The many uses of classification:
Enriched thesauri
as knowledge sources

Dagobert Soergel

College of Information Studies
University of Maryland
ds52@umail.umd.edu

Classification everywhere

concept maps in education

e-commerce (B2B, B2C)

Corporate portals

Agent-to-agent communication

Database schema correlation for interoperability

Text summarization and other NLP applications

IEEE upper ontology working group

The Semantic Web

Ontologies, Taxonomies

Classification under any other name is still classification

Is SIG/CR everywhere?

Is NKOS everywhere?

Expanded functions of thesauri
(SIG/CR 2000 Theme 1)

- Convey meaning, orientation, and structure. Definitions

- Provide rich relationships. Give facts

- Support exploration and browsing, creativity, problem solving

- Knowledge-based assistance for indexing and searching, behind the scenes or collaboratively with the user

- Linkage to thesaurus entries from text. Linkages among thesauri. Integrated access system

- Assistance to users in maintaining their own thesauri. Collaborative development and maintenance of thesauri

Exploit the possibilities
of the new medium

- Data structures of adequate complexity for rich content

- Searchability and selectivity

- Flexibility of display

- Processing power and inference

- Linkage

Convey meaning, orientation, and structure

- Assists any user thinking about a problem

- Helps with better query formulation

- Requires good methods for displaying structure. Most thesaurus interfaces provide local views but not views of the structure at large

- Examples

    Meaningful hierarchical display

    Concept graphs

    Facets to elicit query

Convey meaning, orientation, and structure. Continued

- Meaningful arrangement. There is no need for alphabetical arrangement in online environments

- Requires intensive effort in developing meaningful structure

Yahoo Home

Arts & Humanities
Literature, Photography ...

Business & Economy
B2B, Finance, Shopping, Jobs ...

Computers & Internet
Internet, WWW, Software, Games ...

Education
College and University, K-12 ...

Entertainment
Cool Links, Movies, Humor, Music ...

Government
Elections, Military, Law, Taxes ...

Health
Medicine, Diseases, Drugs, Fitness
...

News & Media
Full Coverage, Newspapers, TV...

Recreation & Sports
Sports, Travel, Autos, Outdoors ...

Reference
Libraries, Dictionaries, Quotations ...

Regional
Countries, Regions, US States ...

Science
Animals, Astronomy, Engineering ...

Social Science
Archaeology, Economics, Languages
...

Society & Culture
People, Environment, Religion ...

Yahoo Home. Meaningful arrangement

Reference and General Interest	Subjects
	Science Animals, Astronomy, Engineering
	Health Medicine, Diseases, Drugs, Fitness ...
Reference Libraries, Dictionaries, Quotations	
	Social Science Archaeology, Economics, Languages ...
Computers & Internet Internet, WWW, Software, Games ...	**Society & Culture** People, Environment, Religion ...
	Government Elections, Military, Law, Taxes ...
	Business & Economy B2B, Finance, Shopping, Jobs ...
News & Media Full Coverage, Newspapers, TV...	**Education** College and University, K-12 ...
Entertainment Movies, Music, Humor, Cool Links ...	**Arts & Humanities** Literature, Photography ...
Recreation & Sports Sports, Travel, Autos, Outdoors...	

Regional Countries, Regions, US States ...

Yahoo Home > Health

Alternative Medicine	Men's Health
Business to Business@	Mental Health
Chats and Forums	Midwifery
Children's Health	News and Media
Conferences	Nursing
Death and Dying@	Nutrition
Dentistry@	Organizations
Disabilities@	Pet Health@
Diseases and Conditions	Pharmacy
Education	Procedures and Therapies
Emergency Services	Public Health and Safety
Employment	Reference
Environmental Health	Reproductive Health
First Aid	Senior Health
Fitness	Sexuality@
General Health	Shopping and Services@
Health Administration	Teen Health
Health Care	Traditional Medicine
Health Sciences	Travel Health and Medicine
Hospitals and Medical Centers	Web Directories
Institutes	Weight Issues
Law@	Women's Health
Long Term Care	Workplace (6
Medicine	

Home > Health. Meaningful arrangement

Reference
　Reference
　Web Directories
　Chats and Forums
　News and Media

Health Sciences Fields
　Health Sciences
　Medicine
　Dentistry@
　Nursing
　Midwifery
　Pharmacy

　Traditional Medicine
　Alternative Medicine

Individual health condition
　Diseases and Conditions
　Disabilities@
　Fitness
　Nutrition
　Weight Issues
　Reproductive Health
　Sexuality@
　Death and Dying@
　Mental Health

Procedures and Therapies

Health by place
　General Health
　Public Health and Safety
　Environmental Health
　Workplace
　Travel Health and Medicine

Health by population group
　Human Health
　　Human Health by Age
　　　Children's Health
　　　Teen Health
　　　Senior Health
　　Human Health by Gender
　　　Women's Health
　　　Men's Health
　Animal Health
　　Pet Health@

Health Care
　Emergency Services
　First Aid
　Long Term Care

Health care organization
　Hospitals and Medical Centers
　Institutes
　Organizations
　Conferences
　Health Administration

　Shopping and Services@
　Business to Business@
　Law@

　Education
　Employment

AAT. Art genres

<art genres>
  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
      cave art
  serial art
  sofa art
  street art

Art genres. Meaningful arrangement

art genres

. **art genres by content or other intrinsic characteristics**
. . figurative art
. . . fantastic art
. . . apocalyptic art
. . nonrepresentational art
. . cybernetic art
. . serial art
. . crafts

. **art genres by standard**

. . academic art
. . folk art
. . dissident art

. **art genres by type of artist or origin**

. . amateur art
. . naive art
. . art brut
. . children's art
. . computer art
. . ethnic art
. . primitive art

. **art genres by audience, purpose, or display context**
. . sofa art
. . court art
. . **public art**
    SN   Art whose purpose is to beautify and enrich public places.
. . . **community art**
    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
. . . cave art [prehistoric, esp. paleolithic]

. **art genres by audience, purpose, or display context**
. . sofa art
. . court art
. . **public art**
SN   Art whose purpose is to beautify and enrich public places.

. . . **community art**
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
. . . cave art [prehistoric, esp. paleolithic]

. . didactic art
. . commercial art
. . funerary art

Concept map climate change

Concept map Instructional design

Concept map information studies

Concept map information studies 2

Definitions

- A thesaurus should give full definitions, not just usage notes

- Multiple definitions

- Links to document segments that elaborate on the concept

People understand concepts
    from examples
    while situated in a given **context** and considering a given **function** or use.

Definitions (and classificatory structure, for that matter) thus need to provide

    examples, context, and functions

JC   **basic prevention categories**

JC2   . **prevention by timing of the intervention**

JC2.2   . . primary prevention

JC2.4   . . secondary prevention

JC2.6   . . tertiary prevention

JC4   . **prevention by scope of recipient group**

JC4.2   . . universal prevention

> SN Directed at the general public or a population group that has not been identified on the basis of members' risk. The intervention is desirable for everyone in that group.

JC4.4   . . targeted prevention

> SN Targeted at subgroups of the population or at individuals who are at high or very high risk. There are two subordinate categories which are distinguished by the specificity of targeting (the precision of selection into the recipient group), the degree of risk, and the warranted cost per recipient.

JC4.4.2   . . . selective prevention

> SN Targeted at all members of a subgroup, for example, public service spots on a TV channel watched predominantly by teenagers.

JC4.4.4   . . . indicated prevention

> SN Targeted at individuals (often members of a subgroup) who have been identified through screening as being at high risk.

JC4.6   . . prevention directed at groups

> SN Includes JC4.2 universal prevention and JC4.4.2 selective prevention as opposed to JC4.4.4 indicated prevention, which is directed at individuals.

Convey meaning, orientation, and structure. Continued

To build a classification/thesaurus
(Knowledge Organization Structure)
that conveys meaning,
follow principles of instructional design:

(1)   Discern the underlying structure of
          the domain

(2)   Find a suitable external expression
          or representation of that structure

For (1) the principles of classification
structure (facet analysis, hierarchy) often
help to achieve a level of clarity not
achieved by domain experts alone.

Rich relationships. Give facts

- Examples

    Cancer *combine-with* Body part

        When cancer is indexed or searched, the system posts a reminder about body part

    Bromocriptin *treats* Alcohol withdrawal

        Now shown, if at all, as

        Alcohol withdrawal agents NT Bromocriptin

    Early behavior disorder *is-risk-factor-for* Alcohol or other drug disorder

    Alcohol *causes* Liver disease

Rich relationships. Give facts.
Cont.

Sample relationship types from the UMLS
Semantic Network

    functionally_related_to
      affects
         manages
         treats
         disrupts
         complicates
         interacts_with
         prevents
      brings_about
         produces
         causes
      performs
         carries_out
         exhibits
         practices
      occurs_in
         process_of
      uses
      manifestation_of
      indicates
      result_of
   temporally_related_to
      co-occurs_with
      precedes

The UMLS contains a few statements using
these relationship types

Rich relationships. Give facts.

Note:

Systems for automated reasoning need differentiated relationships even if for use in retrieval a courser grain is sufficient.

Example:

In a thesaurus for retrieval, it is just fine, for most purposes, to treat isa relationships and part-whole simply as hierarchical relationships (as in body systems or geography).

A system for automated reasoning needs to distinguish

Baltimore *is-part-of* Maryland

Baltimore *is-instance-of* city

city *is-subclass-of* settlement

Rich relationships. Give facts.
Cont.

Rich relationships expand a conventional thesaurus to a knowledge source in its own right. Conceptual and terminological knowledge is embedded in a complex database with many other relationships,

The relationships can be used

to answer factual queries and

to find search terms or indexing terms

Synergy in the combined use of many relationships

Rich relationships. Give facts.

Cont.

Standards such as RDF and Topic Maps support rich relationships.

NISO is thinking about an expanded thesaurus standard that would replace Z39.19.

Harmonization of relationship types is necessary for interoperability

Standard

Relationship repository

Rich relationships. Give facts.
Cont.

Problems

- For the user: The very richness of information will be overwhelming; too many types of relationships, too many relationships for any one term (there can be 50 or more risk factors)

- Solution: Flexible display. User can select information to be displayed by

    type of relationship and

    priority of relationship

Rich relationships. Give facts.
Cont.

Problems

- For the system builder:
  Cost of initial development
  Cost of maintenance

- Solution: ?
  Enormous human labor (e.g.,CYC)
  Collaboration supported by appropriate infrastructure
  Automatic extraction of facts (and definitions) from text
  Pattern discovery in large databases (for example, large patient databases), machine learning, data mining

Support exploration and browsing, creativity, problem solving
(SIG/CR 2000 Theme 4)

Most classifications deal with (static) domain knowledge

Additional approaches are needed to support users, such as

- Problem schemas as organizing principle

- Functions as organizing principle

- Classification of cases for case-based reasoning or for education and learning

Reinforces the theme of rich relationships

We need to learn how to build such new tools

Problem schemas
as organizing principle

A classification of problems by problem type, such as

> *fix a device* (*fix a car*, *fix a washing machine*),
> *buy something*,
> *write a computer program*,

giving for each problem a schema that specifies aspects to be considered in solving the problem:

> information, people, material needed for solving the problem

> procedural steps for solving the problem.

Functions as organizing principle

Classify technical components by all the functions they could serve
Describe the functions abstractly to help users to think out of the box and find novel uses

Classify business cases by all the concepts they illustrate or all the courses they could be used in

Conventionally done: Classify employees by all their skills or all the jobs they could perform in the organization

Knowledge-based assistance for indexing and searching, behind the scenes or with the user

Searching

- Expand use of common techniques:

    Synonym expansion (query term mapping)

    Hierarchic expansion

- Knowledge-based elicitation of user requirements

- Knowledge-based clustering of search results

Knowledge-based assistance, continued

Indexing

- Example: MedIndex

  Can be used for assisting human indexers and for improved automated indexing

- Natural language processing using tools that combine linguistic dictionary information with hierarchy and other thesaurus information.
  Example: UMLS and its Specialist Lexicon

Example: MedIndex (Susanne Humphrey, NLM)

Indexer enters **Bone Neoplasms**

System displays the Neoplasms frame which shows the facets to be considered when indexing a document on neoplasms. The frame is already specialized for bone neoplasms:

Bone Neoplasms - Current Frame

ANATOMICAL STRUCTURE
   Bone and Bones

SECONDARY-FROM

ETIOLOGY

COMPLICATION

PROCEDURE

PROCESS

HISTOLOGIC TYPE

Indexer decides to work further on ANATOMICAL STRUCTURE, clicks on it, and is presented with a hierarchy.

Body Areas
.	Back
.	Extremities
.	.	Arm
.	.	Leg
.	.	.	Foot
.	.	.	Knee
.	Head
.	.	Face
.	Neck
.	Pelvis
.	Thorax
Bone and Bones
.	Facial Bones
.	.	.Palate
.	Leg Bones
.	.	**Femur**
.	.	Fibula
.	.	Tibia
etc.

Indexer selects **Femur**

System checks its knowledge base, sees that there is a MeSH descriptor under *Bone neoplasms* with *Femur* as the ANATOMICAL STRUCTURE and responds

Femur not permitted.
The correct MeSH heading is

 Femoral Neoplasm

The underlying knowledge structure (frame)

Femoral Neoplasm

INHERITS FROM
    Bone neoplasms
 	
ANATOMICAL STRUCTURE
    Femur
 	
SECONDARY-FROM
 	
etc.

Search assistance

Example 1

A medical information system for health consumers

The system displays a picture of the body where the user can click to identify the body part or system where she has a problem (facet 1)

The system then displays a menu of symptoms that might occur for that body part (based on relationships between body part and symptoms it knows about) (facet 2)

The system can then ask more specific questions to pin down the problem further

Just described: An expert system for medical diagnosis

Search assistance

Example 2

e-commerce

Consumers need help with selecting a specific make and model that suits their needs

The system must know for each type of item (camera, washing machine, TV set, car, etc.) what the characteristics are - a facet frame for each of type of item

It can use this knowledge to ask the consumer questions designed to match his needs with desirable values of each characteristic and then find the items that more or less fit the bill

Linkage to thesaurus entries from text.

- Assist readers in understanding text by seeing a definition or seeing a concept in its hierarchical context.

- See a subject descriptor recorded in a metatag in the context of the scheme it comes from.

  This would require a thesaurus registry with URIs for thesauri.

Linkages among thesauri
Integrated access system

- Useful for cross-database searching

- Integrated access useful for getting more information.

- Ideally: A "Virtual Thesaurus" that would provide transparent access to multiple thesauri, dictionaries, and other lexical resources and provide an integrated display of the information about a concept or term.

  The challenge: Do this integration automatically

Assistance to users in maintaining
their own thesauri

Collaborative development and maintenance of thesauri

The breadth of the field

The classification researcher must be a renaissance person. Doing research about and building classification requires knowledge of many fields, many of which both contribute to knowledge about classification and use classifications.

- Principles of classification and knowledge representation

- Philosophy, esp. ontology and epistemology

- Cognitive psychology, the workings of the human mind

- Artificial intelligence

- Linguistics

- Instructional design, document design, interface design. Information architecture

- Markup languages and data structures and their standards (XML, RDF, Topic Maps, thesaurus standards, lexicographic standards) and how they interact with display

- Software considerations for thesaurus-building systems

- Last, but not least, domain knowledge, often in multiple domains