University of Maryland

College of Library and Information Services

LBSC 775

Construction and Maintenance of Index Languages and Thesauri

Dagobert Soergel

Fall 2003

Students who have a disability and wish to discuss academic accommodations should contact the instructor right away.

Students who for any reason (esp. religious observance) cannot attend an examination or other required class activity should contact the instructor to arrange for a make-up time.

Objectives

Major objectives

1. After completion of the course, the student should be able to design an index language or thesaurus (taxonomy, ontology) and construct it using manual methods with computer assistance.

This entails:

1.1 A good understanding of the design options available in the different features of an index language or thesaurus and of the criteria to be applied in the selection decision.

1.2 A thorough familiarity with the procedures to be used and the effort involved.

1.3 The ability to plan and implement an actual project for the construction of an index language or thesaurus (estimate effort, prepare time-schedule and assignments for personnel involved).

2. The student should be able to evaluate existing index languages and thesauri and make an informed judgment whether a particular index language or thesaurus is applicable in a given situation (or select from several index languages the one that is most suitable).

Objectives 1 and 2 imply that the student has a thorough understanding of the structure of index languages and their functions in information storage and retrieval systems.

Additional objectives

3. The student should have a good understanding of the possibilities of computer assistance for the handling of clerical tasks in the construction of index languages and thesauri.

4. The student should have some idea about automated and semi-automated methods in the construction of index languages and thesauri.

5. The student should have a good understanding of the problems involved in and the procedures used for updating index languages and thesauri.

6. The student should have a good understanding of the problems of convertibility of index languages and the implications for sharing the results of subject indexing.

7. The student should have an understanding of the role of thesauri and ontologies in organizing materials on the Web.

8. The student should have an understanding of “enriched thesauri”, ways of representing thesaurus data such as Topic Maps and RDF, and the standards involved

Content and Learning/Teaching methods

The course consists of two components:

1. Reading about theory and methods of thesaurus building and discussion of these readings in class.

2. Practical work in thesaurus construction, divided into

   (2.1) Introductory mini project (see separate description)

   (2.2) Major main project done in groups (see separate description)

3 Lectures on Thesaurus evaluation, thesaurus software, and other thesaurus-related topics.

Accordingly, the lecture periods will be used in two ways:

1. Discussion of theory and methods based on the assigned reading and the student's own experience in the project work. There will be no lecturing repeating the textbook. Reading must be done beforehand.

2. Discussion of problems arising from practical work, in the class as a whole for the mini-project and usually in the groups for the group project.

Lecture materials and readings will be distributed during the semester. There will be a charge for these materials at the end of the semester not to exceed $25.

Readings

Prerequisite

Soergel, Dagobert, 1985
Organizing Information. Principles of data base and retrieval systems.
Orlando: Academic Press, 1985. 450 p. Chapters 1, 3, 5, 8, 9, 12-15.

Text

Soergel, Dagobert, 1974
Construction and maintenance of indexing languages and thesauri
New York: Wiley, 1974.

Other helpful books

Vickery, Bryan C.
Faceted classification.
London: Aslib, 1970.
LibSch Z696.V7 1970 (multiple copies on reserve)
(Read this if you do not have a background in faceted classification).

Lancaster, F. Wilfrid, 1972
Vocabulary control for information retrieval. 1.ed.
Washington, D.C.: Information Resources Press, 1972.
LibSchStacks Z699.L355
(2. ed. not as good)

Aitchison, Jean; Gilchrist, Allen; Bawden, David
Thesaurus construction and use: a practical manual.
London: ASLIB; 2000. 218 p.
LibSchStacks Z695.A47 2000

Further readings (a few others will be added)

D. Soergel, A. McCann, J.A.T. Pennington, J.H. Holden, E.C. Smith, and R.C. Wiley.
Improving access to food and nutrition data. 2. A language for the description of foods in databases
With appendix: **Entity-relationship schema for a moderately detailed Interlinked Food Description (IFD) database with emphasis on thesaurus structure**
Distributed as part of the course materials

Soergel, **Functions of a thesaurus / classification / ontological knowledge base** (from 670)

Soergel, D. Et al. **ASIST SIG/CR 2000. Classification for user support and learning.** Knowledge Organization. 2000; 27(3): 165-172 and ASIST Bulletin. 2001 April; 27(4):

Sources for further literature, lists of thesauri, and thesauri on the Web

www.clis.umd.edu/faculty/soergel/dlthestut.html

www.alexandria.ucsb.edu/~lhill/nkos/index.html

www.willpower.demon.co.uk/thesbibl.htm

Example of a thesaurus on the Web: http://etoh.niaaa.nih.gov/AODVol1/Aodthome.htm

Bibliographies of thesauri

The following bibliographies are all rather old; nothing newer exists. See the URLs on the previous page

Gerstenkorn, A. 1985 ed.; Rolland, M. Th. ed.; et al.
Thesaurus guide. Analytical directory of selected vocabularies for information retrieval.
Amsterdam: Elsevier; 1985. 748p.
UM LibSch PC Soergel

International Classification and Indexing Bibliography. Vol. I: Classification Systems and Thesauri 1950-1982. ICIB 1. 160 pages, DIN A4, DM 48.80; ISBN 3-88672-300-3; FID-Publ.610. Frankfurt/M: Indeks Verlag; 1982.
This comprehensive bibliography of all universal and special classification systems and thesauri which could be found in the literature as well as in libraries, listing some 2300 titles from the time 1950-1982, is the first volume to appear of a series of four volumes covering the classification and indexing literature.
UM LibSch PC Soergel

Chan, Lois Mai; Pollard, Richard.
Thesauri used in online databases: an analytical guide.
New York: Greenwood; 1988. 268 p.
UM LibSch Z695.A1C47 1988

Chiyoowa, C.; Hafner,C.
Bestandsverzeichnis Thesauri. Stand: 01.07.1974. Frankfurt/M: Deutsche Gesellschaft fuer Dokumentation; 1974. 6 S., 135 Lit.

Bibliography of United Nations Thesauri, Classifications, Nomenclatures. Inter-Organization Board for Information Systems.
Geneva 1979

Stolp Nobile, Gertrude.
Dictionaries and vocabularies in the terminology and reference library 1966-1977. 5th edition. Document GIP/Bib/7. Rome: Food and Agriculture Organization of the United Nations; 1978. 170 p.

MacCafferty, Maxine.
Thesauri and thesauri construction.
London: Aslib; 1977. 190p., 825 ref. ISBN 085142-102-4.
UM LibSch 695.T42 folio

Aa, H.J.v.d., ed.

Thesaurus: bibliography.
Amsterdam: Studicentrum NOVI; 1977. 800 ref. 75 Dutch gilders. ISBN 90-6298-073-2.

This bibliography brings together a wide range of references, including sources from the United Kingdom, Russia, West Germany, East Germany, France, Austria, Belgium, Hungary, Czechoslavakia, the Netherlands, Italy and the United States. Books and articles for the most part are available from the Library of the Netherlands Center for Informatics. A subject index is provided. (NFAIS Newsletter)

Bibliographic Bulletin of the Clearinghouse of IINTE: Supplement 1976. Warsaw: Institute of Scientific, Technical and Economic Information (IINTE); 1977. 93p.

The 1976 supplement, as well as the previous issues, covers both current and retrospective material. The contents of the 1976 supplement provide new bibliographic items, relating to thesauri, descriptor lists, subject-heading and keyword lists, indexes and schedules of classification. The Bulletin is supplemented annually, and compiled with the financial assistance of Unesco.

Hoppe, K.; Levy, F.

Liste de quelques instruments de classification en France.
Bull. France 15 (1970) Nr. 5, S.243-257, 1 Index.
Tables comparing 97 thesauri used in French information centers.

Kutten, A., ed. **Thesauri bibliography.** 1975. Elyachar Library, Technion-Israel Institute of Technology, Technion City, Haifa, Israel. 62p. Ref. Indexes.

This bibliography is an attempt to identify the existing thesauri. Arranged alphabetically under subject headings in four sections: general thesauri; general thesauri in Hebrew; general thesauri in science and technology; engineering. It also includes an index of names and a subject index. (BDCTA)
UM LibSch Z695.T43

Laureilhe, Marie-Therese.

Bibliographie des thesauri et index par matieres parus depuis 1960. [Bibliography of thesauri and indexes published since 1960.]
Paris: Bibliotheque Nationale; 1975. 22p.

This bibliography is the result of the merging of six bibliographies which appeared in the Bulletin des bibliotheques de France from 1969 to 1974. It covers thesauri, many of which are multigraphed or typewritten documents designed for internal use, and sometimes even computer listings. Also covers major subject indexes published since 1960. Entries are arranged according to the Universal Decimal Classification scheme. Thesauri and indexes on the shelves of the Bibliotheque Nationale are also listed. (UNBLA)
UM LibSch Z695.A1P3.1975

Pope, Nolan F.; et al.

Thesauri used by SLA Documentation Division members.
1977. 27 p. 115 ref. ED 156 188

Bibliographies of dictionaries are also useful.

Some Standards for Thesaurus construction

International Organization for Standardization.
Documentation--guidelines for the establishment and development of monolingual thesauri. 2. ed. Geneva: International Organization for Standardization; 1986.
International Standard ISO 2788-1986(E).
This standard is based on Austin 1981.9.
SAME AS BS 5723

National Information Standards Organization (U. S.)
Guidelines for construction, format, and management of monolingual thesauri: An American National Standard developed by the National Information Standards Organization.
Bethesda, MD: NISO press; 1994. 69p.
ANSI/NISO Z39.19-1993 R1998
UM LibSch Z695.N36 1994

British Standards Institution.
British standard guide to the establishment and development of monolingual thesauri.
London: British Standards Institution; 1987

International Organization for Standardization.
Documentation--guidelines for the establishment and development of multilingual thesauri. Geneva: International Organization for Standardization; 1985.
International Standard ISO 5964-1985(E).
This standard is based on the Unesco guidelines.

Assignments to be done by students who did not do these in 670

Do in the first week

Conceptual analysis and synthesis

13.1, Semantic factoring
13.2, Building a hierarchy of elemental concepts
13.3, Hierarchy from facet combination

Analysis of conceptual structure

Assignment 13

Conceptual analysis and synthesis

If possible, do 13.1-13.3a before you come to the tutorial, but do not spend a lot of time when you get stuck. Having tried will help you understand the tutorial.

Purpose

To 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:

1. To practice semantic factoring (Practical application is the only way to grasp the idea.)
2. To practice the approach to hierarchy building, applying the pragmatic definition of A is broader than B in a limited set of elemental concepts.
3. To apply the principle of interaction between concept combination and hierarchy.

Each part of the assignment is concerned with one step in the process:

Assignment 13.1	Semantic factoring (results in a list of elemental concepts).
Assignment 13.2	Arranging the elemental concepts in a well-structured hierarchy.
Assignment 13.3	Fit compound concepts into the framework of the hierarchy (if compound concepts need to be dealt with explicitly) Assignment 13.3a is an exercise in facet combination unconnected to the set of concepts from Assignment 13.1. Assignment 13.3b returns to the set of concepts from Assignment 13.1, applying the principles learned from Assignment 13.3a

Note: The list of concepts given for Assignment 13.1 deliberately covers two domains, *Medicine* and *Transportation*. There might be elemental concepts that apply to both!

Assignment 13.1

Semantic factoring

If possible, do before you come to the tutorial.

Time: 1.5 hours

Task:

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**.

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.

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 *cancer* as the semantic factor, not the more general elemental concept *disease*.

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 *cancer*, which occurs several times as a semantic factor. This list will be the basis for Assignment 13.2.

Deliverables

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.)

2 A list of the elemental concepts used. (This will be the basis for Assignment 13.2.)

Definitions you may need

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
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 13.1. Concept list for semantic factoring

You can write your answer on these sheets and hand them in or use the template.

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 13.2

Building a hierarchy of elemental concepts

If possible, do before you come to the tutorial

Time: 1.5 hours

Task:

In Assignment 13.1 you started from a list of concepts, most of which are compound, and derived elemental concepts through semantic factoring. In this assignment you start from the list of **just the elemental concepts** that you prepared in Assignment 13.1 and establish the hierarchical relationships among them. (For example, do **not** include *Leukemia* or *Leukemia - drug therapy* into this hierarchy; you will have a chance to do that later, in Assignment 13.3b.)

Be sure to introduce all hierarchical relationships that are useful for searching and/or the checklist technique of indexing. 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?

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. You may need to introduce additional broad concepts to make for a more logical, more easily understood hierarchy. In many places additional specific concepts will suggest themselves from the logic of your hierarchy; you need not add those, but you may add a few examples.

It is easiest to represent the hierarchy in a linear arrangement (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.

Your hierarchy may include one part for *Medicine*, one part for *Transportation*, and one part for concepts needed in both domains (if any) or not fitting in any of the two domains.

In Assignment 13.3b, you will build a hierarchy that includes all the concepts used in the semantic factoring assignment in a well-structured arrangement.

Deliverable

A hierarchy of elemental concepts **shown as a linear arrangement with indention** (with cross-references as needed)

Assignment 13.3a

Hierarchy from facet-combination. General practice

Before the tutorial, read through this assignment and do as much (or as little) of it as you can do easily. After the tutorial, complete or revise your answer.

Time: 2 hours

Task: Construct the hierarchically ordered set generated by the two facets which are given on the following sheet and represent this hierarchy graphically. Then represent it as a linear arrangement with indention and cross-references. Repeat this choosing a different possibility for the linear sequence. (The graphical representation can be drawn on the assignment 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. Note: 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.

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 the tutorial.

Note: 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 13.1. You deal only with the concepts given on the following page and their combinations; do not introduce any other concepts.

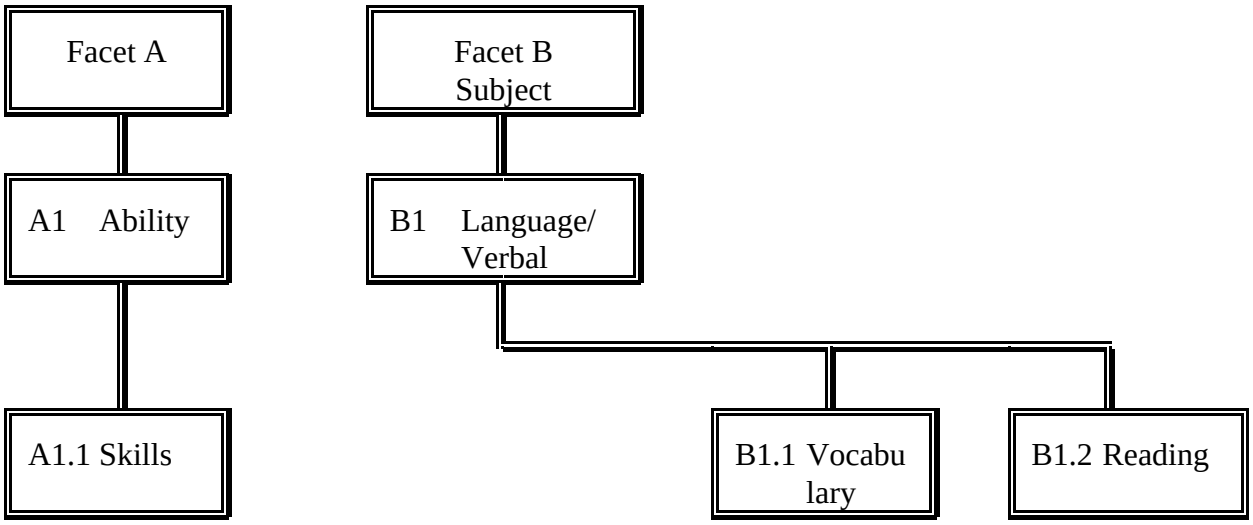
Deliverables

1 Hierarchy in graphical arrangement (may be drawn on the assignment sheet).

2 Hierarchy in linear arrangement with indention, with cross-references.

3 Hierarchy in another linear arrangement with indention, with cross-references.

Assignment 13.3a



Assignment 13.3b

Hierarchy from facet-combination. Application to the set of concepts from Assignment 13.1

Do after the tutorial

Time: 2 hours

Task

You will now apply the principles you have learned from Assignment 13.3a to the set of concepts from the Assignment 13 domain. Arrange all the concepts from Assignments 13.1 and 13.2 (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** nor particularly useful (there are too many concepts).

Start from the hierarchy of elemental concepts that you developed in Assignment 13.2 and fit the compound concepts from Assignment 13.1 into the structure. A compound concept has more than one broader concept; choose one place for the arrangement and make a cross-reference from the other. **You do not need to create all possible combinations of elemental concepts (as you did in Assignment 13.3a); just use the compound concepts actually listed in Assignment 13.1**, 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 handwriting. You can just copy the text from Assignment 13.2 and insert the compound concepts.

Write out each term so it can stand on its own outside the hierarchy

Not	**But**
Lung cancer . Therapy	Lung cancer . Lung cancer - Therapy

Do not introduce cross-references for hierarchical relationships that can be seen from the arrangement. Cross-references are used to show **additional** hierarchical relationships.

Deliverable

A hierarchy in linear arrangement with indention and cross-references.

Analysis of conceptual structure

Purpose: 1. Task 1: To apply Soergel 1985, Chapters 12-15, to a small practical task)

2. Task 2: To solidify the understanding of facet analysis and synthesis

Task 1: 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.

Task 2: Organize the following list of terms for purposes of information retrieval.

U.S. Congress

State Court

County administration

State legislature

Federal court

U.S. Senate

House of Representatives

State administration

State senate

State assembly

Ambitious people may add the terms from the next page, but only after completing the entire task for the terms given here.)

Procedure (facet analysis and synthesis)

Step 1: decompose concepts into semantic factors

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.

This task requires that you follow strictly the procedure given.

Additional terms for task 2 (Finish task 2 for the basic set of terms first!)

UN

Foreign Office

British Parliament

United Nations

Prime minister

House of Commons

House of Lords

Supreme Court

UN Secretary-General

Department of State

Security Council

UN General Assembly

State Department

High Court

Description of Mini-Project

Purpose:

To go through the complete process of thesaurus construction using a limited set of terms. This will lead to a much better understanding of the process than reading alone and thus give the student a better perspective when carrying out the steps in the major project.

Task:

Construct a piece of a thesaurus starting from the two sources that are attached. Carry out steps F1-F4 and part of F5 as described in the book. The result will be:

1. A classified index;

2. A deck of cards (thesaurus forms) for preferred terms in classified order. (Attach the cards that were eliminated in the process, too.) (Cards are available in the CLIS library.)

Note:

In the initial recording of terms on cards, use blank cards for five terms (take the back of a card, if necessary). This should bring out the importance of having printed forms.

Additional remarks on problems that occur often

1. Be sure to use proper source indications, e.g.,

   Legal aid (CPL: VIII.24)

   Be sure to transfer sources when merging.

2. Do as much semantic factoring and facet analysis as possible in the example (the main project will tax your ability to perform this step considerably more).

3. Use the hierarchy level indicator line in the classified index (see example in text, p. 274).

The following criteria will be used in evaluating the mini-project.

- Correct transfer of information from source on cards (all info transferred? Source indications in correct format?)
- Correct transfer of information (including source indications) from one card to another in merging.
- Structure of classified index.
- Cross references on cards and in classified index.

Term lists for mini-project

1. Table of contents of San Francisco People Yellow Pages (bibSchRefAG527.52.1972. Take **all** main terms (printed bold) listed. Source Code: PYP This is an extract.

Abortion 1
see also Pregnancy tests & Counseling, Pregnant teens

Alcoholism 2,3
see also Halfway houses, Drug abuse treatment centers, Counseling, Mental health

Art 7-10
see also Art and craft supplies, Macrame, Leather, Pottery, Sign painting

Art and craft supplies 11
see also Beads, Leather, Yarn

Birth control 15-18
see also Health, Vasectomies

Child care 23-25

Clothes 25,26
see also Second-hand, Free stuff, Shops

Consumers 31

Dance 34,35

Doctors
see Health

Drug abuse treatment centers 38-41
see also counseling, Alcoholism, Mental health, Switchboards, Youth

Food 50-53

Food conspiracies 51

Food stamps 50

Free stuff 53
see also Clothes, Education, Food, Housing, Places to go, Universities, Showers

Health 58-63
see also Birth control, Chiropractic, Drug abuse, Treatment Centers, Ears, Eyes, Feet, Infertility, Mental health, Nutrition, Prepared childbirth, Prenatal care, Venereal disease, Vasectomies, X-rays

Information and referral 65
see also Switchboards

Jobs
see Labor, Work

Labor 67
see also Work

Legal assistance 67-70

Nutrition 82
see also Food

Organic foods
see Honey, Natural food stores

Personal growth 84,85
see also Education, Spiritual growth, Universities, Cosmic consciousness

Pregnancy testing and counseling 91,92
see also Abortion, Pregnant teens, Health

Pregnant teens 90
see also Child care, Pregnancy testing and counseling, Single parents

Senior citizens 106,107

Switchboards 112,113
see also Information and referral

Unemployment insurance 118

Work 127,128
see also Labor, Youth

"Code sheet" of the Public Information Center at Cleveland Public Library.

Take only terms marked with *
Source code: CPL

CODE	Barriers
*01	Inadequate Orientations to Work
02	Inadequate Education & Training
*03	Lack of job
04	Handicapping Effects of Mental Physical or Social Disabilities
05	Discrimination
*06	Lack of Adequate Income
07	Lack of Insurance Protection
08	Inadequate Home & Family Management
*09	Inadequate Food
*10	Inadequate Clothing
*11	Inadequate housing
12	Inadequate Living Arrangements
13	Hazardous Living Conditions
14	Lack of Transportation
15	Lack of Legal Aid
16	Delinquency
17	Physical Problems
18	Mental Problems
19	Illness of Family Member
*20	Unwanted Pregnancy
21	Lack of Child Care
22	Child Behavior Problems
23	Marital or Family Conflict
24	Neglect, Abuse or Exploitation
25	Lack of Knowledge of Parental Functioning
*26	Loss of Social Contact
*27	Isolation
*28	Inadequate Inter-personal Adjustment
29	Lack of Education or Cultural Opportunities
30	Lack of Recreation
31	Individual or Community Attitudes
32	Individual or Restrictive Admission Policies
33	Lack of Information About Available Community Services

CODE	Services
*I.	**Employment services**
*01	Pre-Job Guidance
*02	Job Training
*03	Job Placement & Referral
*04	Spec. Employment Serv. for Disadvantaged
*05	Spec. Employment Serv. for Aging and Handicapped
II	**Income maintenance services**
06	Social Insurance Services
	Financial Aid Services
07	Aid to the Blind
08	Aid to Permanently and Totally Disabled
09	Aid to Families of Dependent Children
10	Old Age Assistance
11	Emergency Welfare Assistance
12	General Assistance
13	Special Benefits for Persons Age 72 and Over
III	**Consumer protection & safety services**
14	Consumer Education
15	Consumer Resource
IV	**Food and nutrition services**
*16	Government Food Services
*17	Home Meals, Mobile Meals or Congregate Meals
V	**Clothing & apparel services**
VI	**Housing services**
18	Relocation & Allocation Under Renewal
19	Housing Assistance
20	General Housing Search & Location Services

VII Transportation services

21 General Transportation Services
22 Spec. Transportation Needs of Selected Groups

VIII Public protection, justice & safety services

Justice Services

23 Law Enforcement
*24 Legal Aid
25 Detention of Law Violators & Alleged Law Violators
26 Corrections
27 General Public Services
28 Public Disaster Services

IX Health (physical) maintenance & services

29 Community Health Maintenance Services
30 Medical Care Services

X Mental health maintenance & care services

31 Inpatient, Outpatient & Emergency Psychiatric Care
32 Residential Treatment of Emotionally Disturbed
33 Transitional Care (Halfway Houses)
*34 Alcoholism Treatment
*35 Drug Abuse & Narcotics Addiction Treatment

XI Mental retardation services

36 Special Day Care of Mental Retardation
37 Residential Care Services for Mentally Retarded

XII Rehabilitation services

38 Therapeutic Services for Handicapped

XIII Education services

39 Formal Education Services
40 Informal Educational Services
41 Supplementary Educational Services
42 Special Educational Serv. for Gifted and Retarded

XIV Individual & family life services

Family Preservation & Strengthening

43 Counseling
44 Chore Services
45 Homemaker
*46 Family Growth Control & Planning

Family Substitute Series

47 Adoption
*48 Child Care
49 Day Care for Adults
50 Foster Home Care
51 Group Home
52 Nursing Homes

XV Small crisis intervention & protective services

53 Suicide Prevention & Protection
54 Protection from Neglect, Abuse and Exposure
55 Supportive Services to Individuals and Families

XVI Social adjustment, social development and growth

56 Recreational Services
57 Social Group Services
58 Intergroup Relations Services

XVII Cultural enrichment services

*59 Artistic & Cultural Opportunities Services

XVIII Mobilization of people services

60 Community Organization Service
61 Political Organizations Service
62 Volunteer Services

XIX Equal opportunity services

Description of main project

Construction of a thesaurus (or parts thereof) in a given area

Purpose of the main project

The construction of a thesaurus can only be learned by actually doing it. The practical experience gained in the project will be valuable in the construction of an actual thesaurus. Note that it is **not** the purpose of the project to construct a complete thesaurus that can actually be used; this is not possible in the framework of a course. It is better anyway to learn on a model and then proceed to an actual life situation. **So nobody should be disappointed if at the end the thesaurus developed is not perfect.**

The **main part of the project is done in groups** of generally three students for two reasons:

(1) For many tasks in thesaurus construction, group discussion is the preferable technique.

(2) The work required for covering a reasonably broad subject is too much for one person.

Scope of the main project

The thesaurus should cover a – possibly interdisciplinary – subject field and be designed for a specific scenario – a user group and a library, online database, Web site, semantic Web component, or other information system. Parameters such as manual or computerized ISAR system, number of searches per month, and collection size can be specified.

Product of the main project

A thesaurus fragment consisting of the following parts:

(1) Introduction to the thesaurus, describing the thesaurus and its use and including a full bibliography of sources used in developing the thesaurus.

(2) A conceptual schema for the domain

(3) Outline of the subdivisions of the entire chosen subject field, e.g., International politics, Economics, Environmental studies, Social psychology, Ornithology, etc.

(4) Classified index (Quick Hierarchy) worked out for x narrow subdivisions of the broad subject area of approximately five double-spaced typewritten pages each. x = number of students in group.

(5) 20 x terms fully worked out with scope note and cross-references (Annotated Hierarchy)

(6) Alphabetical index

These documents will be produced from proper input by the thesaurus software provided.

Sample projects from earlier courses are kept in the Wasserman Library behind the desk. A look at the thesauri on reserve and in the Wasserman Library (in the Cataloging Lab section) may further clarify the nature of the project

Individual term paper

An individually prepared paper is required of each student; this paper is a major basis for the grade in the course. It consists of three parts

(1) Indexing of 15 documents or other entities according to the thesaurus with document-by-document comments on any difficulties encountered in indexing (for example, missing descriptors). Each document should have its own sheet with a title and brief description (such as an abstract obtained from an abstracting service). Write descriptors and document-specific comments on this sheet.

(2) An analysis of what was learned about the thesaurus from the indexing exercise, with conclusions on how the thesaurus could be improved.

(3) A discussion of the methods used, reasons for design decisions made, problems encountered and how they were solved, what would you do different the next time around (in a real project). This will be critical reflection on what you have learned in the course

In order to do this report, each student must **keep a project diary**. Include for each step the number of terms processed and the time it takes so you have some basis for estimating costs in a real-life situation.

Lecture Notes

Lecture 1

Introduction

Objectives, prerequisites, learning/teaching methods.

Mini-project and main project.

Knowledge organization systems (KOS):
Thesauri, taxonomies, ontologies and the semantic Web

- Example thesauri
- From thesauri to full-fledged ontologies (report from NKOS workshop)
- Types of KOS

First discussion of forming groups for main project: select a subject

Thesaurus construction: general framework and overall organization (A)

Finding thesauri. Thesauri on the Internet. See resource pages and email

Insert flowchart

Lecture 2

The entity-relationship approach to thesaurus construction

Introduction to the entity-relationship approach: University database

Semantic networks / frame databases with hierarchical inheritance
and the nature of compound concepts

Examples from UMLS

Insert Univ DB diagram

UB diagram 2

Semantic networks / frame databases with hierarchical inheritance and the nature of compound concepts

Service-001 Alexandria Women's Clinic

isa	Service	
run by	Health Initiatives	
is located at	703 Fern St, Alexandria, VA	
has hours	M-Th 10-12	
serves clientele	Elderly	
serves clientele	Female	Group-001
serves clientele	Poor	
is service type	Direct assistance	
has service area	Health	

Service-002 Newman Women's Center

isa	Service	
run by	Catholic Charities	
is located at	11240 Wilson Blvd, Arlington, VA	
has hours	MWF 3-5pm	
serves clientele	Elderly	
serves clientele	Female	Group-001
serves clientele	Poor	
is service type	Direct assistance	
has service area	Health	

Service-003 Falls Church Women's Clinic

isa	Service	
run by	Health Initiatives	
is located at	770 Broad St., Falls Church, VA	
has hours	M-T 2-4pm	
serves clientele	Elderly	
serves clientele	Female	Group-001
serves clientele	Poor	
is service type	Direct assistance	
has service area	Health	

Introduce a new node in the semantic network. The new node wraps common properties for use as a bundle, creating a new compound concept.

Group-001 Elderly poor women

is a kind of	Group
has characteristic	Elderly
has characteristic	Female
has characteristic	Poor

Strictly speaking, *has characteristic* refers to persons. Group-001 includes all persons who have these characteristics.

Now we can rewrite the Service nodes as follows:

Service-001 Alexandria Women's Clinic

isa	Service
run by	Health Initiatives
is located at	703 Fern St, Alexandria, VA
has hours	M-Th 10-12
serves clientele	Group-001
is service type	Direct assistance
has service area	Health

Service-002 Newman Women's Center

isa	Service
run by	Catholic Charities
is located at	11240 Wilson Blvd, Arlington, VA
has hours	MWF 3-5pm
serves clientele	Group-001
is service type	Direct assistance
has service area	Health

Service-003 Falls Church Women's Clinic

isa	Service
run by	Health Initiatives
is located at	770 Broad St., Falls Church, VA
has hours	M-T 2-4pm
serves clientele	Group-001
is service type	Direct assistance
has service area	Health

Introduce another new node

ServiceType-001 Health Clinic

is a kind of	Service
is service type	Direct assistance
has service area	Health

This means that saying

Service-001 *isa* ServiceType-001

is the same as making the three separate statements on Service-001.

We can now further rewrite the Service nodes as follows:

Service-001 Alexandria Women's Clinic

isa	ServiceType-001
run by	Health Initiatives
is located at	703 Fern St, Alexandria, VA
has hours	M-Th 10-12
serves clientele	Group-001

Service-002 Newman Women's Center

isa	ServiceType-001
run by	Catholic Charities
is located at	11240 Wilson Blvd, Arlington, VA
has hours	MWF 3-5pm
serves clientele	Group-001

Service-003 Falls Church Women's Clinic

isa	ServiceType-001
run by	Health Initiatives
is located at	770 Broad St., Falls Church, VA
has hours	M-T 2-4pm
serves clientele	Group-001

Finally, we can introduce the node

ServiceType-002 Health clinic for elderly poor women

is a kind of	ServiceType-001
serves clientele	Group-001

and rewrite the nodes still further

Service-001 Alexandria Women's Clinic

isa	ServiceType-002
run by	Health Initiatives
is located at	703 Fern St, Alexandria, VA
has hours	M-Th 10-12

Service-002 Newman Women's Center

isa	ServiceType-002
run by	Catholic Charities
is located at	11240 Wilson Blvd, Arlington, VA
has hours	MWF 3-5pm

Service-003 Falls Church Women's Clinic

isa	ServiceType-002
run by	Health Initiatives
is located at	770 Broad St., Falls Church, VA
has hours	M-T 2-4pm

Of course, we could introduce

ServiceType-003 Health clinic for elderly poor women run by Health Initiatives

isa	ServiceType-002
run by	Health Initiatives

and rewrite nodes Service-001 and Service-002. If there are many such clinics operated by this organization, it would be efficient to do that.

Templates for entering services into the system

Service-00
isa Service
run by
is located at
has hours
serves clientele
serves clientele
serves clientele
is service type
has service area

Service-00
isa Service
run by
is located at
has hours
serves clientele
serves clientele
serves clientele
is service type
has service area

Service-00
isa Service
run by
is located at
has hours
serves clientele
serves clientele
serves clientele
is service type
has service area

UMLS examples here

Lecture 3

Mini-project and workflow

Discussion of mini-project. Any questions that might come up either on procedure or semantic structure of the subject matter, especially the structure of the conceptual schema using the entity-relationship approach and the structure of a facet frame.

Thesaurus software (TermMaster) and work flow.

Final formation of groups.

Lecture 4. Tutorial

Hierarchy from facets

Objective: Understand complex hierarchies that result from combining hierarchically structured facets.

Information system of instructional materials. Two facets, only between-facet combinations

Process

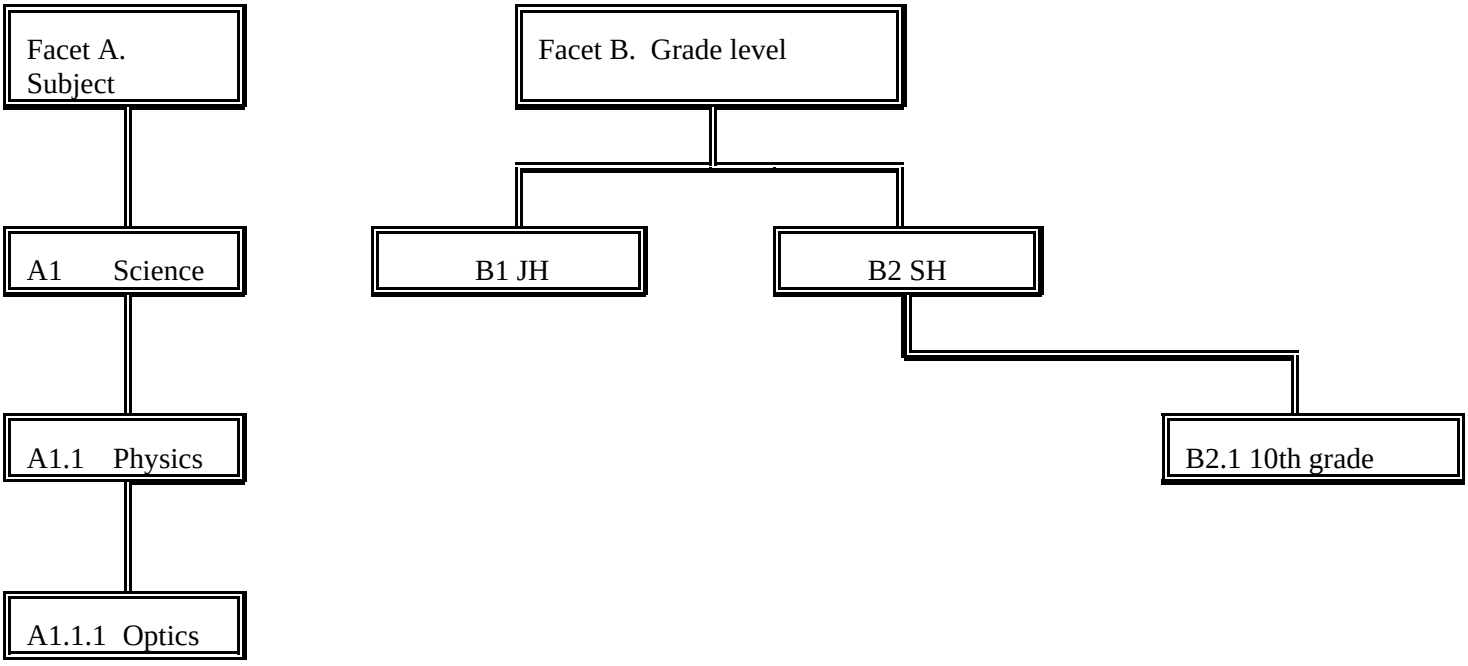
Step 1:	Form all possible between-facet combinations (do not combine with facet heads).
Step 2:	Find all hierarchical relationships. (Specifying all BT one level up defines the hierarchy completely.)
Step 3:	Represent the hierarchy graphically.
Step 4:	Represent the hierarchy as a linear arrangement with indention plus cross-references.

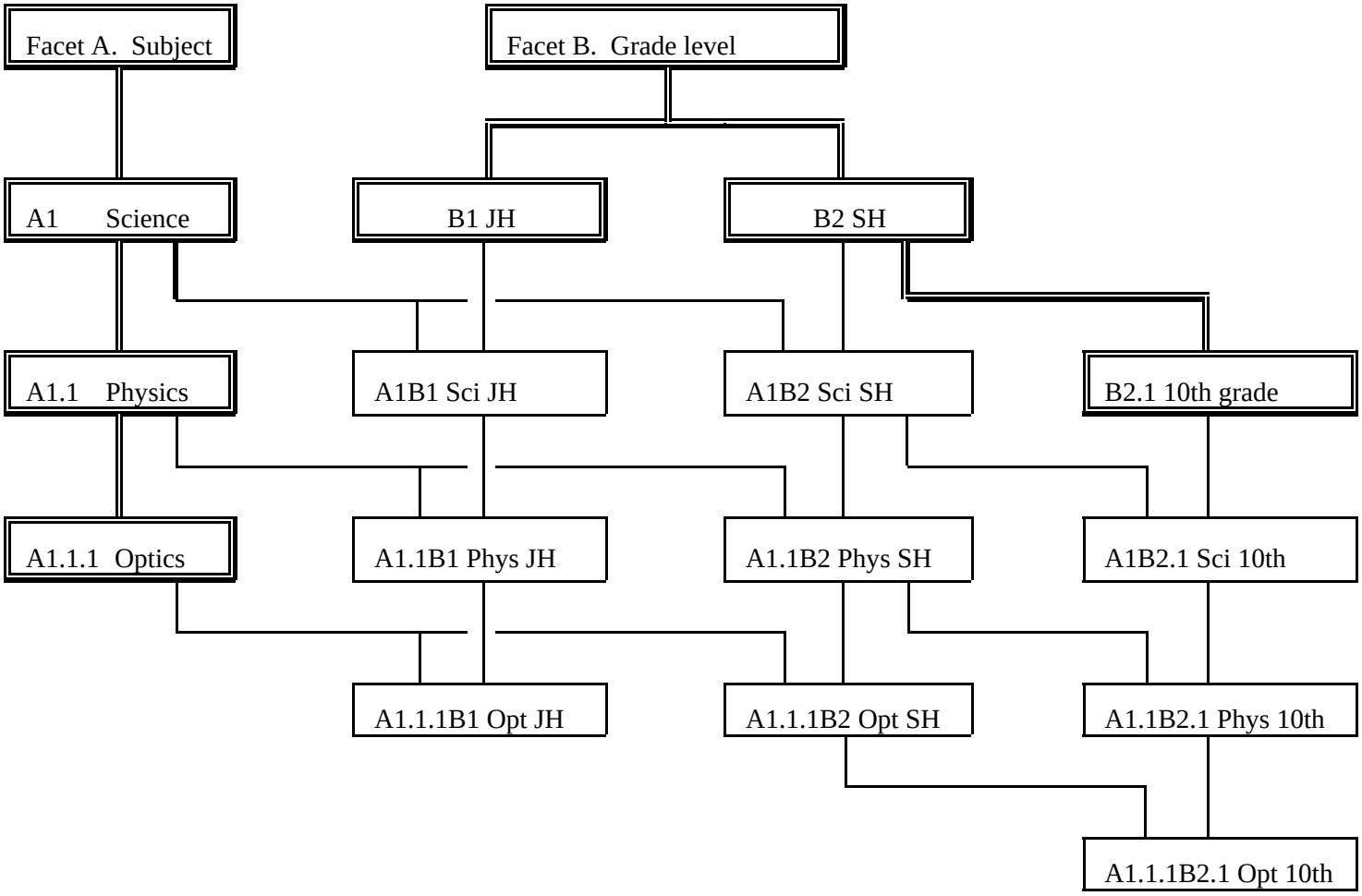
Application to retrieval

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 JH NT A1.1B1; BT B1
. . A1B2 Science SH NT A1.1B2; BT B2
. . . A1B2.1 Science 10th grade NT A1.1B2.1; BT B2.1
. . A1.1 Physics
. . . A1.1B1 Physics JH NT A1.1.1B1; BT A1B1
. . . A1.1B2 Physics SH 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 JH BT A1.1B1
. . . . A1.1.1B2 Optics SH BT A1.1B2
. . . . . A1.1.1B2.1 Optics 10th grade BT A1.1B2.1

B Facet B. Grade level

. B1 JH NT A1B1
. B2 SH 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 JH
. . B1 A1 JH Science BT A1
. . . B1 A1.1 JH Physics BT A1.1
. . . . B1 A1.1.1 JH Optics BT A1.1.1
. B2 SH
. . B2 A1 SH Science NT B2.1A1; BT A1
. . . B2 A1.1 SH Physics NT B2.1A1.1; BT A1.1
. . . . B2 A1.1.1 SH 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