg School child, pupil | | SAME LEC DDC |
| Rag School education and systems | | SAME LEC DDC |
| Lus Library | | SAME LEC DDC |
| Jed Discipline (psychological aspects of relationship) | | SAME LEC DDC |
| Hab Management of education | | SAME LEC DDC |
| Get Librarian, teacher librarian | | SAME LEC DDC |
| Fal Role, responsibility | | SAME LEC DDC |
| | | SAME LEC DDC |
| | No. of concepts | |

| Exhaustivity | | Specificity | |
|-------------------------|--|-------------------------------|--|
| LEC elemental concepts | | LEC same as DDC | |
| DDC elemental concepts | | LEC more specific | |
| | | DDC more specific | |
| More exhaustive: | | Overall more specific: | |

Briefly discuss what you saw about exhaustivity and specificity from these examples:

►J

Do not do Task 5. It will be done in class / included in the discussion of Assignment 13.1

| | |
|---------------------------|--|
| 30 min. | 5. Further probe into the structure of DDC |
| More DDC structure | <p>Analyze instructions in the DDC schedules. Discuss each instruction briefly.</p> <p>(1) with respect to combination order (in Case 1, is the order Level of education – Subject or the other way around?)</p> <p>(2) with respect to the effect on the (2.1) exhaustivity and (2.2) specificity of indexing (how many of the aspects for which a document is relevant are represented in the Dewey class (precombined concept) to be used; at what level of specificity); and</p> <p>(3) with respect to the effect on retrieval.</p> |

More examples are in the supplement

In Case K and Case L you will analyze these problems in the area of education, exploring the effects of the following instruction with 370 Education (slightly edited):

Case K: Level of education versus subject

| | |
|-----|---|
| (a) | <p>Class special education in a specific subject in 371.9 This takes precedence, (b) and (c) apply only if not special education</p> |
| (b) | <p>Class elementary education in a specific subject in 372.3-372.8. Example: a book on <i>Physics experiments for third grade</i> would be classed under 372.35 Elementary education in specific subjects > Science and technology</p> |
| (c) | <p>Class works on secondary, higher, and adult education in a specific subject with the subject plus the appropriate number under 071 from Table 1 Standard subdivisions Example: The book <i>12th grade physics</i> would be classed under 530 .0712 (530 Physics; 0712 Secondary education, from Table 1) <i>Freshmen physics</i> under 530 .0711 (0711 Higher education, from Table 1)</p> |

Case K deals with (b) and (c), Case L (supplement) takes up (a), special education

K Analyze Case K; write your answers in the proper slots. This will not be handed in.

(1) Combination order (Level – *Subject* or *Subject* — Level)

(a) books on the elementary level:

(b) secondary or higher level:

(2.1) Exhaustivity of indexing — which facets are represented in the class?

(a) elementary level:

(b) secondary or higher level:

(2.2) Specificity of indexing — how specific is the concept from each facet?

(a) elementary level: Level: *Subject:*

(b) secondary or higher level: Level: *Subject:*

(3) Effect on retrieval (recall and discrimination)

Consider the query topics

Physics AND Elementary school versus

Physics AND Secondary education

Write a very brief analysis

Case L is in the supplement

WebDewey Exploration

| | |
|---------------------|--|
| Introduction | <p>WebDewey allows for</p> <ul style="list-style-type: none"> A. navigation in the classification and B. for specific search by class number or word. <p>This provides power not available with the printed version. However, the printed version has its own advantages. Compare!</p> <p>You should know your way in Windows (e.g., know how to minimize a window or how to copy and paste text from WebDewey and save some time).</p> <p>You may use WebDewey for any part of the Dewey assignment, but you should also gain some experience with using the printed volumes so that you can compare. You may find the printed volumes in your public or academic library. (Edition 23 is newest, but 22 or 21 will do also.)</p> |
| To start | <p>You can access WebDewey on OCLC Connexion as follows</p> <p>http://connexion.oclc.org/</p> <p>Authorization (Name): 100062747 Password: sils2</p> <p>Click on Logon</p> <p>In the bar on the top, click on ⇒DeweyServices</p> <p>This opens a new browser window. You may want to maximize it.</p> |
| Note | <p>The prompt numbers you see may be different from the ones given in the following instructions. The prompt numbers may change from session to session.</p> |

| A2 The Browse interface | |
|--------------------------------------|--|
| Browse full hierarchy display | <p>On the top, click the orange Browse button <i>Browse Dewey Numbers (with Captions)</i> is the default</p> <p>In the box, type 372-374, press Enter</p> <p>A multi-page display of all the classes under that appears The orange symbol at the left of a class means the class is built; it is not a class you would find in the printed schedule.</p> <p>Find 372.357 and click on it. This takes you to the walk-up-and-down hierarchy display</p> <p>Use the browser Back button to get back the Browse screen</p> <p>For another example of a built class, in the Browse display find 372.1262 Standardized tests--elementary education</p> <p>Click on that. You are now back in the walk-up-and-down hierarchy display.</p> <p>Note that 372.1262 Standardized testing is a built class.</p> <p>Click on 372.11-372.18 to see the rule for building the classes under 372.1</p> |

B Query-based search of the classification

In the navigation bar on the top click on **Advanced Search**.

In Search you specify

a **starting point**, a word or class number typed in a text box labeled **Search for** and a **relationship type to follow**, here expressed as a field in which to search, selected from the drop-down box labeled **in**.

| B0 Fields that can be searched | |
|---------------------------------------|--|
| Searchable fields | All Fields (al:) (default) All Dewey (ad:) Dewey Numbers (dd:) Captions (class name or text) (cp:) Relative Index (ri:) Notes (nt:) LCSH (lc:) |

| B1 Search for class number | |
|-----------------------------------|---|
| Class number search | In the <i>In Box</i> , select <i>Dewey Numbers</i> In the <i>Search for Box</i> , type 620, press Enter This takes you into the walk-up-and-down hierarchy display Briefly explore this area of the classification |

| B2 All Fields search for words | |
|---------------------------------------|--|
| Word search | Query: <i>computer* and education*</i> (* truncates) In the In box, select All Fields In the text box labeled for , paste the query, press [Enter]. click on <i>372.35 Science and technology</i> From the notes section you can see why this class was found |

| B3 All Fields search for words and <u>class numbers</u> | |
|--|---|
| Word and class number search | <p>Query: <i>computer* and 37*</i></p> <p>Get back to the search screen. In the first text box, type <i>computer*</i> make sure in is All Fields (default),</p> <p>In the second text box, type <i>37*</i> , set in to Dewey number and press [Enter].</p> <p>The results show class numbers not found in the previous search. For example, click on number 6 next to <i>371.33466 Computer graphics--instructional use</i> The word <i>education</i> does not appear there, so B1 did not find it.</p> |

| B4 Captions Only search for words | |
|--|--|
| Word search in Captions | <p>Query: <i>computer* and education*</i></p> <p>In the top bar, click on Advanced Search.</p> <p>In the first text box, type <i>computer*</i>, set the In to Captions.</p> <p>In the second text box, type <i>education*</i>, set the In to All Fields [Enter].</p> <p>Examine results. Note that 372.35 is not found this time. Why?</p> <p>Now go Back to the search screen and set the second in to Captions, [Enter].</p> |

| B5 All Fields search for class numbers | |
|---|---|
| Class number search | <p>Query: 372*</p> <p>In the toolbar on top, click Search.</p> <p>In the first text box, type 372*, make sure In is All Fields, make sure the second text box is empty, and press [Enter].</p> <p>Lots of classes are found. Note: C in this display indicates a <i>centered heading</i>.</p> <p>On results page 4, click on 268.432 <i>Children</i> to display the full record. Why was this record found? Note: This is not <i>children in general</i> but rather <i>Christian religious education of children</i></p> <p>Use the browser Back button to get back to the result list</p> <p>Explore some of the other classes found, such as (on p. 11) 649.58 <i>Reading and related activities</i></p> |

C Now complete query F from the Query Forms and, optionally, the query "Helicopters"

List **all** classes (or as many as you can do in a reasonable time) where one should look on the shelves (broad class implies subclasses). Note that this is easier on WebDewey as compared to the printed version because you can search and then copy and paste.

Hint: If you paste into a word processor document, it is best to use
Edit > Paste special > Unformatted text

| | |
|--|--|
| <p>Query F Canals</p> | <p>Do an "All Fields" search for <i>canal*</i>:</p> <p>Scroll through the entire results list and copy and paste classes relevant to the query into a word processor document. Note: Highlight the class no. and caption, use shortcut key Ctrl-C to copy, Ctrl-V to paste.</p> <p>Did you search for the concept <i>canal (waterway)</i> or simply for the word <i>canal</i> in all its meanings? Clearly you are not searching a concept-based descriptor-find index but a word-based index.</p> <p>Click on no. <i>10</i> next to <i>386.244 Freight services</i> to see its full record. Look at the Relative Index terms to see why it was found?</p> <p>Do a Captions search for <i>canal*</i>. Look at the difference in results.</p> |
| <p>Query Helicopters (optional)</p> | <p>Do an All Fields search for <i>helicopter*</i>.</p> <p>In the result list, click on <i>87.73352387.73352</i> (a built number) to see the full record. Scroll down to see the last Relative Index term, <i>Helicopters - transportation services</i>; it gives the meaning of the built class. Compare the number to <i>629.133352</i> (under <i>629.133Aircraft types</i>)</p> <p>Click on <i>387.732-387.733 Specific types of aircraft</i> Notice the instructions for building the class number for the classes under it; some of the numbers that can be built are listed in WebDewey.</p> |

This ends the exercise. You may want to explore some more on your own or close WebDewey.

Outline for the analysis of Knowledge Organization Systems

For some items, a section number from Soergel, Organizing information is given in ()

1. Purpose

1.1 Information system or type of information system in which to be used

Bibliographic information system. Intended for public and school libraries.

1.2 Intended for controlled vocabulary indexing , or query term expansion G (Ch. 12, Introduction)

1.3 Type of file and search mechanism for which originally designed

Shelving , Card catalog / printed index G Online system G (Now promoted for Web subject directories)

2. Coverage and designation of concepts. Coverage and format of terms

2.1 Concepts: Scope, breadth of coverage. Recency of concepts

Universal — covers all of knowledge. But focus on Western culture, esp. US.

2.2 Concepts: Specificity, depth of coverage. (Section 16.2.2). Coverage at each level of specificity.

Medium specificity. Would need closer analysis by subject area. Geography table quite specific.

2.3 Are all needed facets included? Concepts formed in semantic factoring and facet analysis? (S.a. 3.1)

Some general concepts included in the general tables and the in-schedule tables. Many others not included by themselves but only as components in one or more precombined descriptors. Completeness of explicit and implied facets? Answer would require extensive analysis.

2.4 Nature of notation (if none, state that). (Section 15.5.2) *Decimal, highly expressive (with some exceptions).*

2.5 Terms: Completeness of coverage (completeness of lead-in vocabulary). Recency of terms

Some lead-in synonyms included in the alphabetical index. How complete? Would need extensive analysis!

2.6 Form of terms: Consistency, adherence to common usage. *Terms seem appropriate. Many classes cannot be expressed by a simple term but need a phrase devised by the editor.*

3. Terminological and conceptual analysis and conceptual structure.

3.1 Quality of conceptual structure (14): Facet analysis. Types and degree of differentiation of conceptual relationships included. For each type indicate the completeness of inclusion. (Fill in 3.1.1 - 3.1.3)

3.1.1 Expression of concepts through elemental concepts (closely related to definition)

For enumerated compound concepts: Sometimes done implicitly in the relative index. For precombined descriptors constructed according to DDC rules: Done by the indexer.

3.1.2 Hierarchical relationships (polyhierarchy) (Shown by arrangement or Broader Term / Narrower Term X-ref)

Monohierarchical. A few additional BT/NT through cross-references. Many hierarchical relationships implied by the relative index (Example: The classes shown under Blind, as seen from the example query formulation).

3.1.3 Associative relationships. (Implied by physical proximity in the arrangement or explicit Related Term X-ref)

Some explicit cross-references

3.2 Quality of definitions, explications, scope notes (correctness, detail, clarity).

Many notes throughout the schedules and in the Manual. Mostly usage notes explaining the difference between classes or instructions on how to form new precombined descriptors. A few definitions

3.3 Completeness of terminological relationships: Does the vocabulary contain terms that are synonymous or quasi-synonymous without indicating the relationship? *Not a problem in a classification like DDC.*

4. **Use of precombination in the index language** (concerns both 2 and 3) (14, 15, esp. 15.4)

4.1 To what degree are descriptors precombined?

DDC can be used with a medium to high degree of precombination, depending on how many new precombined descriptors the indexer builds.

4.2 To what extent are precombined descriptors enumerated and/or given in the alphabetical index?

Medium degree of enumeration in the schedules, some addl. precombined descriptors in the index.

To what extent can the indexer build additional precombined descriptors?

To a large extent. Libraries differ in their use of this option.

Are precombined descriptors designated by an independent symbol or a string of symbols? Combination order free or fixed? To what extent do the components of a precombined descriptor determine its place in the arrangement? (Relates also to 5) (Section 15.5.2)

Enumerated precombined descriptors have their own independent symbol (which sometimes is constructed using notation components from tables). Combination order is fixed. The components completely determine the place of a precombined descriptor built by the indexer.

5. **Access and display. Format of presentation of the vocabulary**

Consider for each format access/retrieval by concepts versus access/retrieval by terms.

Access can be provided through arrangement in a printed document or through a computer search system.

5.1 **Format of printed document** (Fill in 5.1.1- 5.1.3)

5.1.1 Overall format: Thesaurus parts and information given in each, connections between them. Is the overall format clear and helpful for finding the appropriate concepts and terms or notations in indexing and query formulation?

*Introduction (v.1), Tables (v.1), Schedules (v.2+3), Relative Index (v.4), Manual (v.5)
Need to go back forth between schedules and manual, otherwise reasonably helpful.*

5.1.2 Display of conceptual relationships (Broader Term, Narrower Term, Related Term)

- through linear arrangement or graphical display (Section 15.5.2)

In the tables and schedules.

- through cross-references (Section 14.1)

In the tables and schedules.

- through descriptor-find index (Section 15.5.1)

The relative index combines the functions of an alphabetical index and a descriptor find index of sorts.

How well does the display reflect the conceptual analysis, e.g., sequence of concepts on the same hierarchical level (sequence of the children of a concept, that is, the concepts one level further down).

Usually the sequence of classes makes good sense.

5.1.3 Display of terminological relationships (Synonymous Term)

Terminological relationships are displayed only in the relative index, which gives the lead-in term and points to the appropriate class number.

5.2 **Access through computer system.** Navigation. Format of on-line displays

This would be an analysis of Dewey for Windows. Not required here.

Instructions for using the Dewey Decimal Classification

| | |
|---------------------------------|---|
| <p>Indexing</p> | <p>Give one and only one class per document; if another class is a strong contender, list it as an alternate and give reasons for selecting the class you chose. (If there is no other class that would fit the book, leave the Alternate class blank.) For each class give the text as a hierarchical chain, starting with the main class, e.g.</p> <p style="text-align: center;">388.34 232 Ground transportation > Vehicular transportation > Vehicles > Taxicabs</p> <p>Note: The symbol setting off components of a built number is not part of the “official” Dewey number. Nevertheless, in this assignment you should include the for your own understanding.</p> |
| <p>Query formulation</p> | <p>Try to list all classes where one should look for relevant documents on the shelves; if there are more than 7 classes, just give a representative sample (enough to demonstrate that you how to find all classes throughout all the DDC main classes).</p> <p>Note: A broad class implies all the narrower classes; no need to list these narrower classes, they can be readily seen from the schedules.</p> <p>The query formulation is the OR combination of all the classes in your list. (AND combinations are out: for shelving, only one class is assigned; while more than one class could be assigned for a classified catalog, DDC is not designed for combination searching.)</p> <p>You may use the print version or Dewey for Windows for documents A - C and queries D - E. Query F, <i>canals</i> is part of the Dewey for Windows exercise. If you wish, you may try them in the print version as well for comparison.</p> |

Examples of queries in DDC

| Query topic: <i>Blind people</i> | |
|---|--|
| Look under the following Dewey classes: | |
| 027.663 081 61 | Libraries for persons with disabilities > Persons with impaired vision (027.663 is given in the Relative Index under blind; this number, built using Table 8, is more specific) |
| 305.908 161 | Occupational and miscellaneous groups > Persons with impaired vision (A number built according to an instruction using Table 7; this number is given in the Relative Index) |
| 362.41 | Problems of and services to people with physical disabilities > Persons with impaired vision |
| 371.911 | Blind and partially sighted students |
| 614.599 7 | Incidence of . . . diseases > Eye diseases Any class built with the following numbers from general tables. |
| T1/ -0871 | History and description with respect to kinds of persons > Blind and partially sighted persons |
| T7/ -081 61 | Persons with impaired vision |

| Query topic: <i>Teaching of science at all levels</i> | |
|--|---|
| 372.35 | Elementary education > Specific elementary school subjects > Science and technology |
| 5xx.x...x 071x...x | Any built number where the first component is a subdivision of 500 <i>Natural sciences and mathematics</i> (except 510 <i>Mathematics</i> and its subdivisions) and the second component is a subdivision of T1/ -071 <i>Schools and courses</i> (as found in the schedules or as further built by adding geographical area as instructed) Note: The precise form of these class numbers depends further on specific instructions on how to append standard subdivisions to numbers from the schedules. Sometimes one must use 00 or even 000 instead of plain 0. |

XXX update for DDC 23. Partially updates

The following pages show

First summary: the ten top-level main classes

Also a comparison of the overall outline of the Dewey Decimal Classification, the Library of Congress Classification, and the Yahoo Classification

Second summary: The 110 classes at the top level and one level below

A sample page from the third summary (for 300 Social sciences)

Sample pages from the classification schedules

Sample pages from the Relative Index (alphabetical index)

Dewey Decimal Classification

Edition 23

Summaries

Full headings from v. 2 and 3, with small additions from OCLC's Web site

First Summary
The Ten Main Classes

Tables (auxiliary descriptors)

| | |
|------------|--|
| 0 | Computer science, information & general works |
| 100 | Philosophy & psychology |
| 200 | Religion |
| 300 | Social sciences |
| 400 | Language |
| 500 | Science |
| 600 | Technology |
| 700 | Arts & recreation |
| 800 | Literature |
| 900 | History & geography |

| Dewey Decimal | Library of Congress | Yahoo |
|---|--|--|
| 000 Computers, information, & general reference 000 Computers, Internet, & systems 010 Bibliography 020 Library and information sci. 070 News media | A General works QA Math, incl. computer science Z Bibliography and library science | Computers & Internet Reference News & Media |
| 100 Philosophy & psychology | B-BJ Philosophy. Psychology. | A & H > Humanities > Philosophy |
| 200 Religion | BL-BX Religion | Society and Culture > Religion and Spirituality |
| 300 Social sciences 330 Econ, 380 Commerce 320 Pol. sci., 350 Pub. admin 340 Law 370 Education | H Social sciences HB-HJ Economics J Political science K Law L Education | Social Science Society & Culture Business & Economy Government Government > Law Education |
| 400 Language | P-PM Language | Social Science > Linguistics and Human Languages |
| 500 Natural sciences & math. | Q Science | Science (parts of) |
| 600 Technology (Applied sciences) 610 Medical sciences and medicine 630 Agriculture Most of 600 | R Medicine S Agriculture T Technology | Health Science > Agriculture Science (parts of) |
| 700 Arts & recreation | M Music and books on music N Fine arts GV Recreation, leisure | Arts & Humanities Entertainment Recreation & Sports |
| 800 Literature & rhetoric | PN-PZ Literature | Arts & Humanities > Humanities > Literature |
| 900 Geography & history | C-F History, G Geography | A & H > Humanities > History Regional |
| | U Military science V Naval science | Government > Military |

Dewey, Library of Congress and Yahoo classification compared

Note: Comparison is easier at lower levels of the hierarchy since different schemes may agree on specific narrower fields but group them differently at the top level of the hierarchy

Second Summary. *The Hundred Divisions*

Tables

| | |
|-----------------|---|
| Table 1. | Standard Subdivisions (p. 3) |
| Table 2. | Geographical Areas, Historical Periods, Persons (p. 34) |
| Table 3. | Subdivisions f. the Arts, for Individual Literatures, for Specific Literary Forms (p. 405) |
| Table 4. | Subdivisions of Individual Languages and Language Families (p. 437) |
| Table 5. | Ethnic and National Groups (p. 444) |
| Table 6. | Languages (p. 464) |

| | | | |
|------------|--|------------|---|
| 000 | Computer sci., info. & gen. works | 500 | Natural sciences & mathematics |
| 000 | Computer science, knowledge & systems | 510 | Mathematics |
| 010 | Bibliographies | 520 | Astronomy & allied sciences |
| 020 | Library & information sciences | 530 | Physics |
| 030 | Encyclopedias & books of facts | 540 | Chemistry & allied sciences |
| 040 | [Unassigned] | 550 | Earth sciences |
| 050 | Magazines, journals & serials | 560 | Paleontology Paleozoology |
| 060 | Associations, organizations & museums | 570 | Life sciences Biology |
| 070 | News media, journalism & publishing | 580 | Plants |
| 080 | Quotations | 590 | Animals |
| 090 | Manuscripts & rare books | | |
| 100 | Philosophy and Psychology | 600 | Technology (Applied sciences) |
| 100 | Philosophy | 601-9 | General technology |
| 110 | Metaphysics | 610 | Medical sciences Medicine |
| 120 | Epistemology | 620 | Engineering & allied operations |
| 130 | Parapsychology & occultism | 630 | Agriculture & related technologies |
| 140 | Philosophical schools of thought | 640 | Home economics & family living |
| 150 | Psychology | 650 | Management & auxiliary services |
| 160 | Logic | 660 | Chemical engineering |
| 170 | Ethics (moral philosophy) | 670 | Manufacturing |
| 180 | Ancient, medieval, Oriental philosophy | 680 | Manufacture for specific uses |
| 190 | Modern Western philosophy | 690 | Buildings |
| | | 700 | The Arts Fine and decorative arts |
| | | 710 | Civic & landscape art |
| | | 720 | Architecture |
| | | 730 | Plastic arts Sculpture |
| | | 740 | Drawing & decorative arts |
| | | 750 | Painting & paintings |
| | | 760 | Graphic arts Printmaking & prints |
| | | 770 | Photography & photographs |
| | | 780 | Music |
| | | 790 | Recreational & performing arts |
| 200 | Religion | 800 | Literature & rhetoric |
| 210 | Philosophy & theory of religion | 810 | American literature in English |
| 220 | Bible | 820 | English & Old English literatures |
| 230 | Christianity Christian theology | 830 | Literatures of Germanic languages |
| 240 | Christian moral & devotional theology | 840 | Literatures of Romance languages |
| 250 | Christian orders & local church | 850 | Italian, Romanian, Rhaeto-Romanic |
| 260 | Social & ecclesiastical theology | 860 | Spanish & Portuguese literatures |
| 270 | History of Christianity & Christian church | 870 | Italic literatures Latin |
| 280 | Christian denominations & sects | 880 | Hellenic literatures Classical Greek |
| 290 | Comparative religion & other religions | 890 | Literatures of other languages |
| 300 | Social sciences | 900 | Geography & history |
| 301 | Sociology and anthropology | 909 | World History |
| 310 | Collections of general statistics | 910 | Geography & travel |
| 320 | Political science | 920 | Biography, genealogy, insignia |
| 330 | Economics | 930 | History of the ancient world to ca. 499 |
| 340 | Law | 940 | General history of Europe |
| 350 | Public administration & military science | 950 | General history of Asia Far East |
| 360 | Social problems & services; association | 960 | General history of Africa |
| 370 | Education | 970 | General history of North America |
| 380 | Commerce, communications, transportation | 980 | General history of South America |
| 390 | Customs, etiquette, folklore | 990 | General history of other areas |
| 400 | Language | | |
| 410 | Linguistics | | |
| 420 | English & Old English | | |
| 430 | Germanic languages German | | |
| 440 | Romance languages French | | |
| 450 | Italian, Romanian, Rhaeto-Romanic | | |
| 460 | Spanish & Portuguese languages | | |
| 470 | Italic languages Latin | | |
| 480 | Hellenic languages Classical Greek | | |
| 490 | Other languages | | |

DDC sample pages ending even

Assignment 13.2 Yahoo**Assigned: July 16****Due: July 23****Yahoo classification**

| | |
|-------------------------------|--|
| Objectives | Inherited from Assignment 13 (page 121) |
| Deliverables to submit | Inherited from Assignment 13 (page 121) |
| Tasks | Inherited from Assignment 13 (page 121) The Yahoo worksheet starts on the back of this page. |
| Materials | <p>0 The Yahoo classification itself as available on the Yahoo Web site Use http://dir.yahoo.com/ as a shortcut to the Yahoo directory Or www.yahoo.com, click on <u>More Yahoo! Services</u>, then on Directory Plan to be online to Yahoo while doing the worksheet, even though some of it needs just the attached materials. You need the Yahoo Web site for indexing and query formulation.</p> <p>You can prepare your worksheet answers on a word processor, which gives you the opportunity to copy and paste examples from Yahoo.</p> <p>Figures included in print.</p> <p>1 Yahoo Home (first summary): top level classes (called <i>categories</i> in Yahoo) in the original Yahoo alphabetical arrangement and in a revised meaningful arrangement.</p> <p>2 First-level breakdown of <i>Health</i> in the original Yahoo alphabetical arrangement and in a revised meaningful arrangement.</p> <p>3 Second summary (first two levels of the hierarchy).</p> <p>4 Excerpts from the Yahoo classification designed to illustrate the structure of the classification, particularly its treatment of <i>Education</i> and <i>Transportation</i>.</p> |
| Time | Time: 6 hours (first half - 3.5 hours; second half - 2.5 hours) |

Yahoo Worksheet

Name:

Starts on next page

| | | | | | | | |
|------------------------------------|---|-------------------------|-------------------|----------------|---------------------|----------------------------|-------|
| 60 min. | 1. General layout of the classification. Formal structure | | | | | | |
| General layout | <ul style="list-style-type: none"> • Look at the top level on the Yahoo Directory home page and click down to <i>Education</i>. (http://dir.yahoo.com/) • Look at Figs. 1 and 2 and skim Fig. 3. the Yahoo classific. second summary. • Skim through Figure 4, the excerpt from the Yahoo classification; note the many places in which <i>education</i> and <i>transportation</i> concepts appear. | | | | | | |
| Context dependency of terms | <p>In Yahoo, as in the Library of Congress Classification, the meaning of a category (class) is always defined by its total context. Thus in</p> <p style="padding-left: 40px;">Education</p> <p style="padding-left: 60px;">. K-12</p> <p style="padding-left: 80px;">. . Academic Competitions</p> <p style="padding-left: 100px;">. . . Debate</p> <p style="padding-left: 120px;">. . . . Clubs, Teams, and Societies</p> <p>The last category means</p> <p style="padding-left: 40px;"><i>Clubs, Teams, and Societies</i> that engage in <i>Debates</i> staged as <i>Academic Competitions</i> for students in the <i>K-12</i> level of <i>Education</i></p> <p>On the top of the screen, the active category is shown with its full caption:</p> <p style="padding-left: 40px;">Directory > Education > K-12 > Academic Competitions > Debate > Clubs, Teams, and Societies</p> | | | | | | |
| Precombination | What is the degree of precombination in Yahoo? | | | | | | |
| Search for categories | <p>A search for one or more words in Yahoo returns</p> <p>(1) categories whose caption contains the word(s) and</p> <p>(2) individual sites.</p> <p>So Yahoo functions as a kind of descriptor find index. But the search is based on the words in the full caption; while these words often reflect the conceptual components of the category, there are many cases where they do so only incompletely. See the facing page for an example. A search for <i>vehicles</i> does not <i>ships</i> or <i>aircraft</i>.</p> <p>Some good searches to try to explore the structure of the Yahoo classification. You can restrict your search to a category.</p> <table border="1" data-bbox="451 1724 1300 1875" style="margin-left: auto; margin-right: auto;"> <tr> <td style="padding: 5px;">CDs, Records, and Tapes</td> <td style="padding: 5px;">Mexican Americans</td> </tr> <tr> <td style="padding: 5px;">Transportation</td> <td style="padding: 5px;">Bridge (a homonym)</td> </tr> <tr> <td style="padding: 5px;">Teaching and Learning Aids</td> <td style="padding: 5px;">Humor</td> </tr> </table> | CDs, Records, and Tapes | Mexican Americans | Transportation | Bridge (a homonym) | Teaching and Learning Aids | Humor |
| CDs, Records, and Tapes | Mexican Americans | | | | | | |
| Transportation | Bridge (a homonym) | | | | | | |
| Teaching and Learning Aids | Humor | | | | | | |

Example: Search for categories (descriptor find index): Words vs concepts

| | |
|---|--|
| <p>Search 1: law education Canada (implied AND) Yahoo! Category Matches (1 - 4 of 4)</p> <p>Regional > Countries > <u>Canada</u> > Provinces and Territories > Saskatchewan > Cities > Saskatoon > Education > College and University > University of Saskatchewan > Departments and Programs > College of Law [Note: <i>College</i> rather than <i>School</i>]</p> <p>Regional > Countries > <u>Canada</u> > Provinces and Territories > Ontario > Counties and Regions > Frontenac > Cities > Kingston > Education > College and University > Queen's University > Departments and Programs > School of Law</p> <p>Regional > Countries > <u>Canada</u> > Provinces and Territories > British Columbia > Counties and Regions > Capital > Cities > Victoria > Education > College and University > University of Victoria > Departments and Programs > Law [Note: Simply a <i>department</i>]</p> <p>Regional > Countries > <u>Canada</u> > Provinces and Territories > Alberta > Counties and Districts > Edmonton > Cities > Edmonton > Education > College and University > University of Alberta > Departments and Programs > Law [Note: Simply a <i>department</i>]</p> | <p>Search 2: law school Canada Yahoo! Category Matches (1 - 2 of 2)</p> <p>Regional > Countries > <u>Canada</u> > Provinces and Territories > Ontario > Counties and Regions > Frontenac > Cities > Kingston > Education > College and University > Queen's University > Departments and Programs > School of Law</p> <p>Regional > Countries > <u>Canada</u> > Government > Law > Law Schools</p> |
|---|--|

In the Yahoo categories, all Search 1 retrievals happen to be also relevant for Search 2. (There could be one or more categories on *law education in Canada* generally, but there are none.). All Search 2 retrievals are by definition relevant for Search 1.

When running the same searches with *US*, Search 1 finds 62 categories as follows

- 44 have both *law* and *school* and are thus also found by Search 2
- 15 have *law* and *college* (These are relevant for Search 2 but not found by it)
- 3 have *law* and some other term, such as *center*

Search 2 finds 44 categories; they all happen to have *education*, and thus were found in Search 1 as well.

Searching for *law school* without restrictions would find categories not found by *law education* (such as Directory > Government > Law > Law Schools; this category has under it categories that say *College of Law*) and vice versa (such as Directory > Government > Law > Continuing Legal Education).

Note: Searching for *United States* finds no categories; you need to search for *US*.

1. General layout of the classification. Formal structure. Continued

The Yahoo “multi-tree”

As you know from Chapters 14 and 15, a compound concept fits in many places in a hierarchy. Put differently, in the Yahoo subject directory, a precombined category should be reachable through multiple paths down. How does Yahoo handle this problem? Probe the following example:

Click down to

Directory > Education > K-12
Schools

Click on Christian@

Try Directory > Education > K-12 > By Region > Countries > France > Cities

Probe some on your own

Summarize your observations. What does the @ mean?

► **Answer A**

Nature of subordinate categories

In Example 1 on the facing page, why are the subordinate categories in the first group narrower than *Canada*, why the subordinate categories in the second group? What is the difference? (Hint: Remember *concepts narrower due to autonomous subdivision* and *concepts narrower due to combination*, p. 264 and 270 in *Organizing Information*)

In Example 2 (which is not quite as clear cut),

mark with T the subordinate categories that are transportation-specific
mark with G the subordinate categories that are the broad concept of *transportation* combined with some general concept not specific to the transportation domain.

Note your observations

► **Answer B**

Nature of subordinate categories. Example 1

Directory > Regional > Countries >

Canada

- . Cities (1697)
- . Provinces and Territories (89205)
 - . Alberta (7683)
 - . British Columbia (19882)
 - . Manitoba (3680)
 - . . .
- . Arts and Humanities (2425)
- . Business and Economy (20557)
- . Computers and Internet (333)
- . Country Guides (21)
- . Education (1236)
- . . .

Nature of subordinate categories. Example 2

Directory > Business and Economy >

Transportation

- . Auto-Free Transportation (23)
- . Aviation (513)
- . Buses (26)
- . Companies@
- . Employment (5)
- . Government Agencies (62)
- . Highways and Roads (127)
- . History (5)
- . Institutes (44)
- . Intelligent Transportation Systems (25)
- . Libraries (7)
- . Mass Transit (59)
- . Web Directories (3)

Figure W-2. Nature of subordinate categories

The remaining questions deal with content aspects of the Yahoo classification

| | |
|-------------------------------|--|
| 30 min. | 2. Develop a meaningful arrangement of the categories one level below <i>Education</i> |
| Meaningful arrangement | <p>You can use the meaningful arrangement of the subcategories of <i>Health</i> (Figure 2b) as a general model. You are not expected to completely work out the solution for <i>Education</i>; just listing some facets with sample terms under them would be an acceptable answer.</p> <p>► Answer C</p> |

Directory > Education

Yahoo! Education - K-12 Schools - Colleges - Online Degrees - Reference

CATEGORIES

Top Categories

Adult and Continuing Education (284)
Browse by Region (168)
By Culture or Group (317)

By Subject (1012)
Higher Education (17081) **NEW!**
K-12 (56360)

Additional Categories

Academic Competitions (99)
Bibliographies (5)
Bilingual (18)
Business to Business@
Career and Vocational (311)
Chats and Forums (22)
Conferences (35)
Correctional@
Disabilities@
Distance Learning (658)
Early Childhood Education (96)
Employment@
Equity (23)
Financial Aid (407)
Government Agencies (76)
Graduation (59)
History (17)

Instructional Technology (256)
Journals (40)
Legislation (11)
Literacy (21)
News and Media (77)
Organizations (2772)
Policy (54)
Programs (342)
Reform (87)
Shopping and Services@
Special Education (175)
Standards and Testing (140)
Statistics (6)
Teaching (75)
Theory and Methods (660)
Web Directories (4)

Figure W-2. Yahoo Education

| | |
|---|--|
| 15 min. | 3. Compare Yahoo with Dewey and Library of Congress Classification |
| Comparison of Yahoo with DDC and LCC | <p>Look over Figure W-3 on the facing page and get a sense of how different topics are treated in these classifications.</p> <p>Why is <i>literature</i> given more prominence in DDC and LCC than in Yahoo? Can you find a general principle that would explain the differences in emphasis in DDC and LCC on the one hand and the Yahoo Classification on the other?</p> <p>Note: Comparison is easier at lower levels of the hierarchy since different schemes may agree on specific narrower fields but group them differently at the top level of the hierarchy.</p> <p>► Answer D</p> |

| Dewey Decimal | Library of Congress | Yahoo |
|---|--|--|
| 000 Computers, information, & general reference 000 Computers, Internet, & systems 010 Bibliography 020 Library and information sci. 070 News media | A General works QA Math, incl. computer science Z Bibliography and library science | Computers & Internet Reference News & Media |
| 100 Philosophy & psychology | B-BJ Philosophy. Psychology. | A & H > Humanities > Philosophy |
| 200 Religion | BL-BX Religion | Society and Culture > Religion and Spirituality |
| 300 Social sciences 330 Econ, 380 Commerce 320 Pol. sci., 350 Pub. admin 340 Law 370 Education | H Social sciences HB-HJ Economics J Political science K Law L Education | Social Science Society & Culture Business & Economy Government Government > Law Education |
| 400 Language | P-PM Language | Social Science > Linguistics and Human Languages |
| 500 Natural sciences & math. | Q Science | Science (parts of) |
| 600 Technology (Applied sciences) 610 Medical sciences and medicine 630 Agriculture Most of 600 | R Medicine S Agriculture T Technology | Health Science > Agriculture Science (parts of) |
| 700 Arts & recreation | M Music and books on music N Fine arts GV Recreation, leisure | Arts & Humanities Entertainment Recreation & Sports |
| 800 Literature & rhetoric | PN-PZ Literature | Arts & Humanities > Humanities > Literature |
| 900 Geography & history | C-F History, G Geography | A & H > Humanities > History Regional |
| | U Military science V Naval science | Government > Military |

Figure W-3. Dewey, Library of Congress and Yahoo classification compared

| | |
|---|--|
| 15 min. | 4. Compare Yahoo Directory Home, <i>State</i> subdivision and <i>City</i> subdivision |
| Home, City, and State subdivisions | <p>Look over Figure W-4 on the facing page and briefly describe the differences you see between Yahoo Home and the <i>State</i> subdivision and the <i>State</i> subdivision and the <i>City</i> subdivision.</p> <p>► Answer E</p> |

| Yahoo Directory Home | Subdivision of states | Subdivision of cities |
|---|--|--|
| Regional | Massachusetts Locations | |
| | Metropolitan Areas | |
| | Counties and Regions | |
| | Cities | |
| (Yellow Pages) (Maps) | | Find Businesses Yellow Pages, Maps Driving Dir |
| | | City Guides |
| (Classifieds) | | Classifieds |
| | Real Estate | Real Estate (category, general) Local Real Estate (listings search) |
| | Employment | Employment (category, general) Local Jobs (job listings search) |
| | | Online Community |
| Reference | | |
| Computers & Internet | Computers & Internet | |
| News & Media | News & Media | News & Media |
| Entertainment | Entertainment | Entertainment & Arts |
| Arts & Humanities | Arts & Humanities | (See Entertainment & Arts below) |
| Education | Education | Education |
| Recreation & Sports | Recreation & Sports Travel & Transportation | Recreation & Sports Travel & Transportation |
| Business & Economy (Auctions) (Shopping) | Business & Economy | Business & Shopping |
| Government | | |
| Society & Culture | Community & Culture | Community |
| Social Science | Social Science | |
| Health | Health | Health |
| Science | Science | |

Figure W-4. Yahoo Directory Home compared with *State* subdivision and *City* subdivision

My arrangement (compare Figure 1b). The items in () in column 1 are not categories in Yahoo Directory Home, but links under the search box; there are many other such links. Most of the subdivisions under states and cities are accessed through a drop-down box titled State/Local Web Sites.

| | |
|----------------------------------|---|
| 30 min. | 5. Examine some principles Yahoo uses when designing subdivisions |
| Principles of subdivision | <p>Figure W-5 (starting on the facing page) gives a number of examples of category subdivisions. Where there are two groups, can you tell the difference between them?</p> <p>Write your observations on any two of the examples or state a general principle.</p> <p>► Answer F</p> |

Figure W-5. Subdivisions of selected Yahoo categories for comparison and analysis
Arranged by ease of analysis.

Figure W-5a

| | |
|---|---|
| Directory > Reference > Libraries | |
| Categories (divided into two groups) | |
| Conferences (9) Countries (27) History (6) Librarians (29) Library and Information Science (306) | Organizations (59) Professional Resources (112) Serials (6) Web Directories (19) |
| Academic Libraries (451) Archives@ Arts Libraries@ Business Libraries@ Commercial Library Services@ Dance Libraries@ Digital Libraries (74) Education Libraries@ Environmental Libraries@ Government Documents@ Health Libraries@ Intellectual Property Libraries@ Internet Filtering in Libraries@ Law Libraries@ Lesbian, Gay and Bisexual@ Libraries for the Blind@ Literary Libraries@ Literature@ Map Libraries@ Masonic Libraries@ | Military Libraries@ Music Libraries@ National Libraries (33) Native American Libraries@ Performing Arts Libraries@ Philatelic Libraries@ Physics Libraries@ Presidential Libraries@ Prison Libraries@ Public Libraries (3494) Religious Libraries@ School Libraries (35) Science Libraries@ Social Science Libraries@ Special Collections (41) Sports Libraries@ Theological Libraries@ Transportation Libraries@ U.S. State Libraries (49) |

Figure W-5. Subdivisions of selected Yahoo categories for comparison and analysis, cont.**Figure W-5b**

| | |
|--|---|
| Directory > | |
| News and Media | |
| Categories (divided into two groups) | |
| By Region (21222) Columns and Columnists (286) Commercial Services@ Content Ratings@ Industry Information (824) Internet Broadcasts (400) Journals (33) | Magazines (3835) Newspapers (8419) Photojournalism@ Radio (9418) Television (15900) Web Directories (100) |
| Arts and Humanities@ Automotive@ Business (133) College and University (844) Computers and Internet@ Crime@ Cultures and Groups (15) Disabilities@ Education@ Entertainment@ Environment and Nature@ Good News (9) Government@ Health@ History@ Home and Garden@ Humor (219) | Law@ Outdoors@ Personalized News (14) Philanthropy@ Politics@ Real Estate@ Religion@ Science@ Sexuality@ Sports@ Technology (69) Traffic and Road Conditions@ Transportation@ Travel@ Weather (1087) Weird News (20) World (72) |

Figure W-5. Subdivisions of selected Yahoo categories for comparison and analysis, cont.**Figure W-5c**

| | |
|---|-----------------------------------|
| Directory > Entertainment > Humor | |
| Categories (divided into two groups) | |
| Archives (122) | Lists (65) |
| Chats and Forums (34) | News and Media@ |
| Columns and Columnists (84) | Poetry@ |
| Comedy@ | Quizzes and Tests@ |
| Comic Strips@ | Quotations@ |
| Companies@ | Stories (21) |
| Jokes (371) | Web Directories (20) |
| Advertising (41) | Job Humor (90) |
| Advice (127) | Military (3) |
| Animals (176) | Movies and Film (62) |
| Begging (40) | Murphy's Laws (6) |
| Bitterness (38) | Music (93) |
| Bizarre (248) | Names (25) |
| Boredom (4) | Parenting (20) |
| Bubblewrap (4) | Parody (350) |
| Cars (40) | People (19) |
| Clean Humor (12) | Philosophy (18) |
| Codes (8) | Politics (16) |
| Computers and Internet (631) | Procrastination (13) |
| Cultures and Groups (177) | Rants (75) |
| Distorted Pictures (14) | Religion (216) |
| Drugs and Addictions (16) | Science (109) |
| Duct Tape (11) | Science Fiction and Fantasy (235) |
| Education (43) | Sex (130) |
| Fights (9) | Sports (11) |
| Food and Drink (139) | Stupidity (70) |
| Furniture (5) | Tasteless (370) |
| Gender Wars (37) | Traffic Cones (2) |
| Hair (27) | Useless Pages (123) |
| Health and Medicine (22) | Warning Labels (3) |
| Holidays and Observances (109) | Wedding and Marriage (13) |
| Horoscopes (19) | |

Figure W-5. Subdivisions of selected Yahoo categories for comparison and analysis, cont.

Figure W-5d

| | |
|--|--------------------------|
| Directory > Entertainment > Comics and Animation | |
| Categories (not divided into groups) | |
| Animation (1821) | Editorial Cartoons (110) |
| Artists@ | Graphic Novels (15) |
| Characters (163) | Magazines (6) |
| Chats and Forums (18) | Organizations (22) |
| Comic Books (1781) | People (6) |
| Comic Strips (1023) | Shopping and Services@ |
| Conventions (22) | Small Press Comics (15) |
| Cultures and Groups (21) | Web Directories (8) |

Figure W-5. Subdivisions of selected Yahoo categories for comparison and analysis, cont.**Figure W-5e**

| | |
|--|---|
| Directory > Government > | |
| Law | |
| Categories (divided into two groups) | |
| Alternative Dispute Resolution (38) Attorneys General@ Booksellers@ Cases (287) Continuing Legal Education (24) Countries (43) District Attorneys@ Employment Resources (48) Events (3) Firms and Services@ History (27) Journals (101) Jury Duty (15) | Law Enforcement@ Law Schools (296) Lawyer Jokes@ Legal Ethics (4) Legal Research (98) News and Media (49) Organizations (340) Self-Help (15) Software Companies@ U.S. Judiciary and Supreme Court@ U.S. States (50) Web Directories (40) |
| Administrative (9) Business (65) Constitutional (175) Consumer (25) Criminal Justice (110) Disability (29) Dog Breed Specific@ Elder (10) Employment (24) Entertainment (9) Environmental (72) Estate and Probate (10) Federal (16) Health (12) | Immigration and Naturalization (106) Indigenous Peoples@ Intellectual Property (167) International (48) Lesbian, Gay and Bisexual@ Privacy (35) Property (70) Sexuality@ Tax (54) Technology (104) Trade@ Women's Resources@ Usenet (9) |

Figure W-5. Subdivisions of selected Yahoo categories for comparison and analysis, cont.**Figure W-5f**

| | |
|--|---|
| Directory > Government > Politics Categories (not divided into groups, except for the fairly standard By Region) | |
| By Region (7647) U.S. Politics@ | |
| Activism Resources (47) Chats and Forums (33) Civic Participation@ Elections (3151) General Information (3) Humor@ News and Media (24) Organizations (78) | Parties (134) Political Consulting@ Political Issues (46) Political Opinion (218) Political Science@ Political Theory@ Regional Conflicts (33) Web Directories (8) |

| | |
|------------------------|--|
| 60 min. | 6. Overall facet analysis of the Yahoo classification |
| Facets in Yahoo | <p>Identify entity types / facets that occur throughout the Yahoo classification, preferably with some frequently occurring concepts under each. You can also mention concepts that occur as components in many places but that you cannot assign to a facet. Your listing would be the beginning of a faceted <i>core classification</i> for Yahoo (see Organizing Information, p. 299 and Section 15.6, p. 322-323).</p> <p>It would be interesting to find out how many elemental concepts are in this core classification and how many precombined categories Yahoo has in its <i>extended classification</i>..</p> <p>Answer G</p> |

Outline for the analysis of Knowledge Organization Systems

For some items, a section number from Soergel, Organizing information is given in ()

1. **Purpose**

1.1 **Information system** or type of information system in which to be used

Bibliographic information system: Organize Web pages

1.2 **Intended for** controlled vocabulary indexing , or query term expansion G (Ch. 12, Introduction)

1.3 **Type of file and search mechanism** for which originally designed

Shelving G Card catalog / printed index G Online system , (*Web subject directories: Similar to shelving, but multiple locations for each class and multiple entries for each page*)

2. **Coverage and designation of concepts. Coverage and format of terms**

2.1 **Concepts: Scope**, breadth of coverage. Recency of concepts

Universal — covers all of knowledge. But focus on Western culture, esp. US, on topics in HTML Web pages.,

2.2 **Concepts: Specificity**, depth of coverage. (Section 16.2.2). Coverage at each level of specificity.

Medium specificity. Would need closer analysis by subject area. Geographic names quite specific.

2.3 Are all needed **facets** included? Concepts formed in semantic factoring and facet analysis? (S.a. 3.1)

Answer would require extensive analysis. Many general concepts are visible as one looks at subdivisions that repeat in many places, but no list of these. Bound to English language; if there is no English term for a general concept, that concept is unlikely to show up explicitly in the Yahoo Classification.

2.4 **Nature of notation** (if none, state that). (Section 15.5.2) *None.*

2.5 **Terms:** Completeness of coverage (completeness of lead-in vocabulary). Recency of terms

Only descriptors, no lead-in terms. Many terms are quite recent (very fast update).

2.6 **Form of terms:** Consistency, adherence to common usage. *Terms seem appropriate. Most terms appear to be taken directly from generally used language, except for a few phrases like Arts and Humanities.*

3. **Terminological and conceptual analysis and conceptual structure.**

3.1 **Quality of conceptual structure** (14): Facet analysis. Types and degree of differentiation of conceptual relationships included. For each type indicate the completeness of inclusion. (Fill in 3.1.1 - 3.1.3)

3.1.1 Expression of concepts through elemental concepts (closely related to definition)

Category names are formed by stringing together terms that designate the category's conceptual components; to that extent, compound concepts are expressed through elemental concepts. However, compounds expressed through an accepted term in English, such as Ship, are not expressed through their components.

3.1.2 Hierarchical relationships (polyhierarchy) (Shown by arrangement or Broader Term / Narrower Term X-ref)

Polyhierarchical; a category has a home place but may appear in many places in the hierarchy. From any place one get to the home place by a click, but the other places where the category appears are not indicated

3.1.3 Associative relationships. (Implied by physical proximity in the arrangement or explicit Related Term X-ref)

None by the format. However, some of the hierarchical relationships should in fact be associative.

3.2 **Quality of definitions**, explications, scope notes (correctness, detail, clarity).

No definitions.

3.3 Completeness of terminological relationships: Does the vocabulary contain terms that are synonymous or quasi-synonymous without indicating the relationship?

Yahoo does not attempt to cover terminology beyond the category names.

4. **Use of precombination in the index language** (concerns both 2 and 3) (14, 15, esp. 15.4)

4.1 To what degree are descriptors precombined?

Yahoo categories are highly to very highly precombined.

4.2 To what extent are precombined descriptors enumerated and/or given in the alphabetical index?

Precombined descriptors are enumerated in the Yahoo directory. It is not known whether Yahoo indexers have schedules of just the elemental concepts to index from.

To what extent can the indexer build additional precombined descriptors?

Probably new categories built from existing components are added all the time either by indexers or by an editor based on indexer suggestions.

Are precombined descriptors designated by an independent symbol or a string of symbols? Combination order free or fixed? To what extent do the components of a precombined descriptor determine its place in the arrangement? (Relates also to 5) (Section 15.5.2)

By a string of symbols, the terms for the individual components. Exception: Words/terms like Ship that designate a compound concept. Combination order is free, with some apparent rules. The components completely determine the place of a precombined descriptor built by the indexer.

5. **Access and display. Format of presentation of the vocabulary**

Consider for each format access/retrieval by concepts versus access/retrieval by terms.

Access can be provided through arrangement in a printed document or through a computer search system.

5.1 **Format of printed document** (Fill in 5.1.1- 5.1.3) *No print version. Comments here refer to online display.*

5.1.1 Overall format: Thesaurus parts and information given in each, connections between them. Is the overall format clear and helpful for finding the appropriate concepts and terms or notations in indexing and query formulation?

A menu tree, walking down one level at a time.

5.1.2 Display of conceptual relationships (Broader Term, Narrower Term, Related Term)

- through linear arrangement or graphical display (Section 15.5.2)

In the subject directory "multi-tree", in which each descriptor (category) can appear in multiple places

- through cross-references (Section 14.1)

No cross-references, other than a category appearing in multiple places.

- through descriptor-find index (Section 15.5.1)

The search function provides a descriptor find index of sorts.

How well does the display reflect the conceptual analysis, e.g., sequence of concepts on the same hierarchical level (sequence of the children of a concept, that is, the concepts one level further down).

The sequence of categories at each level is usually strictly alphabetical, sometimes divided into groups based on meaningful criteria.

5.1.3 Display of terminological relationships (Synonymous Term)

Terminological relationships are not included..

5.2 **Access through computer system.** Navigation. Format of on-line displays

Described in 5.1.

Instructions for using the Yahoo Classification

| | |
|---------------------------------|---|
| <p>Indexing</p> | <p>Give up to three categories per document. Yahoo does assign multiple categories to a document since a subject directory does not have the one-place-only constraints imposed by the physical arrangement of documents. Still only very important categories should be assigned (high-threshold, low-exhaustivity indexing). (I do not know the rules for number of categories that is given to Yahoo indexers.) Use the most specific category in each case. Give the categories in the format generally accepted on the Web (you can omit Directory):</p> <p style="text-align: center;">Directory > Education > K-12 > Schools > Elementary Schools</p> <p>Browse through the subject directory or use search to identify the categories under which a document should be found and under which it should therefore be indexed.</p> |
| <p>Query formulation</p> | <p>Try to list all categories where one should look for relevant documents; if there are more than 10 categories, just give a representative sample (enough to demonstrate that you know how to find all categories throughout the Yahoo classification).</p> <p>Note: A broad category implies all the narrower categories; no need to list these narrower categories, they can be readily seen from the Yahoo multi-tree.</p> <p>The query formulation is the OR combination of all the categories in your list. AND combinations of categories would run counter to the subject directory approach to searching; whether this approach makes best use of computer capabilities is a different question.</p> <p>You can formulate each query in terms of elemental concepts and than use these to find categories (see Worksheet, Task 6).</p> |

| | |
|--|---|
| <p>Arts & Humanities Literature, Photography ...</p> <p>Business & Economy B2B, Finance, Shopping, Jobs ...</p> <p>Computers & Internet Internet, WWW, Software, Games ...</p> <p>Education College and University, K-12 ...</p> <p>Entertainment Cool Links, Movies, Humor, Music ...</p> <p>Government Elections, Military, Law, Taxes ...</p> <p>Health Medicine, Diseases, Drugs, Fitness ...</p> | <p>News & Media Full Coverage, Newspapers, TV...</p> <p>Recreation & Sports Sports, Travel, Autos, Outdoors ...</p> <p>Reference Libraries, Dictionaries, Quotations ...</p> <p>Regional Countries, Regions, US States ...</p> <p>Science Animals, Astronomy, Engineering ...</p> <p>Social Science Archaeology, Economics, Languages ...</p> <p>Society & Culture People, Environment, Religion ...</p> |
|--|---|

Figure 1a. **Yahoo classification. Directory Home** (first summary)

| | |
|---|--|
| <p>Reference and General Interest</p> <p>Reference Libraries, Dictionaries, Quotations ...</p> <p>Computers & Internet Internet, WWW, Software, Games ...</p> | <p>Subjects</p> <p>Science Animals, Astronomy, Engineering ...</p> <p>Health Medicine, Diseases, Drugs, Fitness ...</p> <p>Social Science Archaeology, Economics, Languages ...</p> <p>Society & Culture People, Environment, Religion ...</p> <p>Government Elections, Military, Law, Taxes ...</p> <p>Business & Economy B2B, Finance, Shopping, Jobs ...</p> |
| <p>News & Media Full Coverage, Newspapers, TV...</p> <p>Entertainment Movies, Music, Humor, Cool Links ...</p> <p>Recreation & Sports Sports, Travel, Autos, Outdoors...</p> | <p>Education College and University, K-12 ...</p> <p>Arts & Humanities Literature, Photography ...</p> |
| <p>Regional Countries, Regions, US States ...</p> | |

Figure 1b. **Yahoo Classification. Directory Home. Meaningful arrangement**

Directory > Health**Categories**

| | |
|------------------------------------|---------------------------------|
| Alternative Medicine (480) | Men's Health (30) |
| Business to Business@ | Mental Health (682) |
| Chats and Forums (52) | Midwifery (60) |
| Children's Health (153) | News and Media (201) |
| Conferences (19) | Nursing (431) |
| Death and Dying@ | Nutrition (207) |
| Dentistry@ | Organizations (21) |
| Disabilities@ | Pet Health@ |
| Diseases and Conditions (7392) | Pharmacy (1096) |
| Education (39) | Procedures and Therapies (292) |
| Emergency Services (236) | Public Health and Safety (740) |
| Employment (108) | Reference (93) |
| Environmental Health (194) | Reproductive Health (659) |
| First Aid (14) | Senior Health (79) |
| Fitness (166) | Sexuality@ |
| General Health (81) | Shopping and Services@ |
| Health Administration (65) | Teen Health (13) |
| Health Care (356) | Traditional Medicine (179) |
| Health Sciences (26) | Travel Health and Medicine (16) |
| Hospitals and Medical Centers (38) | Web Directories (50) |
| Institutes (34) | Weight Issues (77) |
| Law@ | Women's Health (153) |
| Long Term Care (116) | Workplace (6) |
| Medicine (4955) | |

Figure 2a. **Yahoo classification. Health.**

Directory > Health**Categories**

| | |
|--|---|
| <p>Reference</p> <ul style="list-style-type: none"> . Reference . Web Directories . Chats and Forums . News and Media <p>Health Sciences Fields</p> <ul style="list-style-type: none"> . Health Sciences . Medicine . Dentistry@ . Nursing . Midwifery . Pharmacy . Traditional Medicine . Alternative Medicine <p>Individual health condition</p> <ul style="list-style-type: none"> . Diseases and Conditions . Disabilities@ . Fitness . Nutrition . Weight Issues . Reproductive Health . Sexuality@ . Death and Dying@ . Mental Health <p>Procedures and Therapies</p> | <p>Health by place</p> <ul style="list-style-type: none"> . General Health . Public Health and Safety . Environmental Health . Workplace . Travel Health and Medicine <p>Health by population group</p> <ul style="list-style-type: none"> . Human Health <ul style="list-style-type: none"> . Human Health by Age <ul style="list-style-type: none"> . Children's Health . Teen Health . Senior Health . Human Health by Gender <ul style="list-style-type: none"> . Women's Health . Men's Health . Animal Health <ul style="list-style-type: none"> . Pet Health@ <p>Health Care</p> <ul style="list-style-type: none"> . Emergency Services . First Aid . Long Term Care <p>Health care organization</p> <ul style="list-style-type: none"> . Hospitals and Medical Centers . Institutes . Organizations . Conferences . Health Administration . Shopping and Services@ . Business to Business@ . Law@ . Education . Employment |
|--|---|

Figure 2b. Yahoo classification. Health. Meaningful arrangement.

Figure 3. Yahoo classification second summary**Arts & Humanities**

- SN Literature, Photography . . .
- . By Region (131)
- . Art History (742)
- . Artists (3474)
- . Arts Therapy@
- . Awards (16)
- . Booksellers@
- . Censorship (17)
- . Chats and Forums (45)
- . Companies@
- . Crafts (796)
- . Criticism and Theory (30)
- . Cultural Policy@
- . Cultures and Groups (483)
- . Design Arts (5177)
- . Education (631)
- . Employment (44)
- . Events (198)
- . Humanities (45830)
- . Institutes (38)
- . Museums, Galleries, & Cntrs (1016)
- . News and Media (260)
- . Organizations (360)
- . Performing Arts (6065)
- . Reference (35)
- . Thematic (456)
- . Visual Arts (12134)
- . Web Directories

Business & Economy

- SN B2B, Finance, Shopping, Jobs. . .
- . Business to Business (268905)
- . Shopping and Services (378101)
- . Business Libraries (23)
- . Business Schools@
- . Chats and Forums (24)
- . Classifieds (3632)
- . Consortia (42)
- . Consumer Advocacy and Information@
- . Conventions and Conferences (38)
- . Cooperatives (24)
- . Directories (347)
- . Economics@
- . Education (809)
- . Electronic Commerce (215)
- . Employment and Work (1742)
- . Ethics and Responsibility (46)
- . Finance and Investment (1804)
- . Global Economy (287)
- . History (20)
- . Intellectual Property@
- . Labor (725)
- . Law@
- . Magazines (130)
- . Management Science (203)
- . Marketing and Advertising (350)
- . News and Media@
- . Organizations (11880)
- . Quality Standards@
- . Real Estate (391)
- . Small Business Information (299)
- . Statistics and Indicators (5)
- . Taxes@
- . Television@
- . Trade (422)
- . Transportation (2094)

Computers & Internet

- SN Internet, WWW, Softw., Games . . .
- . Business to Business@
- . Shopping and Services@
- . Art@
- . Bibliographies (6)
- . Communications & Netw. (1128)
- . Computer Science@
- . Contests (26)
- . Conventions and Conferences@
- . Countries, Cultures, & Groups (38)
- . Cyberculture@
- . Data Formats (380)
- . Desktop Customization@
- . Desktop Publishing (53)
- . Dictionaries (32)
- . Employment@
- . Ethics (18)
- . Games@
- . Graphics (308)
- . Hardware (2292)
- . History (85)
- . Humor@
- . Industry Information@
- . Internet (5999)
- . Magazines@
- . Mobile Computing (64)
- . Multimedia (673)
- . Music@
- . News and Media (203)
- . Operating Systems@
- . Organizations (93)
- . People (119)
- . Personal Computers@
- . Product Reviews (2974)
- . Programming Languages (1515)
- . Science and Technology Policy@
- . Security and Encryption (589)
- . Semiconductors@
- . Software (6133)
- . Standards (45)
- . Supercomputing and Parallel Computing@
- . Technical Guides and Support (45)
- . Telecommunications@
- . Training@
- . User Groups@
- . Web Directories (14)
- . World Wide Web@
- . Year 2000 Problem (248)
- . Cnet
- . ZDNet

Education

- SN College and University, K-12
- ...
- . Browse by Region (170)
- . By Culture or Group (404)
- . By Subject (12)
- . Academic Competitions (80)
- . Adult and Continuing Education (330)
- . Bibliographies (4)
- . Bilingual (23)
- . Career and Vocational (234)
- . Chats and Forums (40)
- . Companies@
- . Conferences (50)
- . Correctional@
- . Disabilities@
- . Distance Learning (491)
- . Early Childhood Education (92)
- . Employment (146)
- . Equity (27)
- . Financial Aid (396)
- . Government Agencies (78)
- . Graduation (52)
- . Higher Education (16594)
- . Instructional Technology (341)
- . Journals (38)
- . K-12 (54618)
- . Literacy (12)
- . News and Media (83)
- . Organizations (3094)
- . Policy (52)
- . Programs (335)
- . Reform (73)
- . Special Education (172)
- . Standards and Testing (59)
- . Statistics (6)
- . Teaching (63)
- . Theory and Methods (672)
- . Web Directories (47)

Entertainment

- SN Cool Links, Movies, Humor, Music ...
- . Actors and Actresses (11355)
- . Amusement & Theme Parks (396)
- . Books and Literature@
- . Chats and Forums (92)
- . Comedy (794)
- . Comics and Animation (4979)
- . Consumer Electronics (588)
- . Contests, Surveys, and Polls (424)
- . Cool Links (1837)
- . Employment (366)
- . Entertainment and Media Production@
- . Events (214)
- . Food and Drink@
- . Gambling@
- . Games@
- . Genres (1363)
- . History (15)
- . Humor (4857)
- . Magic (296)
- . Movies and Film (19920)
- . Music (82025)
- . News and Media (394)
- . Organizations (11)
- . Performing Arts@
- . Radio@
- . Randomized Things (76)
- . Reviews (39)
- . Shopping and Services@
- . Television@
- . Trivia (109)
- . Virtual Cards (1019)
- . Web Directories (36)
- . Webisodes (94)
- . X of the Day, Week, etc. (171)

Government

- SN Elections, Military, Law, Tax
- . Countries (147)
- . Chats and Forums (7)
- . Civic Participation (27)
- . Conventions and Conferences (17)
- . Documents (26)
- . Embassies and Consulates (99)
- . Ethics (14)
- . Intelligence (125)
- . International Organizations (531)
- . Law (2671)
- . Military (867)
- . National Symbols and Songs (50)
- . News and Media (11)
- . Politics (11446)
- . Public and Civil Service (8)
- . Research Labs (26)
- . Statistics (40)
- . Student Government@
- . Taxes (53)
- . US Government (11649)
- . Web Directories (14)

Health

- SN Medicine, Diseases, Drugs, Fitness ...
- . Alternative Medicine (515)
- . Business to Business@
- . Chats and Forums (56)
- . Children's Health (186)
- . Conferences (18)
- . Death and Dying@
- . Dentistry@
- . Disabilities@
- . Diseases and Conditions (8505)
- . Education (62)
- . Emergency Services (536)
- . Employment (118)
- . Environmental Health (200)
- . First Aid (12)
- . Fitness (201)
- . General Health (90)
- . Health Administration (66)
- . Health Care (347)
- . Health Sciences (27)
- . Hospitals and Medical Centers (44)
- . Institutes (34)
- . Law@
- . Long Term Care (109)
- . Medicine (5071)
- . Men's Health (37)
- . Mental Health (753)
- . Midwifery (56)
- . News and Media (199)
- . Nursing (459)
- . Nutrition (218)
- . Organizations (21)
- . Pet Health@
- . Pharmacy (1231)
- . Procedures and Therapies (458)
- . Public Health and Safety (2207)
- . Reference (98)
- . Reproductive Health (713)
- . Senior Health (85)
- . Sexuality@
- . Shopping and Services@
- . Teen Health (23)
- . Traditional Medicine (197)
- . Travel Health and Medicine (24)
- . Web Directories (54)
- . Weight Issues (90)
- . Women's Health (174)
- . Workplace (67)

News & Media

- SN Full Coverage, Newspapers, TV . . .
- . By Region (21280)
- . Columns and Columnists (276)
- . Commercial Services@
- . Content Ratings@
- . Industry Information (821)
- . Internet Broadcasts (396)
- . Journals (33)
- . Magazines (3845)
- . Newspapers (8447)
- . Photojournalism@
- . Radio (9428)
- . Television (15866)
- . Web Directories (99)
- . Arts and Humanities@
- . Automotive@
- . Business (133)
- . College and University (1003)
- . Computers and Internet@
- . Crime@
- . Cultures and Groups (15)
- . Disabilities@
- . Education@
- . Entertainment@
- . Environment and Nature@
- . Good News (9)
- . Government@
- . Health@
- . History@
- . Home and Garden@
- . Humor (218)
- . Law@
- . Outdoors@
- . Personalized News (14)
- . Philanthropy@
- . Politics@
- . Real Estate@
- . Religion@
- . Science@
- . Sexuality@
- . Sports@
- . Technology (69)
- . Traffic and Road Conditions@
- . Transportation@
- . Travel@
- . Weather (1088)
- . Weird News (20)
- . World (71)
- . ABC News
- . BBC News
- . CNN
- . Fox News
- . MSNBC

Recreation & Sports

- SN Sports, Travel, Autos, Outdoors . . .
- . Amusement and Theme Parks@
- . Automotive (5657)
- . Aviation (810)
- . Booksellers@
- . Chats and Forums (7)
- . Cooking@
- . Dance@
- . Employment (7)
- . Events (10)
- . Fitness@
- . Gambling (315)
- . Games (17999)
- . Hobbies (3044)
- . Home and Garden (760)
- . Magazines (63)
- . Motorcycles@
- . Outdoors (9884)
- . Pets@
- . Sports (47097)
- . Television@
- . Toys (948)
- . Travel (112681)

Reference

- SN Libraries, Dictionaries, Quotations . . .
- . Acronyms and Abbreviations (25)
- . Almanacs (13)
- . Arts and Humanities@
- . Ask an Expert (556)
- . Bibliographies (7)
- . Booksellers@
- . Calendars (81)
- . Codes (24)
- . Dictionaries (153)
- . Directories (3)
- . Encyclopedia (23)
- . English Language Usage@
- . Environment and Nature@
- . Etiquette@
- . FAQs (18)
- . Finance and Investment@
- . Flags (26)
- . General (13)
- . Geographic Name Servers@
- . Health@
- . Journals@
- . Libraries (4803)
- . Maps@
- . Measurements and Units@
- . Music@
- . Parliamentary Procedure (13)
- . Patents@
- . Phone Numbers and Addresses (166)
- . Postal Information (21)
- . Quotations (235)
- . Research Papers@
- . Searching the Net@
- . Standards (75)
- . Statistics (27)
- . Thesauri (22)
- . Time@
- . Web Directories (16)
- . World Population Counts@

Regional

- SN Countries, Regions, US States . . .
- . U. S. States
- . Countries
- . Regions (8344)
- . Geography@
- . Web Directories (28)

Science

- SN Animals, Astronomy, Engineering. . . .
- . Acoustics (66)
- . Agriculture (2054)
- . Alternative (1047)
- . Amateur Science (18)
- . Animals, Insects, and Pets@
- . Anthropology and Archaeology@
- . Artificial Life (129)
- . Ask an Expert (21)
- . Astronomy (2519)
- . Aviation and Aeronautics (236)
- . Bibliographies (6)
- . Biology (16535)
- . Booksellers@
- . Chats and Forums (45)
- . Chemistry (1275)
- . Cognitive Science (94)
- . Complex Systems (23)
- . Computer Science (1516)
- . Dictionaries (27)
- . Earth Sciences (2831)
- . Ecology (746)
- . Education (549)
- . Employment (44)
- . Energy (554)
- . Engineering (4659)
- . Events (35)
- . Forensics (59)
- . Geography (3410)
- . Geology and Geophysics@
- . History (83)
- . Humor@
- . Hydrology@
- . Information Technology (71)
- . Institutes (58)
- . Journals (31)
- . Libraries (33)
- . Life Sciences (17)
- . Mathematics (1934)
- . Measurements and Units (219)
- . Medicine@
- . Meteorology@
- . Museums and Exhibits (146)
- . Nanotechnology (51)
- . News and Media (140)
- . Oceanography@
- . Organizations (160)
- . Paleontology@
- . People (52)
- . Physics (1647)
- . Psychology@
- . Religion and Science@
- . Research (161)
- . Science and Technology Policy
- . Science on Postage Stamps (8)
- . Space (1292)
- . Sports@
- . Web Directories (43)

Social Science

- SN Archaeology, Economics, Languages . . .
- . Anthropology and Archaeology (1145)
- . Area Studies (698)
- . Bibliographies (13)
- . Books@
- . Chats and Forums (12)
- . Communications (1700)
- . Conferences (17)
- . Critical Theory@
- . Disability Studies (8)
- . Economics (1113)
- . Education (25)
- . Employment (4)
- . Environmental Studies@
- . Ethnic Studies (173)
- . Futures Studies (25)
- . Gender Studies (21)
- . Genealogy@
- . Geography@
- . Gerontology (36)
- . History@
- . Humanities@
- . Institutes (83)
- . Journals (29)
- . Law@
- . Lesbian, Gay, & Bisexual Studies (62)
- . Libraries (13)
- . Library and Information Science@
- . Linguistics & Human Languages (2814)
- . Migration and Ethnic Relations (37)
- . Organizations (29)
- . Peace and Conflict Studies (129)
- . Political Science (1125)
- . Popular Culture Studies@
- . Psychology (1346)
- . Recreation and Leisure Studies (82)
- . Rural Development (36)
- . Science, Technology, and Society Studies (100)
- . Sexology (33)
- . Social Research (62)
- . Social Work (158)
- . Sociology (421)
- . Urban Studies (324)
- . Web Directories (13)
- . Women's Studies (178)

Society & Culture

- SN People, Environment, Religion . . .
- . Advice (108)
- . Bibliographies (11)
- . Chats and Forums (41)
- . Crime (4537)
- . Cultural Policy (11)
- . Cultures and Groups (13258)
- . Death and Dying (477)
- . Disabilities (1482)
- . Environment and Nature (6972)
- . Etiquette (36)
- . Events (35)
- . Families (878)
- . Fashion@
- . Firearms (155)
- . Food and Drink (5147)
- . Gender (35)
- . Holidays and Observances (2062)
- . Issues and Causes (3921)
- . Journals (4)
- . Magazines (219)
- . Museums and Exhibits (5366)
- . Mythology and Folklore (976)
- . People (46774)
- . Pets@
- . Relationships (447)
- . Religion and Spirituality (36093)
- . Reunions (370)
- . Sexuality (1470)
- . Social Organizations (440)
- . Web Directories (10)
- . Weddings (283)

Figure 4

Excerpts from the Yahoo Classification

These excerpts from the Yahoo classification are designed to illustrate the structure of the classification, particularly its treatment of *Education* and *Transportation*.

This list is by no means complete with respect to Yahoo classes that deal with *Education* or *Transportation*. Only the areas labeled “(complete)” list all the subdivisions or at least enough subdivisions to give a sense of the overall structure and content of the area. In other areas the subdivisions have been selected to *Education* or *Transportation* or otherwise make a point about the structure. The list does illustrate patterns that are repeated throughout the classification.

Note: Some Yahoo classes do not have the number of associated Web pages. I sometimes used a simplified method for copying from the Yahoo displays that did not carry the number with it. Numbers were also deleted if they ran into a second line.

Symbols used

- Category falls conceptually under *Education* (this may or may not be shown in Yahoo).
- ▶ Category falls conceptually under *Transportation* (this may or may not be shown in Yahoo).

SN Scope Note. If SN is bolded, you should read it because it illustrates a general point.

At the top of each left page the hierarchical context is given by repeating the preceding levels of the hierarchy in italics, e.g. *Education*. Across a page spread, you can always find the hierarchical context.

The top level categories are **large and bold**, the categories on level 1 are **bold**. Sometimes important subdivisions further down are bolded to make the structure easier to see.

Arts & Humanities

- SN Literature, Photography ...
- . **Design Arts** (5187)
 - . . Architecture (1497)
 - . . . Buildings and Structures (474)
 - ▶Bridges@
 - •Education (158)
 - . . •Education (73)
- . **•Education**
 - . . Art History@
 - . . Art Schools (132)
 - . . College and University (252)
 - . . Companies@
 - . . Courses (12)
 - . . Design Arts@
 - . . Humanities@
 - . . K-12 (90)
 - . . . Curriculum Standards (28)
 - . . . Drama@
 - . . . Lesson Plans (17)
 - . . . Schools (35)
 - . . . Usenet (2)
 - . . Non-Degree Programs (47)
 - . . Organizations (32)
 - . . Performing Arts@
 - . . Teaching (6)
 - . . Workshops
- . **Humanities**
 - . . History
 - . . . •Education
 - Art History@
 - College and University (217)
 - Courses (2)
 - K-12 (71)
 - Courses (16)
 - Curriculum Standards (5)
 - Fairs and Competitions (3)
 - Social Studies@
 - Teacher Resources (26)
 - Organizations (4)
 - U.S. History
 - By Time Period
 - 19th Century
 - ▶Transcontinental Railroad
 - . . Literature (16329)
 - . . . Authors (9971)
 - ▶Travel Writers (86)
 - . . . Genres (2194)
 - Nonfiction (130)
 - ▶Travel Writing (16)
 - . **Performing Arts**
 - . . Dance
 - . . . •Education
 - College and Univ. Depts (24)
 - K-12 Curriculum Standards
 - Schools (138)
 - Summer Programs & Festivals

Business & Economy

- SN B2B, Finance, Shopping, Jobs
- . **Business to Business** (268905)
 - . . Construction (36758)
 - . . . General Contractors (448)
 - Commercial (344)
 - ▶Railroads (15)
 - ▶Automotive (2053)
 - •Training and Development (32)
 - Professional Driving Schools
 - . . •Education (complete)
 - . **By Subject** (909)
 - Adult Basic Education (8)
 - Art@
 - Business@
 - Computers & Technology (38)
 - Environment and Nature (11)
 - Health and Fitness@
 - Languages (242)
 - Directories (1)
 - Professional Development (2)
 - Teaching & Learning Aids
 - Booksellers@
 - CDs, Records, & Tapes
 - Curriculum (3)
 - Publishers@
 - Software (35)
 - Specific Languages (163)
 - Amharic (1)
 - Arabic (4)
 - Bahasa Indonesia (1)
 - Chinese (11)
 - English 2nd Language
 - Farsi (1)
 - French (15)
 - German (4)
 - Greek (4)
 - Hawaiian (1)
 - Italian (2)
 - Japanese (13)
 - Korean (3)
 - Navajo (1)
 - Russian (7)
 - Scots English (1)
 - Sign Language (12)
 - Spanish (16)
 - Swedish (1)
 - Tagalog (2)
 - Thai (2)
 - Vietnamese (1)
 - Welsh (1)
 - Videos (7)
 - Life Skills (61)
 - Mathematics (133)
 - Media Studies (2)
 - Music (11)

- Reading and Writing (176)
- Religion@
- Science (166)
- Social Studies (39)
- Special Education (6)
- Vocational Arts (12)
- Academic Competitions (8)
- Administration (198)
- Computers and Technology (81)
- Consulting (62)
- Directories (5)
- Financial Services (14)
- Fund Raising@
- International (9)
- Marketing and Recruiting (14)
- Presenters (13)
- Professional Development (55)
- School Management (42)
- Supplies and Equipment (83)
- . . . **Teaching & Learning Aids** (493)
 - By Subject (49)
 - Adult Basic Education@
 - Adult Literacy@
 - Curriculum (4)
 - Software (3)
 - Software (4)
 - Adult Literacy@
 - Art (8)
 - Curriculum (3)
 - Software (2)
 - Videos (2)
 - Business (6)
 - Computers & Technology@
 - Curriculum (2)
 - Logo Programming
 - Language@
 - Software (4)
 - Videos (2)
 - Titles (1)
 - Environment and Nature@
 - Health and Fitness (12)
 - Curriculum (6)
 - Sex Education (2)
 - Languages@
 - Booksellers@
 - Arabic (6)
 - Chinese (5)
 - French (5)
 - German (5)
 - Greek (4)
 - Japanese (8)
 - Titles (1)
 - Polish (4)
 - Portuguese (7)
 - Publishers@
 - Russian (7)
 - Spanish (17)

| | | |
|-----------------------------------|----------------------------------|----------------------------------|
| Antique, Rare, & Used Titles (4) | Sign Language (12) | Adult Literacy@ |
| Tamil (3) | Software (2) | Handwriting (2) |
| Titles (7) | Videos (8) | Speed Reading@ |
| Dictionaries@ | Titles (4) | Spelling@ |
| Japanese@ | Spanish (15) | Titles (7) |
| Spanish@ | Booksellers@ | Speed Reading@ |
| CDs, Records, & Tapes | Software (6) | Software (4) |
| English as a second Language @ | Swedish (1) | Spelling (7) |
| Curriculum (3) | Tagalog (2) | Games (2) |
| Publishers@ | Thai (2) | Software (3) |
| English as a second Language (5) | Vietnamese (1) | Videos (4) |
| Software (34) | Welsh (1) | Vocabulary (11) |
| Chinese@ | Videos (7) | Religion (10) |
| English as a second Language@ | English as a second Language@ | Christian Home Schooling@ |
| Japanese@ | Sign Language@ | Science@ |
| Resellers (5) | Titles (1) | CDs, Records, & Tapes |
| Sign Language@ | Life Skills@ | Curriculum (12) |
| Spanish@ | Character Education (11) | Environment & Nature@ |
| Titles (2) | Curriculum (3) | Games (1) |
| Specific Languages (157) | Curriculum (13) | Lab Equipment&Supplies (23) |
| Amharic (1) | Character Education@ | Manipulatives (6) |
| Arabic (4) | Software (3) | Online Subscription Services (3) |
| Bahasa Indonesia (1) | Videos (8) | Posters and Charts (3) |
| Chinese (11) | Titles (2) | Software (43) |
| Booksellers@ | Mathematics@ | Chemistry (4) |
| Software (7) | Books@ | Physics (6) |
| English as a second Language (57) | Algebra (3) | Titles (4) |
| CDs, Records,&Tapes (5) | Calculus (4) | Videos (8) |
| Flash Cards (2) | Chaos (3) | Social Studies@ |
| Magazines@ | Statistics (1) | Curriculum (8) |
| Online Subscription Services (2) | Curriculum (5) | Online Subscription Serv. |
| Publishers@ | Games (2) | Software (5) |
| Software (17) | Manipulatives (11) | Videos (4) |
| TOEFL Preparation@ | Online Subscription Services (1) | Titles (1) |
| Videos (4) | Posters and Charts (1) | Special Education@ |
| Farsi (1) | Software (77) | Publishers@ |
| French (14) | Testing & Assessment Software@ | Vocational Arts@ |
| Booksellers@ | Titles (2) | Software (4) |
| German (4) | Titles (16) | Academic Competitions (3) |
| Greek (4) | Videos (6) | Books@ |
| Hawaiian (1) | Titles (2) | Retail@ |
| Italian (2) | Media Studies@ | Publishers@ |
| Japanese (13) | Music@ | Academic@ |
| Booksellers@ | Reading and Writing@ | College & University Presses@ |
| Titles (1) | Adult Literacy (9) | Organizations (1) |
| Software (9) | Curriculum (4) | Textbooks@ |
| Korean (2) | Software (3) | Booksellers@ |
| Navajo (1) | Books@ | Home Schooling Materials (21) |
| Russian (6) | Curriculum (24) | Language Education@ |
| Booksellers@ | Adult Literacy@ | |
| Scots English (1) | Flash Cards (5) | |
| | Games (4) | |
| | Manipulatives (3) | |
| | Software (49) | |

Business & Economy**. Business to Business****. . . •Education****. . . Teaching & Learning Aids (493)**

. Special Education Books

. Supplementary Materials

. Forensics & Debate

. Teacher Aids (12)

. Textbooks (88)

. Higher Education (53)

. K-12 (17)

. Vocational Training

. Books (4)

. Textbooks (43)

. College & University

. Bookstores@

. Textbooks@

. Science@

. Used (18)

. Titles (95)

. Administration (2)

. College & University

. Admissions (10)

. Athletic Recruiting@

. Financial Aid (6)

. University Life (4)

. Distance Learning (3)

. Educating Your Child@

. Home Schooling (8)

. Reform (3)

. School Directories (4)

. Study Guides (7)

. Teaching (20)

. Reading (8)

. Test Preparation (12)

. Textbooks (8)

. Titles@

. Christian Home Schooling@

. Curriculum (38)

. By Subject (9)

. Christian Home Schooling@

. Preschool (9)

. School Reform (2)

. Textbooks@

. Learning Skills (21)

. Memory Improvement (10)

. Software (2)

. Study Skills (2)

. Books@

. Publishers@

. Resellers (23)

. Reviews (2)

. Supplementary Materials

. CDs, Records, & Tapes (7)

. Languages@

. English 2nd Language@

. Science@

. Distance Learning (35)

. Online Subscription Serv.

. Email (2)

. Art@

. English 2nd Language@

. Mathematics@

. Science@

. Social Studies@

. Flash Cards (10)

. English as a second

. Language@

. Reading and Writing@

. Software (5)

. Games (7)

. Mathematics@

. Reading and Writing@

. Science@

. Software@

. Manipulatives (5)

. Mathematics@

. Reading and Writing@

. Science@

. Montessori (10)

. Newspapers & Magazines

. Posters and Charts (9)

. Mathematics@

. Science@

. Publishers@

. Software (149)

. By Subject (10)

. Adult Basic Education@

. Art@

. Computers & Technology@

. Languages@

. Life Skills@

. Mathematics@

. Reading and Writing@

. Science@

. Social Studies@

. Vocational Arts@

. Flash Cards@

. Games (1)

. Resellers (18)

. Shareware (2)

. Software Reviews@

. Teaching Tools (15)

. Admin. Functions@

. Classroom Organizers

. Consulting (1)

. Curriculum Planning

. Discipline Tracking

. Financial (5)

. Grading & Evaluation

. Resellers (1)

. Titles (3)

. Home Schooling (5)

. Info. Mangmnt Systems

. School-to-Wrk Progr.

. Student Information

. Inventory Management

. Registration (1)

. Resellers (3)

. Scheduling (8)

. Special Education

. Testing & Assessment

. Classroom Computer

. Control (4)

. Test Preparation@

. College Entrance@

. Titles (4)

. Mathematics@

. Reading and Writing@

. Videos (44)

. By Subject (13)

. Art@

. Computers & Technol.@

. Languages@

. Life Skills@

. Mathematics@

. Reading and Writing@

. Science@

. Social Studies@

. Titles (5)

. Computers & Tech..@

. Languages@

. Sign Language@

. Life Skills@

. Mathematics@

. Social Studies@

. Test Preparation@

. Titles (2)

. By Subject@

. Test Preparation@

. Teaching Methods (1)

. Testing and Assessment (27)

. **Emergency Services (803)**

. ▶Air Ambulance Services (34)

. Supplies and Equipment (159)

. ▶Vehicles (20)

. Ambulances (12)

. Fire Trucks@

. Rescue Watercraft@

. . . Labor

. . . Unions

. . . ▶Railroad@

. . . ▶**Transportation (3755)**

. . . Aerospace@

. . . Automotive@

. . . Aviation (1627)

. . . Aerial Advertising@

. . . Aerospace@

. . . Agricultural@

. . . Air Cargo Services@

. . . Air Shows (8)

- Aircraft (900)
- Airlines@
- United Airlines
- •Education
- Airports (110)
- Aviation Weather@
- Avionics@
- Aviophobia@
- Books@
- Classifieds@
- Consulting (71)
- Directories (28)
- In-Flight Entertainment (3)
- Insurance@
- Navigation Systems@
- Organizations (80)
- Software (76)
- Supplies and Equipment (107)
- Trade Magazines (19)
- Training (189)
- Videos@
- . . . Buses (74)
- . . . Consulting (73)
- . . . Directories (10)
- . . . Fleet Services (92)
- . . . Intelligent Transportation Syst.
- . . . Limousines and Shuttles (23)
- . . . Maritime (307)
- Boat Transport (25)
- Boating@
- Cargo Services@
- Construction@
- Consulting (5)
- Conventions &Trade Shows (6)
- Directories (10)
- Insurance@
- Navigation@
- Organizations (17)
- Parts and Accessories (39)
- Port Authorities@
- Publications (4)
- Shipbuilding (94)
- Ships (33)
- Software (7)
- Submarines (9)
- Trade Magazines (2)
- . . . Movers@
- . . . Organizations@
- . . . Road Maintenance (30)
- . . . Software (64)
- . . . Trade Magazines (10)
- . . . Traffic Control (165)
- . . . Trains and Railroads (126)
- . . . Trolleys (8)
- . . . Trucks
- Trucking (594)
- •Driving Schools (35)
- . **Shopping and Services** (378101)
- . . ▶Automotive (24751)
- . . . •Driving Schools (524)
- By Region (491)
- U.S. States (256)
- Massachusetts (5)
- SN Boston Metro not here
- New York (42)
- Cities (14)
- New York@ **see *****
- . . . Motorcycles (1054)
- •Driving Schools (5)
- . . . Publishers (3285)
- . . . •Education (277)
- . . . Languages (35)
- •Language Education (17)
- English as a second Language
- . **Education**
- . . Business Schools@
- . . College and University (775)
- . . K-12 (26)
- . . Curriculum Standards (4)
- . . Organizations (22)
- . . Organizations (6)
- . ▶**Transportation** (2094) (complete)
- . . SN No subcategory Education
- . . Auto-Free Transportation (23)
- . . Bicycle Advocacy@
- . . Mass Transit@
- . . Organizations (8)
- . . Pedestrian Advocacy & Safety@
- . . Aviation (513)
- . . Accidents (51)
- . . . SN LCC has X-ref from Health
- . . . Grief Support@
- . . . Specific Crashes (37)
- . . . Aerospace Companies@
- . . . Aviation@
- . . . Consulting (45)
- . . . Development (6)
- . . . Electronic Warfare@
- . . . Engines (16)
- . . . Equipment (214)
- . . . Industry Information (28)
- . . . Research and Design (32)
- . . . Simulations (15)
- . . . Software (25)
- . . . Spacecraft
- . . . Air Traffic Control (14)
- . . . Air Travel@
- . . . Aircraft (119)
- . . . Art@
- . . . Aviation and Aeronautics@
- . . . Aviation & Aerospace Medcin@
- . . . Aviation Weather@
- . . . Aviators (55)
- . . . Aviophobia@
- . . . Classifieds@
- . . . Companies@
- . . . History (149)
- . . . Military@
- . . . Museums@
- . . . Pictures (46)
- . . . Recreational Aviation@
- . . . Safety (32)
- . . . Women
- . . Buses (26)
- . . Canals (22)
- . . Commuting (30)
- . . Companies@
- . . Employment (4)
- . . Events (2)
- . . Freight (8)
- . . Government Agencies (62)
- . . Highways and Roads (127)
- . . . Automated Highway Systems@
- . . . Bridges (52)
- . . . Government Agencies (9)
- . . . Organizations (13)
- . . . Regional Information (38)
- . . . Road Maintenance@
- . . . Roman Roads@
- . . . Scenic Highways and Byways@
- . . . Traffic and Road Conditions@
- . . . Traff. Signs, Signals, & Lamps@
- . . History (5)
- . . Institutes (44)
- . . Intelligent Transportation Systems
- . . Libraries (7)
- . . Limousines and Shuttles@
- . . Maritime (237)
- . . Boats@
- . . Companies@
- . . Cruise Lines@
- . . Employment (8)
- . . Maritime History@
- . . Merchant Marines (6)
- . . Passenger Ferries (51)
- . . Port Authorities (128)
- . . Ships (28)
- . . Submarines@
- . . Mass Transit (58)
- . . Museums (17)
- . . News and Media (247)
- . . . Traffic & Road Conditions (234)
- Traffic Cams (15)
- . . Organizations (67)
- . . Port Authorities@
- . . Safety (14)
- . . Statistics (7)

Business & Economy

- . ▶**Transportation**
- . . Streetcars, Trolleys, & Trams (45)
- . . Taxis (17)
- . . Traffic and Road Conditions@
- . . Trains and Railroads (390)
- . . . Companies@
- . . . High Speed Rail (11)
- . . . History (81)
- . . . Magazines (10)
- . . . Model Trains and Railroads@
- . . . Monorails (9)
- . . . Museums (64)
- . . . Organizations (27)
- . . . Railbiking@
- . . . Rails-to-Trails (27)
- . . . Railway Enthusiasts (55)
- . . . Railwayana (5)
- . . . Safety (17)
- . . . Steam Locomotives (8)
- . . . Subways (5)
- . . . Travel@
- . . . Web Directories (10)
- . . . Usenet
- . . Transportation Engineering@
- . . Trucking (46)
- . . Tunnels (32)
- . . Web Directories (3)
- . . Usenet (4)

Computers & Internet

- SN Internet, WWW, Software, Games ...
- . **Internet** (5999)
- . . Devices Connected to the Internet (1316)
- . . . Web Cams (1233)
- Outdoor Cams (233)
- ▶Traffic Cams@
- . . •Online Teaching and Learning@
- . **Multimedia** (675)
- . . Audio (200)
- . . . Formats (133)
- MP3 (118)
- Players@
- ▶Automotive (9)
- . **Software**
- . . Reviews
- . . . •Education
- . . . Scientific
- Math
- •Educational

•Education (complete, except for regional subdivisions)

- SN College and University, K-12
- SN No subcategory Transportation either directly or indirectly
- . **By Region** (170)
- . . Regions (6)
- . . . Africa@
- . . . Asia@
- Browse By Country (38)
- Armenia@
- Azerbaijan@
- Bahrain@
- Bangladesh@
- Higher Education (10)
- Organizations (3)
- Primary and Secondary (3)
- Cities@
- Complete List@
- Brunei@
- Cambodia@
- China@
- India@
- Indonesia@
- Iran@
- Government Agencies (1)
- Higher Education (19)
- Colleges & Universities
- Cities@
- Isfahan@
- Shiraz@
- Tehran@
- Complete List@
- Teaching (1)
- Iraq@
- Israel@
- Japan@
- Jordan@
- Kazakstan@
- Korea, South@
- Krygyzstan@
- Kuwait@
- Lebanon@
- Macau@
- Malaysia@
- Mongolia@
- Myanmar@
- Nepal@
- Oman@
- Pakistan@
- Philippines@
- Qatar@
- Russia@
- Saudi Arabia@
- Singapore@
- Sri Lanka@
- Taiwan@

- Thailand@
- Turkey@
- United Arab Emirates@
- Vietnam@
- Yemen@
- Companies@
- Conferences (1)
- Higher Education (1)
- Organizations (3)
- Primary and Secondary (3)
- Complete List@
- Guidance Counseling
- Student Resources
- . . . Europe@
- . . . Latin America@
- . . . Oceania@
- . . . Pacific Rim@
- . . Countries (113)
- . . . Andorra@
- . . . Argentina@
- Distance Learning (1)
- Higher Education (28)
- Colleges and Universities(29)
- Cities@
- Buenos Aires@
- Cordoba@
- Lujan@
- Moron@
- Rosario@
- Salta@
- Provinces and Regions@
- Buenos Aires@
- Cordoba@
- Lujan@
- Moron@
- Rosario@
- Salta@
- Complete List@
- Organizations (1)
- Primary and Secondary (20)
- Cities@
- Buenos Aires@
- Cordoba@
- Provinces and Regions@
- Buenos Aires@
- Cordoba@
- Complete List@
- Schools (1)
- Programs (1)
- Spanish Language Schools@
- . . . Armenia@
- . . . Australia@
- . . . Austria@
- . . . Bahrain@
- . . . Bangladesh@
- . . . Belarus@
- . . . Belgium@

- . . . Belize@
- . . . Bolivia@
- . . . Bosnia and Herzegovina@
- . . . Brazil@
- . . . Brunei@
- . . . Bulgaria@
- . . . Canada@
- . . . Chile@
- . . . China@
- . . . Colombia@
- . . . Congo, Democr. Republic of@
- . . . Costa Rica@
- . . . Croatia@
- . . . Cuba@
- . . . Cyprus@
- . . . Czech Republic@
- . . . Côte d'Ivoire@
- . . . Denmark@
- . . . Dominican Republic@
- . . . Ecuador@
- . . . Egypt@
- . . . El Salvador@
- . . . Estonia@
- . . . Federal Republic of
Yugoslavia@
- . . . Fiji@
- . . . Finland@
- . . . France@
- . . . Germany@
- . . . Ghana@
- . . . Greece@
- . . . Grenada@
- . . . Guatemala@
- . . . Haiti@
- . . . Honduras@
- . . . Hungary@
- . . . Iceland@
- . . . India@
- . . . Indonesia@
- . . . Iran@
- . . . Iraq@
- . . . Ireland@
- . . . Israel@
- . . . Italy@
- . . . Jamaica@
- . . . Japan@
- . . . Jordan@
- . . . Kenya@
- . . . Korea, South@
- . . . Kuwait@
- . . . Latvia@
- . . . Lebanon@
- . . . Liechtenstein@
- . . . Lithuania@
- . . . Luxembourg@
- . . . Macedonia, Former Yugoslav @
- . . . Malaysia@

- . . . Malta@
- . . . Marshall Islands@
- . . . Mexico@
- . . . Micronesia, Fed. States of@
- . . . Monaco@
- . . . Morocco@
- . . . Myanmar@
- . . . Nepal@
- . . . Netherlands@
- . . . New Zealand@
- . . . Nicaragua@
- . . . Norway@
- . . . Pakistan@
- . . . Panama@
- . . . Papua New Guinea@
- . . . Paraguay@
- . . . Peru@
- . . . Philippines@
- . . . Poland@
- . . . Portugal@
- . . . Romania@
- . . . Russia@
- . . . Saint Vincent and The
Grenadines@
- . . . Saudi Arabia@
- . . . Singapore@
- . . . Slovakia@
- . . . Slovenia@
- . . . South Africa@
- . . . Spain@
- . . . Sri Lanka@
- . . . Sudan@
- . . . Sweden@
- . . . Switzerland@
- . . . Taiwan@
- . . . Thailand@
- . . . Tonga@
- . . . Trinidad and Tobago@
- . . . Tunisia@
- . . . Turkey@
- . . . Uganda@
- . . . Ukraine@
- . . . United Arab Emirates@
- . . . United Kingdom@
- . . . Uruguay@
- . . . Venezuela@
- . . . Vietnam@
- . . . Zambia@
- . . . Zimbabwe@
- . . . U.S. States (51)
- . **By Culture or Group** (396)
- . . African American (17)
- . . African American Studies@
- . . . History@
- Amistad@
- Black History Month (18)
- Buffalo Soldiers (14)

- Civil Rights Movement@
- Civil War Units (4)
- Genealogy (15)
- Harlem Renaissance@
- Juneteenth@
- Lynching@
- Museums & Memorials (13)
- People@
- Science and Technology (4)
- Slavery@
- Sports@
- Timelines (3)
- Web Directories (3)
- . . . Institutes (31)
- . . . Journals (2)
- . . . Libraries (2)
- . . . Literature@
- . . . Theorists and Critics (7)
- . . Historically Black Colleges &
Universities@
- . . Lesbian, Gay, and Bisexual (207)
- . . Migrant (9)
- . . Native American (44)
- . . Religious@
- . . Rural (11)
- . . Seniors (8)
- . . U.S. Hispanic and Latino (9)
- . . Women (90)
- . **By Subject** (11)
- . . Art@
- . . . Art History@
- Art Historians@
- College and University
Departments (75)
- Courses (3)
- . . . Art Schools (126)
- . . . College and University (253)
- . . . Companies@
- . . . Courses (12)
- . . . Design Arts@
- . . . Humanities@
- . . . K-12 (88)
- . . . Non-Degree Programs (47)
- . . . Organizations (31)
- . . . Performing Arts@
- . . . Teaching (6)
- . . . Workshops (46)
- . . Business@
- . . Environment and Nature@
- . . Health@
- . . Humanities@
- . . Languages@
- . . Mathematics@
- . . Music@
- . . News and Media@
- . . Science@
- . . Social Science@.

Education

. Academic Competitions (77)

- . . . College and University (32)
- College Bowl (13)
- Teams (11)
- Debate@
- Teams (16)
- . . . Companies@
- . . . Forensics@
- . . . K-12 (39)
- . . . Teams (2)
- . **Adult & Continuing Education**
- . . Career Specific Training@
- Apparel@
- Auctioneering@
- ▶Automotive@
- ▶Aviation@
- Bartending@
- ▶Bicycle Mechanic Schools@
- Biomedical@
- Bootmaking@
- Brewing@
- Broadcasting@
- Building Inspection Services@
- Clock and Watch Repair@
- Commercial Diving@
- Computer@
- Construction@
- Cosmetology@
- Culinary@
- Customer Service@
- Dog Grooming@
- •Education@
- Electronics@
- Emergency Services@
- Engineering@
- Environment@
- Facilities Management@
- Financial Services@
- Fitness@
- Floral Design@
- Funeral Service@
- Gambling@
- Gunsmithing@
- Health Care@
- Hospitality Industry@
- Human Resources@
- Interior Design@
- Investigative Services@
- Jewelry and Gemstones@
- Law@
- Makeup Artist Training@
- Manufacturing@
- Museums@
- Music Production@
- Neuro-Linguistic Programm@

- . . . Real Estate@
- . . . Religious@
- . . . Security and Law Enforcement@
- ▶Travel@
- ▶Truck Driving@
- Welding@
- Writing and Editing@
- . . . Chautauqua Movement (6)
- . . . Companies@
- . . . Distance Learning@
- . . . GED@
- . . . Institutes (203)
- . . . Literacy (66)
- . . . Media Education@
- . . . Organizations (27)
- . . . Publications (2)
- . . . Special Education (4)
- . **Bibliographies** (4)
- . . Character Education@
- . **Bilingual** (24)
- . . English as a Second Language@
- . . . Bilingual Education@
- Bilingual Education@
- . . . Chats and Forums (9)
- . . . College and University Departments (84)
- . . . Commercial Products@
- . . . Conferences (1)
- . . . IELTS (5)
- . . . Language Schools (143)
- . . . Lessons & Tutorials Online (45)
- . . . Magazines (8)
- . . . Organizations (2)
- . . . Student Projects (9)
- . . . Teaching (52)
- . . . TOEFL (1)
- . . . TOEIC (1)
- . . . Web Directories (15)
- . **Career and Vocational** (232)
- . . Career Planning (127)
- . . Career Specific Training@
- . . Institutes (11)
- . . Occupational Standards (11)
- . . Organizations (28)
- . . School to Work (27)
- . . Schools (23)
- . **Chats and Forums** (40)
- . . Chat (4)
- . . •Educational MOOs@
- . . Mailing Lists (7)
- . . Message Boards (7)
- . . Usenet (16)
- . **Companies@**
- . . Admissions (74)
- . . Athletic Recruiting@
- . . Business to Business@
- . . Career Training@

- . . Counseling (12)
- . . Financial Aid (79)
- . . Home Schooling (52)
- . . International (19)
- . . Internet Services (1)
- . . Learning Centers (9)
- . . School Reports (3)
- . . School Supplies (17)
- . . Teaching & Learning Aids@
- . . Test Preparation (169)
- . . Tour Operators@
- . . Tutoring (40)
- . **Conferences** (52)
- . . Distance Learning (2)
- . . Home Schooling@
- . . Instructional Technology@
- . . K-12@
- . . Languages@
- . . Math Education@
- . . Past Conferences (20)
- . . Science Education@
- . **Correctional@**
- . . Organizations (1)
- . . Regional Agencies (8)
- . **Disabilities@**
- . . Blindness@
- . . College Support & Resources (31)
- . . Deafness@
- . . Disability Studies@
- . . Organizations (8)
- . . Special Education@
- . **Distance Learning** (462)
- . . Adult & Continuing Education
- . . Colleges and Universities (245)
- . . Conferences@
- . . Courses About (2)
- . . Courses Online (11)
- . . K-12 (63)
- . . Language Schools@
- . . Online Teaching and Learning@
- . . Teacher Education (11)
- . . Telementoring@
- . . Television (22)
- . . Vocational Schools (15)
- . . Web Directories (11)
- . **Early Childhood Education** (84)
- . . Child Care@
- . . Institutes (9)
- . . Organizations (33)
- . . Schools (7)
- . . Teaching (11)
- . **Employment** (138)
- . . English as a Second Language@
- . . Individual Resumes (66)
- . . Jobs (69)
- . . Recruiting and Placement@
- . . Unions@

- . **Equity** (27)
 - .. Gender Equity@
 - .. Government Agencies (4)
 - .. Organizations (11)
 - .. Research Centers (2)
- . **Financial Aid** (386)
 - .. College Aid Offices (163)
 - .. Companies@
 - .. Grants (57)
 - .. K-12 School Funding@
 - .. Loans (11)
 - .. Organizations (13)
 - .. Regional Resources (23)
 - .. Savings & Investment Planning@
 - .. Scholarship Programs (88)
- . **Government Agencies** (77)
 - .. Canada@
 - .. Equity@
 - .. United States (53)
- . **Graduation** (58)
 - .. Clip Art and Graphics (6)
 - .. Graduation Poems (3)
 - .. Speeches (42)
 - .. Virtual Cards@
- . **Higher Education** (17643)
 - .. Academic Competitions@
 - ... College Bowl (13)
 - ... Teams (11)
 - ... Debate@
 - Clubs, Teams, & Societies (29)
 - Teams (16)
 - ... College Bowl@
 - Debate@
 - Forensics@
 - .. Books@
 - ... Admissions (10)
 - ... Athletic Recruiting@
 - ... Financial Aid (6)
 - ... University Life (4)
 - .. **College Entrance** (441)
 - ... Admissions Offices (299)
 - ... Books@
 - ... Business Schools@
 - ... Companies@
 - Athletic Recruiting@
 - Baseball@
 - Books@
 - Football@
 - Soccer@
 - Books@
 - College Tour Operators (2)
 - Counseling (29)
 - Graduate School (4)
 - International Students (9)
 - Essays and Applications (25)
 - International Students (5)
 - Counseling@
 - Online Applications (3)
 - Software (2)
 - Videos (3)
 - Educational Standards & Testing@
 - ACT (1)
 - GED (2)
 - GED Prep. Companies@
 - GMAT (3)
 - GRE (3)
 - IELTS@
 - By Region (52)
 - Canadian Provinces (8)
 - U.S. States (44)
 - By Subject (13)
 - K-12 Curriculum Standards@
 - LSAT (4)
 - MCAT (5)
 - Companies@
 - Occupational Standards@
 - Australian Qualifications Framework@
 - U.K. Natl Vocatl Qualific.@
 - U.S. National Skill Standards
 - SSAT (1)
 - Test Preparation Companies@
 - Books@
 - Career Fields (79)
 - Civil Service (5)
 - Customs Broker (2)
 - •Education (10)
 - TEFL/TESL (9)
 - Electrician (2)
 - Engineering (2)
 - Finance (29)
 - CFA (11)
 - CPA (8)
 - EA (1)
 - Insurance (2)
 - NASD (3)
 - Health (18)
 - Medicine (11)
 - Boards (6)
 - USMLE (5)
 - Mental Health (1)
 - Nursing@
 - Law (6)
 - Bar Examination (6)
 - Military (1)
 - Social Work
 - College Entrance (26)
 - GED (5)
 - Graduate School Entrance
 - Online Subscription Serv. (7)
 - Software (10)
 - TOEFL (7)
 - Videos (2)
 - Testing Companies@
 - Computer-Based Testing Services (7)
 - Online (5)
 - Software@
 - Math (3)
 - Titles (2)
 - Resellers (1)
 - TOEFL@
 - TOEIC@
 - Voluntary National Testing (4)
 - Financial Aid@
 - College Aid Offices (163)
 - Companies@
 - Books@
 - Counseling (4)
 - Lenders (33)
 - International (1)
 - Loan Guarantors (4)
 - Loan Servicers (6)
 - Scholarship Search Serv. (16)
 - Secondary Markets (5)
 - Grants (57)
 - Web Directories (4)
 - K-12 School Funding@
 - Organizations (7)
 - Technology Funding@
 - Technology Funding (16)
 - Organizations (7)
 - Programs (5)
 - Loans (11)
 - Lenders@
 - International (1)
 - Organizations (13)
 - Regional Resources (23)
 - Savings & Investment Planning@
 - Education Savings Plans@
 - Scholarship Programs (88)
 - Search Services@
 - Web Directories (4)
 - Search Services@
 - Web Directories (4)
 - Online Applications (83)
 - Individual Schools (76)
 - School Rankings (12)
 - Graduate (7)
 - Law School@
 - Test Preparation Companies@
 - Books@
 - Career Fields (79)
 - Civil Service (5)
 - Customs Broker (2)
 - •Education (10)
 - TEFL/TESL (9)
 - Electrician (2)
 - Engineering (2)

Education**. Higher Education**

- . . . College Entrance
- Test Preparation Companies@
- Career Fields (79)
 - Finance (29)
 - CFA (11)
 - CPA (8)
 - EA (1)
 - Insurance (2)
 - NASD (3)
 - Health (18)
 - Medicine (11)
 - Boards (6)
 - USMLE (5)
 - Mental Health (1)
 - Nursing@
 - Law (6)
 - Bar Examination (6)
 - Military (1)
 - Social Work (2)
 - College Entrance (26)
 - AP (2)
 - SAT (13)
 - SAT II (1)
 - Software (3)
 - GED (5)
 - Graduate School Entrance (25)
 - GMAT (4)
 - GRE (3)
 - LSAT (4)
 - MCAT (4)
 - Online Subscription Serv. (7)
 - Software (10)
 - College Entrance@
 - TOEFL (7)
 - Videos (2)
 - Web Directories (4)
- . . . **Colleges and Universities** (16861)
 - By Region (15221)
 - SN A long list as above
 - Argentina@
 - Armenia@
 -
- . . . **Distance Learning@**
 - Business@
 - Graduate Programs@
 - Community Colleges (20)
 - Consortia (10)
 - Credit for Prior Experience (6)
 - Engineering@
 - Graduate Programs (9)
 - Graduate Programs (24)
 - Business@
 - Engineering@
 - Web Directories (2)

. . Graduate Education (74)

- By Subject (49)
 - Accounting and Auditing@
 - American (United States) Studies@
 - Archaeology@
 - Arts@
 - Asian Studies@
 - Business Schools@
 - Classics@
 - Comparative Literature@
 - Conflict Resolution@
 - Cultural Studies@
 - Economics@
 - European Studies@
 - Film@
 - Finance@
 - First Nations Studies (Canada)@
 - Gender Studies@
 - History@
 - History & Philosophy of Science@
 - Human Ecology@
 - International Relations@
 - Irish Studies@
 - Japan Studies@
 - Journalism@
 - Latin American Studies@
 - Law Schools@
 - Library & Information Science@
 - Literature@
 - Marine Archaeology@
 - Mathematics@
 - Medical Schools@
 - Mexican American Studies@
 - Middle East Studies@
 - Peace and Conflict Studies@
 - Philosophy@
 - Physics@
 - Political Economy@
 - Political Science@
 - Psychology@
 - Public Administration@
 - Public Policy@
 - Recreation and Leisure Studies@
 - Rhetoric@
 - Russian & East European Studies@
 - Scandinavian Studies@
 - Science, Technology, & Society Studies@
 - Sexology@
 - Slavic Studies@
 - Social Work@

- Women's Studies@
- Distance Learning@
- Business@
- Engineering@
- Electronic Theses & Dissertations (ETDs)@
- Graduate Record Exam@
- Organizations (6)
- Rankings@
- Law School
- Web Directories (7)
- . . . **Guidance** (64)
 - Academic Advising (2)
 - College Entrance@
 - Admissions Offices (299)
 - Books@
 - Business Schools@
 - Companies@
 - Educational Standards & Testing@
 - Financial Aid@
 - Online Applications (83)
 - School Rankings (12)
 - Test Preparation Companies@
 - Web Directories (4)
 - Medical School (34)
 - MCAT@
 - Companies@
 - Medical Schools@
 - MCAT@
 - Medical Schools@
 - Admissions & Guidance@
 - MCAT@
 - Medical Schools@
 - Student Organizations@
 - Web Directories (2)
 - Caribbean Med. Schools (8)
 - Dental Schools@
 - Orthodontic@
 - Student Organizations@
 - Interns and Residents (3)
 - Optometry@
 - Web Directories (2)
 - Student Organizations@
 - Web Directories (2)
- . . . Honors Programs (46)
- . . . **News and Media@**
 - Individual Schools (419)
 - SN Long list of countries and US states
 - Australia@
 - Canada@
 -
 - Alabama (4)
 - Arizona (3)
 -
 - Magazines (174)

- Individual Schools (139)
- Alumni (40)
- Humor@
- Literary (52)
- Newspapers (436)
- Business Schools (8)
- Medical Schools (2)
- Newswires@
- Web Directories (2)
- Radio Stations@
- Countries (13)
- Australia@
- Belgium@
- Cities@
- Provinces@
- Complete List@
- Brazil@
- Canada@
- Denmark@
- Ireland@
- New Zealand@
- Singapore@
- South Africa@
- Sweden@
- Switzerland@
- United Kingdom@
- Vietnam@
- Eastern United States (168)
- Internet Broadcasts@
- Western United States (110)
- . . Organizations (96)
- . . . Alumnae/I Associations@
- . . . Community College (9)
- . . . Graduate Education@
- . . . Transfer Student (3)
- . . Policy Research Centers@
- . . Seminaries@
- . . . Christian@
- . . . Jewish@
- . . Student Life (25)
- . . . Books@
- . . . Classifieds@
- . . . Organizations@
- . . Web Directories (14)
- . . . College Entrance@
- . . . Colleges and Universities@
- . . . Graduate Programs@
- . **Instructional Technology (327)**
- . . Conferences (23)
- . . . Online Teaching and Learning@
- . . . Past Events (2)
- . . Institutes (47)
- . . . College & University Departments (34)
- . . Journals (9)
- . . Online Teaching & Learning (167)
- . . . Conferences (8)

- . . . Corporate Programs (10)
- . . . Courses (9)
- . . . Educational MOOs (7)
- . . . Magazines (3)
- . . . Organizations (14)
- . . . Papers (4)
- Telementoring@
- Projects (20)
- Higher Education (4)
- Regional (4)
- Virtual Field Trips@
- . . . Research Institutes (3)
- . . . Teacher Resources (43)
- Companies@
- Email (2)
- English as a 2nd Language@
- Mathematics@
- Science@
- Social Studies@
- . . . Telementoring (12)
- . . . Papers (2)
- . . . Web Directories (1)
- . . . Web Directories (8)
- . . . Telementoring@
- . . WebQuests (19)
- . . Collections (8)
- . . Science@
- . . Social Studies@
- . . School Technology Funding@
- . . Organizations (7)
- . . Programs (5)
- . . Software Reviews@
- . . Titles (64)
- . . . College Preparatory (6)
- Math@
- Reading (23)
- Reader Rabbit (7)
- . . Web Directories (6)
- . **Journals (33)**
- . . •Educational Theory@
- . . Instructional Technology@
- . . Music@
- . **K-12 (52293)**
- . . **By Region (27417)**
- . . . Countries (7635)
- . . . Regions (9)
- . . . U.S. States (19773)
- . . **Academic Competitions@**
- . . . Debate@
- . . . Clubs, Teams, & Societies (25)
- Lincoln-Douglas@
- Institutes (2)
- Policy Debate (7)
- 1998-99 Debate Topic (3)
- 1999-2000 Debate Topic (1)
- . . History@
- . . Mathematics@

- International (6)
- . . . Science@
- International (4)
- Mathematics@
- Mathematics@
- International (6)
- Projects and Ideas (20)
- Science Project Books@
- Web Directories (2)
- . . . Teams (13)
- Debate@
- Lincoln-Douglas@
- Forensics@
- . . Arts@
- . . Curriculum Standards (27)
- . . . Dance@
- . . . Drama@
- . . . Lesson Plan (4)
- . . . School Departments (10)
- . . . Lesson Plans (17)
- . . . Theater@
- . . . Schools (35)
- . . . Departments (5)
- . . . Usenet (2)
- . . Conferences (10)
- . . **Countries (41)**
- . . . Australia@
- . . . Belgium@
- . . . Brazil@
- . . . Canada@
- . . . Chile@
- . . . China@
- . . . Columbia@
- . . . Costa Rica@
- . . . Cyprus@
- . . . Denmark@
- . . . Egypt@
- . . . Estonia@
- . . . Ethiopia@
- . . . France@
- . . . Germany@
- . . . Ghana@
- . . . Greece@
- . . . Hong Kong@
- . . . India@
- . . . Ireland@
- . . . Israel@
- . . . Italy@
- . . . Japan@
- . . . Korea, South@
- . . . Macau@
- . . . Malaysia@
- . . . Malta@
- . . . Mexico@
- . . . Netherlands@
- . . . New Zealand@
-

•Education

• Organizations (2945)

- . . . Alumnae/I Associations (459)
- . . . Colleges and Universities (207)
- . . . High Schools (239)
- Reunions@
- . . . Reunions@
- College and University (8)
- High Schools (274)
- . . . Arts@
- . . . College and University@
- . . . Business@
- . . . Case Research (4)
- . . . College and University@
- . . . K-12@
- . . . Christian@
- . . . Catholic@
- . . . Home Schooling@
- . . . Student (171)
- Baptist@
- Catholic@
- Church of Christ@
- Fraternities and Sororities (20)
- Presbyterian@
- Presbyterian Church in America@
- United Methodist Church@
- . . . Early Childhood Education@
- . . . Equity@
- . . . Financial Aid@
- . . . **Higher Education@**
- . . . Alumnae/I Associations@
- . . . Community College (9)
- . . . Graduate Education@
- . . . Transfer Student (3)
- . . . **K-12@**
- . . . Administrators@
- . . . Alternative (7)
- . . . Alumnae/I Associations@
- . . . Business@
- . . . Charter School@
- . . . Computer Clubs@
- . . . Home Schooling@
- Christian@
- . . . International Schools@
- . . . Lesbian, Gay, and Bisexual@
- . . . Parent@
- . . . Reading@
- Literacy@
- . . . School Associations (24)
- . . . School Board Associations (13)
- . . . School Funding@
- Technology Funding@
- . . . Social Studies@
- . . . Violent Crime Prevention@
- . . . Languages@

- . . . English as a Second Language@
- Student (1)
- Teaching@
- . . . Lesbian, Gay and Bisexual@
- . . . Literacy@
- . . . Math@
- . . . Music@
- . . . Teaching@
- Kodaly Method@
- . . . Online Teaching and Learning@
- . . . Policy@
- . . . Professional (197)
- . . . Administrators (23)
- . . . Consultants (2)
- . . . Faculty (19)
- . . . Unions (77)
- . . . Reading (8)
- . . . Reform@
- . . . Rural Education@
- . . . Science@
- . . . Engineering@
- . . . Student (2131)
- . . . Animal Rights@
- . . . Anthropology@
- . . . Architecture@
- . . . Community Service@
- . . . Computer Science@
- . . . Cultural (361)
- African (2)
- African-American@
- Arabic (5)
- Armenian (8)
- Asian American (23)
- Azerbaijani (1)
- Bangladesh (4)
- Bosnian (2)
- Brazilian (2)
- Cambodian (1)
- Caribbean (3)
- Chinese (23)
- Cypriot (2)
- Egyptian (2)
- Filipino (6)
- Finnish (1)
- Haitian (2)
- Hawaiian (3)
- Hellenic (27)
- Hispanic and Latino (32)
- Law@
- Indian (21)
- Indonesian (14)
- Italian (3)
- Japanese (2)
- Korean (24)
- Lebanese (3)
- Malaysian (16)
- Mexican (3)

- Multicultural (7)
- Native American@
- Pakistani (9)
- Romanian (8)
- Scandinavian (1)
- Singaporean (29)
- Sri Lankan (2)
- Taiwanese (12)
- Thai (21)
- Turkish (12)
- Vietnamese (22)
- . . . Economics@
- Accounting and Auditing@
- . . . Engineering@
- Materials Science@
- Mechanical Engineering@
- . . . Environmental@
- . . . European Union@
- Austria@
- Germany@
- Government (1)
- Netherlands@
- . . . Fraternities and Sororities (1443)
- Apparel@
- Christian@
- Collectibles@
- Directories (4)
- Hazing (2)
- Professional (128)
- Music@
- Service (138)
- Social (1154)
- Interfraternity Councils (14)
- Usenet (3)
- . . . Government (90)
- Graduate (13)
- High School (4)
- . . . Homelessness@
- . . . Honor Societies (74)
- . . . Law@
- Hispanic and Latino (7)
- Pre-Law Societies (2)
- Women@
- . . . Lesbian, Gay, and Bisexual@
- Alumni Associations (9)
- Campus Support Offices@
- . . . Medical@
- Interns and Residents (3)
- Optometry@
- . . . Nursing@
- . . . Political (24)
- Democratic Party@
- Federalist Society@
- Libertarian Party@
- Reform Party@
- Republican Party@
- Young Americans for Freedom

- . . . Religious (5)
 - Christian@
 - Hindu@
 - Islamic@
 - Jewish@
 - Sikh@
- . . . Residence Hall Associations (9)
- . . . Web Directories (2)
- . . Student Affairs (15)
- . . . Web Directories (1)
- . . Trade Associations (3)
- . . Vocational Education@
 - . . . Private Industry Councils@
 - . . . Training & Enterprise Councils@
- . . Women@
 - . . . Law@
- . **Policy** (52)
 - . . College & Univ. Dpts & Programs
 - . . News and Media (3)
 - . . Organizations (6)
 - . . Research Centers (23)
 - . . . Higher Education (6)
 - . **Programs** (309)
 - . . Co-operative Programs (8)
 - . . K-12@
 - . . . Environment and Nature@
 - . . . National Guard Youth Challenge Program@
 - STARBASE (3)
 - Student Exchange@
 - Study Abroad@
 - Companies@
 - Student Exchange@
 - Admissions@
 - Counseling@
 - Credential Evaluation (8)
 - Financial Aid@
 - Summer Programs (56)
 - Learning Disabilities@
 - Math@
 - Science (22)
 - Math@
 - . . . Model Congress@
 - . . . Model United Nations@
 - . . . College and University (35)
 - Events (12)
 - Events (2)
 - College and University@
 - High School@
 - High School (33)
 - Events (16)
 - . . . National Writing Project (27)
 - . . . Student Exchange (78)
 - . . . Study Abroad (157)
 - . . . Summer Programs (18)
 - K-12
 - . **Reform** (67)
 - . . Books@
 - . . Class Size (6)
 - . . Organizations (35)
 - . . School Choice (9)
 - . **Special Education** (167)
 - . . Companies@
 - . . . Administrative Software@
 - . . . Teaching & Learning Aids (5)
 - Publishers@
 - . . Conductive Education (19)
 - . . . Organizations (4)
 - . . . Schools and Institutes (14)
 - . . Employment@
 - . . Inclusion (7)
 - . . Institutes (29)
 - . . . College and Univ. Departments
 - . . . Conductive Education
 - . . . Learning Disabilities@
 - Attention Deficit Disorder@
 - Companies@
 - Conferences (1)
 - Organizations (8)
 - Web Directories (2)
 - Companies@
 - Attention Deficit Disorder@
 - Dyslexia@
 - Dyslexia@
 - Organizations (12)
 - Shopping and Services@
 - . . . Dysphagia@
 - . . . Dyspraxia@
 - . . . Gifted & Learning Disabled (16)
 - . . . Institutes@
 - College & Univ. Departments
 - Conductive Education@
 - Organizations (20)
 - Dyslexia@
 - Schools (31)
 - K-12@
 - Summer Programs (3)
 - U.S. Boarding Schools (9)
 - Organizations (18)
 - Conductive Education@
 - Schools (69)
 - Adult & Continuing Education@
 - Colleges and Universities (3)
 - Conductive Education@
 - Early Education (16)
 - Schools for the Blind@
 - Schools for the Deaf@
 - K-12 (46)
 - Learning Disabilities
 - Schools for the Blind@
 - Schools for the Deaf@
 - Schools for the Blind@
 - Schools for the Deaf@
 - . . Web Directories (3)
 - . . Usenet
 - . **Standards and Testing** (63)
 - . . ACT (1)
 - . . GED (2)
 - . . . GED Preparation Companies@
 - . . GMAT (3)
 - . . GRE (3)
 - . . IELTS@
 - . . K-12 Curriculum Standards@
 - . . . By Region (55)
 - . . . By Subject (14)
 - Agriculture@
 - Arts@
 - Business@
 - Dance@
 - English Language Arts@
 - Health@
 - History@
 - Languages@
 - Library & Information Literacy@
 - Mathematics@
 - Music@
 - Physical Education@
 - Science@
 - Social Studies@
 - . . LSAT (4)
 - . . MCAT (5)
 - . . . Comapnies@
 - . . Occupational Standards@
 - . . . Australian Qualific Framework@
 - . . . U.K. Natl Vocational Qualific.@
 - . . . U.S. National Skill Standards (5)
 - . . SSAT (1)
 - . . . Test Preparation Companies@
 - . . . Testing Companies@
 - . . . Computer-Based Testing Serv.
 - . . . Software@
 - . . TOEFL@
 - . . TOEIC@
 - . . Voluntary National Testing (4)
 - . **Statistics** (6)
 - . **Teaching** (91)
 - . . Arts@
 - . . **Curriculum** (9)
 - . . . Companies@
 - By Subject (9)
 - Art@
 - Computers and Technology@
 - Health and Fitness@
 - Languages@
 - Life Skills@
 - Mathematics@
 - Reading and Writing@
 - Science@
 - Social Studies@

•Education

- . **Teaching** (91)
 - . . . Arts@
 - . . . Curriculum (9)
 - . . . Companies@
 - Christian Home Schooling@
 - Preschool (9)
 - School Reform (2)
 - Textbooks@
 - Retail
 - Used (3)
 - . . . Home Schooling@
 - . . . K-12 Lesson Plans@
 - Arts@
 - History@
 - Literature@
 - Math@
 - News and Current Events (5)
 - Reading@
 - Science@
 - Social Studies@
 - Television Curriculum@
 - Theater@
 - WebQuests@
 - Writing@
 - . . Early Childhood Education@
 - . . English@
 - . . English as a Second Language@
 - . . Evaluation (4)
 - . . Forums (3)
 - . . International Teaching (7)
 - . . English as a Second Language@
 - . . International Schools@
 - . . Internet@
 - . . Companies@
 - . . **K-12@**
 - . . . Lesson Plans (51)
 - Arts@
 - Theater@
 - History@
 - Literature@
 - Math@
 - News and Current Events (5)
 - Reading@
 - Science@
 - Social Studies@
 - History
 - Webquests (15)
 - Television Curriculum@
 - Theater@
 - WebQuests@
 - Collections (8)
 - Science@
 - Social Studies@
 - Writing@
 - . . . School Library Resources@

- Children's Literature@
- Young Adult Literature@
- . . . Substitute Teaching (6)
- . . . Teacher Certification (53)
- U. S. States (50)
- . . . Web Directories (24)
- . . Math@
- . . . Organizations (7)
- . . . Professional Development Services (@)
- . . Music@
- . . . Organizations (40)
- . . . Web Directories (3)
- . . Online Teaching and Learning@
- Conferences (8)
- Corporate Programs (10)
- Courses (10)
- Educational MOOs (7)
- Magazines (3)
- Organizations (14)
- Papers (4)
- Projects (20)
- Research Institutes (3)
- Teacher Resources (45)
- Telementoring (12)
- Web Directories (8)
- WebQuests (19)
- . . Science@
- . . . Science@
- . . . Biology@
- Lesson Plans (32)
- WebQuests (9)
- Space@
- . . Teacher Education (24)
- . . College & Univ. Departments@
- . . . Distance Learning@
- . . . Montessori Method@
- . . . Professional Development Services@
- . . . Teacher Certification@
- . . College & University Departments@
- . . . Distance Learning@
- . . . Montessori Method@
- . . . Professional Dev. Serv.@
- . . . Teacher Certification@
- . **Theory and Methods** (644)
- . . Block Scheduling (6)
- . . Character Education (18)
- . . Articles and Papers (4)
- . . Bibliographies (3)
- . . Companies@
- . . Institutes (1)
- . . Charter Schools (15)
- . . . Organizations@
- . . Critical Thinking@
- . . . Logic@

- . . Experiential Education (4)
- . . Feminist Pedagogy (6)
- . . Home Schooling (113)
- . . Institutes (240)
- . . . College & Univ. Departments
- . . . Early Childhood Education@
- . . . K-12@
- . . Journals (6)
- . . Libraries (24)
- . . Montessori Method (146)
- . . . Schools
- . . . Teacher Education
- . . Multiple Intelligences@
- . . Service Learning (10)
- . . Theorists (12)
- . . Dewey, John@
- . . Freire, Paulo (8)
- . . Montessori, Maria (2)
- . . Steiner, Rudolf@
- . . Waldorf Method (50)
- . . . Schools@
- . . WebQuests@
- . **Web Directories** (45)

Entertainment

SN Cool Links, Movies, Humor, Music ...

• Consumer Electronics

- . . Audio
- . . . MP3 Players
- ▶Automotive (9)

• Humor (4865)

- . . ▶Cars (40)
- . . •Education (43)
- . . ▶Traffic Cones (2)

• Music

- . . •Education
- . . . By Instrument (3)
- . . . Camps (29)
- . . . Conferences (4)
- . . . Courses and Lessons (8)
- . . . Ear Training (5)
- . . . Festivals (9)
- . . . Jazz@
- . . . Journals (2)
- . . . K-12 Curriculum Standards (6)
- . . . Kodály Method (18)
- . . . Organizations (17)
- . . . Schools (267)
- . . . Shopping and Services@
- . . . Suzuki Method (10)
- . . . Teaching (50)
- . . . Web Directories (6)
- . . Instruments
- . . . Stringed Instruments
- Guitar
- •Education

Government

- SN Elections, Military, Law, Taxes...
- . **Law**
- SN No narrower category Transportation
- .. •Continuing Legal Education
- ... Companies@
- ... Court Reporting@
- ... Law Schools@. . .
- . **Military** (867)
- .. Weapons and Equipment (269)
- ... ▶Aircraft@
- ... ▶Armored Vehicles (17)
- ... ▶Ships (9)
- . **U.S. Government**
- .. Executive Branch
- ... Departments and Agencies
- ... •Department of Education
- ... ▶Department of Transportation
- .. Military
- ... •Training and Education (18)
- ... Navy
- ... ▶Fleets and Ships (247)
- ... ▶Submarines (83)
- ... •Training and Education
- .. State Government
- ... Massachusetts@
- ... •Department of Education
- ... ▶Department of Transportation

Health

- SN Medicine, Diseases, Drugs, Fitness ...
- . **Alternative Medicine** (499)
- .. Apitherapy (6)
- .. Applied Kinesiology (8)
- .. Aromatherapy (14)
- ... Companies@
- ... Organizations (3)
- .. Ayurveda@
- ... Companies@
- ... Herbal Supplements@
- ... Brand Names (8)
- ... Products@
- .. Biofeedback (8)
- ... Companies@
- .. Booksellers@
- ... Titles (33)
- ... Herbs (3)
- ... Massage (3)
- ... Yoga@
- .. Breathwork (9)
- ... Companies@
- ... Holotropic (3)
- ... Rebirthing (5)
- ... Vivation (1)

- ... Rebirthing (6)
- ... Companies@
- .. Business to Business
- ... Massage Therapy (132)
- ... Equipment and Supplies (33)
- ... Manufacturers (14)
- ... Training and Development (97)
- ... Massage Therapy@
- .. Buteyko@
- ... Companies@
- .. Chinese Medicine
- ... Acupuncture (24)
- ... Companies@
- ... Electro-acupuncture (6)
- ... Software@
- ... Supplies (5)
- ... Organizations (13)
- ... Chinese Herbs@
- ... Brand Names (26)
- ... Kombucha (3)
- ... Products@
- ... Medical Schools (28)
- ... Qigong (23)
- ... Companies@
- ... Falun Gong@
- ... Organizations (16)
- ... International Yan Xin Qigong Association (7)
- .. Chiropractic@
- .. Business to Business@
- ... Equipment and Supplies (13)
- ... Manufacturers (2)
- ... Internet Services (4)
- ... Software (6)
- ... Journals (5)
- ... Professional Organizations (33)
- ... •Schools, Departments, & Programs (17)
- .. Shopping and Services@
- ... By Region (963)
- SN A long list of geographical. entities
- . **•Education**
- .. K-12 Curriculum Standards (24)
- .. Medicine@
- .. Midwifery@
- .. Nursing@
- .. Shopping and Services@
- . **Emergency Services** (476)
- .. Companies@
- ... ▶Air Ambulance Services (34)
- ... Supplies and Equipment
- ... ▶Vehicles
- ... Ambulances
- . **Fitness**

- .. Aerobics (9)
- ... Jazzercise (3)
- .. Books@
- ... Yoga@
- .. Business to Business@
- ... •Career Training (13)
- ... Yoga@
- ... Equipment Distributors (2)
- ... Equipment Manufacturers (30)
- ... Health Club Management Software (16)
- .. Events (6)
- .. Health Clubs@
- ... By Region (1060)
- ... Directories (2)
- .. Indoor Rowing@
- .. Institutes (8)
- .. Magazines (16)
- ... Bodybuilding@
- ... Yoga@
- .. Organizations (36)
- ... Professional@
- .. •Physical Education (PE) (48)
- ... College & Univ. Departments
- ... K-12 (24)
- ... Curriculum Standards (15)
- ... Departments (1)
- ... Organizations (2)
- .. Organizations (2)
- . **Mental Health** (755)
- .. Bereavement (56)
- ... ▶Aircraft Accident Support (1)
- . **Nutrition**
- .. Institutes (49)
- ... •Schools, Departments, & Programs (31)
- . **Public Health and Safety** (2206)
- ... ▶Driving Safety@
- ... Helmets (4)
- .. Injury Prevention (20)
- SN No X-ref to Accidents.
- . **▶Travel Health and Medicine** (24)
- .. Deep Vein Thrombosis@
- .. Jet Lag@
- .. Motion Sickness@
- .. Shopping and Services@
- .. Tropical Diseases@
- ... Cholera@
- ... Dengue Fever@
- ... Ebola@
- ... Giardia@
- ... Hepatitis@
- ... Lassa Fever@
- ... Leprosy@
- ... Malaria@
- ...

News and Media

- SN Full Coverage, Newspapers, TV
- . **Industry Information**
- .. Journalism
- ... •Education
- Career and Continuing (5)
- College and University (96)
- Organizations (4)
- Web Directories (1)
- .. •Media Education
- ... Broadcasting@
- ... Career and Continuing (3)
- ... College and Univ. Departments
- ... Journalism@
- ▶Automotive@
- College and University (898)
- Traffic and Road Conditions@
- ▶Transportation@
- ▶Travel@

Recreation & Sports

- SN Sports, Travel, Autos, Outdoors
- . **▶Automotive**
- .. Alternative Fuel Vehicles (147)
- .. Audio (33)
- .. Auto-Free Transportation@
- .. Booksellers@
- .. British Cars (45)
- .. Bumper Stickers (4)
- .. Business to Business@
- .. Buyer's Guides (44)
- .. Car Art@
- .. Charitable Vehicle Donation (9)
- .. Chats and Forums (51)
- .. Classic Cars (145)
- .. Classifieds@
- .. Clubs and Organizations (177)
- .. Concept Cars (10)
- .. Driving (184)
- ... •Driving Schools@
- .. Dune Buggies (16)
- .. Electric Vehicles@
- .. Emissions (21)
- .. Employment (2)
- .. Engineering@
- .. Events and Shows (58)
- .. Famous Cars (30)
- .. Financing (8)
- .. Four Wheel Drive (111)
- .. Hearses (16)
- .. History (88)
- .. Humor@
- .. Industry Information@
- .. Kit Cars (22)
- .. Lemon Law@
- .. License Plates (39)
- .. Licensing and Registr. Agencies

- .. Lowriders (63)
- .. Maintenance (13)
- .. Makes and Models (2968)
- .. Model Cars@
- .. Motorcycles (1012)
- ... Pictures
- .. Museums@
- .. News and Media (113)
- .. Police Vehicles@
- .. Racing@
- .. Recreational Vehicles@
- .. Shopping and Services@
- .. Software (2)
- .. Special Needs@
- .. Sport Utility Vehicles (48)
- .. Station Wagons (28)
- .. Technicians (3)
- .. Theft@
- .. Travel@
- .. Trucks (49)
- .. Web Directories (16)
- .. Women (8)
- .. Wrecks (4)
- .. FAQs (3)
- . **Dance@**
- .. •Education
- . **Hobbies (3034)**
- .. Models (615)
- ... ▶Aircraft (216)
- ... ▶Armored Fighting Vehicles
- ... ▶Boats and Yachts (57)
- ... ▶Cars (73)
- ... ▶Motorcycles (1)
- ... ▶Radio-Controlled (13)
- ... ▶Trains and Railroads (175)
- . **▶Motorcycles@**
- . **Sports**
- .. •College and University
- .. •Physical Education@
- .. Science
- ... •Education
- . **▶Travel**
- .. Air Travel
- ... Airlines@
- United Airlines
- •Education

Reference

- SN Libraries, Dictionaries, Quotations ...
- .. Libraries
- ... Library & Information Science
- •Education
- College and University (55)
- ... •Education Libraries@
- ... ▶Transportation Libraries@

Regional

- SN Countries, Regions, US States
- . **U.S. States**
- .. **Massachusetts**
- ... Massachusetts Locations
- Metropolitan Areas
- Boston Metro
- Business and Shopping
- Shopping and Services
- ▶Automotive
- •Driving Schools
- Counties and Regions
- Cities
- ... **Boston**
- Local Web Directory
- Business & Shopping
- Community
- •Education
- Adult, Career, & Continuing (16)
- Business to Business@
- Child Care Centers & Preschools@
- College & University
- Government Agencies
- K-12 (22)
- Language Schools (11)
- Organizations (25)
- Shopping & Services@
- Employment
- Entertainment & Arts
- Health
- News & Media
- Real Estate
- Recreation & Sports
- ▶Travel & Transportation
- Airports (4)
- Bicycle Advocacy@
- Car Rentals@
- Lesb., Gays, & Bisex.@
- Limos & Shuttles@
- Local Cruises@
- Local Guides (58)
- Lodging (71)
- Maps and Views (12)
- Mass Transit (8)
- Points of Interest (6)
- Restaurants@
- Taxi Services@
- Tour Operators@
- Traffic & Road Cond.
- Transportation Agencies
- Travel Agents@
- Travelogues (2)
- Marketplace
- Yellow Pages

| | | |
|---|--|---|
| <ul style="list-style-type: none"> ... State Web Directory ... Area Guides ... Arts & Humanities ... Business & Economy ... Community & Culture ... Computers & Internet ... •Education ... Business to Business@ ... Career and Vocational (3) ... College and University (96) ... K-12 (11) ... Organizations (6) ... Shopping and Services@ ... Employment ... Entertainment ... Government ... Health ... News & Media ... Real Estate ... Recreation & Sports ... Science ... Social Science ... ▶Travel & Transportation ... Cities@ ... Counties and Regions@ ... Complete List@ ... Airports (3) ... Car Rentals@ ... Destination Guides (13) ... Highways and Roads (5) ... Lodging (8) ... Maps and Views (4) ... Mass Transit (3) ... Transportation Agencies (2) ... Transportation Organizations ... Travel Safety (1) ... New York ... Cities ... New York ... Business and Shopping ... Shopping and Services ... ▶Automotive ... •Driving Schools *** ... •Education ... Adult, Career, & Continuing ... Business to Business@ ... Child Care Centers & Preschools@ ... College & University (633) ... Events (2) ... Government Agencies (3) ... K-12 (115) ... Language Schools (18) ... English 2nd Language ... German (2) ... Italian (1) ... Spanish (2) | <ul style="list-style-type: none"> ... Organizations (47) ... Primary and Secondary (1) ... Programs (7) ... Shopping and Services@ ... ▶Travel & Transportation ... Washington, D.C. ... •Education ... Adult, Career, & Continuing ... Business to Business@ ... Child Care Centers & Preschools@ ... College and University (427) ... Events (4) ... Government Agencies (1) ... K-12 (467) ... Language Schools (13) ... English as a 2nd Language ... French (2) ... Literacy (1) ... Organizations (98) ... Programs (3) ... Shopping and Services@ ... Special Education (13) ... ▶Travel & Transportation ... Countries ... Canada ... •Education ... Academic Competitions (8) ... Adult and Continuing Education (22) ... Bilingual (2) ... By Culture or Group (41) ... By Subject (11) ... Career and Vocational (47) ... Companies@ ... Conferences (5) ... Correctional@ ... Distance Learning (23) ... Early Childhood Education (8) ... Employment (8) ... Financial Aid (17) ... Government Agencies (12) ... Higher Education (239) ... Instructional Technology (19) ... Literacy (6) ... News and Media (9) ... Organizations (142) ... Policy (2) ... Primary and Secondary (498) ... Cities@ ... Metropolitan Areas@ ... Provinces and Territories@ ... Complete List@ ... Academic Competitions@ ... Alternative (8) ... Arts@ ... Curriculum Standards (9) | <ul style="list-style-type: none"> ... Distance Learning@ ... Environment and Nature@ ... Gifted Youth (1) ... Home Schooling@ ... Mathematics@ ... Newspapers@ ... Organizations (20) ... Parental Involvement@ ... Physical Education@ ... Programs (23) ... Reading (4) ... School Funding (3) ... Schools (373) ... Science@ ... Social Science@ ... Social Studies (7) ... Student Resources (10) ... Teaching (22) ... Teaching & Learning Aids@ ... Web Directories (2) ... Programs (34) ... Reform (1) ... Special Education (11) ... Standards and Testing (8) ... Teaching (11) ... Theory and Methods (38) ... Web Directories (5) ... ▶Travel & Transportation |
|---|--|---|

Science

- SN Animals, Astronomy, Engineering ...
- . **Engineering** (4691)
 - . . ▶Aerospace Engineering@
 - . . ▶Automotive Engineering (36)
 - . . Civil Engineering (475)
 - . . . ▶Transportation (31)
 - Companies@
 - Institutes (25)
 - ▶Tunnels@
 - . . •Education (304)
 - Courses (4)
 - Distance Learning (11)
 - Institutes (247)
 - Organizations (19)
 - Vocational Schools (11)
 - . . ▶Naval Engineering (3)
 - . **Mathematics**
 - . . •Education
 - . . . Academic Competitions (2)
 - . . . College and University (366)
 - . . . Companies@
 - . . . Conferences (3)
 - . . . Courses (4)
 - . . . Exercises@
 - . . . Gender Equity@
 - . . . K-12 (149)
 - Academic Competitions (22)
 - Courses (1)
 - Curriculum Standards (40)
 - Exercises@
 - Organizations (6)
 - Programs (27)
 - School Departments (8)
 - Teaching (31)
 - Organizations (28)
 - Software@
 - Teaching (32)
 - Web Directories (3)

Social Science

- SN Archaeology, Economics, Languages ...
- . **Anthropology and Archaeology**
 - SN No narrower category
 - Transportation
- . **Linguistics & Human Languages** (2814)
 - . . Languages (2235)
 - . . . •Education (266)
 - Chinese@
 - English@
 - English as a 2nd Language@
 - French@
 - German@
 - Italian@
 - Japanese@
 - Russian@
 - Spanish@
 - College and University Departments (48)
 - Commercial Products@
 - Conferences (2)
 - Courses (19)
 - K-12 (28)
 - Chinese@
 - Curriculum Standards (20)
 - English as a 2nd Language@
 - French@
 - German@
 - Magnet Schools@
 - Spanish@
 - Language Labs (7)
 - Language Schools (125)
 - Organizations (8)
 - Web Directories (9)
 - Specific Languages
 - Chinese
 - •Education
 - French
 - •Education
 - German
 - •Education

Society and Culture

- SN People, Environment, Religion ...
- . **Disabilities**
 - . . Assistive Technology (54)
 - . . . ▶Automotive (3)
 - . . . Universal Design@
 - . . Recreation and Sports
 - . . . ▶Travel
 - Transportation Resources
- . **Environment and Nature** (7070)
 - . . Pollution (204)
 - . . . Air (60)
 - ▶Auto Emissions@
 - . **Food and Drink**
 - . . Cooking
 - . . . •Culinary Education
 - Baking (7)
 - Vegetarian (4)
 - . **Religion and Spirituality**
 - . . Faiths and Practices
 - . . . Christianity
 - Denominations and Sects
 - Catholic
 - •Education
 - Colleges and Universities
 - K-12 (268)
 - Organizations (14)
 - Seminaries (13)
 - . . . Islam
 - Hajj
 - Makkah@
 - ▶Travel and Transportation

Assignment 13.2 LCC
Library of Congress Classification (LCC)

Assigned: July 16
Due: July 23

| | |
|-------------------------------|--|
| Objectives | Inherited from Assignment 13 (page 121) |
| Deliverables to submit | Inherited from Assignment 13 (page 121) |
| Tasks | Inherited from Assignment 13 (page 121) Do the exploration of <i>Classification Plus</i> (which follows the worksheet) as part of the first task in the worksheet; then you can use it for subsequent parts. |
| Materials | Attached Top-level LCC outline, detailed outline, outline of class L Education some sample pages and sample tables, more sample pages with Small Groups 1, Exploration of subject access (lecture notes) LCC outline (2003) www.loc.gov/catdir/cpsolcco/lcco/lcco.html (pdf, blue font) LCC volumes (schedules) (Any edition after 1980 will do). Lockwood Library Call numbers: US LC26.9: Check an academic Library near you Exercise introducing <i>Classification Web</i> attached (1.5 hrs) |
| Readings | Needham, Ch. 8, Schemes of classification , p.163-168 (in reading packet for Lecture 12.1B) Chan, Lois Mai A guide to Library of Congress Classification . 5th ed. Englewood, Colo. : Libraries Unlimited, 1999. Z696.U4C47 1999 in Lockwood Library p. 1-14 The history of the classification (optional) p. 14-19 Focus and use p. 23-47 Principles, structure, and format (skim the examples). For further study, read more from this book |
| Time | 6 hours |

Name:

LCC Worksheet

►A - ►D in class, ►E - ►I on your own

| 1. General introduction | |
|--|---|
| Historical context | <p>Why do you think Military science and Naval science were each given a whole letter of the alphabet in the LC scheme? Write Answer A here.</p> <p>►A</p> |
| Explanation of call nos. | <p>Note: Complete call numbers for books consist of two parts: The class number and a number for the author of the book, such as</p> <p>Z668.R6 Roberts, Library instruction (Z668 Library education, R6 Roberts)</p> <p>These author numbers are called <i>Cutter numbers</i> after their inventor. We will not deal with Cutter numbers in these assignments. The developers of LCC thought of another use for Cutter numbers when they ran out of numbers for specific subdivisions: They simply arranged subdivisions alphabetically (not necessarily the best thing in a <i>classification</i>) and used a Cutter number beginning with the first letter of the subject to extend the class number. For example,</p> <p>Z695.1.E3 Subject cataloging. Subject headings > By subject A-Z > Education</p> <p>Z695.1.E3E34 Thesaurus of ERIC descriptors (E34 for ERIC)</p> |
| Study Class Z | <p>Examine the LCC outline and some volumes (particularly classes H (1997 or 2002), L (1995 or 1998), and Z (1995)). Examine the schedules and the tables (in the back). Examine the alphabetical index in the back of each volume but take heed: As with all other classifications you will use, do not classify books from the alphabetical index itself.</p> <p>Study class Z in the Outline. Then take out a copy of the Z schedule. Flip through it and note the general organization.</p> |
| Find classes that are enumerated in the schedules | <p>Give the class numbers for the following titles</p> <p>►B Bookbinder's Monthly:</p> <p>►C A Bibliography of Publishing through the Ages:</p> |

| 2. Building new classes using tables | |
|---|--|
| Explanation of tables | While LCC is a highly enumerative scheme, it does not enumerate every specific class. Some classes must be built from a main class in the schedule following instructions in a table. As in DDC there are two kinds of tables: |
| General table | A general table , printed at the end of an LCC volume, applies to several classes in that volume as specified in instructions (not to the entire classification, as in DDC) |
| Local table | A local table , printed in the schedules as a note for a specific class, applies only to the children of that class. These are mostly small tables, sometimes up to a page, often used after a general table has been applied. |
| Building new classes w general table | <p>Example 1. Book selling and publishing > By region or country (vol. Z , p. 28)</p> <hr/> <p>Example 1.1 History of Bookselling in France Z303-310 Book selling and publishing > France (Table Z7, p. 373) In Table Z7 we find 2 History. Biography Thus: Z305 Bookselling in France. History (Z303 + 2)</p> <hr/> <p>Example 1.2</p> <hr/> <p>Example 2. Banking (vol. H, 1997, p. 317)</p> <p style="text-align: center;">Banking By region or country</p> <p style="text-align: center;">HG2701-3542.7 Other regions or countries (Table H8) Add number in table toHG2700</p> <hr/> <p>Example 2.1 Banking in Algeria In Table H8 we find 683 <i>Algeria</i> (vol. H, p. 646) Thus, HG3383 Banking in Algeria (HG2700 + 683)</p> <hr/> <p>Example 2.2 Banking in China In Table H8, 631-640 China (a number range) Thus: HG3331-40 Banking in China (HG2700 + 631-640)</p> <p>See next page on how to determine the specific number in this range.</p> |

| General tables | Table Z7 | Table Z8 |
|-----------------------|---|---|
| | <p>Add the appropriate number from this table to the first number of the classification number span to which the table applies.</p> | <p>Add the appropriate number from this table to the first number of the classification number span to which the table applies.</p> |
| | <p>0 General Works</p> | <p>0 General Works</p> |
| | <p>1 Bibliography</p> | <p>0.2 Bibliography</p> |
| | <p>2 History. Biography</p> | <p>0.3 History. Biography</p> |
| | <p>3 Special lines of business (not A-Z)</p> | <p>0.4 Special lines of business (not A-Z)</p> |
| | <p>4 Directories</p> | <p>0.5 Directories</p> |
| | <p>5 Handbooks, manuals, etc.</p> | <p>0.6 Handbooks, manuals, etc.</p> |
| | <p>6 Periodicals. Societies. Congresses</p> | <p>0.7 Periodicals. Societies. Congresses</p> |
| | <p>7 Collections Local</p> | <p>0.8 Collections Local</p> |
| | <p>7.3.A-Z By state or region, A-Z</p> | <p>0.83.A-Z By state or region, A-Z</p> |
| | <p>7.6.A-Z By city, A-Z</p> | <p>0.86.A-Z By city, A-Z</p> |

| Note on adding numbers from tables | Note on adding numbers from tables in LCC |
|---|--|
| | <p>If the table numbering starts with 0, just add the number from the table to the base number in the schedules.</p> |
| | <p>If the table numbering starts with 1, add the number from the table and subtract 1.</p> |
| | <p>Either way, the first number built is the first number of the range given in the schedule.</p> |

| | |
|---|---|
| <p>Building new classes with local table</p> | <p>In Example 2.2, we did not arrive at an actual class number, but rather a range. We need to determine the subject more specifically to determine the specific class number. The instruction at HG2701-3542.7 actually goes on to say</p> <p style="padding-left: 40px;">Under each, and then gives the local table shown opposite (reformatted here for easier reading)</p> <p>This table allows for subdividing each <i>Banking in country</i> class further. Not all countries are treated as equals. On some countries (such as France) there is a lot of material and LCC wants to provide for subdivisions with short numbers, so it uses a range of 20 numbers for such countries; on other countries (such as Algeria) there is little material and longer numbers are ok, so LCC uses just 1 number. In between are “10-number countries” and “5-number countries”. Thus I have arranged the table in four columns.</p> <p>We can now construct the class for <i>History of Banking in China</i>:</p> <p style="padding-left: 40px;">HG3331-40 Banking in China from above</p> <p style="padding-left: 40px;">The table says to use the fourth number in this range, thus</p> <p style="padding-left: 40px;">HG3334 History of banking in China</p> <p style="padding-left: 40px;"><i>Recent history of banking in China</i> would fall under the same class</p> <p>On the other hand, consider the number for</p> <p style="padding-left: 40px;"><i>Recent history of banking in France</i></p> <p style="padding-left: 40px;">In Table H8, France is 321-340, thus</p> <p style="padding-left: 40px;">HG3021-40 Banking in France</p> <p style="padding-left: 40px;">HG3026-8 History of banking in France</p> <p style="padding-left: 40px;">HG3028 Recent history of banking in France</p> <p style="padding-left: 40px;"><i>China. Banking in foreign countries</i> is HG3325, but</p> <p style="padding-left: 40px;"><i>Algeria. Banking in foreign countries</i> is simply HG3383, same as <i>Banking in Algeria</i></p> <p>Analyze these examples from the perspective of exhaustivity and specificity of indexing and the effects on retrieval.</p> <p>•D</p> |
|---|---|

Local table to build numbers under HG2701-3542.7

| Banking By region or country --Continued | | | | |
|--|---|----------|---------|---------|
| HG2701-3542.7 | Other regions or countries (Table H8) | | | |
| | Add [use a calculator, DS] number in table to HG2700 Under each: | | | |
| | 20 nos. | 10 nos. | 5 nos. | 1 no. |
| Periodicals. Serials | 1 | 1 | 1 | .A1-.A4 |
| Societies, see HG 1507-HG1515 | | | | |
| Yearbooks | 4 | 3 | | |
| Directories | 5 | | | .A5 |
| History and Policy | 6-8 | 4 | 2 | .A6 |
| General Works | 6 | | | |
| Biography, see HG 1552 | | | | |
| Early e.g., Great Britain to 1844; date may vary for different countries | 7 | | | |
| Recent | 8 | | | |
| Statistics (Monographs) | 11 | | | |
| Banking in foreign countries, foreign branches | 12 | 5 | | |
| Central bank, national bank, banks of issue | 14-16 | 6 | 3 | .A7 |
| General works, History and description, including reports | 14 | | | |
| Policy | 15 | | | |
| Administration | 16 | | | |
| Other banks, A-Z | 18.A-Z | 8.A-Z | 4.A-Z | .A8A-Z |
| By region, A-Z | 19.A-Z | 9.A-Z | 4.5.A-Z | .A85A-Z |
| By city, A-Z | 20.A-Z | 10.A-Z5 | 5.A-Z5 | .A9-.Z5 |
| Under each city: | | | | |
| Yearbooks | .x | .x | .x | .x |
| General works | .x2 | .x2 | .x2 | .x2 |
| Including history | | | | |
| Policy, etc. | .x3 | .x3 | .x3. | .x3 |
| Individual banks, A-Z | x4A-Z | .x4A-Z | .x4A-Z | .x4A-Z |
| Savings and loans associations | 20.Z9A-Z | 10.Z6A-Z | 5.Z6A-Z | .Z6A-Z |

2. Building new classes using tables, continued

| | |
|-------------------------------------|---|
| Build your own class numbers | <p>The Proofreader's Handbook</p> <p>E</p> <p>Prison Libraries</p> <p>F</p> <p>Children's Book Publishing:</p> <p>G</p> <p>History of Bookselling in Germany:</p> <p>H</p> <p>History of Bookselling in Poland:</p> <p>I</p> |
|-------------------------------------|---|

Tables in DDC, LCC, and LCSH compared

| | Dewey | LCC | LCSH |
|--|---|---|--|
| Global table (DDC and LCSH) General table (LCC) | Global tables in v.1 Apply to all classes | At end of each volume Apply only to classes within one main class | Standard subdivisions |
| Local table | In the schedule Apply to classes in a small section | In the schedule Usually apply to all classes built with a global table | |
| Apply subdivision from other part of the classification (add from) | In the schedule Apply to one place | Not used | Subheadings under a "pattern heading" |

Library of Congress Classification and Subject Headings on Classification Web

| | |
|---------------------|---|
| Introduction | <p>Classification Web is an electronic version of the Library of Congress Classification schedules and the Library of Congress Subject Headings. You should follow the navigation on the following pages to get a feel for the system, and use it to complete the queries on work sheets H and I. You are welcome to complete this assignment in groups.</p> <p>You may use Classification Web for any part of the LCC and LCSH assignments, but you should also gain some experience with using the printed volumes so that you can compare the two forms of presentation.</p> <p>To learn more, you can look at the Quick Start Tutorial at http://www.loc.gov/catdir/cps/classwebtutorial/1intro.html</p> <p>A comparison between Dewey for Windows and Classification Web is instructive from a document design perspective.</p> |
| Logon | <p>Go to the McKeldin Reference Desk and ask for the Cataloger's Workstation.</p> <p>You did receive an email with instructions on how to sign up for a trial. If you did sign up, you can do this from any computer.</p> |

Reference

- H** Display classification hierarchy
- R** Display classification or LCSH record. Click Close to return to the hierarchy display.
- B** Display bibliographic records from selected OPACs for this class (in separate Window).
- L** Display LC subject headings linked in LC authority records.
- D** Display Dewey classes corresponding to this LC class.
- S** Display subject headings frequently used with this class.
- C** Display in classification browser
- T** Display secondary table
- N** Edit the local notes for this record (only shown if user is authorized).
- E** Edit the authority record for the class (only shown if user is authorized).

A Navigation-based search of the Classification

| A1 General | |
|-------------------|--|
| General | <p>Select Enhanced Classification Browser.</p> <p>In the box LC Class #, type L and press [Enter].</p> <p>Schedule <i>L Education</i> is displayed, towards the top of the screen.</p> <p>Using the box LC Class #, go to LB2411.</p> <p>The box just above the main hierarchy display is called the Hierarchy Frame. It shows the hierarchical chain down to the highlighted class.</p> <p>(See the Reference for explanation of the letters in () or just try them)</p> <p>Click on R to see the record for this class. Click on Close.</p> <p>Try out B (Use Syracuse).</p> <p>Use the browser Back button to go back to the list of libraries, scroll to the end and click Close.</p> <p>Try out L, D, and S.</p> <p>Click Close or Cancel after each.</p> |

| A2 Hyperlinks to cross-references | |
|--|---|
| Hyperlinks to X-ref | <p>Look at <i>Surveys of college graduates</i> (before LB 2420, number range LB2420-2430. Double click on the cross-reference <i>HD6277</i> (blue).</p> <p>In LCSH you can also jump to a cross-referenced heading by a simple click.</p> <p>Examine schedule H around <i>HD6277</i>.</p> |

A3 The Hierarchy Browser (Click down the hierarchy)

Hierarchy Browser

You may need to click on Home before proceeding.

Click on **Menu**.

Click on **Hierarchy Classification Browser**.

Using the box LC Class #, go to *L*.

Click on ▼ to see more of the outline for L, again to see still more.

Click on ▲ until you get back to the original screen (approximately).

Click on LB1-3640 *Theory and practice of education*.

Use ▼ to find

LB1554.2-1602 Elementary or public school education.
Elementary school teaching

and click on it

Click on *LB1572-1602 Special branches*, check out that part of the hierarchy.

If you like, you can repeat this little exercise starting with class H.

Double-click on *HE1-9900 Transportation and communications*.

Find *HE380.8-971 Water transportation* and double-click it.

Double-click *HE380.8-560 Waterways*.

Click in the Hierarchy Frame on *Water transportation* to go back to that level.

B Query-based search of the classification

In WebDewey there are a number of different fields to search, corresponding to different relationships between the search starting point (term or class number) and the classes to be found. In Classification Web, the only index that is really useful is the **Keyword index**. (The Help information says otherwise.) All examples in this exercise use **Boolean search**.

| B1 Start: establish settings | |
|-------------------------------------|---|
| To start | <p>Click on Menu, then on Classification Search.</p> <p>The Help is very helpful. If you want to learn more on how to search this system and about searching in general, read it!</p> <p>Turn on the radio button for boolean search.</p> <p>Under display options click entire hierarchy. (Unfortunately, the resulting display is somewhat hard to read since the entries are not separated by a blank line, but still the most useful.)</p> <p>Under Number of results select 100.</p> |

| B2 Example search: <i>computers in education</i> | |
|---|--|
| Classification search | <p>In the Keyword box, type "<i>computer*</i>" and "<i>education</i>" (* truncates).</p> <p>Click on Search at the bottom of the screen (not Browse).</p> <p>Now try "<i>computer*</i>" and ("<i>education</i>" or "<i>instruction</i>" or "<i>teach*</i>" or "<i>learning</i>").</p> <p>Instead of 47 classes you now find 76.</p> <p>By clicking on C you can jump into the hierarchy around the class found.</p> |

| B3 Subject heading search | |
|----------------------------------|--|
| Subject heading search | <p>Click on Menu, then on LC Subject Headings.</p> <p>If you use the Keyword index, it works just like the classification search.</p> |

| B4 More search possibilities | |
|-------------------------------------|--|
| Explore more | There are many more possibilities in this system. Experiment and/or read the help if you like. One feature that is particularly helpful is restricting the search to a part of the classification either by entering a class number in the Classification number box or by selecting a Subset (subset selection is effective for further searches until you remove all check marks in the subset selection screen. You might want to explore building numbers automatically (see the Quick Start Tutorial also accessible from the bottom of the menu page). |

C Now complete query F Canals from the Query Forms

Do on your own.

List at least 7 important LCC classes from at least 3 different areas of the classification where one should look on the shelves (broad class implies subclasses) and 5 LCSH headings under which one should search.

Copy and paste into a word processor document; it is best to use
Edit > Paste special > Unformatted text

This ends the exercise. Explore some more on your own or close WebDewey.

See next page for further comments

Notes on Classification Web

The display of the hierarchy is very poor. It is even more difficult than in the printed schedule to follow the hierarchical levels by indentation. Major divisions should be bolded.

Some lines are not classes that can be assigned but merely headings to organize the classification. These lines have no class numbers. It would be helpful to have numbers or number ranges in [].

Should use > instead of -- (easier for people who are used to Yahoo); also, the customary use of -- is in Main heading -- Subheading.

Displays do not scroll as expected.

The Hierarchy Browser does not use the standard Explorer-type interface for such displays.

The search for captions searches only the lowest level caption which for most cases is not what is most useful. The Keyword index is the only index that searches the entire chain down to the class (along with any other word in the record for the class, which may sometimes retrieve a bit more than is wanted.) (The **Caption** index or the **Index term** may serve some specialized uses at LC but most users can safely ignore them.)

Even though the unusual mandatory use of quotation marks around single words in Boolean searching suggests it, one cannot search for phrases.

The search results list gives the class number only the lowest level caption as the default; this is not helpful since the lowest level caption does not reveal the hierarchical context. Somebody familiar with LCC will have an idea what the class is about, but other users must click on R for every class found to get the picture.

The results list does not seem to be sorted in any particular order. It should be sorted by class number as the default.

Outline for the analysis of Knowledge Organization Systems

For some items, a section number from Soergel, Organizing information is given in ()

| | |
|----|--|
| 1. | <p>Purpose</p> <p>1.1 Information system or type of information system in which to be used <i>Bibliographic information system. Intended for academic and research libraries and large public libraries.</i></p> <p>1.2 Intended for controlled vocabulary indexing , or query term expansion G (Ch. 12, Introduction)</p> <p>1.3 Type of file and search mechanism for which originally designed Shelving , Card catalog / printed index G Online G</p> |
| 2. | <p>Coverage and designation of concepts. Coverage and format of terms</p> <p>2.1 Concepts: Scope, breadth of coverage. Recency of concepts <i>Universal - covers all knowledge. But focus on Western culture, esp. U. S.</i></p> <p>2.2 Concepts: Specificity, depth of coverage (Section 16.2.2). Coverage at each level of specificity. <i>Level of specificity varies (even among similar entities - e.g. 5 number countries vs. 20-number countries)</i></p> <p>2.3 Are all necessary facets included? Concepts formed in semantic factoring and facet analysis? (S.a. 3.1) <i>Some facets are reflected in general and local tables. Whether all facets for a subject, such as education, are included is impossible to say without extensive analysis because the facets are not explicit.</i></p> <p>2.4 Nature of notation (if none, state that) (Section 15.5.2) <i>Alphanumeric notation, expressive on first two levels (two letters), then simply ordinal, at the very bottom it is again expressive (as in 675.P6 Prison libraries)</i></p> <p>2.5 Terms: Completeness of coverage of terminology (completeness of lead-in vocabulary). Recency of terms <i>No attempt at covering terminology. Terms often old.</i></p> <p>2.6 Form of terms: Consistency, adherence to common usage. <i>No attempt to use common terminology. Terms often created for precise meaning of class.</i></p> |
| 3. | <p>Terminological and conceptual analysis and conceptual structure.</p> <p>3.1 Quality of conceptual structure (14): Facet analysis. Types and degree of differentiation of conceptual relationships included. For each type indicate the completeness of inclusion. (Heading for 3.1 - 3.3)</p> <p>3.1.1 Expression of concepts through elemental concepts (closely related to definition) <i>Explicit only where tables used or where the semantic factors are obvious from the hierarchy</i></p> <p>3.1.2 Hierarchical relationships (polyhierarchy). (Shown by arrangement or Broader Term / Narrower Term X-ref) <i>Monohierarchy with very few cross-references (cf.) for hierarchical relationships not shown by arrangement</i></p> <p>3.1.3 Associative relationships (Implied by physical proximity in the arrangement or explicit Related Term X-ref) <i>Very few Related Term cross-references, subsumed under designation cf.</i></p> <p>3.2 Quality of definitions, explications, scope notes (correctness, detail, clarity). <i>Very few usage notes</i></p> <p>3.4 Completeness of terminological relationships: Does the vocabulary contain terms that are synonymous or quasi-synonymous without indicating the relationship? <i>Not a goal in LCC.</i></p> |

4. **Use of precombination in the index language** (concerns both 2 and 3) (Ch. 14, 15, esp. 15.4)

4.1 To what degree are the final descriptors assigned to documents, whether enumerated or built by the indexer, precombined? *Highly precombined*

4.2 To what extent are precombined descriptors enumerated and/or given in the alphabetical index?
Highly enumerated (more than Dewey)

To what extent can the indexer build additional precombined descriptors?

Indexer can build more precombined classes by adding components from local and global tables.

Are precombined descriptors designated by an independent symbol or a string of symbols? Combination order free or fixed? To what extent do the components of a precombined descriptor determine its place in the arrangement? (Relates also to the arrangement of a classification) (Section 15.5.2)

Precombined descriptors have their own independent symbols. Combination order actually used by the editor varies. Often standard sequence of countries is used, with exceptions.

5. **Access and display. Format of presentation of the vocabulary**

For each format consider access/retrieval by concepts versus access/retrieval by terms.

Access can be provided through arrangement in a printed document or through a computer search system.

5.1 **Format of printed document** (Heading for 5.1.1- 5.1.3)

5.1.1 Overall format: Thesaurus parts and information given in each, connections between them. Is the overall format clear and helpful for finding the appropriate concepts and terms or notations in indexing and query formulation?

Divided into individual volumes by main class. Each volume has outline, main schedule of classes, sometimes global (volume-wide) tables, and an alphabetical index. Has outline to the entire classification, but no overall alphabetical index.

5.1.2 Display of conceptual relationships (Broader Term, Narrower Term, Related Term)

- through linear arrangement or graphical display (Section 15.5.2) *Almost exclusively*

- through cross-references (Section 14.1) *Very few*

- through descriptor-find index (Section 15.5.1) *No*

How well does the display reflect the conceptual analysis, e.g., sequence of concepts on the same hierarchical level (sequence of the children of a concept, that is, the concepts one level further down).

Concepts are arranged in a meaningful sequence in the major hierarchical sequence, but on the lowest level there is often alphabetical arrangement of subjects. In the Bibliographies by subject part in Z, subjects are arranged alphabetically.

5.1.2 Display of terminological relationships (Synonymous Term) *No lead-vocabulary included*

5.2 **Access through computer system.** Navigation. Format of on-line displays

Classification Plus. Searchable and cross-references are hyperlinks. (More detailed analysis omitted.)

Instructions for using the Library of Congress Classification

| | | | | | | | | | | | | | | | | | |
|---------------------------------|--|-----------------|-----------------------------|---------|---------------|-----------|--------------------------------|------------|------------|------------|-------|------------|--------|------------|-----------------------|------------|----------|
| <p>Indexing</p> | <p>Give one and only one class per document; if another class is a strong contender, list it as an alternate and give reasons for selecting the class you chose. (If there is no other class that would fit the document, leave the Alternate class blank.) For each class give the text as a hierarchical chain, starting with the main class, e.g.</p> <p style="padding-left: 40px;">E814.M5 History/US/Elements in the population/Elements, A-Z/ Mexicans</p> | | | | | | | | | | | | | | | | |
| <p>Query formulation</p> | <p>Try to list all classes where one should look for relevant documents on the shelves; if there are more than 10 classes, list a representative sample (enough to demonstrate that you how to find all classes throughout all the LCC main classes).</p> <p>Note: A class implies all its narrower classes; no need to list these narrower classes, they can be readily seen from the schedules. Consider this example from volume H (1997), Section HD <i>Economic history and conditions</i>, p. 158 (reproduced in the materials for Assignment 11):</p> <table style="margin-left: 40px; border: none;"> <tr> <td style="padding-right: 20px;">HD 9710-9710.37</td> <td>Automobiles. Motor Vehicles</td> </tr> <tr> <td>HD 9710</td> <td>General works</td> </tr> <tr> <td>HD 9710.3</td> <td>Automobile supply ... industry</td> </tr> <tr> <td>HD 9710.33</td> <td>Limousines</td> </tr> <tr> <td>HD 9710.34</td> <td>Buses</td> </tr> <tr> <td>HD 9710.35</td> <td>Trucks</td> </tr> <tr> <td>HD 9710.37</td> <td>Recreational vehicles</td> </tr> <tr> <td>HD 9710.38</td> <td>Trailers</td> </tr> </table> <p>Once you have specified HD9710-9710.37 <i>Automobiles. Motor Vehicles</i>, you need not individually specify the narrower classes HD9710.3 <i>Automobile supply ... industry</i>, HD9710.33 <i>Limousines</i>, HD9710.34 <i>Buses</i>, etc. The query formulation is the OR combination of all the classes or headings you list. (AND combinations are out: for shelving, only one class is assigned; while more than one class could be assigned for a classified catalog, LCC is not designed for combining class numbers in searching.)</p> <p>You may use the print version or the electronic version for all documents and queries. Questions H, <i>canals</i>, and I, <i>helicopters</i> (optional) are part of the Classification Web exercise. If you wish, you may try them in the print version as well for comparison.</p> | HD 9710-9710.37 | Automobiles. Motor Vehicles | HD 9710 | General works | HD 9710.3 | Automobile supply ... industry | HD 9710.33 | Limousines | HD 9710.34 | Buses | HD 9710.35 | Trucks | HD 9710.37 | Recreational vehicles | HD 9710.38 | Trailers |
| HD 9710-9710.37 | Automobiles. Motor Vehicles | | | | | | | | | | | | | | | | |
| HD 9710 | General works | | | | | | | | | | | | | | | | |
| HD 9710.3 | Automobile supply ... industry | | | | | | | | | | | | | | | | |
| HD 9710.33 | Limousines | | | | | | | | | | | | | | | | |
| HD 9710.34 | Buses | | | | | | | | | | | | | | | | |
| HD 9710.35 | Trucks | | | | | | | | | | | | | | | | |
| HD 9710.37 | Recreational vehicles | | | | | | | | | | | | | | | | |
| HD 9710.38 | Trailers | | | | | | | | | | | | | | | | |

Example of query formulation with LCC

Query topic: *Teaching of science at all levels*

LB1532 Education > Theory and practice of education > Primary education > Branches of study > Special > Nature study. Science

LB1585-1585.7 Education > Theory and practice of education > Elementary or public school education > Special branches > Nature study. Science

Q160-161.2 Science > Science (general) > Textbooks

Q181.A1-183.4 Science > Science (general) > Study and teaching

Q190-197 Science > Science (general) > Audiovisual aids in science teaching

QB61-62.7 Science > Astronomy > Study and teaching. Research

QC20.8-20.82 Science > Physics > Mathematical physics > Study and teaching. Research

QC23 Science > Physics > **Elementary** textbooks

QC30-48 Science > Physics > Study and teaching. Research

etc.: Under every field of science (and some of their subfields), there are subdivisions such as *Textbooks* and *Study and teaching. Research*

Note on the example: As in Dewey, the combination order is as follows:

For primary and elementary education: Education – Subject

For secondary or higher education: Subject – Education

(there are no subject-specific classes in LB1603-169.6
Secondary education. High schools).

But LCC has some exceptions to this general rule, as in the example of QC23.

Your own notes on query formulation with LCC

Library of Congress Classification

Broad Outline (Main classes)

| | |
|-----|---|
| A | General works |
| B | Philosophy. Psychology. Religion |
| C | Auxiliary sciences of history |
| D | History: General and outside the Americas |
| E-F | History of America |
| E | History: America General and United States General |
| F | History: United States local, Canada, and Latin America |
| G | Geography |
| H | Social sciences |
| J | Political science |
| K | Law |
| L | Education |
| M | Music and books on music |
| N | Fine arts |
| P | Language and literature |
| Q | Science |
| R | Medicine |
| S | Agriculture |
| T | Technology |
| U | Military science |
| V | Naval science |
| Z | Bibliography and library science |

The following pages give first a detailed outline and then examples of classes dealing with or relevant to *transportation and traffic*.

In the detailed outline, each group of classes indicated by [on the left margin is in one volume of LCC.

Library of Congress Classification. Detailed Outline

A General works

- + AC Collections. Series. Collected works
- * AE Encyclopedias (General)
- * AG Dictionaries and other General reference works
- * AI Indexes (General)
- * AM Museums (General). Collectors and collecting (General)
- * AN Newspapers
- * AP Periodicals
- * AS Academies and learned societies (General)
- * AY Yearbooks. Almanacs. Directories
- . AZ History of scholarship and learning. The humanities

B Philosophy. Psychology. Religion

B-BJ Philosophy, incl. BF Psychology

- + B Philosophy (General)
- * BC Logic
- * BD Speculative philosophy
- * BF Psychology. Parapsychology. Occult sciences
- * BH Aesthetics
- . BJ Ethics. Social usages. Etiquette
- BL-BX Religion**
- + BL Religions. Mythology. Rationalism
- * BM Judaism
- * BP Islam. Bahaism. Theosophy
- . BQ Buddhism
- BR-BX Christianity
- + BR Christianity
- * BS The Bible
- * BT Doctrinal theology
- . BV Practical theology
- [BX Christian denominations

C Auxiliary sciences of history

- + C Auxiliary sciences of history (General)
- * CB History of civilization
- * CC Archaeology (General)
- * CD Diplomatics. Archives. Seals
- * CE Technical chronology. Calendar
- * CJ Numismatics
- * CN Inscriptions. Epigraphy
- * CR Heraldry
- * CS Genealogy
- . CT Biography [General]

D History: General and Old World

- + D History (General). Europe (General)
- * DA Great Britain
- * DAW Central Europe
- * DB Austria, Hungary, Czech Republic, Slovakia
- * DC France
- * DD Germany
- * DE Mediterranean region. Greco Roman World
- * DF Greece
- * DG Italy
- * DH Netherlands (low Countries). Belgium, Luxemburg
- . DJ Netherlands (Holland)
- + DJK Eastern Europe
- . DK Russia and former Soviet republics. Poland
- + DL Northern Europe. Scandinavia
- * DP Spain. Portugal
- * DQ Switzerland
- . DR Balkan peninsula
- [DS Asia
- + DT Africa
- * DU Oceania (South Seas)
- * [Australia. New Zealand]
- . DX Roma (Gypsies)

E-F History of America

- + E1-143 America (General)
- * E151-857 United States (Gen.)
- * F1-957 United States: States and Local
- * F1001-1140 Canada
- . F1201- Other individual countries [mostly Latin America]

G Geography

G - GF Geography

- + G Geography (General). Atlases. Maps
- * GA Mathematical geography. Cartography
- * GB Physical geography
- * GC Oceanography
- * GE Environmental sciences
- * GF Human ecology. Anthropogeography
- * GN Anthropology
- * GR Folklore
- * GT Manners and customs (General)
- . GV Recreation. Leisure

H Social sciences

- + H Social sciences (General)
- * HA Statistics
- HB-HJ Economics**
- * HB Economic theory. Demography
- * HC- Economic history and conditions
- * HD conditions
- * HE Transportation and communication
- * HF Commerce
- * HG Finance
- * HJ Public finance
- HM-HX Sociology**
- * HM Sociology (General and theoretical)
- * HN Social history. Social problems. Social reform
- * HQ The family. Marriage. Woman
- * HS Societies: secret, benevolent, etc. Clubs
- * HT Communities. Classes. Races
- * HV Social pathology. Social and public welfare. Criminology
- * HX Socialism. Communism. Anarchism

J Political science

- + J General legislative and executive papers
- * JA-JC Political science
- * JA Collections and general works
- * JC Political theory
- * JF-JQ Political institutions and public administration
- * JF General works. Comparative works
- * JK United States
- * JL Brit. America. Latin America
- * JN Europe
- * JQ Asia. Africa. Australia. Oceania
- * JS Local government
- * JV Colonies and colonization. Emigration and Immigration
- . JX International law. International relations
- No longer used at LC

K Law

- [K Law (General)
- [KD United Kingdom and Ireland
- [KDZ America. N.Am. (w/ KG,KH)
- [KE Canada
- [KF United States
- + KG Central America, Caribbean
- . KH South America
- [KJ-KKZ Europe (3 vols., KJV-KJW France, KK-KKC Germany)
- [KL-KWX [The rest of the world]

L Education

- + L Education (General)
- * LA History of education
- * LB Theory and practice of educ.
- * LC Special aspects of education
- * LD-LG Individual institutions
- * LD United States
- * LE America except United States
- * LF Europe
- * LG Asia, Africa, Oceania
- * LH Coll. & school mag. & papers
- * LJ Student fraternities and societies in the United States
- . LT [Multi-subject] Textbooks

M Music and books on music

- + M Music [instrumental and vocal]
- * ML Literature of music
- . MT Musical instruction and study

N Fine arts

- + N Visual arts (General)
- * NA Architecture
- * NB Sculpture
- * NC Drawing. Design. Illustration
- * ND Painting
- * NE Print media
- * NK Decorative arts. Applied arts. Decoration and ornament
- . NX Arts in General

P Language and literature (l&l)

- + P Philology and linguistics (Gen.)
- . PA Classical l&l (w. supplement)
- . PB-PH Modern European lang.
- + PB Celtic languages and literature
- * PC Romance languages
- * PD-PF Germanic languages
- * PD Scandinavian. North Germanic
- * PE English
- * PF West Germanic
- * PG Slavic. Baltic. Albanian l&l
- . PH Finno-Ugrian. Basque l&l
- . PJ-PL Oriental languages & lit.
- + PJ Oriental. Semitic
- . PK Indo-Iranian
- + PL L&l E. Asia, Africa, Oceania
- . PM Hyperborean, Indian, and artificial languages
- [Index to languages & dialects
- . **PN-PZ Literature**
- + PN Literary history and collections
- [PQ French & Romance lit. (1&2)
- / PR English literature
- / PS American literature
- [PT German & Germanic lit. (1&2)
- . PZ Children's literature

Q Science

- + Q Science (General)
- * QA Mathematics.
- * [Computer science]
- * QB Astronomy
- * QC Physics
- * QD Chemistry
- * QE Geology
- * QH-QR Biology
- * QH Natural history (General). Biology (General)
- * QK Botany
- * QL Zoology
- * QM Human anatomy
- * QP Physiology
- . QR Microbiology

R Medicine

R-RL Medicine

- + R Medicine (General)
- * RA Public aspects of medicine
- * RB Pathology
- * RC Internal medicine. Practice of Medicine
- * RD Surgery
- * RE Ophthalmology
- * RF Otorhinolaryngology
- * RG Gynecology and obstetrics
- * RJ Pediatrics
- * RK Dentistry
- * RL Dermatology
- * **RM-RZ Allied disciplines**
- * RM Therapeutics. Pharmacology
- * RS Pharmacy and materia medica
- * RT Nursing
- * RV Botanic, Thomsonian, and eclectic medicine
- * RX Homeopathy
- . RZ Other systems of medicine [Chiropractic. Osteopathy. Mental healing]

S Agriculture

- + S Agriculture (General)
- * SB Plant culture
- * SD Forestry
- * SF Animal culture
- * SH Aquaculture. Fisheries. Angling
- . SK Hunting

T Technology

- + T Technology (General)
- * **TA-TH General engineering and civil engineering**
- * TA General
- * TC Hydraulic and ocean eng.
- * TD Environmental technology, sanitary engineering
- * TE Highway engineering
- * TF Railroads
- * TG Bridge engineering
- * TH Buildings
- * **TJ-TL Mechanical group**
- * TJ Mechanical engineering
- * TK Electrical engineering. Nuclear engineering
- * TL Motor vehicles. Aeronautics. Astronautics
- * **TN-TR Chemical group**
- * TN Mining, metallurgy
- * TP Chemical technology
- * TR Photography
- * **TS-TX Composite group**
- * TS Manufactures
- * TT Arts and crafts. Handicrafts
- . TX Home economics

U Military science

- + U Military science (General)
- * UA Armies: Organization, description, facilities, etc
- * UB Military administration
- * UC Maintenance and transportation
- * UD Infantry
- * UE Cavalry, armor
- * UF Artillery
- * UG Military engineering. Air forces. Air warfare
- . UH Other services

V Naval science

- + V Naval science (General)
- * VA Navies: Org., descr., fac., etc
- * VB Naval administration
- * VC Naval maintenance
- * VD Naval seamen
- * VE Marines
- * VF Naval ordnance
- * VG Minor services of navies
- * VK Navigation. Merchant marine
- . VM Naval engineering. Ship building. Marine Engineering

Z Bibliography and library science.

- + Z4-115 Books (General). Writing. Paleography
- * Z116-659 Book industry & trade
- * Z662-1000 Libraries. [Library science. Information science]
- . Z1001-8999 Bibliography

LCC Education outline

LCC sample pages inserted here

LCC sample pages end on even

Assignment 13.2 DDC 2
Dewey Decimal Classification (DDC 2)

Assigned: July 16
Due: July 23

| | |
|-------------------------------|---|
| Objectives | Inherited from Assignment 13 (page 121) |
| Deliverables to submit | Not inherited. The sheets with the task results |
| Tasks | Not inherited For each topic give the Dewey classes (as expressed by its number) under which one should look to find relevant documents. Give the full class caption, such as 388.34 232 Ground transportation > Vehicular transportation > Vehicles > Taxicabs Also comment briefly on how you found those classes, how difficult it was (some of the classes may be easy to find, others more difficult) Comment on how easy or difficult it would be for a student to find the proper place(s) on the shelves; distinguish by level: elementary school / middle school / high school . |
| Materials | DDC volumes in the student lounge and Baldy 14A Web Dewey through Connexion |
| Readings | None |
| Time | 6 hours |

Dewey Decimal Classes for School Library Media topics

For each topic, find the applicable Dewey class. Some classes may need to be “built” to get more specific.

- 1. Drawing books**
- 2. Cars**
- 3. Cows**
- 4. Record books (like the Guinness World Record books)**
- 5. Army/military books**
- 6. Dinosaurs**
- 7. Halloween or scary books**
- 8. Volcanoes outside the United States**
- 9. Gluten-free cookbooks for kids**
- 10. The roles women had during the Civil War**
- 11. Continents**
- 12. Rivers**

More questions

Easier to identify in my collection

A1 Iwo Jima

A2 D-Day

A3 Hiroshima

A4 Freedom Riders (civil rights)

A5 Counterculture movement

More challenging

A6 Teens in the Holocaust (depends on the collection, for me this is more in Easier to identify)

A7 Invasion of Poland

A8 Pop culture trends during 60s/70s

A9 Effect of Star Trek on the original generation

A10 Warren Commission (why formed, its report, and its effect)

A11 AND book projects for same class started this week
Criteria choose a book (fiction or nonfiction, alternative history) for 1900-1940 (not including WWII)

A12 The gifted teacher at my school had kids studying major rivers of the world.
Kids developed questions and then came to the LMC to try to find answers to things Like:

how deep is this river at it's deepest?

What vegetation can be found along this river?

How did the river impact the civilization.

And more (this was a 4thgrade gifted group) it was a nightmare for all and of course the students thought the LMC had no resources to help them.

Some I answered last year:

A13 Weapons in ancient Egypt.

A14 Compare types of transportation during the 19th Century.

A15 The World Trade Center before it collapsed?

A16 Chupacabra

A17 La Llorona. (My school was predominantly Hispanic!)

A18 I have a funny to share, it just happened yesterday. A 3rd grade girl came up to me and whispered, "I need your help." When I asked what she needed, she told me that she was interested in history. When I asked her what part of history she was interested in, she told me "The history of fairies." I tried to explain to her that I could probably find her a book on fairies, it would be in the fiction section, not the non-fiction area. In the meantime, I got called away. She went to the media paraprofessional and asked the same question, basically got the same reply. When I returned, she got rather loud and said, "Seriously, you don't have a book on the history of fairies?"

A19 Community helper books.

A25 Science experiment books.

A26 I need all of your gingerbread man books. (Some are in E, FIC, as well as nonfic.)

A27 One question I had no idea where to find, where can you find books about critiques of authors?

From teachers:

A28 Animal adaptations

A29 State history

A30 Culture for second grade readers (this one stumped me for a minute!)

Outline for the analysis of Knowledge Organization Systems

For some items, a section number from Soergel, Organizing information is given in ()

1. **Purpose**

- 1.1 **Information system** or type of information system in which to be used
- 1.2 **Intended for** controlled vocabulary indexing G or query term expansion G (Ch. 12, Introduction)
- 1.3 **Type of file and search mechanism** for which originally designed
Shelving G Card catalog / printed index G Online system G

2. **Coverage and designation of concepts. Coverage and format of terms**

- 2.1 **Concepts: Scope**, breadth of coverage. Recency of concepts
- 2.2 **Concepts: Specificity**, depth of coverage (Section 16.2.2). Coverage at each level of specificity.
- 2.3 Are all necessary **facets** included? Concepts formed in semantic factoring and facet analysis? (S.a. 3.1)
- 2.4 **Nature of notation** (if none, state that) (Section 15.5.2)
- 2.5 **Terms**: Completeness of coverage of terminology (completeness of lead-in vocabulary). Recency of terms
- 2.6 **Form of terms**: Consistency, adherence to common usage.

3. **Terminological and conceptual analysis and conceptual structure.**

- 3.1 **Quality of conceptual structure** (14): Facet analysis. Types and degree of differentiation of conceptual relationships included. For each type indicate the completeness of inclusion. (Fill in 3.1.1 - 3.1.3)
 - 3.1.1 Expression of concepts through elemental concepts (closely related to definition)
 - 3.1.2 Hierarchical relationships (polyhierarchy). (Shown by arrangement or Broader Term / Narrower Term X-ref)
 - 3.1.3 Associative relationships (Implied by physical proximity in the arrangement or explicit Related Term X-ref)
- 3.2 **Quality of definitions**, explications, scope notes (correctness, detail, clarity).
- 3.3 Completeness of terminological relationships: Does the vocabulary contain terms that are synonymous or quasi-synonymous without indicating the relationship?

4. **Use of precombination in the index language** (concerns both 2 and 3) (Ch. 14, 15, esp. 15.4)

4.1 To what degree are the final descriptors assigned to documents, whether enumerated or built by the indexer, precombined?

4.2 To what extent are precombined descriptors enumerated and/or given in the alphabetical index?

To what extent can the indexer build additional precombined descriptors?

Are precombined descriptors designated by an independent symbol or a string of symbols? Combination order free or fixed? To what extent do the components of a precombined descriptor determine its place in the arrangement? (Relates also to the arrangement of a classification) (Section 15.5.2)

5. **Access and display. Format of presentation of the vocabulary**

For each format consider access/retrieval by concepts versus access/retrieval by terms.

Access can be provided through arrangement in a printed document or through a computer search system.

5.1 **Format of printed document** (Fill in 5.1.1- 5.1.3)

5.1.1 Overall format: Thesaurus parts and information given in each, connections between them. Is the overall format clear and helpful for finding the appropriate concepts and terms or notations in indexing and query formulation?

5.1.2 Display of conceptual relationships (Broader Term, Narrower Term, Related Term)

- through linear arrangement or graphical display (Section 15.5.2)

- through cross-references (Section 14.1)

- through descriptor-find index (Section 15.5.1)

How well does the display reflect the conceptual analysis, e.g., sequence of concepts on the same hierarchical level (sequence of the children of a concept, that is, the concepts one level further down).

5.1.3 Display of terminological relationships (Synonymous Term)

5.2 **Access through computer system.** Navigation. Format of on-line displays

Assignment 13.3 *Assigned: July 23*
Due: July 30
Library of Congress/Sears Subject Headings

| | |
|-------------------------------|--|
| Objectives | Inherited from Assignment 13 (page 121) |
| Deliverables to submit | The filled-in document forms and query forms (worksheet is just a reading) The filled-in <i>Outline for the analysis of Knowledge Organization Systems</i> |
| Tasks | Inherited from Assignment 13 (page 121) Complete the worksheet for LCSH after indexing and query formulation with LCSH. |
| Materials | <p>Some sample pages dealing with <i>Education</i> are included. Many sample pages for <i>Transportation</i> included with Lecture 8.1. Explorations in subject access.</p> <p>Library of Congress Subject Headings. 33. ed. 2011, 6 large red volumes. In Lockwood Library Gov Docs. Can use earlier editions back to 1990 in Baldy 14A</p> <p>Sears List of Subject Headings. 20. ed., 2010. Much smaller and simpler than LCSH. If you are interested in school media centers or public libraries, you can use SSH instead of LCSH for indexing and query formulation. Online: www.ebscohost.com/academic/sears-list-of-subject-headings (free trial)</p> <p>We have already examined LCSH in Small Groups 1, Explorations in Subject Access. Classification Web (part of Cataloger's Desktop, Intro. in Ass. 13.2 LCC) provides access to LCC and LCSH.</p> <p>Or search for LCSH in OCLC Connexions (see Ass. 13.1, WebDewey). In the top navigation bar, click on Authorities; you are on your own.</p> |
| Readings | <p>Beforehand: Chan, Cataloging and classification, Chapter 8 on LCSH Needham, Ch. 10, The alphabetic subject catalog, p. 199-223 (optional)</p> <p>For further study</p> <p>Chan, Lois Mai 1995 Library of Congress Subject Headings. principles of structure and application. 3. ed. Englewood, CO: Libraries Unlimited; 1995. Z695.Z8 L5226 1995 in Lockwood Library</p> <p>Perreault, Jean M. 1979 Library of Congress Subject Headings: A New Manual. International Classification 1979 November; 6(3):158-169. Extensive review of an earlier version of Chan's book. Gives a good feel for some of the problems in LCSH.</p> |
| Time | 5 hours |

Worksheet for LCSH: Comparison example (LCSH and DDC or LCC)

This worksheet is only a reading; just read and understand. This is only for LCSH, not SSH

The following example will give you a sense of the nature of the relationship (or lack of it) between the subject headings and classification schemes.

In the example, I searched WorldCat (see Assignment 3) for the sample query *Teaching of science at all levels* and looked at the call numbers for relevant documents found to see whether I could find LC classes beyond the ones found by examining the LC classification.

Here is a list of class numbers found by examining LC classification for this topic.

Query topic: *Teaching of science at all levels*

LB1532 Primary education > Branches of study > Special > Nature study. Science

LB1585-1585.7 El. or public school educ. > Special branches > Nature study. Science

Q160-161.2 Science (general) > Textbooks

Q181.A1-183.4 Science (general) > Study and teaching

Q190-197 Science (general) > Audiovisual aids in science teaching

QB61-62.7 Astronomy > Study and teaching. Research

QC20.8-20.82 Physics > Mathematical physics > Study and teaching. Research

QC23 Physics > **Elementary** textbooks

QC30-48 Physics > Study and teaching. Research

etc.: Under every field of science (and some of their subfields), there are subdivisions such as *Textbooks* and *Study and teaching. Research*

Since this covers only general science and physics, I first searched for

(su:Science OR su:Physics OR su:Astronomy)
AND su:study AND su:teaching [to get the subheading *study and teaching*]

Relevant documents found were assigned the following class numbers (new ones are **bold**):

LB 1065 Educational psychology > Learning > Interest. Attention. Motivation
LB 1585
Q 161.2
Q 181 (3x)
Q 182.3 (note: book about elementary level!) (2x)
Q 183.4
QA 76.88 Mathematics > Instruments and machines > Calculating machines > Electronic computers. Computer science > Supercomputers. High performance computing.
QB 501.5 Astronomy > Descr. astronomy /Solar system > Study and teaching. Research
QH 541.14 Biology (General) > Ecology > Juvenile works (note: book is on primary ed.)
QP 251.5 Physiology > Urinary and reproductive organs > Reproduction. Physiology of sex > Juvenile works (note: book is on primary level)

Then, to find more class numbers, I searched under

(su:Chemistry OR su:Biology OR su:Geology)
AND su:study AND su:teaching

This turns up many more class numbers, all new:

GC31.3 Oceanography > Study and teaching > Audiovisual aids
LD6241.7 1999 Individual institutions > U S > Universities. Colleges > Wright State Univ., Dayton, OH
QD31.2 Chemistry > General works, treatises, and advanced textbooks > 1970 -
QD40 Chemistry > Study and teaching. Research > General works
QD43 Chemistry > Study and teaching. Research > Experiments
QH308.7 (2x) Biology > Textbooks > **Elementary** > 1970 - [elementary educ. by subject usually in LB]
QH315 (2x) Biology > Study and teaching. Research > General works
QH316.4 Biology > Study and teaching. Research > Activity programs
QH362 Biology > Evolution > Study and teaching. Research
QH541.15.L35 Biology > Ecology > Special aspects of the subject as a whole > Landscape ecology

Outline for the analysis of Knowledge Organization Systems

For some items, a section number from Soergel, Organizing information is given in ()

1. Purpose

- 1.1 **Information system** or type of information system in which to be used
- 1.2 **Intended for** controlled vocabulary indexing G or query term expansion G (Ch. 12, Introduction)
- 1.3 **Type of file and search mechanism** for which originally designed
Shelving G Card catalog / printed index G Online system G

2. Coverage and designation of concepts. Coverage and format of terms

- 2.1 **Concepts: Scope**, breadth of coverage. Recency of concepts
- 2.2 **Concepts: Specificity**, depth of coverage (Section 16.2.2). Coverage at each level of specificity.
- 2.3 Are all necessary **facets** included? Concepts formed in semantic factoring and facet analysis? (S.a. 3.1)
- 2.4 **Nature of notation** (if none, state that) (Section 15.5.2)
- 2.5 **Terms**: Completeness of coverage of terminology (completeness of lead-in vocabulary). Recency of terms
- 2.6 **Form of terms**: Consistency, adherence to common usage.

3. Terminological and conceptual analysis and conceptual structure.

- 3.1 **Quality of conceptual structure** (14): Facet analysis. Types and degree of differentiation of conceptual relationships included. For each type indicate the completeness of inclusion. (Fill in 3.1.1 - 3.1.3)
 - 3.1.1 Expression of concepts through elemental concepts (closely related to definition)
 - 3.1.2 Hierarchical relationships (polyhierarchy). (Shown by arrangement or Broader Term / Narrower Term X-ref)
 - 3.1.3 Associative relationships (Implied by physical proximity in the arrangement or explicit Related Term X-ref)
- 3.2 **Quality of definitions**, explications, scope notes (correctness, detail, clarity).
- 3.3 Completeness of terminological relationships: Does the vocabulary contain terms that are synonymous or quasi-synonymous without indicating the relationship?

4. **Use of precombination in the index language** (concerns both 2 and 3) (Ch. 14, 15, esp. 15.4)

4.1 To what degree are the final descriptors assigned to documents, whether enumerated or built by the indexer, precombined?

4.2 To what extent are precombined descriptors enumerated and/or given in the alphabetical index?

To what extent can the indexer build additional precombined descriptors?

Are precombined descriptors designated by an independent symbol or a string of symbols? Combination order free or fixed? To what extent do the components of a precombined descriptor determine its place in the arrangement? (Relates also to the arrangement of a classification) (Section 15.5.2)

5. **Access and display. Format of presentation of the vocabulary**

For each format consider access/retrieval by concepts versus access/retrieval by terms.

Access can be provided through arrangement in a printed document or through a computer search system.

5.1 **Format of printed document** (Fill in 5.1.1- 5.1.3)

5.1.1 Overall format: Thesaurus parts and information given in each, connections between them. Is the overall format clear and helpful for finding the appropriate concepts and terms or notations in indexing and query formulation?

5.1.2 Display of conceptual relationships (Broader Term, Narrower Term, Related Term)

- through linear arrangement or graphical display (Section 15.5.2)

- through cross-references (Section 14.1)

- through descriptor-find index (Section 15.5.1)

How well does the display reflect the conceptual analysis, e.g., sequence of concepts on the same hierarchical level (sequence of the children of a concept, that is, the concepts one level further down).

5.1.3 Display of terminological relationships (Synonymous Term)

5.2 **Access through computer system.** Navigation. Format of on-line displays

Instructions for using Library of Congress/Sears Subject Headings

| | |
|---|--|
| <p>Indexing with LCSH/SSH</p> | <p>Give all appropriate subject headings. If applicable, list other candidate headings and give reasons why you chose the headings you did. According to general practice, a subject heading should represent the content or relevance of the document as a whole. The number of subject headings assigned to a document varies between 1 and 5, with 2 or 3 being most common.</p> |
| <p>Query formulation with LCSH/SSH</p> | <p>Try to list all subject headings where one should look for relevant documents; if there are more than 7 subject headings, list a representative sample (enough to demonstrate that you know how to find all applicable subject headings).</p> <p>Note: Listing a main heading implies all the subordinate main heading - subheading combinations (which can be readily seen from the subject heading list). If you give just a sample fo subject headings to be used, prefer headings from different sections of the alphabet. For example, in the <i>canal</i> search, give just one heading starting with <i>canal</i>, make a note that there are many headings starting with <i>canal</i>, and find relevant headings elsewhere.</p> <p>The query formulation is the OR combination of all the headings you list. (AND combinations are out: while multiple subject headings can be assigned, LCSH/SSH is not designed for combination searching, and many online catalogs do not allow for it.)</p> <p>You may use the print version or the electronic version for documents A-C and queries D-E. Query F, <i>canals</i>, is part of the Assignment 13.4LCC exercise on Classification Web. You may try them in the print version as well for comparison.</p> |

Example of query formulation with LCSH (XXX construct some examples for SSH)

| Query topic: <i>Teaching of science at all levels</i> | |
|--|---|
| Query in online catalog | <p>In an online catalog in which one can search for phrases in subject headings and use Boolean AND, the following query formulation would find a large portion of the relevant documents:</p> <p>(Science OR Nature study OR Physics OR Chemistry OR Biology OR ...) AND Study and teaching</p> <p>One might further look under</p> <p>Nature study — Activity programs Science students Science teachers</p> <p>For a list of illustrative full headings, see the next box</p> |
| A sampling of illustrative LC subject headings | <p>Science – Study and teaching Science – Study and teaching – Activity programs Science – Study and teaching – Aids and devices Science – Study and teaching – Supervision Science – Study and teaching (Early childhood) Science – Study and teaching (Elementary) Science – Study and teaching (Graduate) Science – Study and teaching (Higher) Science – Study and teaching (Internship) Science – Study and teaching (Preschool) Science – Study and teaching (Primary) Science – Study and teaching (Secondary) Science – Study and teaching (Undergraduate)</p> <p>This pattern of subdivision is used with branches and sub-branches of science, for example</p> <p>Biology – Study and teaching plus many more possible Chemistry – Study and teaching Nature study – Study and teaching Physics – Study and teaching</p> <p>One needs to compile a complete list of all branches and sub-branches of science. Following several steps of NT cross-reference starting from <i>Science</i> helps with this.</p> |

Sample pages from LCSH here

Sample pages end on even page

Assignment 13.4

Assigned: July 23
Due: July 30

ERIC Thesaurus

| | |
|-------------------------------|--|
| Objectives | Inherited from Assignment 13 (page 121) |
| Deliverables to submit | Inherited from Assignment 13 (page 121) |
| Tasks | Complete the worksheet (use attached sample pages for this task). For indexing and query formulation you will need the full thesaurus. |
| Materials | Sample pages for the ERIC thesaurus are attached. Full thesaurus available in Lockwood. (14th ed. is the newest, print version seems no longer to be updated, sample pages are from an earlier edition.) You can search ERIC online at http://www.eric.ed.gov |
| Time | 3 hours |

Name:

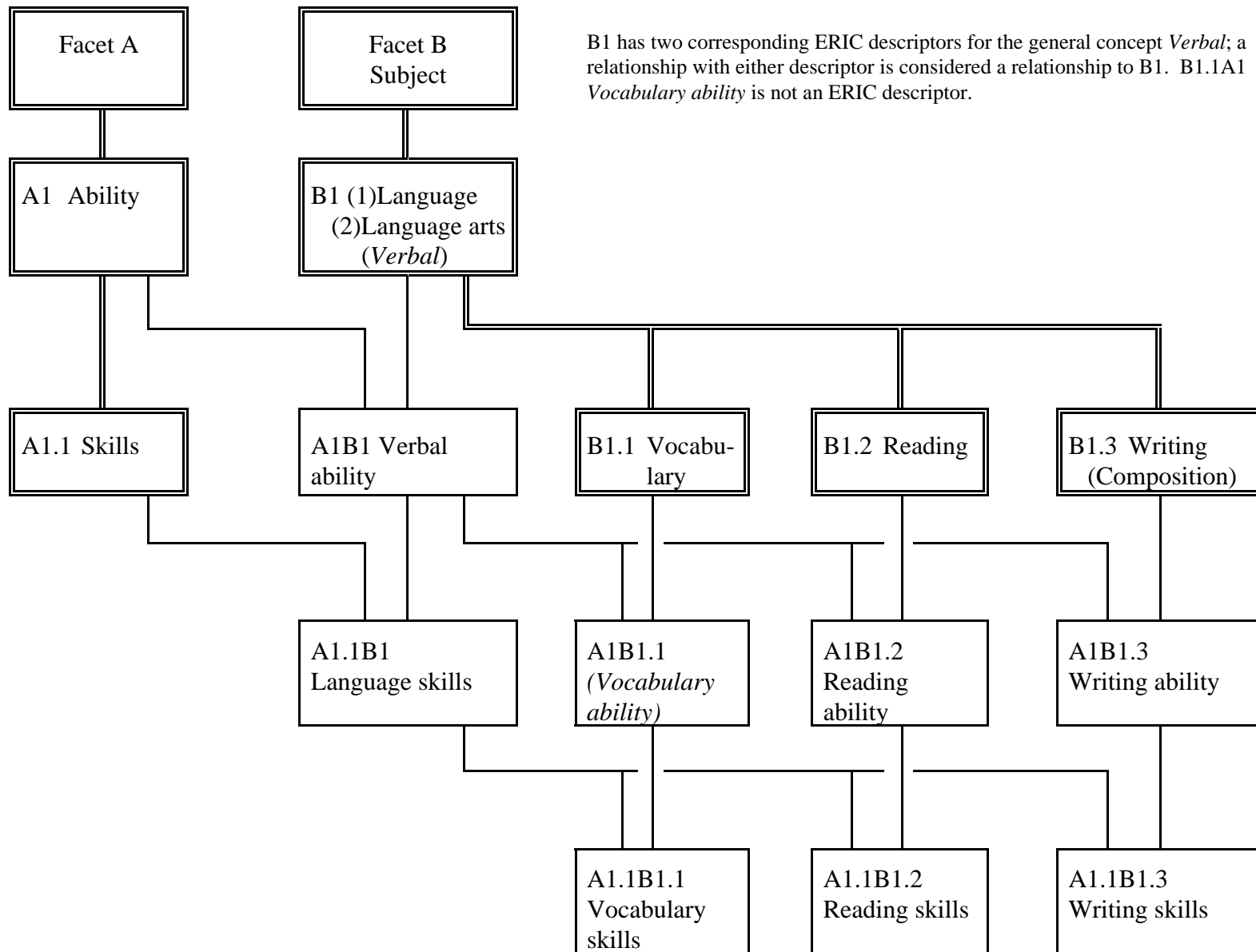
Mailbox:

ERIC worksheet

| | |
|------------------------|--|
| Objectives | Understand the application of facet analysis to the analysis of concepts and concept relationships in a thesaurus. |
| Task background | <p>On the next page is an expanded version of the answer sheet for Assignment 12.3a, a hierarchy in graphical representation. The terms in the boxes are actual ERIC descriptors unless indicated otherwise; these descriptors make up the study set for this worksheet. On pages 178 - 180 (or thereabouts) is an excerpt from the ERIC Thesaurus that shows these descriptors with their cross-references. In the long lists of cross-references, descriptors from the study set are bold so they can be located easily.</p> <p>Note. Concept B1 <i>Verbal</i> in the diagram conflates two ERIC descriptors that address basically the same concept:</p> <p style="padding-left: 40px;">(1) <i>Language</i> and (2) <i>Language arts</i>.</p> <p>A relationship to or from either one of these is considered a relationship to or from B1.</p> <p>A number of other ERIC descriptors that fall into this “conceptual grouping” have not been included in the example to keep it simple. Some of these concepts are <i>Language proficiency</i>, <i>Linguistic competence</i>, <i>Communicative competence (languages)</i>.</p> |
| Tasks | <ol style="list-style-type: none"> 1 For each hierarchical relationship shown in the graph, find out whether it is given in the ERIC Thesaurus, either as BT/NT or as RT, and label the line in the graph accordingly. If a hierarchical relationship does not appear in the ERIC Thesaurus at all, label the line NO. 2 Write a brief essay (max. ½ page) on how the ERIC Thesaurus handles these relationships. |
| Result | <ol style="list-style-type: none"> 1 The labeled hierarchy graph. 2 A brief essay on relationships in the ERIC Thesaurus |

Analysis of ERIC descriptors and relationships using facets

B1 *Verbal* has two corresponding ERIC descriptors; a relationship with either descriptor is considered a relationship to B1. B1.1A1 *Vocabulary ability* is not an ERIC descriptor.



Outline for the analysis of Knowledge Organization Systems

For some items, a section number from Soergel, Organizing information is given in ()

| | |
|----|--|
| 1. | <p>Purpose</p> <p>1.1 Information system or type of information system in which to be used</p> <p>1.2 Intended for controlled vocabulary indexing G or query term expansion G (Ch. 12, Introduction)</p> <p>1.3 Type of file and search mechanism for which originally designed Shelving G Card catalog / printed index G Online system G</p> |
| 2. | <p>Coverage and designation of concepts. Coverage and format of terms</p> <p>2.1 Concepts: Scope, breadth of coverage. Recency of concepts</p> <p>2.2 Concepts: Specificity, depth of coverage (Section 16.2.2). Coverage at each level of specificity.</p> <p>2.3 Are all necessary facets included? Concepts formed in semantic factoring and facet analysis? (S.a. 3.1)</p> <p>2.4 Nature of notation (if none, state that) (Section 15.5.2)</p> <p>2.5 Terms: Completeness of coverage of terminology (completeness of lead-in vocabulary). Recency of terms</p> <p>2.6 Form of terms: Consistency, adherence to common usage.</p> |
| 3. | <p>Terminological and conceptual analysis and conceptual structure.</p> <p>3.1 Quality of conceptual structure (14): Facet analysis. Types and degree of differentiation of conceptual relationships included. For each type indicate the completeness of inclusion. (Fill in 3.1.1 - 3.1.3)</p> <p>3.1.1 Expression of concepts through elemental concepts (closely related to definition)</p> <p>3.1.2 Hierarchical relationships (polyhierarchy). (Shown by arrangement or Broader Term / Narrower Term X-ref)</p> <p>3.1.3 Associative relationships (Implied by physical proximity in the arrangement or explicit Related Term X-ref)</p> <p>3.2 Quality of definitions, explications, scope notes (correctness, detail, clarity).</p> <p>3.3 Completeness of terminological relationships: Does the vocabulary contain terms that are synonymous or quasi-synonymous without indicating the relationship?</p> |

4. **Use of precombination in the index language** (concerns both 2 and 3) (Ch. 14, 15, esp. 15.4)

4.1 To what degree are the final descriptors assigned to documents, whether enumerated or built by the indexer, precombined?

4.2 To what extent are precombined descriptors enumerated and/or given in the alphabetical index?

To what extent can the indexer build additional precombined descriptors?

Are precombined descriptors designated by an independent symbol or a string of symbols? Combination order free or fixed? To what extent do the components of a precombined descriptor determine its place in the arrangement? (Relates also to the arrangement of a classification) (Section 15.5.2)

5. **Access and display. Format of presentation of the vocabulary**

For each format consider access/retrieval by concepts versus access/retrieval by terms.

Access can be provided through arrangement in a printed document or through a computer search system.

5.1 **Format of printed document** (Fill in 5.1.1- 5.1.3)

5.1.1 Overall format: Thesaurus parts and information given in each, connections between them. Is the overall format clear and helpful for finding the appropriate concepts and terms or notations in indexing and query formulation?

5.1.2 Display of conceptual relationships (Broader Term, Narrower Term, Related Term)

- through linear arrangement or graphical display (Section 15.5.2)

- through cross-references (Section 14.1)

- through descriptor-find index (Section 15.5.1)

How well does the display reflect the conceptual analysis, e.g., sequence of concepts on the same hierarchical level (sequence of the children of a concept, that is, the concepts one level further down).

5.1.3 Display of terminological relationships (Synonymous Term)

5.2 **Access through computer system.** Navigation. Format of on-line displays

Instructions for using the ERIC Thesaurus

| | |
|------------------------------------|---|
| Indexing with ERIC | Give all appropriate ERIC descriptors. According to common practice an ERIC descriptor can represent a topic treated in part of the document, but still somewhat extensively. 15 - 20 descriptors for a document is not uncommon. |
| Query formulation with ERIC | Try to list all categories where one should look for relevant documents (if more than 10, give representative examples); the query formulation might consist of just one concept or of two or more concepts ANDed. In ERIC, a concept may need to be represented by an OR combination of descriptors. Remember in particular that in ERIC a hierarchically expanded search requires ORing a descriptor with all its narrower descriptors. Do the best you can do in a reasonable time. |

Example of query formulation with the ERIC Thesaurus

| | |
|--|---|
| Query topic: <i>Teaching of science at all levels</i> OR combination of | |
| Science education Science education history Science instruction Science curriculum General science Science programs Science activities Science projects Science course improvement projects Science process skills | Science tests Elementary science Secondary school science College science Summer science programs Science teachers Science teaching centers Aerospace education Marine education Environmental education |
| <p>If indexing is done properly, this should find most everything (the list could be made more complete). But if an indexer has a document <i>The physics curriculum</i> and assigns simply <i>Physics</i> and <i>Curriculum</i> (instead of the correct <i>Science curriculum</i>), we are out of luck. We could compensate for such indexing errors by running a second search as follows:</p> <p style="padding-left: 40px;">(OR combination of all <i>science</i> descriptors) AND (OR combination of all <i>curriculum</i> and <i>teaching methods</i> descriptors)</p> <p>To find all science descriptors, we would start from <i>Natural sciences</i> and follow several steps of cross-references; same principle for <i>curriculum</i>. This second search would add a few relevant and many irrelevant documents.</p> | |

Excerpts from the ERIC Thesaurus 13th ed. 1995

You can search the ERIC thesaurus on the web under <http://www.eric.ed.gov>

ABILITY

- SN The degree of actual power present in an organism or system to perform a given physical or mental act (note: use a more specific term if possible)
- NT Academic ability
Cognitive ability
Competence
Language proficiency
Leadership
Nonverbal ability
Skills
Spatial ability
Verbal ability
- RT Ability grouping
Ability identification
Achievement
Aptitude
Aspiration
Difficulty level
Disabilities
Gifted
Performance
Productivity
Qualifications
Readiness
Talent

LANGUAGE

- SN Systematic means of communicating ideas and feelings through the use of signs, gestures, words, and/or auditory symbols (note: for natural languages and language families, see "languages")
- NT Artificial languages
Child language
Figurative language
Interlanguage
Language of instruction
Language universals
Languages for special purposes
Official languages
Oral language
Programming languages
Second languages
Sign language
Symbolic language
Tone languages
Uncommonly taught languages
Unwritten languages
Urban language
Written language
- RT Artificial speech
Code switching (language)
Dialects
Language acquisition
Language arts
Language attitudes

- Language enrichment
Language impairments
Language patterns
Language planning
Language processing
Language proficiency
Language research
Language rhythm
Language role
Language skill attrition
Language skills
Language styles
Language tests
Language universals
Language usage
Language variation
Languages
Linguistics
Onomastics
Semiotics
Social dialects
Speech
Speech communication
Verbal communication
Word frequency
Word order

LANGUAGE ARTS

- NT Debate
Handwriting
Listening
Outlining (discourse)
Reading
Rhetoric
Speech
Spelling
Story telling
Writing (composition)
- RT Child language
Communication (thought transfer)
Comprehension
Discourse modes
Dramatics
Editing
English curriculum
Journalism education
Language
Language experience approach
Language skills
Lexicology
Literature
Reading writing relationship
Self expression
Speech communication
Speech curriculum
Translation
Verbal ability
Verbal communication
Vocabulary
Whole language approach

LANGUAGE SKILLS

- BT **Skills**

- NT Audiolingual skills
Communicative competence (languages)
Reading skills
Vocabulary skills
Writing skills
- RT Basic skills
Cloze procedure
Communication skills
Daily living skills
Dictation
English (second language)
Error analysis (language)
Expressive language
Inferences
Interpreters
Interpretive skills
Language
Language acquisition
Language aptitude
Language arts
Language dominance
Language fluency
Language processing
Language proficiency
Language skill attrition
Languages
Linguistic competence
Linguistic performance
Listening comprehension
Metalinguistics
Monolingualism
Psycholinguistics
Receptive language
Second language learning
Sentence combining
Thinking skills
Translation
Verbal ability
Whole language approach
Word study skills

READING

- BT **Language arts**
Literacy
- NT Basal reading
Beginning reading
Content area reading
Corrective reading
Critical reading
Directed reading activity
Early reading
Functional reading
Independent reading
Individualized reading
Music reading
Oral reading
Reading aloud to others
Recreational reading
Remedial reading
Silent reading
Speed reading
Story reading
Sustained silent reading

- RT Advance organizers
Bibliotherapy
Braille
Cloze procedure
Context clues
Decoding (reading)
Diacritical marking
Informal reading inventories
Initial teaching alphabet
Inner speech (subvocal)
Language processing
Miscue analysis
Pattern recognition
Phoneme grapheme correspondence
Proofreading
Readability
Readability formulas
Reader response
Reader text relationship
Reading ability
Reading achievement
Reading assignments
Reading attitudes
Reading centers
Reading comprehension
Reading consultants
Reading diagnosis
Reading difficulties
Reading failure
Reading games
Reading habits
Reading improvement
Reading instruction
Reading interests
Reading material selection
Reading materials
Reading motivation
Reading processes
Reading programs
Reading rate
Reading readiness
Reading readiness tests
Reading research
Reading skills
Reading strategies
Reading teachers
Reading tests
Reading writing relationship
Tachistoscopes
Verbal communication
Vocabulary
Writing (composition)

READING ABILITY

- BT **Verbal ability**
NT **Reading skills**
RT **Reading**
Reading achievement
Reading attitudes
Reading difficulties
Reading readiness

READING SKILLS

- BT **Language skills**

Reading ability
 NT Reading comprehension
 Reading rate
 RT Adult literacy
 Basic skills
 Cloze procedure
 Content area reading
 Decoding (reading)
 Eye voice span
 Functional literacy
 Functional reading
 Inferences
 Literacy
 Literacy education
 Minimum competencies
 Phonics
 Reader response
Reading
 Reading centers
 Reading habits
 Reading improvement
 Reading instruction
 Reading processes
 Reading strategies
 Reading tests
 Thinking skills
Vocabulary skills
 Word recognition
 Word study skills

SKILLS
 BT **Ability**
 NT Agricultural skills
 Basic skills
 Business skills
 Communication skills
 Daily living skills
 Decision making skills
 Home economics skills
 Homemaking skills
 Information skills
 Interpretive skills
 Job skills
Language skills
 Locational skills (social studies)
 Map skills
 Mathematics skills
 Mechanical skills
 Minimum competencies
 Parenting skills
 Psychomotor skills
 Research skills
 Salesmanship
 Science process skills
 Study skills
 Teaching skills
 Thinking skills
 Visual literacy
 RT Competence
 Difficulty level
 Familiarity
 Interpersonal competence
 Mastery learning
 Mastery tests
 National competency tests

Process education
 Qualifications
 Skill analysis
 Skill centers
 Skill development

VERBAL ABILITY
 BT **Ability**
 NT **Reading ability**
Writing ability
 RT Academic ability
 Basic skills
 Communication skills
 Language aptitude
Language arts
Language skills
 Linguistic competence
 Linguistic input
 Linguistic performance
 Nonverbal ability
 Speech skills
 Verbal development
 Verbal learning
 Verbal operant conditioning
 Verbal tests

VOCABULARY
 NT Aviation vocabulary
 Banking vocabulary
 Basic vocabulary
 Chemical nomenclature
 International trade vocabulary
 Jargon
 Keywords
 Kinship terminology
 Mathematical vocabulary
 Medical vocabulary
 Sight vocabulary
 Subject index terms
 Word lists
 RT Adjectives
 Adverbs
 Code switching (language)
 Definitions
 Glossaries
 Glottochronology
Language arts
 Lexicology
 Linguistic borrowing
 Multilingual materials
 Nouns
Reading
 Thesauri
 Verbs
 Vocabulary development
Vocabulary skills
 Word frequency

VOCABULARY SKILLS
 BT **Language skills**
 RT Basic skills
 Basic vocabulary
 Context clues
Reading skills
Vocabulary

WRITING (COMPOSITION)
 BT **Language arts**
 Literacy
 NT Abstracting
 Basic writing
 Childrens writing
 Content area writing
 Creative writing
 Descriptive writing
 Expository writing
 Free writing
 Freshman composition
 Journal writing
 Local color writing
 News writing
 Notetaking
 Paragraph composition
 Parallelism (literary)
 Playwriting
 Proposal writing
 Technical writing
 Writing for publication
 RT Audience analysis
 Coherence
 Cohesion (written composition)
 Discourse modes
 Handwriting
 Language processing
 Letters (correspondence)
 Literary devices
 Literary styles
 Narration
 Outlining (discourse)
 Paragraphs
 Persuasive discourse
 Plagiarism
 Poetry
 Prewriting
 Prose
Reading
 Reading writing relationship
 Revision (written composition)
 Rhetoric
 Rhetorical invention
 Sentences
 Spelling
 Story grammar
 Student writing models
 Verbal communication
Writing ability
 Writing achievement
 Writing apprehension
 Writing assignments
 Writing attitudes
 Writing difficulties
 Writing evaluation
 Writing exercises
 Writing improvement
 Writing instruction
 Writing laboratories
 Writing processes
 Writing readiness
 Writing research
Writing skills

Writing strategies
 Writing teachers
 Writing tests
 Writing workshops
 Written language

WRITING ABILITY
 BT **Verbal ability**
 NT **Writing skills**
 RT Handwriting
Writing (composition)
 Writing achievement
 Writing attitudes
 Writing difficulties
 Writing readiness

WRITING SKILLS
 BT **Language skills**
Writing ability
 RT Adult literacy
 Audience awareness
 Basic skills
 Basic writing
 Capitalization (alphabetic)
 Childrens writing
 Cohesion (written composition)
 Content area writing
 Essay tests
 Functional literacy
 Grammar
 Handwriting
 Language styles
 Language tests
 Literacy
 Literacy education
 Minimum competencies
 Notetaking
 Outlining (discourse)
 Paragraph composition
 Parallelism (literary)
 Prewriting
 Proofreading
 Punctuation
 Revision (written composition)
 Rhetorical invention
 Sentence combining
 Spelling
 Story grammar
 Student journals
 Student writing models
 Text structure
 Thinking skills
Writing (composition)
 Writing evaluation
 Writing exercises
 Writing improvement
 Writing instruction
 Writing laboratories
 Writing processes
 Writing strategies
 Writing tests
 Writing workshops

Sample pages from ERIC

Ending on even page

Indexing forms Documents A - C

Query forms Queries D - F

Index 3 documents (A-C, all on education). For each document, a copy of the title page, the cover or jacket blurb, and the table of contents are provided (very end of Assignment Packet). Copies of the two books can be examined in Baldy 14A; the third document is a Web site, <http://forum.swarthmore.edu/mathmagic>.

For each document there is a *document indexing form*.
For each scheme, there are instructions for indexing.

Formulate 3 queries (D-F, 2 on education, 1 on transportation).

For each query there is a *query form*.
For each scheme, there are instructions for query formulation.

The indexing and query forms provide space for writing down the descriptors from all schemes for ease of comparison. Write down the class/descriptor number (if any) and the class/descriptor text, for example,

DDC: 371.5 Education > Elementary ed. > El. ed. in special subjects > Computers, science, technology, health > Science and tech.

ERIC: Elementary School Science

LCSH: Science - Study and teaching (elementary)

Yahoo: Science > Education > K-12 [el. school not available]

LCC: LB1535 Education > Theory and practice of education >
Elementary or public school education > Special branches > Nature study. Science >
General works

Add comments or observations on each scheme in the space for the scheme, spilling over to the comment area at the bottom if needed. Add comments or observations on the comparison of two or more schemes at the bottom

A Document indexing form. **Differentiating the teaching staff.** / James Lewis

Yahoo precombined class(es):

Elemental indexing concepts (components of the class):

DDC class:

Alternate class:

Hierarchical chain (in words):

Components (if class was built):

LCC class:

Alternate class:

Hierarchical chain (in words):

Components (if class was built):

LCSH or **SSH** headings (not more than 5)

ERIC descriptors (not more than 10)

Comments (on any of the schemes or on comparisons):

B Document indexing form. The open classroom reader. / Charles E. Silberman, editor 301

Yahoo precombined class(es):

Elemental indexing concepts (components of the class):

DDC class:

Alternate class:

Hierarchical chain (in words):

Components (if class was built):

LCC class:

Alternate class:

Hierarchical chain (in words):

Components (if class was built):

LCSH or **SSH** headings (not more than 5)

ERIC descriptors (not more than 10)

Comments (on any of the schemes or on comparisons):

C Document indexing form. **MathMagic** (The Web site <http://forum.swarthmore.edu/mathmagic>)
303

Yahoo precombined class(es):

Elemental indexing concepts (components of the class):

DDC class:

Alternate class:

Hierarchical chain (in words):

Components (if class was built):

LCC class:

Alternate class:

Hierarchical chain (in words):

Components (if class was built):

LCSH or **SSH** headings (not more than 5)

ERIC descriptors (not more than 10)

Comments (on any of the schemes or on comparisons):

D Query form. (1) Mexicans in (2) American schools

305

(1) Mexican-Americans and aliens from Mexico. (2) American is understood, so just look for schools

Yahoo List 5-7 classes or give a Boolean query formulation using elemental concepts from Yahoo

DDC List 5 - 7 representative classes from across the classification

LCC List 5 - 7 representative classes from across the classification

LCSH or SSH List 5 - 7 representative subject headings from across the alphabet

ERIC Give a Boolean query formulation using applicable ERIC descriptors
(Some concepts may require an OR combination of descriptors similar in meaning.)

Comments (on any of the schemes or on comparisons):

E Query form. Mexicans in American society (a topic broader than D) 307
(Mexican-Americans and aliens from Mexico)

(For Yahoo, DDC, LCC, and LCSH/SSH, all the classes or headings for query D must be used in searching; no need to repeat them. Just give a few of the many additional classes or headings for illustration)

Yahoo List 5 - 7 illustrative additional classes or give a Boolean query formulation using elemental concepts from Yahoo

DDC List 5 - 7 illustrative additional classes

LCC List 5 - 7 illustrative additional classes

LCSH or SSH List 5 - 7 illustrative additional subject headings

Comments (on any of the schemes or on comparisons):

F Query form. All aspects of canals (for transportation)

(This query illustrates that the main classes in DDC and LCC are based on disciplines, not on concrete objects. Observe the "relative" nature of the alphabetical index in DDC which brings together all the different places where **Canal** occurs in the scheme.)

Yahoo List 5 - 7 representative classes or give a Boolean query formulation using elemental concepts from Yahoo

DDC List 5 - 7 representative classes from across the classification (use Dewey for Windows)

LCC List 5 - 7 representative classes from across the classification (use Cataloger's Desktop)

LCSH or SSH List all applicable subject headings (use electronic version)

Comments (on any of the schemes or on comparisons):

Document materials

for the three documents to be indexed

The actual books are in Baldy 14A

- A. Differentiating the teaching staff.** / James Lewis

- B The open classroom reader.** / Charles E. Silberman, editor

- C MathMagic** (The Web site <http://forum.swarthmore.edu/mathmagic>)

