Knowledge Organization Systems in Digital libraries

Tutorial

Note: The pdf does not include all pages that are in the paper tutorial booklet

European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2009)
Corfu, Greece
Sept. 27, 2009
Abstract

This introductory workshop is intended for anyone concerned with subject access to digital libraries. It provides a bridge by presenting methods of subject access as treated in an information studies program for those coming to digital libraries from other fields. It will elucidate through examples the conceptual and vocabulary problems users face when searching digital libraries. It will then show how a well-structured thesaurus or other KOS (Knowledge Organization System) can be used as the knowledge base for an interface that can assist users with search topic clarification (for example through browsing well-structured hierarchies and guided facet analysis) and with finding good search terms (through query term mapping and query term expansion — synonyms and hierarchic inclusion). It will touch on cross-database and cross-language searching as natural extensions of these functions. The workshop will cover the KOS structure needed to support these functions: Concept-term relationships for vocabulary control and synonym expansion, conceptual structure (semantic analysis, facets, and hierarchy) for topic clarification and hierarchic query term expansion). It will introduce a few sample thesauri and other KOS to illustrate these principles. Lastly the workshop will give a checklist for evaluating thesauri and other KOS.

Course objectives

Participants should appreciate the complexity of subject access and understand the problems that a thesaurus or other KOS (Knowledge Organization System) can help solve.

Participants should understand the principles of KOS structure.

Participants should be able to apply KOS structure to solving subject access problems.

Participants should be able to identify and evaluate KOS suitable for a specific situation defined by a user community and by the collection of a digital library.

Brief biography of the Instructor

Dagobert Soergel holds an MS equivalent in mathematics and physics (1964) and a PhD in political science (1970), both from the University of Freiburg, Germany. He is Professor of Information Studies, University of Maryland, where he teaches courses in information retrieval, thesaurus development, expert systems, and information technology, and an information systems consultant. He has been a visiting professor at the universities of Western Ontario, Chicago, and Konstanz, Germany. Among other books, he has authored Organizing Information (1985), which received the American Society of Information Science Best Book Award, Indexing Languages and Thesauri. Construction and Maintenance (1974) and numerous papers. He has designed several thesauri, most recently the Alcohol and Other Drug Thesaurus (for which he chairs the advisory committee) and is developing TermMaster, a thesaurus management software package. In 1997 he received the American Society of Information Science Award of Merit.

9:00 - 10:00  KOS functions

9:00 - 9:05  Introduction. Challenges for digital libraries

9:05 - 9:10  Why thesauri: a first look with examples

9:10 - 9:20  User orientation in a concept space; avoid vocabulary confusion

9:20 - 9:25  What is a KOS? A first look with examples

9:25 - 10:00  KOS functions

KOS functions

10:00 - 10:40  KOS structure

9:50 - 10:00  Concept-term relationships

10:00 - 10:40  Conceptual structure: Semantic analysis and facets. Hierarchy

10:40 - 12:00  Implementation, evaluation, resources

10:40 - 11:00  Implementing KOS functions

Implementing KOS functions

11:00 - 11:30  Break

11:30 - 11:40  Implementing KOS functions, continued

11:40 - 11:55  Evaluation of thesauri. Yahoo classification as an example

11:55 - 12:00  Resources

12:00 - 1:00  Examples of classifications and thesauri after p.262, blue divider

Alcohol and Other Drug Thesaurus (AOD Thesaurus)

US National Institute on Alcohol Abuse and Alcoholism (NIAAA)

Medical Subject Headings (MeSH) and
Unified Medical Language System (UMLS)

US National Library of Medicine (NLM)

Art and Architecture Thesaurus (AAT). Getty Foundation

Dewey Decimal Classification. US Libr. of Congress & OCLC/Forest Pr

WordNet. Princeton University, George Miller

CYC Ontology
Part 2. Design and development. Outline

<table>
<thead>
<tr>
<th>Time</th>
<th>Subject</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:30 - 14:35</td>
<td>Introduction and overview</td>
<td>160</td>
</tr>
<tr>
<td></td>
<td>KOS development principles</td>
<td>162</td>
</tr>
<tr>
<td>14:35 - 15:05</td>
<td>The process of KOS development</td>
<td>164</td>
</tr>
<tr>
<td>13:35 - 13:40</td>
<td>The overall process of KOS development</td>
<td>165</td>
</tr>
<tr>
<td>13:40 - 13:55</td>
<td>Sources of concepts, terms, relationships, definitions</td>
<td>166</td>
</tr>
<tr>
<td></td>
<td>Methods of data collection</td>
<td></td>
</tr>
<tr>
<td>13:55 - 14:05</td>
<td>Merging data from many sources</td>
<td>171</td>
</tr>
<tr>
<td>15:05 - 16:00</td>
<td>Developing the conceptual structure</td>
<td>173</td>
</tr>
<tr>
<td>14:05 - 14:30</td>
<td>Facet analysis 1: Education (starting with classes from DDC)</td>
<td>174</td>
</tr>
<tr>
<td>14:30 - 14:40</td>
<td>More facet examples: Job titles</td>
<td>176</td>
</tr>
<tr>
<td>14:40 - 14:50</td>
<td>Guidelines for meaningful arrangement</td>
<td>179</td>
</tr>
<tr>
<td>14:50 - 15:00</td>
<td>Rules for selection of preferred terms</td>
<td>187</td>
</tr>
<tr>
<td></td>
<td>Rules for selection of concepts as descriptors</td>
<td></td>
</tr>
<tr>
<td>16:00 - 16:30</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>16:30 - 17:15</td>
<td>Developing the conceptual structure, continued</td>
<td>177</td>
</tr>
<tr>
<td>16:30 - 17:15</td>
<td>Facet exercise (in pairs)</td>
<td></td>
</tr>
<tr>
<td>17:15 - 18:15</td>
<td>The structure and processing of KOS data</td>
<td>189</td>
</tr>
<tr>
<td>17:15 - 17:30</td>
<td>Interoperability of thesauri/ontologies. Crosswalks</td>
<td>189</td>
</tr>
<tr>
<td>17:30 - 17:45</td>
<td>The structure of a KOS/ontology database</td>
<td>193</td>
</tr>
<tr>
<td>17:45 - 17:55</td>
<td>The many forms of Knowledge Organization Systems (KOS) and their standards</td>
<td>203</td>
</tr>
<tr>
<td></td>
<td>SKOS (Simplified Knowledge Organization Systems) example</td>
<td>217</td>
</tr>
<tr>
<td></td>
<td>OWL (Web Ontology Language) example</td>
<td>223</td>
</tr>
<tr>
<td>17:55 - 18:15</td>
<td>KOS software and its evaluation</td>
<td>209</td>
</tr>
<tr>
<td></td>
<td>Appendix. The KOS development process</td>
<td>257</td>
</tr>
</tbody>
</table>
Challenges for digital libraries

- Improve retrieval effectiveness to handle the sheer mass of material
- Provide unified access to materials in different media (esp. access to non-text materials)
- Provide knowledge-based support for end users who access networked information without the benefit of an intermediary
- Support creation and maintenance of personal or work-group information systems
- Support information seeking as an integral part of problem solving, learning, and intellectual work
- Support collaborative work:
  - Scholarly communication as computer-supported multi-party conversation

KOS (Thesauri, ontologies, taxonomies, ...) must support these functions

Support information seeking as an integral part of problem solving, learning, and intellectual work

- Help users to explore ideas in conjunction with exploring information
- Support fine-grained retrieval and assimilation of information
- Support processing of information along with or after retrieval

Support collaborative work

- Make users full participants in the continuing improvement of information systems through feedback and other contributions
- Establish linkages between people
Why thesauri /KOS:
A first look with examples

Problems

Vocabulary confusion

User orientation in a concept space

Queries illustrating these problems
Queries:

Synonym expansion and Hierarchic expansion

Query 1. Drug use by teenagers

Query 1.1 teenage* AND drug*

Query 1.2 Synonym expansion for teenage*

(teenage* OR teen OR teens OR youth* OR adolescent* OR kid* OR "high school") AND drug*

Query 1.3 In addition, synonym expansion and hierarchic expansion for drug*

(teenage* OR teen OR teens OR youth* OR adolescent* OR kid* OR "high school") AND (drug* OR substance* OR alcohol OR nicotine OR smoking OR cigarette* OR mari*una OR cocaine OR crack OR heroin)

Query 1.4 Query more narrowly focused

(teenage* OR teen OR teens OR youth* OR adolescent* OR kid* OR "high school") AND (cocaine OR crack OR heroin)
Query 1.1. teenage* AND drug* (AltaVista)

About 30 documents match your query.

1. CEIDA Druglinks - Info Centre - PARENTS TALKING TO TEENAGERS ABOUT DRUGS
   What do parents want from their teenagers? Basically, parents want: To know your kids are alright and not in danger. To know your kids think you're OK...

2. CEIDA Druglinks - Info Centre - PARENTS TALKING TO TEENAGERS ABOUT DRUGS
   Better Ways of Communicating. Different points of view Communication is the key to resolving problems, if they exist. Or to finding out if they exist....

3. Testimony of Donna E. Shalala, Secretary of HHS on Teenage Drug Use
   Testimony of Donna E. Shalala, Secretary of Health and Human Services on Teenage Drug Use. Testimony of. Donna E. Shalala. Secretary of Health and Human...

4. Statement of Senator Richard C. Shelby on Teenage Drug Use

5. Testimony of John P. Walters on Teenage Drug Use
   Testimony of John P. Walters, President of The New Citizenship Project, on Teenage Drug Use. Testimony by. John P. Walters* President of the New...

6. Drug Use Rises for Teenagers

7. CEIDA Druglinks - Info Centre - PARENTS TALKING TO TEENAGERS ABOUT DRUGS
Query 1.2. Synonym expansion of teenager

(teenage* OR teen OR teens OR youth OR adolescent* OR kid* OR "high school")
AND drug*

About 249 documents match your query.

1. Adolescent Drug Abuse Treatment Outcome
   Adolescent Drug Abuse Treatment Outcome. Executive Summary. This is a report on the evaluation of an inpatient adolescent drug abuse treatment program in..
   http://www.cbc.med.umn.edu/~andy/drugabuse/adoltx.htm - size 3K - 28-Sep-96 - English

2. Poll finds parents overestimate communication with kids on drugs
   03/03/97 - 07:26 PM ET - Click reload often for latest version. Poll finds parents overestimate communication with kids on drugs. NEW YORK - Most parents..

3. Albany Youth Futures shows kids alternatives to drugs, alcohol

4. IPRC Version - Keeping Youth Drug-Free - Exercise #3

5. Online NewsHour: Teen Drug Use Doubling -- August 20, 1996
   THEY'RE NOT SAYING "NO" AUGUST 20, 1996. TRANSCRIPT. Two new and deeply troubling reports have just been released showing that drug abuse among 12 to 17...

6. Kmart: HOTNEWS/Kmart Kids Race Against Drugs Race Results
   Kmart Kids Race Against Drugs. And the winner is... On Saturday, January 18. Jamie Barreiro of Port St. Lucie, FL, Joshua Brown of Willingboro, NJ and...

11. OMH-RC Database Record: Drug Abuse Among Minority Youth: Methodological Issues
   Office of Minority Health Resource Center Database Record. When available, information on where these materials may be obtained has been listed below...
   http://wwwomhrc.gov/mhr2/docs/95D2315.htm - size 3K - 1-May-97 - English
Query 1.3. Plus synonym and hierarchic expansion of “drug*”

(teenage* OR teen OR teens OR youth* OR adolescent* OR kid* OR "high school")
AND (drug* OR substance* OR alcohol OR nicotine OR smoking OR cigarette*)
About 409 documents match your query.

1. **Smoking is NOT for kids!**
   We believe smoking is for adults only. We therefore require that you be at least 18 years of age in order to view this site. Click below to enter the...
   http://www.smokers.org/ - size 820 bytes - 20-Apr-97 - English

2. **Adolescent Drug Abuse Treatment Outcome**
   Adolescent Drug Abuse Treatment Outcome. Executive Summary. This is a report on the evaluation of an inpatient adolescent drug abuse treatment program in..
   http://www.cbc.med.umn.edu/~andy/drugabuse/adoltx.htm - size 3K - 28-Sep-96 - English

3. **Poll finds parents overestimate communication with kids on drugs**
   03/03/97 - 07:26 PM ET - Click reload often for latest version. Poll finds parents overestimate communication with kids on drugs. NEW YORK - Most parents..
   http://cgi.usatodaycomieelect/eq/eqJ7&htm - size 2K - 21-May-97 - English

4. **Albany Youth Futures shows kids alternatives to drugs, alcohol/TITLE>**

5. **IPRC Version - Keeping Youth Drug-Free - Exercise #3**

6. **Smoking still increasing among teens**
   Despite a chorus of ignorance one woman wanted to dance... To all of those people who say that national role models are a thing of the past, I want to...

7. **Online NewsHour: Teen Drug Use Doubling -- August 20, 1996**
   THEY'RE NOT SAYING "NO" AUGUST 20, 1996. TRANSCRIPT. Two new and deeply troubling reports have just been released showing that drug abuse among 12 to 17...

8. **KCEOC SUBSTANCE ABUSE/YOUTH PROGRAM**
   KCEOC SUBSTANCE ABUSE/YOUTH PROGRAM. Address: 1611 First Street. Phone Number: 336-5310. FAX Number: 336-5303. Contact Person: Robert Cubit. Target Group..
9. Kmart: HOTNEWS/Kmart Kids Race Against Drugs Race Results
Kmart Kids Race Against Drugs. And the winner is... On Saturday, January 18, Jamie Barreiro of Port St. Lucie, FL, Joshua Brown of Willingboro, NJ and...
http://wwwkmart.coiri/hotnews/hotnews.stm - size 7K - 21-May-97 - English

10. Connecticut Kidslink - Substance Abusing Mothers and Their Children
Inter-agency Committee on Substance Abusing Mothers and Their Children in Connecticut: A Summary of Problems and Solutions. Report Summary by Andy Dodge...
http://statlab.stat.yale.edu/cityroom/kidslink2/welffire/texts/9603-03.html - size 9K - 7-Nov-96 - English

11. OMH-RC Database Record: Drug Abuse Among Minority Youth: Methodological Issues
Office of Minority Health Resource Center Database Record. When available, information on where these materials may be obtained has been listed below...
http://wwwomhrc.gov/mhr2/docs/95D2315.htm - size 3K - 1-May-97 - English

12. Browne for President - Release - teenage smoking
NEWS FROM THE BROWNE FOR PRESIDENT CAMPAIGN. FOR IMMEDIATE RELEASE August 23, 1996. Clinton's new "War On Teenage Smoking" is moral grandstanding, charges.
http://wwwharrybrowne96.org/release-teenage-smoking.html - size 4K - 24-Aug-96 - English

13. Teacher Talk, 3(3), Alcohol and Adolescents
Alcohol and Adolescents: Prevention, Intervention, Treatment, Aftercare Volume 3, Issue 3 A Publication Just for Secondary Teachers. 1996 Indiana...
http://education.indiana.edu/cas/tt/v3i3/v3i3toctext.html - size 2K - 6-Jun-96 - English

14. White House Conference on Youth Drug Use
White House Conference on Youth Drug Use. (from March/April 1996 Marijuana Policy Report) In a further attempt to defuse criticism of being soft on drugs..

15. Anti-Smoking Software Installed at Bronx High School of Science
Anti-Smoking Software Installed at Bronx High School of Science. March 6, 1997: The Alumni Association of the prestigious Bronx High School of Science has.

16. Optum: Live Event! Talking to Kids about Alcohol and Drugs
Optum: What is happening This Month at Optum? Check here and find out.
Query 1.4. Drug component more specific

(teenage* OR teen OR teens OR youth OR adolescent* OR kid* OR "high school")
AND (cocaine OR crack OR heroin)

2 documents match your query

1. Teenage "Huffing" - Worse Than Cocaine
   Teenage “Huffing” - Worse Than Cocaine. May 22, 1996. MEEUWSEN: Imagine substances experts call deadlier than heroin or cocaine. Imagine that...

2. Teen is arrested with a kil of crack cocaine
   Teen is arrested with a kilo of crack cocaine. STROUDSBURG, Pa. (AP) - A 14-year-old New York City girl was busted during a bus trip through here last...
Queries: Homonyms and polysemes

Query 2. wordnet (homonym: 6 meanings)

Query 3. classification (polyseme)

Query 3.1. classification AND security
Query 2. Wordnet (homonym: 6 meanings)

3. WordNet: A Lexical Database for English
   Lexical Resources for Human Language Technology. Princeton University.
   DARPA/ITO
   http://www.ito.darpa.mil/Summaries95/B370--Princeton.html - size 12K - 12-Sep-96 -

4. VDI - Racal WordNet Networks

6. WordNet lexical database
   http://www.grafnetix.com/thesaurus/QueryExpansionIntro/node1.html - size 6K -

8. WORDNET, the new generation of digital communications recorders
   Digital communications logging recorder.
   http://www.abds.net/dss/wordnet.htm - size 2K - 30-Jan-97 - English

13. WordNet's Christian Links
    Christian Web Sites. Below is your passport to a wider Christian on-line community. Some contain links to many other Christian sites
    http://www.wordnet.co.uk/links.html - size 3K - 23-May-97 - English

18. The Wordnet Story
    Wordnet Productions. Jesus, the Divine Word, casts his net, the Good News, to all through mass media. Wordnet is a Catholic television ministry dedicated..
    http://www.rlagroup.com/wordnet/wrdntstr.htm - size 2K - 6-Feb-97 - English

30. Tesi di Laurea WordNet
    Linguaggio Naturale. Proposta per Tesi di Laurea: WordNet. WordNet è una base di conoscenza lessicale per l'inglese, disponibile gratuitamente su..
    http://ecate.itec.it:1024/cirave/wordnet.html - size 2K - 30-Sep-96 - Italian

48. WORDNET Language Translation Service
    WORDNET is a team of language experts specializing in foreign language translation, typesetting and printing. In recent years, we have helped a number of..

52. Consortium of the EuroWordNet project
    All Rights reserved by Computer Centrum Letteren University of Amsterdam. Coordinator: builder of...
Query 3. classification (polyseme)

Examples from AltaVista search

1. GNWT Administrative Records Classification System BUILDINGS AND PROPERTIES

   BUILDINGS AND PROPERTIES - DAMAGES 2063. Records relating to damages incurred by government buildings, facilities and structures. It includes...

2. LC Classification: U - Military Science

   http://www.library.yorku.ca/lc/u.html - size 6K - 13-Nov-96 - English

7. Table Tennis Classification Procedures

   International Paralympic Committee. Sports Science | Medical | Sports | Secretariat | General. Table Tennis Classification Procedures. A. Purposes. 1) To..
   http://info.lboro.ac.uk/research/paad/ipc/table-tennis/class-proc.html - size 7K - 2-Jul-96

8. MPW Public Highways (Road Classification)

   ROAD CLASSIFICATION. Law No. 13 of the year 1980 (UU 13/1980) concerning roads distinguishes the category of road into public and special roads. The...

9. Hurricane and Tropical Storm Classification

   http://www.hiwaay.net/cwbol/scale.html - size 3K - 7-Jul-95 - English

17. DEPARTMENT OF ENERGY FUNDAMENTAL CLASSIFICATION POLICY REVIEW

   Secretary Hazel O'Leary has emphasized the importance of improved public accountability

29. Subject guide to the classification

   Subject guide to the Library of Congress classification. For subjects not listed here please consult the printed, red-bound Subject Index in the entrance..
   http://potter.cc.keele.ac.uk/depts/li/lctable.htm - size 7K - 21-May-97 - English

30. BRYOPHYTES: Hornwort Classification

   Phylum ANTHOCEROTOPHYTA. DENDROCEROS. Gametophyte plant with horn-like sporophyte. copyright ©1996 Southern Illinois...
   http://www.science.siu.edu/bryophytes/anthocerotophyta.html - size 940 bytes - 5-Apr-97
31. **Policy & Planning Support - Staff Level Classification**
   Staff Classification & Level. All staff are assigned a classification on employment. This data element indicates the classification...

34. **Classification Reform Approval**
   March 5, 1996. FOR IMMEDIATE RELEASE. Release No. 14. POSTAL SERVICE APPROVES CLASSIFICATION REFORM RECOMMENDATIONS

57. **Universal Decimal Classification Index 5414**
   NATURAL SCIENCES. MATHEMATICS. 54 CHEMISTRY. MINERALOGY. 541 GENERAL, THEORETICAL, AND PHYSICAL CHEMISTRY. 5414 CHEMICAL.
   http://www.chem.ualberta.ca/~plambeck/udc/u5414.htm - size 827 bytes - 9-May-97 -

61. **Draft Public Guidelines to Department of Energy Classification of Information**

71. **The GNU C Library - Classification of Characters**
   This section explains the library functions for classifying characters.
   http://www.ia.pw.edu.pl/Pl-iso/tex-info/libc/libc_55.html - size 7K - 6-Apr-94 - English

80. **Dewey Decimal Classification System**
   Dewey Decimal Classification System. Overview. 000 Generalities 100 Philosophy & psychology 200 Religion 300 Social sciences 400 Language 500 Natural...
   http://www.tnrddlib.bc.ca/dewey.html - size 38K - 7-Aug-96 - English

88. **Extended Computing Reviews Classification Scheme**
   Extended Computing Reviews Classification Scheme. Computing Reviews Classification System. Copyright 1994, by the Association for Computing Machinery....
   http://www.dpmms.cam.ac.uk/MR/CRclass.html - size 37K - 1-Sep-95 - English

89. **627.440 - Classification of costs.**
   627.440 - Classification of costs. Standard Number: 627.440. Standard Title: Classification of costs. SubPart Number: D. SubPart Title: Administrative...
   http://www.doleta.gov/regs/cfr/20cfr/toc_Part600-699/0627.0440.htm - size 12K -

90. **Pirelli Cumbria Rally 1996 Final Classification**
   Pirelli Cumbria Rally 1996. Final Classification. POSITION OVERALL CLASS NUMBER CREW CLASS TOTAL 1 201 Richard Tuthill/Nick Kennedy Vauxhall Nova 1300...
   http://www.idiscover.co.uk/tcs21/1996/pirelli/c_class.html - size 2K - 5-May-96 - English
117. **Classification of Students**

Classification of Students. Students at Bemidji State University are classified as regular, special, or auditor. Regular: A regular student is one who is..

Query 3. classification

Examples from Lycos search

2) Classification of Signatures
   http://www.seas.gwu.edu/faculty/pbock/SignatureCla [99%]

5) Supervised Classification
   Neural Network Classification of Multispectral Imagery Supervised Classification.
   http://www.ece.arizona.edu/~paola/SupervisedClass. [99%]

139) RESIDENCE CLASSIFICATION
   Residence Classification Residence Classification Nonresident students seeking to become California residents for tuition/fee purposes must petition.
   http://www.reg.uci.edu/REGISTRAR/SOC/rc.html [99%]

152) PRODUCT CLASSIFICATION
   EPA may classify a pesticide product for restricted use if its characteristics warrant special handling. Restricted use pesticide.
   http://hammock.ifas.ufl.edu/txt/fairs/26668 [99%]

426) Dewey Decimal Classification Home Page
   DDC 21 and Dewey for Windows now available! OCLC Forest Press is pleased to announce the publication of DDC 21, the latest edition of the Dewey Decimal.
   http://www.oclc.org/fp/ [99%]

429) Dewey Decimal Classification Web Site
   The Dewey Decimal Classification: Numbers You Can Count On catalog is now available. Use the online form to have.
   http://www.oclc.org/oclc/fp/fptxthm.htm [99%]

634) Library of Congress Classification System Introduction
   Introduction to the LC Classification System Some say Information is Power. Others say Information is the door to Knowledge. Libraries hold the key to.
   http://snoopy.tblc.lib.fl.us/laudem/Introduction.h [99%]
Query 3.1. classification and security

Examples from AltaVista search

Restricts results but also misses a lot.

1. **EXSYS: Specific Applications: Security Classification**
   Nuclear Weapons Security Classification. US Dept. of Energy. Nuclear...

2. **SLATE Application Note --Security Classification and Automatic Page Marking wi**
   Introduction. If your document contains classified information, you can identify the classification by.

3. **Computer Security Classification**
   The Classification. alert Advisories on various security vulnerabilities. dict Dictionaries and word lists. doc Security related documents.
   access_control.
   http://www.cs.purdue.edu/coast/archive/Classification.html - size 8K - 17-Mar-95 - English

4. **355 Security Classification Control (R)**
   Top] -- MARC Field Guides Table of Contents -- 300 - Physical Description Fields. 355 Security Classification Control (R)Contains specifics pertaining to..

5. **Security and Classification**
   By John Pike (johnpike@clark.net) The classification system is designed primarily to protect the confidentiality of certain...

6. **National Security Classification Cost Estimates**
   A report to Congress from the Information Security Oversight Office
What is a thesaurus / KOS? A first look

A dictionary is a listing of words and phrases giving information such as spelling, morphology and part of speech, senses, definitions, usage, origin, and equivalents in other languages (bi- or multilingual dictionary).

A thesaurus is a structure that manages the complexities of terminology and provides conceptual relationships, ideally through an embedded classification/ontology.

A thesaurus may specify descriptors authorized for indexing and searching. These descriptors form a controlled vocabulary (authority list, index language).

A monolingual thesaurus has terms from one language, a multilingual thesaurus from two or more languages.

A classification is a structure that organizes concepts into a hierarchy, possibly in a scheme of facets. The term ontology is often used for a shallow classification of basic categories or a classification used in linguistics, data element definition, or knowledge management or (increasingly) for any classification.

In AI-related contexts, an ontology is a classification with a rich set of semantic relationships that support reasoning.
AOD navigation page here
What is a thesaurus / KOS?

**EF**

route of administration

**EF2** — by scope of drug action
- EF2.2 topical and local administration
- EF2.2.2 topical administration
- EF2.2.4 local drug administration
- EF2.4 systemic administration

**EF4** — by method or body site
- EF4.2 enteral administration
- EF4.2.2 oral enteral administration
- EF4.2.4 rectal enteral administration
- EF4.4 mucosal administration
- EF4.4.2 transdermal administration
- EF4.4.4 inhalation, smoking, sniffing
- EF4.4.4.2 smoking
- EF4.4.4.2.2 smoking w/out inhalation
- EF4.4.4.2.4 smoking with inhalation
- EF4.4.4.4 pulmonary administration
- EF4.4.6 oral mucosal administration
- EF4.4.6.2 buccal administration
- EF4.4.6.4 sublingual administration
- EF4.4.8 rectal mucosal administration
- EF4.6 parenteral administration
- EF4.6.2 intravenous injection
- EF4.6.2.2 intravenous infusion
- EF4.6.4 intra-arterial injection
- EF4.6.6 intraperitoneal administration
- EF4.6.8 intracutaneous injection
- EF4.6.10 administration through skin implant
- EF4.6.12 subcutaneous injection
- EF4.6.14 intramuscular injection
- EF4.6.16 CNS injection
- EF4.6.16.2 intrathecal injection
- EF4.8 skin administration
  - (The full entry shows Narrower Term cross-references to the more specific methods involving the skin: EF4.4.2, EF4.6.8, EF4.6.10, and EF4.6.12)

**EF4.10** oral administration
  - (NT to EF4.2.2, EF4.4.4.2, and EF4.4.6)

**EF4.10** rectal administration
  - (NT to EF4.2.4 and EF4.4.8)

**EF6** drug administration by self vs. others
- EF6.2 self administration of drugs
- EF6.4 drug administration by others

Excerpt from a thesaurus hierarchy
EF route of administration

SN The way in which a substance reaches its site of action in the body. The substance may be administered for therapeutic or psychoactive effects - possibly as part of a human or animal experiment - by a third party or by the subjects themselves, or the subject may be exposed to the substance through the environment or in utero.

The major distinction between routes of administration is not the site where a substance is introduced or applied to the body, or even the way it is introduced or applied, but whether it takes effect merely in the local area where it is applied or whether it reaches its destination through systemic circulation. A further criterion is whether the drug reaches systemic circulation directly or whether it first passes through the liver, where it may be metabolized or excreted (first-pass effect in enteral administration). (Note: Drugs administered into the systemic circulation by any route, excluding intra-arterial injection, are subject to possible first-pass elimination in the lung prior to distribution to the rest of the body.)

Whether administration of a drug results in local or systemic action depends not only on the site and method of administration but also on the properties of the drug; sometimes the drug has both local and systemic action. This is particularly true for application to a mucous membrane, which may be intended for a local action but also may have - sometimes unwanted - systemic action. Furthermore, a drug may be absorbed at several sites (e.g., the mouth and the lung, the rectum and the intestine) in various proportions. To account at least partially for the very complex phenomena of the absorption of drugs into the body, the following classification uses two dimensions, or facets: By scope of drug action and by method or body site of administration. To index route of administration completely, use at least one descriptor from each facet.

ST medication route
ST method of delivery of drugs or food
ST mode of substance administration
ST route of drug application
ST route of drug entry
ST route of exposure
BT +EE12 pharmacokinetics
RT +AA2 AOD use
RT +BS AOD substance by route of administration
RT EE12.2 drug absorption
RT +EE14.4.8 drug effect by location
RT +HR drug therapy
RT MD2.2.2.2 drug paraphernalia

EF2 route of administration by scope of drug action

SN Use one of these descriptors in combination with a descriptor from +EF4 route of administration by method or body site.

EF2.2 topical and local administration
SN The application of a substance to a localized area, chiefly for local effects at this site.
NT HU4.2 local anesthesia
RT GH10.2 chemical injury

EF2.2.2 topical administration
SN The application of a substance on the surface of the skin or on a mucous membrane (incl. the gastrointestinal membrane) so that the substance will take effect on the surface or on a localized layer under the surface.
For example, for the administration of a decongestant spray, use EF2.2.2 topical administration combined with EF4.4.4.4 nasal administration.

EF2.4 systemic administration
SN The introduction of a substance into systemic circulation so that it is carried to the site of effect.
NT +EF4.6.2 intravenous injection
NT EF4.6.10 administration through skin implant
NT HU4.4 general anesthesia
RT +GH10.4 chemical poisoning

Examples of full thesaurus entries
## Multilingual KOS problems

<table>
<thead>
<tr>
<th>simian</th>
<th>Affe</th>
</tr>
</thead>
<tbody>
<tr>
<td>monkey</td>
<td>niederer Affe</td>
</tr>
<tr>
<td>ape</td>
<td>Menschenaffe</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>timepiece</th>
<th>Uhr</th>
</tr>
</thead>
<tbody>
<tr>
<td>clock</td>
<td>Wanduhr, Standuhr, Turmuhr</td>
</tr>
<tr>
<td>wall clock</td>
<td>Wanduhr</td>
</tr>
<tr>
<td>standing clock</td>
<td>Standuhr</td>
</tr>
<tr>
<td>tower clock</td>
<td>Turmuhr</td>
</tr>
<tr>
<td>watch</td>
<td>Taschenuhr, Armbanduhr</td>
</tr>
<tr>
<td>pocket watch</td>
<td>Taschenuhr</td>
</tr>
<tr>
<td>wrist watch</td>
<td>Armbanduhr</td>
</tr>
<tr>
<td>alarm clock</td>
<td>Wecker</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>blanket, rug, carpet</th>
<th>Teppich</th>
</tr>
</thead>
<tbody>
<tr>
<td>blanket</td>
<td>Betteppich</td>
</tr>
<tr>
<td>rug, carpet</td>
<td>Bodenteppich</td>
</tr>
<tr>
<td>rug (or carpet)</td>
<td>loser Bodenteppich</td>
</tr>
<tr>
<td>long, narrow rug</td>
<td>Läufer</td>
</tr>
<tr>
<td>(wall-to-wall) carpet</td>
<td>Teppichfußboden</td>
</tr>
<tr>
<td>hanging rug</td>
<td>Wandteppich</td>
</tr>
</tbody>
</table>

*Italics* denotes terms created to express a concept not lexicalized in English or German, respectively.

Note that most English-German dictionaries would have you believe that the German equivalent for "monkey" is "Affe", but that equivalence holds only in some contexts.

Another difficulty arises when two terms mean almost the same thing but differ slightly in meaning or connotation, such as *alcoholism* in English and *alcoholisme* in French, or *vegetable* in English (which includes potatoes) and *Gemüse* in German, which does not. If the difference is big enough, one needs to introduce two separate concepts under a broader term; otherwise a scope note needs to clearly instruct indexers in all languages how the term is to be used so that the indexing stays, as far as possible, free from cultural bias or reflects multiple biases by assigning several descriptors.
Examples of thesauri and other KOS

Alcohol and Other Drug Thesaurus (AOD Thesaurus)
(US Nat. Inst. of Alcohol Abuse and Alcoholism)
http://etoh.niaaa.nih.gov/AODVol1/Aothome.htm

Medical Subject Headings (MeSH) and Unified Medical Language System (UMLS)
(US National Library of Medicine)

Art and Architecture Thesaurus (AAT)
(Getty Foundation)
http://www.getty.edu/research/tools/vocabulary/aat/index.html

Dewey Decimal Classification
(US Library of Congress and OCLC/Forest Press)
http://www.oclc.org/dewey/about/ddc_21_summaries.htm

WordNet (Princeton University, George Miller)
www.cogsci.princeton.edu/~wn/
www.notredame.ac.jp/cgi-bin/wn (Not reachable on July 6, 2002)

CYC Ontology (CYC Corporation)

Example pages form part 2 of the tutorial materials. They will be examined briefly but are intended primarily for further study.
More thesaurus / KOS examples

A few sample pages included

**Yahoo**  The Yahoo classification.  Web pages [www.yahoo.com](http://www.yahoo.com)

**Bloom**  Anderson, L. W., & Krathwohl, D. R.
A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.
http://www.unesco.org/webworld/ramp/html/r8810e/r8810e0e.htm
http://websites.ntl.com/~james.atherton/learning/bloomtax.htm,
http://sweep.riv.csu.edu.au/td/bloom.html,
http://faculty.washington.edu/~krumme/guides/bloom.html
http://www.uwsp.edu/education/lwilson/curric/newtaxonomy.htm

Original version, 1956

Taxonomy of educational objectives.  The classification of educational goals / by a committee of college and university examiners ; Benjamin S. Bloom, editor.
Prelim. ed.
New York : Longmans, Green, 1954-

**SOC**  Standard Occupational Classification  2000
Bureau of Labor Statistics (BLS) + other agencies
http://stats.bls.gov/soc/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/contstan.html

**ERIC**  Education Resources Information Center Thesaurus.  13th ed. Bibliographic retr.
http://searcheric.org/
Additional 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 Estadistica, Geografia e Informatica 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."


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 employed 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.
  http://www.itis.usda.gov/
  http://www.ucmp.berkeley.edu/help/taxaform.html

- Metadata schemas (such as CSDGM), data element dictionaries, object hierarchies in object-oriented programming
KOS development principles

(1) Determine the right organizational scope or user group(s) to be served

(2) Determine the right function scope

(3) Ensure adoption by supporting many views through an inclusive, flexible KOS database

(4) Build on existing KOS. Reuse, reuse, reuse

(5) Use automation for efficiency

(6) Capitalize on collaborative creation and editing
KOS development principles

Good cost-effective KOS development stands and falls with the observation of six simple principles.

(1) **Determine the right organizational scope or user group(s)** to be served. A broad organizational scope results in higher return on investment but must be balanced against tailoring the KOS to specific requirements (but see principle (3)). The scope can range from a KOS for a single user who needs to organize a collection of documents, Web sites, files, email messages, notes, contacts, appointments, and tasks to a KOS that is used worldwide for organizing many databases in many languages, such as AGROVOC or the Medical Subject Headings. In between would be a KOS for an organization, an online community, or a specific database such as AERS, FDA’s Adverse Events Reporting System database (fda.gov/cder/aers/default.htm). A broad scope of application (many departments in an organization, several online communities with related purposes, databases in the same general subject domain) improves the return on investment into the KOS and fosters semantic interoperability.

(2) **Determine the right function scope.** The more functions are served by a KOS the higher the return on investment. Therefore it is important to identify all possible applications that could profit from KOS support. This includes functions such as retrieval from any kind of database or through a Web search engine and structured output of results, project planning, insurance billing, natural language processing, reasoning in expert systems; see XXX for an extensive list. Understanding all these functions is crucial if one wants to maximize ROI. For each application state

- requirements and the spectrum of users’ level of knowledge.
- required characteristics of the KOS

(3) **Ensure adoption by supporting many views through an inclusive, flexible KOS database.** Different applications need different views for intrinsic reasons that are rooted in the nature of the application (for example, to optimize reasoning over large knowledge bases) or for historical reasons (a given user group may not want to change what they are used to, or a large body of material can not feasibly be re-indexed). The solution is a flexible comprehensive KOS database in which different views can coexist or from which different views can be easily extracted; that way each application’s or group’s view of the domain can be accommodated, and that removes a key barrier to engaging several groups that might otherwise go their own ways (at great expense), even though all groups may agree on 80% and disagree on just 20% of concepts, terms, and relationships. There is no need to force everybody into the same mold; many views can coexist within the same KOS database, where they can be related to each other to the extent possible, capturing commonalities and explicating differences. Through the KOS database, different groups can learn from each other; this may lead to improvements of each group’s view. In that sense, a KOS database

- records present usage by different groups (like a dictionary records usage);
- creates an overarching well-ordered structure for all the concepts from different views;
- through that structure, creates relationships between different views;
- through that structure, recommends improvements for each group’s view.

(4) **Build on existing KOS. Reuse, reuse, reuse.** There is enormous intellectual capital in existing KOS; use it! Identify the KOS presently used, explicitly or implicitly, in the organizational and functional scope. Find other KOS that could meet some or all of the requirements, perhaps with some modification, or that would be useful sources for developing a new KOS (see Box 1. Searching for KOS).

(5) **Use automation for efficiency.** Automation can be used to massage data from existing KOS for the purpose at hand. It can also be used to extract terms, concepts, relationships, and global structure from texts and document collections. While automation cannot entirely replace intellectual effort, a KOS constructed using automatic processing alone is better than no KOS at all.

(6) **Capitalize on collaborative creation and editing** with central expert support and some measure of control. Possibly have a cadre of KOS editors throughout the organization or throughout the world.
Functions of a KOS  
(thesaurus / classification / ontological knowledge base)  
in the context of digital libraries

Support learning and assimilating information.

Assist researchers and practitioners with problem clarification.

Support information retrieval.

Provide knowledge-based support for end-user searching.

Support meaningful information display.

Provide a tool for indexing.

Facilitate the combination of multiple databases or unified access to multiple databases.

Support document processing after retrieval.

Support learning and assimilating information

Support learning about any topic by providing the learner with a coherent, age-appropriate conceptual framework.

Learning as information retrieval. Conceptual framework for asking the right questions.

Assist readers in understanding text.
Assist researchers and practitioners with problem clarification —

provide the conceptual basis for the design of good research and implementation and for good
query formulation. Includes help with

exploring the conceptual context of a research or practical problem — a study, policy,
plan, or implementation project

and with

structuring the problem.

Examples of specific functions:

Present the issues in a field or application area in a coherent framework.

Assist in problem-solving: Assist in the exploration of the dimensions of a problem and
aspects to be considered in its solution; provide a classification of approaches to solving a
specific problem.

Provide classification and consistent definition of variables for research / of evaluation
criteria for practical problems, thus enhancing the comparability of research and
evaluation results and making research more cumulative.
Support information retrieval

Provide knowledge-based support for end-user searching. Support searching in multiple natural languages; free-text searching; searching multiple databases using different index languages.

Elicitation of user needs through a series of menus based on search tree, or through guidance in the conceptual analysis of a search topic (questions based on a facet structure, presentation of a segment of the concept hierarchy for each applicable facet).

Browsing the classification structure to identify useful concepts for a search at the level of specificity desired. Browsing a collection, as in a subject directory.

Mapping from the user's query terms to descriptors used in a database or to the multiple natural language expressions to be used for free-text searching.

Inclusive (hierarchically expanded) searching.

Enhanced ranking algorithms based on concept and term relationships.

Searching multiple databases by mapping the users query terms to the descriptors used in each of the databases, or mapping the descriptors from one database to another databases (switching); common search language.

Support information retrieval, continued

Support information display, especially presentation of search results:

Meaningful arrangement of units (document records, paragraphs, property data on a given substance assembled from several databases), including knowledge-based clustering of records retrieved.

This supports exploration of large retrieved sets and, by extension, exploration of the content of an entire collection or subcollection.

Meaningful arrangement of information within a record (for example meaningful ordering of descriptors assigned).
Support information retrieval, continued

Provide a tool for indexing.

Vocabulary control.

User-centered (request-oriented, problem-oriented) indexing.

Indexing several databases in a field with a common index language and sharing the results of indexing to reduce overall indexing effort.

Mapping indexing descriptors from one system to another.

Support information retrieval, continued

Facilitate the combination of multiple databases or unified access to multiple databases through

mapping the users query terms to the descriptors used in each of the databases;

mapping the query descriptors from one database to another (switching);

providing a common search language from which to map to multiple databases;

providing a common index language for a number of databases in a field;

mapping indexing descriptors from one database to another.
Support document processing after retrieval

For example

**Highlight descriptors responsible for retrieval**, using different colors for different facets.

**Highlight** terms belonging to a given category, for example, **personal names**, again using different colors for different categories.

**Prepare document summaries**, possibly in a different language, taking into account the query topic.

**Translate full documents**.

**Extract facts from text.** Compile and arrange facts extracted from several texts.

The underlying function of a knowledge base on concepts and terminology

**Map out a concept space, relate concepts to terms, and provide definitions**, thus providing orientation and serving as a reference tool.

Provide a **semantic road map and common language** for an individual field and, perhaps more importantly, map the relationships among fields.

**Clarify concepts by putting them in the context of a classification** / typology and to provide a system of definitions.

**Relate concepts and terms across disciplines, languages, and cultures.**
KOS (thesaurus / ontology functions)

Reference list
Functions of a KOS
thesaurus / classification / ontological knowledge base

Overview

Provide a semantic road map to individual fields and the relationships among fields. Map out a concept space, relate concepts to terms, and provide definitions, thus providing orientation and serving as a reference tool.

Improve communication generally. Support learning and assimilating information.

- Support learning through conceptual frameworks. Conceptual framework to help the learner ask the right questions.
- Support the development of instructional materials through conceptual frameworks.
- Assist readers in understanding text by giving the meaning of terms.
- Assist writers in producing understandable text by suggesting good terms.
- Support foreign language learning.

Provide the conceptual basis for the design of good research and implementation.

- Assist researchers and practitioners with problem clarification.
- Consistent data collection, compilation of statistics (related to information analysis)

Provide classification for action. Classification for social and political purposes

- a classification of diseases for diagnosis,
- of medical procedures for insurance billing,
- of commodities for customs.

Support information retrieval and analysis. Organizing and keeping track of goods and services for commerce (esp. ecommerce) and inventory

- Provide a tool for searching, particularly knowledge-based support for end-user searching, including hierarchically expanded searching.
- Provide a tool for indexing.
- Facilitate the combination of or unified access to multiple databases
- Support document processing after retrieval.

Support meaningful, well-structured display of information.

Ontology for data element definition. Data element dictionary.

Conceptual basis for knowledge-based systems.

Do all this across multiple languages

Mono-, bi-, or multilingual dictionary for human use.
The underlying function of a knowledge base on concepts and terminology:

Provide a semantic road map to individual fields and the relationships among and across fields.

Map out a concept space, relate concepts to terms, and provide definitions, thus providing orientation and serving as a reference tool.

Provide a semantic road map and common language for an individual field and, perhaps more importantly, map the relationships among fields.

Clarify concepts by putting them in the context of a classification / typology and to provide a system of definitions.

Relate concepts and terms across disciplines, languages, and cultures.

Many specific functions build on this foundation.

Improve communication generally.
Support learning and assimilating information

Support learning about any topic by providing the learner/reader with a coherent, age-appropriate conceptual framework. Conceptual frameworks help the learner ask the right questions; learning as information retrieval.

Assist readers in understanding text; help them ascertain the proper meaning of a term and placing it in context.

Support the development of instructional materials by providing a conceptual framework to the instructional developer / writer and by suggesting didactically useful arrangements of topics.

Assist writers in producing understandable text by helping them to conceptualize the topic and suggesting from a semantic field the term that best conveys the intended meaning and connotation.

Support foreign language learning.
Provide the conceptual basis for the design of good research and implementation.

Assist researchers and practitioners with problem clarification

Includes help with

exploring the conceptual context of a research or practical problem — a study, policy, plan, or implementation project

and with

structuring the problem and providing a conceptual framework for asking the right questions and devising good query formulations for retrieval.

Examples of specific functions:

Present the issues in a field or application area in a coherent framework.

Assist in problem-solving: Assist in the exploration of the dimensions of a problem and aspects to be considered in its solution; provide a classification of approaches to solving a specific problem (for example, a classification of approaches to drug abuse prevention as a help in designing drug abuse prevention projects).

Provide classification and consistent definition of variables for research / of evaluation criteria for practical problems, thus enhancing the comparability of research and evaluation results and making research more cumulative.

Support the compilation and use of statistics

This is a very important function. The Census Bureau, the Bureau of Labor Statistics, and other statistical agencies are heavily involved in developing classifications and defining concepts.

Support data collection

The concepts in a classification used for statistics not only make the collected data retrievable, they define the very nature of the data.

Support data aggregation

For example, get the value of all electronic goods imported into the US in the year 2000, or the tonnage of green leafy vegetables produced in a given year in the US.

Support retrieval of specific numbers (also part of information retrieval)

Support data tabulation and analysis (Need to have proper variables available)
### Provide classification for action

This list addresses the functions of formal classifications. In a broader perspective, classification is the basis for much of everyday action, where we put people, things, and events in certain categories and, based on these categories, predict the behavior of persons and things and the course and effects of events, determine our attitudes towards them, and plan action accordingly.

For example,

- a classification of diseases for diagnosis,
- a classification of medical procedures for insurance billing,
- a classification of medical outcomes to assist with treatment evaluation,
- a classification of commodities for customs,
- a classification of educational objectives for instructional development,
- a classification of occupations for matching job applicants with job openings and for pay scale;
- a classification of skills for employee task assignments.
- a classification of crimes for determining sentences
- a classification of types of expenses for tax purposes

### Classification for social and political purposes. Socially charged classification

For example

Establishing that a profession has its own knowledge base, thereby enhancing the recognition of the profession (for example, the Nursing Intervention Classification)

Establishing a persons condition or behavior as normal, or as a disease, or as a moral failing or otherwise deviant. Different groups may want the same condition or behavior classified in different ways to further their agenda

**Examples:**

Should homosexuality be classified as a disease?

Is alcoholism or other drug abuse a disease or a moral failing?

Is mental illness a disease on a par with physical illness, and thus covered by health insurance the same way?

Is some levy to be classified as a *tax* or as a *user fee*
Support information retrieval 1:

A tool for searching, particularly knowledge-based support for end-user searching. Support

searching in any kind of database — bibliographic, full-text and hypermedia, directory, numeric, etc.;

searching in any kind of medium — printed indexes, CD-ROM systems, online systems, and the Internet;

searching in multiple natural languages independent of the language used in each database;

free-text searching;

searching multiple databases using different index languages.

Elicitation of user needs through a series of menus based on a search tree, or through guidance in the conceptual analysis of a search topic (questions based on a facet structure, presentation of a segment of the concept hierarchy for each applicable facet).

Browsing the classification structure to identify useful concepts for a search at the level of specificity desired. (The user may not have command of the vocabulary needed.)

Browsing a collection (as on the shelves or in a subject directory)

Mapping from the user's query terms to descriptors used in a database or to the multiple natural language expressions to be used for free-text searching.

Inclusive (hierarchically expanded) searching.

Enhanced ranking algorithms that use concept and term relationships.

Searching multiple databases by mapping the users query terms to the descriptors used in each of the databases, or mapping the descriptors from one database to another databases (switching); common search language.

Support information retrieval 2: Provide a tool for indexing.

Vocabulary control.

User-centered (request-oriented, problem-oriented) indexing.

Indexing several databases in a field with a common index language and sharing the results of indexing to reduce overall indexing effort.

Mapping indexing descriptors from one system to another.
Support information retrieval 3:

Facilitate the combination of multiple databases or unified access to multiple databases through

- mapping the users query terms to the descriptors used in each of the databases;
- mapping the query descriptors from one database to another (switching);
- providing a common search language from which to map to multiple databases;
- providing a common index language for a number of databases in a field;
- mapping indexing descriptors from one database to another.

Support information retrieval 4: Document processing after retrieval

Sample functions that require knowledge-based support:

- Meaningful arrangement of search results (see next box)
- Highlight descriptors responsible for retrieval, using colors to show facets.
- Highlight terms belonging to a given category, for example, personal names, again using different colors for different categories.
- Prepare document summaries, possibly in a different language, taking into account the query topic.
- Translate full documents.
- Extract substantive data from text. Compile and arrange data extracted from several texts.

Support meaningful, well-structured display of information

Meaningful arrangement of units (document records, paragraphs, property data on a given substance assembled from several databases), including knowledge-based clustering of records retrieved. This includes meaningful structure for Web sites and subject directories

This supports exploration of large retrieved sets and, by extension, exploration of the content of an entire collection or subcollection.

Meaningful arrangement of information within a unit (for example meaningful ordering of descriptors within a bibliographic record).
Organizing and keeping track of goods and services for commerce (esp. ecommerce) and inventory

The functions detailed for information retrieval apply to this special case

Organize a store, an inventory, an online merchandise catalog, a yellow page directory so items can be found

Display the inventory in a meaningful arrangement so users can find things (as in a store)

Keep track of inventory

These functions apply both to business-to-consumer and to business-to-business commerce. Classification by function or purpose is especially important here.

Ontology for data element definition.

Data element dictionary.

Consider data processing systems in a multinational corporation

Conceptual basis for knowledge-based systems.

Do all this across multiple languages
Mono-, bi-, or multilingual dictionary for human use.

Printed or machine-readable, such as dictionary on CD-ROM or a thesaurus used in conjunction with a word processor

Dictionary/knowledge base for automated language processing

Machine translation and natural language understanding (data extraction, automatic abstracting/indexing). (It should be noted that parsing natural language requires not only morphological information and information about the possible syntactic roles of a term but also a great deal of semantic information.)

Spell check dictionary

Knowledge base for grammar checking.

Functions of an ontological knowledge base in software development

Assist in the design and implementation of the user interface, esp. choice of terms and icons.

Terms and icons must be chosen with the sometimes conflicting goals of communicating to the intended user group and of adhering to standards.

Assist in the organization and formulation of help messages and of documentation and third-party software books.

Serve as the lexicon for machine translation of interfaces and software-related documents

Assist the user in understanding interfaces and documentation, esp. in a foreign language.

Support retrieval of software for the end user or for software reuse.

Data element definition and standardization and organization of CASE tool databases.

All this functionality must be provided in multiple languages (for example, software localization for end users, CASE tool databases for multinational development teams).

End of reference list of KOS functions
User-centered indexing / request-oriented indexing

Construct a classification/ontology (embedded in a thesaurus) based on actual and anticipated user queries and interests.

Thus provide a conceptual framework that organizes user interests and communicates them to indexers.

Index materials from users' perspective:
Add need-based retrieval clues beyond those available in the document. Increase probability that a retrieval clue corresponding to a query topic is available.

Index language as checklist.
Indexing = judging relevance against user concepts. Relevance rather than aboutness

Implementation:
Knowledgeable indexers
Expert system using syntactic & semantic analysis & inference.
User-centered indexing / request-oriented indexing.

Sample concepts included in the index language due to user interest

Systemic administration

Intergenerational social mobility

Biochemical basis of behavior

Longitudinal study
User-centered / request-oriented indexing.
Sample documents with descriptors

Document
The drug was injected into the aorta
User concept: *Systemic administration*

Document:
The percentage of children of blue-collar workers going to college
User concept: *Intergenerational social mobility*

Document:
CSF studies on alcoholism and related behaviors
User concept: *Biochemical basis of behavior*
User concept: *longitudinal study*
(Longitudinal not mentioned in the document; determined through careful examination of the methods section.)
Design of a classification scheme for fiction based on analysis of actual user-librarian communication

Annelise Mark Pejtersen

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Freq.</th>
<th>Sub-classes</th>
<th>Freq.</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. subject matter</td>
<td>38</td>
<td>a. action and course of events (plot)</td>
<td>10</td>
<td>a. mystery novel, book with action</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. psychological development /</td>
<td>5</td>
<td>b. love story, book with psychology</td>
</tr>
<tr>
<td></td>
<td></td>
<td>description</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. social relations</td>
<td>19</td>
<td>c. family chronicle, not with politics</td>
</tr>
<tr>
<td>2. frame</td>
<td>25</td>
<td>a. time</td>
<td>13</td>
<td>a. historical novel, books from 16th and 17th</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. place (geographical, place</td>
<td>12</td>
<td>b. travelogue, books from the countryside, books</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(geographical, social environment,</td>
<td></td>
<td>about working people</td>
</tr>
<tr>
<td></td>
<td></td>
<td>professions)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. author's intention /attitude</td>
<td>37</td>
<td>a. emotional experience</td>
<td>34</td>
<td>a. humoristic, suspense, amusing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. cognition / information</td>
<td>3</td>
<td>b. philosophical, political, not too many problems</td>
</tr>
<tr>
<td>4. accessibility</td>
<td>34</td>
<td>a. readability</td>
<td>16</td>
<td>a. easy, not complicated, not heavy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. physical characteristics</td>
<td>18</td>
<td>b. typography, modern/old, series, size, volume</td>
</tr>
<tr>
<td>Other formulations</td>
<td>87</td>
<td>a. author's name / title</td>
<td>25</td>
<td>b. something like Emily Bronte</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. author's name / title as example</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. good book</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>d. diverse</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

Sample user concepts for indexing images

Good scientific illustration

Useful for fundraising brochure

Appealing to children

Cover page quality
User-centered /request-oriented / problem-oriented indexing

As summarized in the overheads, user-centered indexing involves analyzing actual and anticipated user queries and interests and constructing a framework, a hierarchically structured controlled vocabulary, that includes the concepts of interest to the users and thus communicates these interests to the indexers or an expert system that can infer user-relevant concepts from text. The indexers then become the "eyes and ears" of the users and index materials from the users' perspective. The indexer uses the structured list of user-relevant concepts as a checklist, applying her understanding of a document (or other object) to judge its relevance to any of these concepts. This process ensures that users will find the documents that they themselves would judge relevant upon examination.

Request-oriented indexing contrasts with document-oriented indexing, where the indexer simply expresses what the document is about or where simply the terms in the text are used. But, as the examples show, a document can be relevant for a concept without being about the concept: a document titled *The percentage of children of blue-collar workers going to college* is not necessarily about *intergenerational social mobility*, but a researcher interested in that topic would surely like to find it, so it is relevant.

Request-oriented indexing is essential for good performance in fiction retrieval and even more so in image retrieval. Image retrieval profit from descriptors that capture imponderables, such as the mood of an image or from descriptors indicating possible uses of an image (such as

This perspective on indexing has implications for cross-language retrieval: The conceptual framework must be communicated in every participating language to allow a meeting of minds to take place, regardless of the languages of the user and the indexer. This is particularly salient in the context of indexing: One needs to make sure that, as far as possible, the term used by the indexer in one language communicates the same mood as the term given to the user in another language for searching.
Thesaurus / KOS structure

Concept-term relationships

Conceptual structure

Semantic analysis and facets

Hierarchy
Concept-term relationships  
(Terminological structure)

Controlling synonyms

<table>
<thead>
<tr>
<th>Term</th>
<th>Preferred synonym</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teenager</td>
<td>Adolescent</td>
</tr>
<tr>
<td>Teen</td>
<td>Adolescent</td>
</tr>
<tr>
<td>Youth (young person)</td>
<td>Adolescent</td>
</tr>
<tr>
<td>Pubescent</td>
<td>Adolescent</td>
</tr>
<tr>
<td>Black</td>
<td>African American</td>
</tr>
<tr>
<td>Afro-American</td>
<td>African American</td>
</tr>
<tr>
<td>Alcoholism</td>
<td>Alcohol dependence</td>
</tr>
<tr>
<td>Inheritance</td>
<td>Heredity</td>
</tr>
<tr>
<td>Ultrasonic cardiography</td>
<td>Echocardiography</td>
</tr>
</tbody>
</table>
Soergel, p. 215, enlarged
Disambiguating homonyms

administration 1 (management)
administration 2 (drugs)

Läufer 1 (Sportler)  
*English*: runner (athlete)

Läufer 2 (Teppich)  
*English*: long, narrow rug

Läufer 3 (Schach)  
*English*: bishop (chess)

discharge 1 (From hospital or program)  
*German*: Entlassung

discharge 2 (From organization or employment)  
*Preferred synonym*: Dismissal  
*German*: Entlassung

discharge 3 (Medical symptom)  
*German*: Absonderung, Ausfluss

discharge 4 (into a river)  
*German*: Ausfluss

discharge 5 (Electrical)  
*German*: Entladung (which also means unloading)
Importance of terminological structure

The terminological structure is equally important in controlled vocabulary systems and in free-text searching.

In free-text searching,

- synonym expansion of query terms is important for recall
- homonym indicators can trigger a question to the user on the intended meaning of the query term.
Conceptual structure

A well-developed conceptual structure *sine qua non* for user-centered indexing very useful for free-text retrieval as well.

The two principles of conceptual structure

- facet analysis
- hierarchy

Facets and hierarchy interact
Facets.
Semantic factoring or feature analysis

Analyzing a concept into its defining components (elemental concepts or features).

Concept frame with facet slots

liver cirrhosis

Pathologic process: inflammation
Body system: liver
Cause: not specified
Substance/organism: not specified

alcoholic liver cirrhosis

Pathologic process: inflammation
Body system: liver
Cause: chemically induced
Substance/organism: alcohol

hepatitis A

Pathologic process: inflammation
Body system: liver
Cause: infection
Substance/organism: hepatitis A virus
Facet principles

A facet groups concepts that fill the same role:
- concepts that fall under the same aspect or feature in the definition of more complex concepts;
- concepts that can be answers to a given question.

In frame terminology:
The facets listed above are slots in a disease frame; a facet groups all concepts that can serve as fillers in one slot.

Using elemental concepts as building blocks for constructing compound concepts

- drastically reduces the number of concepts in the thesaurus / KOS and thus leads to conceptual economy.
- facilitates the search for general concepts, such as searching for the concept dependence, which occurs in the context of medicine, psychology, and social relations.

Facets can be defined at high or low levels in the hierarchy, as illustrated in the next overhead.
Top-level facets

organism
body part
chemical substances by function
chemical substances by structure

Low-level facets

route of administration
  route of administration by scope of drug action
    (local/topical or systemic)
  route of administration by body part
  route of admin. by method of application
    (injection, rubbing on, etc.)

liver
  liver tissue (hepatocyte, Kupffer cell, etc.)
  liver part (hepatic lobule, portal lobule, etc.)
Hierarchy

groups at high risk of drug use

- suicidal or physically or mentally disabled
- persons from unstable or low-cohesion families
- children of alcoholic or other drug-abusing parents
  - SN Adult or still under age
- children of single teenage mothers
- persons subjected to abuse or neglect (now or past)
  - persons subjected to abuse/neglect by parents
  - latchkey children
  - persons subjected to abuse/neglect by spouse
- single teenage mothers
- school dropouts or those at risk of dropping out
- unemployed or in danger of being unemployed
- economically disadvantaged
- homeless
  - runaway youth
- gateway drug users
- persons engaged in violent or delinquent acts

See also examples given previously in
What is a thesaurus / KOS
Hierarchy from combining two facets

**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 Physics 10th grade NT A1.1.1B2.1; BT A1B2.1
- A1.1.1 Optics
  - A1.1B1 Optics MS BT A1B1
  - A1.1B2 Optics HS BT A1B2
  - A1.1B2.1 Optics 10th grade BT A1B2.1

**B Facet B. Grade level**
- B1 MS
  - A1.1 Science NT B1A1, B2A1
  - A1.1 Physics NT B1A1.1, B2A1.1
  - A1.1.1 Optics NT B1A1.1.1, B2A1.1.1
- B2 HS
  - A1.1 MS Physics BT A1.1
  - A1.1.1 MS Optics BT A1.1.1
- B2.1 10th grade

**Linear arrangement 1**

**Linear arrangement 2**
Uses of facet analysis and hierarchy

Help to organize the concept space and establish concept relationships.

Discover concepts, esp. general concepts spanning several disciplines

Assist the user in analyzing and clarifying a search problem:
  elicit the facets involved
  present hierarchical structure within each facet

Facilitate the search for general concepts, such as
  inflammation or
  dependence (which occurs in the context of medicine, psychology, and social relations)

Hierarchic query term expansion

These functions are useful in both controlled vocabulary and free-text searching.
Concept discovery

through facet analysis and hierarchy building

Through facet analysis and hierarchy building, one often discovers concepts that are needed in searching or that enhance the logic of the concept hierarchy. Need to create terms for these concepts.

Consider

train station, bus station, harbor, airport

Common semantic component: traffic station

gin, whiskey, cherry brandy, tequila, etc.

common semantic component: distinct distilled spirits
(counterpart of the already lexicalized neutral distilled spirits)

transactional analysis, dream analysis, insight therapy, Gestalt therapy, reality therapy, cognitive therapy

Umbrella concept for structuring the hierarchy and for retrieval: analytic psychotherapy
(methods that seek to assist patients in a personality reconstruction through insight into their inner selves)

Payment in exchange for some consideration (see above)
Web-based thesaurus / KOS display and incorporation into search functions

Vignette on thesaurus / KOS use in searching a digital library. The director of a drug-free community coalition is faced with developing a prevention project and the funding for it. Signing on to the AOD Digital Library, she begins by browsing the prevention section of the thesaurus / KOS hierarchy to get a structured overview of various prevention approaches. From the thesaurus / KOS scope notes, some of these approaches seem particularly applicable to her community, so she follows the links to more in-depth explanations. She returns to the thesaurus / KOS and follows a link from prevention through education to a funding program announcement. She opens the guidelines for submitting proposals to this program and copies a proposal template into her private space (shown in another window) and fills in some text and copies some text (which is transferred with the proper source). From the program announcement, she follows a link to projects funded previously and further to project reports and evaluations. She comes across the unfamiliar term triangulation and clicks on it to see the thesaurus / KOS entry, which gives an explanation and the hierarchical context. In another document she highlights the phrase prevention program evaluation to initiate a search in the system and one external database. She copies three references with abstracts to her private space. (Later she will return to these, select one for detailed reading, and add more notes and quotes to her emerging proposal.) Returning to the program announcement, she follows a link to relevant research, selects some articles to read, and adds more material to her outline. One of the papers compares the effectiveness of several prevention curricula. She follows a link to the curriculum that came out on top and from there finds further reviews. She also finds some discussion of resources required. She needs some more data — namely, demographics of her community and funding sources for the required local match — so she initiates searches in two external databases, incorporating the results into her proposal. Now she completes the first draft, including the text itself and annotations that explain why a piece is included or why certain language is used. Before submitting the proposal, she emails two board members and a city staff member for comments, giving them access to her private space. The three people read the draft and add their annotations, including suggested wording. The director now revises the draft, requests the final document in PDF format, links to the agency’s submission system, and sends off her proposal.

The next page shows a proposed digital library structure built around a thesaurus / KOS and the pages following show two steps in a search using the thesaurus / KOS.
Figure 2. Portion from an information structure schema
Having at least one concept from the concept box

**Projects**

Self-expression teen theater institute for academic and personal excellence (SIAPE). **Target group:** High risk junior high school students. Trains groups of teens as peer educators; they perform social dramas in schools, churches, and public housing.

Community youth activities program (CYAP). **Target group:** School dropouts. Information for parents, counseling for families and individuals. Information for parents, counseling for families and individuals.

New Horizons/SUCCESS programs. **Target group:** High school students. Prevention through education using a lifestyle risk reduction approach and skills training.

**Documents**

Self-reported health problems and physical symptomatology in adolescent alcohol abusers. Grant: NIAAA-AA-087646

Family influence on alcohol abuse and other problem behaviors among black and white adolescents in a general population sample. Grants: NIAA-R01-AA-06925

Effects of alcohol price policy on youth: A summary of economic research. Grants: NIAAA-5R01-AA-08359

Alcohol problems among adolescents: Current directions in prevention research.

**Figure 3a.** Search for the concept “adolescent” (enhanced with further concepts from the AOD Thesaurus).
Any of
TA12 adolescent  
TE10.6 pregnant teen  
TN8 high school student  
TU8.10.2 run away youth  
TZ2.2 high risk youth

Any of
LG22 family environment  
LG24 family relations

Projects
Community youth activities program (CYAP). Target group: School dropouts. Information for parents, counseling for families and individuals.

Documents
Family influence on alcohol abuse and other problem behaviors among black and white adolescents in a general population sample. Grants: NIAA-R01-AA-06925

Family support decreases influence of deviant peers on Hispanic adolescents’ substance abuse

Figure 3b. Search results after ANDing a second concept related to family.
Web-based thesaurus / KOS display

Requirements

Browsing a hierarchy at different levels of detail

Hyperlinks for following relationships

Searching for compounds containing any combination of elemental concepts

Searching for a word or phrase (full complement of Boolean and adjacency operators).
Search in the combination of the descriptor field and the synonymous term field

For a controlled vocabulary search:
Insert descriptor or descriptor + narrower terms into search form

For a free text search:
Insert descriptor + synonyms or descriptor + synonyms + narrower terms + their synonyms into search form

The following pages have examples of a proposed interface that is very simple but functional.
(Fancy graphics are often more a hindrance than a help.)
Thesaurus interface pages are in file dlthestut2.pdf
Fig. 2a from DL proposal
Fig 2b from DL proposal
Searching with elemental concepts

Example 1. AOD Thesaurus

Search for:

central nervous system AND disorder

Result:

GH6.10.2 brain injury
GX4 CNS disorder

Search for:

central nervous system

Result:

EF4.6.16 CNS injection
EW8 CNS function
GH6.10.2 brain injury
GX4 CNS disorder
XV4.4.4 CNS sensory pathway
XZ central nervous system
Searching with elemental concepts

Example 2. LC Classification

Search for:

buildings, architecture AND acoustics

Result:

NA2800 Architectural acoustics
TH1725 Soundproof construction

Search for:

vehicles AND acoustics

Result:

TL681.S6 Airplanes. Soundproofing
VM367.S8 Submarines. Soundproofing
Searching with elemental concepts. Ex. 3 DDC

Search for: Payment in exchange for some consideration

Result: general concepts containing this component with examples of more specific Dewey classes (many, but not all, in 330 Economics)

Wage
  331.21 Labor economics / Compensation
  658.32 General management / Personnel management / Wage and salary administration

Price/Cost
  338.52 Production economics / Prices
  339.42 Macroeconomics / Cost of living (Prices)
  354.5285 Public administration / Admin. of agriculture / Agricultural price supports

Interest
  332.82 Financial economics / Interest
  336.2426 Public finance / Income taxes / Interest income

Rent
  336.11 Public finance / Non-tax revenue / Revenues fr. rents
Example for a word search in an online thesaurus / KOS

Search for

commercial AND organization

finds the following record

 corporation
   ST business organization
   ST commercial enterprise
   ST company

Search for the German words

Hirn AND Entzuendung

finds the following record

meningitis
  German Hirn-haut-entzuendung
Searching interaction using facets:

Facets for eliciting user needs

User enters subject field of search.
System displays list of facets (limiting aspects).
User indicates first aspect for limiting the search
Subject field of search: **Education**

Indicate limiting aspects to be used:

- Level of education
- Ethnic origin of students
- Giftedness/handicap of students
- Curriculum subject
- Country of education
- Public/private education
User selects *level of education* descriptor

<table>
<thead>
<tr>
<th>Level of education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preschool</td>
</tr>
<tr>
<td>Kindergarten</td>
</tr>
<tr>
<td>Elementary education</td>
</tr>
<tr>
<td>Secondary education</td>
</tr>
<tr>
<td>Higher education</td>
</tr>
</tbody>
</table>
Searching interaction:

Facets for eliciting user needs

System displays query formulated so far. User indicates *Ethnic origin of students* as a limiting aspect
Subject field of search

Education

Indicate limiting aspects to be used

- Level of education  Elementary
- Ethnic origin of students
- Giftedness/handicap of students
- Curriculum subject
- Country of education
- Public/private education
User selects *Ethnic origin of students* descriptor

---

**Ethnic origin of students**

- Latin American / Spanish American
- Mexican American
- Puerto Rican
- African American
- Asian American
- Chinese American
- Japanese American
- Gypsy
Searching interaction:

Facets for eliciting user needs

System displays query formulated so far. User indicates *Curriculum subject* as the next limiting aspect
Subject field of search

**Education**

Indicate limiting aspects to be used

- Level of education  **Elementary**
- Ethnic origin of students  **African American**
- Giftedness/handicap of students
- Curriculum subject
- Country of education
- Public/private education
Searching interaction:

Facets for eliciting user needs

After a few more interactions, the system displays the completed query formulation
Subject field of search

**Education**

Indicate limiting aspects to be used

- Level of education  **Elementary**
- Ethnic origin of students  **African American**
- Giftedness/handicap of students
- Curriculum subject  **Reading**
- Country of education  **U.S.**
- Public/private education
Thesaurus / KOS-supported Web search engines

These sites are still experimental; they come and go

Using synonym expansion

Google (but limited by thesaurus / KOS)

Using homonym disambiguation

www.hotbot.com/ (used to do this at one time)

www.oingo. com (will change name to appliedsemantics.com)
  Finds Open Directory categories (Full Web search with homonym disambiguation is hard)
  Acquired by Google, usage there not known

Using a large enriched thesaurus /ontology

www.seruba.com/ (defunct)

Using facets

www.endeca.com (used in www.tesco.com/winestore/)

www.i411.com (used in www.yellowpages.co.za)
Implementing thesaurus / KOS functions in retrieval with emphasis on cross-language retrieval

Important principle: Let the system do the work. Full exploitation of thesaurus / KOS power cannot rely on users learning how to use a thesaurus / KOS but rather requires a system that gives behind-the-scene knowledge-based assistance with the thesaurus / KOS serving as the knowledge base.

Controlled vocabulary

With a controlled vocabulary there is a defined set of concepts used as descriptors in indexing and searching. The user can browse the thesaurus / KOS hierarchies to identify search concepts; or the user can start from a term or phrase and consult the thesaurus / KOS to find the proper descriptor(s) or let the system do the mapping behind the scene. In either case, the user need not worry about the various ways each search concept is expressed in natural language. In cross-language retrieval this simply means that the user should be able to use a term in his own language to find documents (or whatever the retrieval objects are) indexed by the corresponding descriptor (concept identifier). The user can consult a multilingual thesaurus / KOS that includes for each concept corresponding terms from several languages and has an index for each language. Or the mapping from a user term in any covered language to the corresponding descriptor could be automatic. As an example, consider a library catalog using the Library of Congress Subject Headings, for which French and Spanish translations are available. In the VTLS automated library system, each subject heading is identified by a number that is used in the document records. The authority file includes for each subject heading the preferred term and any synonyms; this information can be included in multiple languages. From any user term in English, French, or Spanish the system can map to the corresponding subject heading number through a free-text search on authority records to find any subject heading for which either the preferred term or any synonym contains the user's query word or phrase in any language.

Whenever the mapping from user terms to descriptors is done "behind the scenes", transparent to the user, the system should ask the user for clarification whenever the query word or phrase has multiple meanings and cannot be disambiguated automatically. Beyond that, showing the user the descriptor(s) the system came up with in their hierarchical context might improve the accuracy of the query formulation and thus retrieval. The success of this type of interaction depends on the quality of the hierarchy and the interface.

If voice input is available, one might even include the spoken form of terms in the thesaurus / KOS to enable voice input of query terms which would then be mapped to the appropriate descriptors.

A controlled vocabulary system must promote correct use of descriptors in indexing. Hierarchy and scope notes assist the indexer in understanding the meaning of a descriptor. Request-oriented indexing assures that important descriptors are not overlooked.. In cross-language retrieval the thesaurus / KOS version in each language must make sure that the indexer in that
language fully understands the meaning of a descriptor that originated from another language; otherwise, the indexing of such a descriptor will not be consistent across the database.

Automated indexing with a controlled vocabulary, particularly if it is to take a request-oriented slant, can be accomplished with a knowledge base that (1) allows recognition of important words and phrases (in spoken form for speech indexing) and allows for homonym disambiguation and (2) gives mapping rules that lead from the (possibly weighted) set of words and phrases identified for a document to a set of descriptors that should be assigned.

Such mapping rules can take many forms. In their simplest form, they specify a direct mapping from text words or phrases to the appropriate descriptors for each word or phrase (and possibly even word or phrase combinations). To increase accuracy, the mapping can be made dependent on context (Hlava 97). A more complex mapping relies on association strengths between terms (words and phrases) and descriptors. Broadly speaking, the association strength between term T and descriptor D could be seen as the predictive probability that the document containing term T should be indexed with descriptor D. Such association strengths can be computed from a training set of indexed documents. This is the approach often taken in automated text categorization, where often, but not always, the goal is to index each document by only one descriptor (assign it to one of a set of non-overlapping categories). An advanced version of this approach is the use of “topic signatures”, profiles consisting of a set of terms with weights; a document is assigned the topic if its terms match the topic signature (Lin 1997). In effect, a topic signature is a query which identifies documents relevant to the topic.

As the foregoing discussion illustrates, the knowledge base needed to support automated indexing is more complex than a Thesaurus / KOS for manual indexing. It must include more terms and term variants so that the words and phrases important for indexing can be recognized in the text, and it must include information needed for the disambiguation of homonyms (which often requires determining the part of speech of a text word).

For indexing and searching, a controlled-vocabulary cross-language retrieval system can be seen as a set of monolingual systems, each of which maps the terms from its language to a common system of concepts used in indexing and searching. For manual indexing and query formulation, this is accomplished through a multilingual thesaurus / KOS, which may in fact consist of multiple monolingual thesauri linked through common descriptor identifiers (such as Dewey Decimal class numbers). Automated indexing in cross-language text retrieval with texts in multiple languages means mapping from each language to the common conceptual structure represented in the controlled vocabulary. The knowledge base component dealing with identification of words and phrases for automated indexing can be developed independently for each language. Mapping rules that are entirely term-based can also be developed independently for each language. However, some mapping rules, for example rules based on context or topic profiles, may include conceptual elements that could be shared across languages.

There are a number of controlled-vocabulary cross-language retrieval systems based on manual indexing in use in bilingual or multilingual areas such as Switzerland, Belgium, Canada, and areas of the US with large Spanish-speaking populations; in international organizations, such as the European Community; and in international collaborative systems, such as AGRIS. These systems are based on the Universal Decimal Classification, which has been translated into many languages (library of the ETH, Zurich); on the Library of Congress Subject Headings (translated into French); on EUROVOC, an EC thesaurus in 9 languages; and AGROVOC, a thesaurus in
six plus languages created by translation from its original English-only version. There are a large number of thesauri that either have been developed as multilingual thesauri or have been translated into several languages.

**Free-text searching**

High-recall (even moderate-recall) free-text searching requires query-term expansion as discussed above. Cross-language free-text searching, finding texts in one language that are relevant for a query formulated in another language without relying on controlled vocabulary indexing, is an extension of this principle: Each query term must be mapped to a set of search terms in the language of the texts, possibly attaching weights expressing the degree to which occurrence of a search term in a text would contribute to the relevance of the text to the query term. To assist with this task, a thesaurus / KOS must include the mapping information. If the thesaurus / KOS includes fine-grained definitions that deal with subtle differences of meaning, distance between such definitions can be used to derive term weights.

A major difficulty of this mapping is that a homonym used in the query gives rise to multiple translations, each corresponding to one of its meanings. The target terms may in turn be homonyms in their language and thus retrieve many irrelevant documents unless text terms are disambiguated. (This problem exists in synonym expansion in one language as well but is exacerbated in cross-language text retrieval.) When the mapping goes to a term that has multiple meanings, the specific meaning should be identified, possibly in interaction with the user. For best retrieval results the terms in the texts should also be disambiguated so that only documents that include the term in the right sense score.

The issue of homonymy in retrieval is not as straightforward as it may seem at first glance (Sanderson 1994). First of all, quite a bit of disambiguation may occur “naturally”, in that a given term may assume only one of its meanings in the specific domain of the collection and therefore in the queries. Second, in a multi-component query, a document that includes a homonymous term from the first query component in a meaning other than that intended in the query is unlikely to also include a term from another query component; thus excluding irrelevant documents may not require disambiguation in either the query or the texts. On the other hand, with single-concept query to a general collection (such as the World Wide Web), disambiguation can be expected to have a beneficial effect on retrieval performance. Failing that, a system might be able to suggest to the user an additional query component that would separate out the documents that include the query term but in a different meaning. Note that information extraction is much more dependent on homonym disambiguation.

In any event, for best support of free-text retrieval a thesaurus / KOS should flag homonyms, give their senses, and include rules for disambiguation.

The greater difficulty of free-text cross-language retrieval stems in no small measure from the fact that one must work with actual usage, while in controlled-vocabulary retrieval one can, to some extent, dictate usage.
Thesauri for knowledge-based search support

Whether searching is by controlled vocabulary or by free text, it is often helpful to the user to browse a well-structured and well-displayed hierarchy of concepts, preferably with the option of including definitions. A more sophisticated system may guide a user through a facet analysis of her topic. These aids provided by the system enable the user to form a better idea of her need and to locate the most suitable descriptors or free-text search terms. The guidance through facets and their hierarchical display must be available in the language of the user. These suggestions are based on the assumption that browsing a hierarchy is natural to most users and that users will appreciate the structure provided. This assumption rests on the belief that people try to make sense of the world and that guided facet analysis and browsing well-structured hierarchies help them do so. There is anecdotal evidence to support this assumption, but it needs to be investigated by building prototype systems and studying users' success (see, for example, Pollitt 1996).

This is one example of using a thesaurus / KOS as a knowledge base to make searching more successful. The assistance provided does not require that the user be an expert in thesauri / KOS. This is even more true for "behind-the-scenes" assistance. There is no need to teach users about following a cross-reference from a synonym to a descriptor if the system searches for the descriptor automatically. There is no need to tell the user to look under narrower terms also if the system can do a hierarchically expanded search. There is no need to tell the user about strategies of broadening the search if the system, in response to a user input that not enough was found, can suggest further descriptors to be searched based on cross-references in the thesaurus / KOS. Sophisticated retrieval software can make the use of thesauri / KOS in retrieval independent of the user’s knowledge and thereby can get much more mileage out of the investment in thesauri / KOS.

Special issues in multilingual thesauri / KOS

A multilingual thesaurus / KOS for indexing and searching with a controlled vocabulary can be seen as a set of monolingual thesauri / KOS that all map to a common system of concepts. With a controlled vocabulary, indexing is concept-based; cross-language retrieval is simply a matter of providing designations for these concepts in multiple languages so that queries can be written in multiple languages. However, as the example given above illustrates, conceptual systems represented in the vocabulary of different languages do not completely coincide.
The crux of the matter, then, is which concepts to include. Ideally, the thesaurus / KOS should include all concepts needed in searching by any user in any of the source languages. Language differences often also imply cultural and conceptual differences, more so in some fields than in others. We need to create a classification that includes all concepts suggested by any of the languages. At a minimum this includes all relevant concepts lexicalized in at least one of the source languages. Also, different languages often suggest different ways of classifying a domain; the system needs to be hospitable to all of these. The problem that has bedeviled many developers of multilingual thesauri / KOS is that a concept lexicalized in one language may not be lexicalized in another and that the terms that do exist often vary slightly in meaning, possibly giving rise to different relationships. Starting from the misguided notion that a thesaurus / KOS should include only concepts for which there is a term in the language and that term meanings cannot be adjusted for purposes of the thesaurus / KOS, they had difficulty making the system of concepts the same for all languages. But, as we have seen, even in a monolingual thesaurus / KOS the lexicographer often discovers concepts needed in searching or to enhance the logic of the concept hierarchy and then needs to create terms for these concepts. In multilingual thesauri / KOS this necessity arises more often, particularly when different languages differ in the hierarchical levels at which they lexicalize concepts.

The principle proposed here is to establish a common conceptual system, which may require an arduous, and expensive, process of negotiation, and then arrange for the terms in all languages to fit, giving proper definitions, of course. In contrast, many multilingual thesauri / KOS have been produced by translating an established monolingual thesaurus / KOS, thus accepting the conceptual system of one language and limiting the number of synonyms (if any) in the other languages. EuroWordNet is based on a more comprehensive, but still limited approach: Rather than developing a conceptual structure based on an analysis of the vocabulary in all participating languages, EuroWordNet accepts the conceptual system of the English language WordNet. On the other hand, EuroWordNet does not simply translate WordNet but develops synonym sets independently in each language and then links them to the concepts (synsets) established in WordNet.

So far we have described a multilingual thesaurus / KOS with a common conceptual system, however constructed, where the terms of each language are linked to a concept but not to each other. Relationships between terms from different languages are established through their relationships to concepts. This simple model will do for most information retrieval. But unless the concepts are exceedingly fine-grained and include in their definition affective components of meaning and usage considerations, this model is too simplistic for natural language processing, especially translation. There one needs direct relationships between terms to enable the proper word choice in translation.

The problems discussed here and illustrated in the example above have major implications for cross-language free-text searching: Each query term should be mapped from the source language to its multiple equivalents in the target language; each of these equivalents may have other meanings in the target language, presenting potential problems for retrieval. The query term may not have a precise equivalent in the target language; one may need to map to broader or narrower terms, distorting the meaning of the original query.
Key issues in multilingual thesauri / KOS

Conceptual systems in different languages differ

What concepts are lexicalized differs from language to language

Translation of an English thesaurus / KOS into French does not make a French thesaurus / KOS

Develop common conceptual structure integrating perspectives from multiple languages. Harmonize concepts where possible, keep concepts where necessary, invent a term if a concept is not lexicalized in a language

Problems of structure: simplified versus real

<table>
<thead>
<tr>
<th>Simplified</th>
<th>Real</th>
</tr>
</thead>
<tbody>
<tr>
<td>English term 1</td>
<td>French term 1</td>
</tr>
<tr>
<td>English term 2 Concept</td>
<td>Concept French term 2</td>
</tr>
<tr>
<td>English term 3</td>
<td>French term 3</td>
</tr>
<tr>
<td>English term 1</td>
<td>French term 1</td>
</tr>
<tr>
<td>English term 2 Concept</td>
<td>Concept French term 2</td>
</tr>
<tr>
<td>English term 3</td>
<td>French term 3</td>
</tr>
</tbody>
</table>
Evaluation of Thesauri

Introductory example: Yahoo classification
Yahoo home page
<table>
<thead>
<tr>
<th>Reference and General Interest</th>
<th>Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reference</strong>&lt;br&gt;Libraries, Dictionaries, Quotations ...</td>
<td><strong>Science</strong>&lt;br&gt;Animals, Astronomy, Engineering ...</td>
</tr>
<tr>
<td><strong>Computers &amp; Internet</strong>&lt;br&gt;Internet, WWW, Software, Games ...</td>
<td><strong>Health</strong>&lt;br&gt;Medicine, Diseases, Drugs, Fitness ...</td>
</tr>
<tr>
<td><strong>News &amp; Media</strong>&lt;br&gt;Full Coverage, Newspapers, TV...</td>
<td><strong>Social Science</strong>&lt;br&gt;Archaeology, Economics, Languages ...</td>
</tr>
<tr>
<td><strong>Entertainment</strong>&lt;br&gt;Movies, Music, Humor, Cool Links ...</td>
<td><strong>Society &amp; Culture</strong>&lt;br&gt;People, Environment, Religion ...</td>
</tr>
<tr>
<td><strong>Recreation &amp; Sports</strong>&lt;br&gt;Sports, Travel, Autos, Outdoors...</td>
<td><strong>Government</strong>&lt;br&gt;Elections, Military, Law, Taxes ...</td>
</tr>
<tr>
<td><strong>Education</strong>&lt;br&gt;College and University, K-12 ...</td>
<td><strong>Business &amp; Economy</strong>&lt;br&gt;B2B, Finance, Shopping, Jobs ...</td>
</tr>
<tr>
<td><strong>Arts &amp; Humanities</strong>&lt;br&gt;Literature, Photography ...</td>
<td><strong>Regional</strong>&lt;br&gt;Countries, Regions, US States ...</td>
</tr>
</tbody>
</table>

Yahoo Classification. Home. Meaningful arrangement
Yahoo education
Yahoo classification. Education. Meaningful arrangement.

Home > Education

Categories

Browse by Region (170)
By Culture or Group (398)
By Subject (11)

<table>
<thead>
<tr>
<th>Information sources</th>
<th>Educational methods</th>
<th>Political and economic aspects</th>
<th>Organizational aspects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliographies (4)</td>
<td>Theory and Methods (659)</td>
<td>Policy (52)</td>
<td>Government Agencies (77)</td>
</tr>
<tr>
<td>Web Directories (47)</td>
<td>Teaching (63)</td>
<td>Reform (70)</td>
<td>Organizations (3008)</td>
</tr>
<tr>
<td>News and Media (83)</td>
<td>Instructional Technology (334)</td>
<td>Equity (27)</td>
<td>Companies@</td>
</tr>
<tr>
<td>Chats and Forums (40)</td>
<td>Distance Learning (476)</td>
<td>Financial Aid (395)</td>
<td>Programs (322)</td>
</tr>
<tr>
<td>Conferences (52)</td>
<td>Standards and Testing (63)</td>
<td>Employment (143)</td>
<td></td>
</tr>
<tr>
<td>Journals (36)</td>
<td>Academic Competitions (79)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistics (6)</td>
<td>Graduation (53)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Education by level
Early Childhood Education (90)
K-12 (53910)
Higher Education (16638)
Adult and Continuing Education (325)

Special students and subjects
Special Education (168)
Disabilities@
Literacy (12)
Bilingual (24)
Career and Vocational (236)
Correctional@
Several pages from Yahoo
Evaluation of Knowledge Organization Systems (KOS)

Characteristics for describing and evaluating KOS

(classifications/ontologies/taxonomies/index languages/thesauri/glossaries/dictionaries)

(For some items, a section number from Soergel, Organizing information (starting with a digit), and/or Soergel, Indexing languages and thesauri (starting with a capital) is given.)

1. **Overall description and evaluation**

1.1 **Purpose**, for example (see full list given earlier)

Providing "conceptual infrastructure"

Mapping out the conceptual structure and providing a common language for a field

Providing classification/typology and concept definitions. Clarifying concepts by putting them into context. Thus providing orientation and serving as a reference tool for individual researchers and practitioners and thereby

Assisting with the exploration of the conceptual context of a research problem and in structuring the problem, thereby providing the conceptual basis for the design of good research, for the consistent definition of variables, and thus the cumulation of research results.

Providing the conceptual basis for the exploration of the various aspects of a program in program planning, in the identification of approaches and strategies, and in the development of evaluation criteria

Information retrieval (IR), database searching, expert systems, and other AI applications

One information system

Several IR systems, switching language. Support the coordination or combination of several databases in the same area to facilitate access to multiple databases.

Assisting readers in understanding text

Assisting writers with conceptualizing a topic and with finding the proper term

Translation

Language learning

In each case specify the intended audience
If purpose is IR specify

Information system(s) in which the vocabulary is to be used
Type of objects to be retrieved (text documents, images, people, etc.)

Use of the vocabulary

- Vocabulary control in indexing and searching (controlled vocabulary)
- Vocabulary control only for searching. Assist with clarifying a search topic and assembling all applicable concepts and terms, whether searching with a controlled vocabulary of free-text.

IR technique(s) (such as: printed index, computer search system). Support of inclusive (hierarchically expanded) searching

Automated vs. manual indexing or query formulation. Approach to indexing to be supported: Request-oriented vs. entity-oriented

Techniques for eliciting user needs (e.g., menu based on search tree; questions based on facet structure)

1.2 Relationship to other KOS, especially standard schemes

1.3 Summary evaluation of the vocabulary's adequacy for the stated purpose based on the more detailed analysis as outlined below.

Note:

In the following, links to two textbooks are given in [] as follows:
chapter or section numbers starting with a digit, such as 14.3, refer to Soergel, Organizing Information
chapter or section numbers starting with a letter, such as F0.4.3, refer to Soergel, Indexing Languages and Thesauri. Construction and Maintenance
2. **Coverage of concepts and terms. Sources, quality of usage analysis.**

2.1 Concepts: scope, breadth of coverage (See also 2.3.1 below)

2.2 Concepts: specificity, depth of coverage

Completeness of coverage at each level of specificity considering all concepts (descriptors and other preferred terms) and descriptors alone [F0.4.3]

Specificity must be adapted to the purpose. Assistance in the choice of terms or the comprehension of text requires many nuances. An IR system for propositions requires high specificity. A bibliographic IR systems may require only low specificity.

2.3 Sources from which concepts and terms are included (natural languages, classifications/thesauri /KOS).

Relationship to other vocabularies, especially standard schemes.

For each source:

2.3.1 Completeness of coverage; all vs. selected concepts; all vs. selected terms for each concept (this includes coverage of synonyms)

2.3.2 Quality of analysis of actual term usage in the source.

2.3.3 Recency

Specifically: Completeness of coverage of the terminology from a given language (English, French, German, Chinese, etc.; the language is the source)

2.4 Augmentation of sources through concepts created in concept analysis [15, C3])

2.4.1 Are all necessary facets included?

2.4.2 Formation of new concepts arising from semantic factoring and other methods of concept analysis. Specifically: Are the concepts applicable across disciplines? Are the concepts applicable across different societies and cultures? (See also 3.4 below)

2.5 Choice of terms

2.5.1 Form of terms - consistency, adherence to common usage.

2.5.2 Appropriateness of selection of preferred terms from among synonyms.

2.5.3 Choice of terms to designate descriptors [F0.4.2] Closeness to user terminology.

2.6 Nature of notation (if none, state that) [D4]
3. Conceptual analysis and conceptual structure. Terminological analysis

3.1 Quality of conceptual structure [14, C1]

Types and degree of differentiation of conceptual relationships included:

3.1.1 Expression of concepts through elemental concepts (closely related to definition)

3.1.2 Hierarchical relationships (polyhierarchy)

3.1.3 Associative relationships

3.1.4 Rich set of semantic relationships (full-fledged ontology)

Completeness and correctness of conceptual relationships included.

3.2 Quality of definitions, explications, scope notes (correctness, detail, clarity). [C3]

3.3 Completeness of terminological relationships.

Does the thesaurus / KOS contain terms that are synonymous or quasi-synonymous without indicating the relationship?

4. Use of precombination in the index language (cuts across 2 and 3) [14, 15, C2]

4.1 To what degree are descriptors precombined?

4.2 To what extent are precombined descriptors enumerated and/or given in the alphabetical index? Built by the indexer? Updating characteristics.

Are precombined descriptors designated by an independent symbol or a string of symbols? Citation order free or fixed? To what extent do the components of a precombined descriptor determine its place in the arrangement? (Relates also to 5)
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 computerized search system.

5.1 **Format of printed document**

5.1.1 **Overall format [D1]**

Thesaurus / KOS 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**
- through arrangement [15.5.2, C2, D3]
- through cross-references [D3.1.1,1]
- through descriptor-find index [15.5.1, D3.6]

How well does the display reflect the conceptual analysis (e.g., sequence of concepts on some hierarchical level) [D3.1.2]

5.1.3 **Display of terminological relationships. Format of alphabetical index [C5]**

5.2 **Access through computer systems. Retrieval of concepts and terms. Navigation. Format of on-line displays**

5.2.1 **Overall format. Available windows and their relationships**

5.2.2 **Display of conceptual relationships, esp. hierarchy. Localized hierarchical chains vs. entire hierarchy. Overviews and total hierarchy. Expandable levels vs. expanded or expand-all option. Graphical displays, concept maps. Are cross-references active hyperlinks? Is there an online descriptor-find index.**

5.2.3 **Access by words and phrases. Is the thesaurus / KOS database searchable? How does the search work. What is searched? Just the term itself, synonyms, scope notes, all cross-references (not good!).**

5.3 **Format of machine-readable form (if any). What standard is followed**

5.4 **Detail of keeping records of the origin of information included in the vocabulary.**

6. **Updating**
Outline for the analysis of KOS. DDC

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 □ [Ch. 12, Introduction]
   1.3 **Type of file and search mechanism** for which originally designed
      Shelving □ Card catalog □ Online system □ (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. (Section16.2.2). 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 **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!
   2.5 **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.
   2.6 **Nature of notation** (if none, state that). [Section 15.5.2] Decimal, highly expressive (with some exceptions).

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 im-
      plied by the relative index (Example: The classes shown under Blind).
   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 / KOS 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 [Section15.5.2]  
  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.
Some points on procedure for evaluating a KOS

What went into the development of a KOS

Check sources used.

Check procedures used in KOS development.

Examine the KOS structure and content

Use knowledge of KOS structure for analysis of structure and internal consistency.

Check against other KOS and against encyclopedias, dictionaries, or other authoritative sources.

In this examination, collect data on all the criteria in parallel by looking through the KOS, probing for examples and following leads as they arise. Interact with the KOS. Keep notes according to the outline of criteria. (Much like anthropological field work, where the investigator observes as events occur, keeping the variables of interest in mind but is always open to aspects not though of beforehand. At some point, the notes are indexed and sorted by the variables of interest.)

Check how well the KOS supports the functions it was designed for

If the KOS purpose is to support finding items (documents, software, products, people, jobs, etc., either indexed by the KOS or searched free-text), try the KOS to see how well search requests can be expressed and how well the aspects under which items should be found can be represented. Conduct indexing and retrieval experiments. Observe users interacting with the KOS online; see how it helps them in formulating a successful query. Study the effects of behind-the-scenes query expansion through the KOS. Such tests can be done by the evaluator - for example, to shed light on completeness - or by real indexers and real searchers, making sure that the indexers have sufficient domain knowledge and training in the use of the KOS. Indexing experiments where several indexers index the same documents can be very useful; disagreements may point out problems in the KOS.

If the purpose of the KOS is to support reasoning, run sample reasoning tasks and check the results. This may require tracing a reasoning process, paying particular attention to the steps where the KOS was used. If the purpose of the KOS is business process reengineering, talk to the people who do the task to see how the KOS helps them. If the purpose of the KOS is to collect and use statistics, check how well the KOS supports consistent data collection (depends on good definition of the concepts) and how well it supports finding data and processing them to derive indices or decision support.
Resources

Directories of thesauri / KOS and dictionaries on the Web.

Printouts from the following Web sites were included in the paper tutorial notebook:

http://nkos.slis.kent.edu/
Networked Knowledge Organization Systems (NKOS).
Has a workshop at ECDL and ECDL

http://www.lub.lu.se/desire/desireIIindex.html#idx
DESIRE I and II. The role of classification schemes in Internet resource description and discovery

www.darmstadt.gmd.de/~lutes/thesoecl.html
Web Thesaurus Compendium (representative list with descriptions)

www.onelook.com OneLook Dictionaries. The Faster Finder

www.yourdictionary.com/

www.strategic-road.com/pratique/dicofr.htm Strategic Road Dictionaries

http://linguistlist.org/sp/Dict.html

www.mikesart.net/giantglossarycom Terminology - Search

http://www.asel.udel.edu/nli/nlp/lrd.html The Language Representation Database Project

XXX

Enterprise search report:

A short bibliography follows
Thesaurus / KOS software Web sites

http://www.willpower.demon.co.uk/thessoft.htm
http://sky.fit.qut.edu.au/~middletm/cont_voc.html
http://www.fbi.fh-koeln.de/fachbereich/labor/Bir/thesauri_new/indexen.htm
http://bak-information.ub.tu-berlin.de/software/term.html (covers a wider range of software, annotations in German)

Concept mapping resources

http://eaa-knowledge.com/ojni/ni/602/strategies.htm
http://www.graphic.org/concept.html
http://www.observeotry.com/conceptmappingvs1.htm

Ontology resources. Ontology editor examples

Protege (client-based and Web-based versions)
http://protege.stanford.edu/, click on Documentation

Biomedical ontologies: www.obofoundry.org/
http://ontology.buffalo.edu/smith/

On description logic see, for example http://potato.cs.man.ac.uk/seanb/publications.php

The URLs given on the standards page are also useful more generally, esp. on RDF and OWL

Topic maps

www.ontopia.net.
For an example application, see www.ontopia.net/omnigator/models/index.jsp (try opera.xtm)

Search terms for a Web search for thesauri / KOS

(ontolog* OR classification* OR Klassifikation* OR taxonom* OR thesaur* OR dictionar* OR dictionnaire OR Woerterbuch OR glossar* OR glossaire OR “word list” OR lexicon OR lexique OR Lexik* OR terminolog* OR vocabulaire OR vocabulary OR “knowledge organization” OR “knowledge structure” OR “authority list”)

Note: "classification" may pull a lot of irrelevant URLs

Possibly add OR concept OR mot-clé OR keyword OR “subject heading” OR definition

It is best to require these terms in the title. Otherwise there will be a lot of irrelevant material retrieved, especially by the term classification.
Short bibliography

Website:  http://www.clis.umd.edu/faculty/soergel/dlthestut

Basic information retrieval and classification concepts

Soergel, Dagobert, 1985
Organizing Information.  Principles of data base and retrieval systems.

Vickery, Bryan C.
Faceted classification.

Thesaurus textbooks

Soergel, Dagobert
Construction and maintenance of indexing languages and thesauri

Lancaster, F. Wilfrid
Vocabulary control for information retrieval.  1.ed.
(2. ed. not as good)

Aitchison, Jean; Gilchrist, Alan; Bawden, David
Thesaurus Construction and Use : A Practical Manual.  4. ed.

Also watch for the Proceedings of the ASIS SIG/CR Classification Research Workshop, published as Advances in Classification Research by Information Today

Standards (use with caution) (see the section on Standards in Part 2 of the tutorial)

National Information Standards Organization
Guidelines for the construction, format, and management of monolingual controlled vocabularies.
http://www.slis.kent.edu/~mzeng/Z3919/9Z3919.htm

International Organization for Standardization.
Documentation—guidelines for the establishment and development of monolingual thesauri.  2. ed.
International Organization for Standardization.  
**Documentation—guidelines for the establishment and development of multilingual thesauri.**  
www.collectionscanada.gc.ca/iso/tc46sc9/standard/5964e.htm

**Machine-Readable Dictionaries and Computational Linguistics Research**

Walker, Don, ed. 1995; Zampolli, A., ed.; Calzolari, N., ed..  
**Automating the Lexicon: Research and Practice in a Multilingual Environment.**  

**Survey of the State of the Art in Human Language Technology.**  
With Chapter 12 Language Resources and Section 12.4 Lexicons.  
http://www.cse.ogi.edu/CSLU/HLTsurvey/HLTsurvey.html

**Preparation of multilingual vocabularies.**  
**Standardizing and Harmonizing Terminology: Theory and Practice.**  

**Other relevant publications by the tutorial instructor**

A universal source Thesaurus as a classification generator.  

Indexing and retrieval performance: The logical evidence.  


Software support for Thesaurus construction and display.  
Proceedings of the 5th ASIS SIG/CR Classification Research Workshop.  
Held at the 57th ASIS Annual Meeting, Oct. 16-20, 1994, Alexandria, VA.  


Dagobert Soergel; Boris Lauser; Anita Liang; Frehiwot Fisseha; Johannes Keizer; Stephen Katz. Reengineering thesauri for new applications. The AGROVOC example Journal of Digital Information, Volume 4 Issue 4, 2004 March, Article No. 257, jodi.tamu.edu/Articles/v04/i04/Soergel/


Bibliographies of thesauri

Basis for the Echo database of thesauri

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,
Includes bibliography of editions in multiple languages of
Universal Decimal Classification (UDC) Library of Congress Classification (LCC)
Dewey Decimal Classification (DDC) Library of Congress Subject Headings (LCSH)


Examples of multilingual thesauri


Food and Agriculture Organization of the United States. AGROVOC multilingual agricultural Thesaurus. Second edition, English version; 798 pages; APIMONDIA, Rome; 1992. (Not latest)


Knowledge Organization Systems in Digital Libraries

Tutorial

Part 2

Design, evaluation, and development
### Part 2. Design and development. Outline

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:30 - 14:35</td>
<td><strong>Introduction and overview</strong></td>
<td>160</td>
</tr>
<tr>
<td></td>
<td>KOS development principles</td>
<td>162</td>
</tr>
<tr>
<td>14:35 - 15:05</td>
<td><strong>The process of KOS development</strong></td>
<td>164</td>
</tr>
<tr>
<td>13:35 - 13:40</td>
<td>The overall process of KOS development</td>
<td>165</td>
</tr>
<tr>
<td>13:40 - 13:55</td>
<td>Sources of concepts, terms, relationships, definitions</td>
<td>166</td>
</tr>
<tr>
<td></td>
<td>Methods of data collection</td>
<td></td>
</tr>
<tr>
<td>13:55 - 14:05</td>
<td>Merging data from many sources</td>
<td>171</td>
</tr>
<tr>
<td>15:05 - 16:00</td>
<td><strong>Developing the conceptual structure</strong></td>
<td>173</td>
</tr>
<tr>
<td>14:05 - 14:30</td>
<td>Facet analysis I: Education (starting with classes from DDC)</td>
<td>174</td>
</tr>
<tr>
<td>14:30 - 14:40</td>
<td>More facet examples: Job titles</td>
<td>176</td>
</tr>
<tr>
<td>14:40 - 14:50</td>
<td>Guidelines for meaningful arrangement</td>
<td>179</td>
</tr>
<tr>
<td>14:50 - 15:00</td>
<td>Rules for selection of preferred terms</td>
<td>187</td>
</tr>
<tr>
<td></td>
<td>Rules for selection of concepts as descriptors</td>
<td></td>
</tr>
<tr>
<td>16:00 - 16:30</td>
<td><strong>Break</strong></td>
<td></td>
</tr>
<tr>
<td>16:30 - 17:15</td>
<td><strong>Developing the conceptual structure, continued</strong></td>
<td>177</td>
</tr>
<tr>
<td>16:30 - 17:15</td>
<td>Facet exercise (in pairs)</td>
<td></td>
</tr>
<tr>
<td>17:15 - 18:15</td>
<td><strong>The structure and processing of KOS data</strong></td>
<td>189</td>
</tr>
<tr>
<td>17:15 - 17:30</td>
<td>Interoperability of thesauri/ontologies. Crosswalks</td>
<td>189</td>
</tr>
<tr>
<td>17:30 - 17:45</td>
<td>The structure of a KOS/ontology database</td>
<td>193</td>
</tr>
<tr>
<td>17:45 - 17:55</td>
<td>The many forms of Knowledge Organization Systems (KOS) and their standards</td>
<td>203</td>
</tr>
<tr>
<td></td>
<td>SKOS (Simplified Knowledge Organization Systems) example</td>
<td>217</td>
</tr>
<tr>
<td></td>
<td>OWL (Web Ontology Language) example</td>
<td>223</td>
</tr>
<tr>
<td>17:55 - 18:15</td>
<td>KOS software and its evaluation</td>
<td>209</td>
</tr>
<tr>
<td></td>
<td>Appendix. The KOS development process</td>
<td>257</td>
</tr>
</tbody>
</table>
Introduction
Developing a good Knowledge Organization System (KOS) – ontology, taxonomy, classification, thesaurus – for a given domain means to
- evolve the conceptual structure of the domain;
- discern and crystallize its key concepts and their relationships; and then
- represent this structure (1) externally in a form that readers can easily grasp and make their own and (2) internally in a form that computer programs can use.
This is an intellectually challenging task that requires considerable resources.

Introduction and overview

Scope:
"Thesaurus / KOS" or just KOS is used for Knowledge Organization Systems (KOS)
Includes Thesauri, classifications, ontologies, taxonomies, concept maps, dictionaries, etc.

Main objective:
Participants should be able to crystalize the conceptual structure of a domain

Outline

KOS development principles

The process of KOS development
Developing the conceptual structure

The structure and processing of KOS data
KOS development principles

(1) Determine the right organizational scope or user group(s) to be served

(2) Determine the right function scope

(3) Ensure adoption by supporting many views through an inclusive, flexible KOS database

(4) Build on existing KOS. Reuse, reuse, reuse

(5) Use automation for efficiency

(6) Capitalize on collaborative creation and editing
KOS development principles

Good cost-effective KOS development stands and falls with the observation of six simple principles.

(1) **Determine the right organizational scope or user group(s)** to be served. A broad organizational scope results in higher return on investment but must be balanced against tailoring the KOS to specific requirements (but see principle (3)). The scope can range from a KOS for a single user who needs to organize a collection of documents, Web sites, files, email messages, notes, contacts, appointments, and tasks to a KOS that is used worldwide for organizing many databases in many languages, such as AGROVOC or the Medical Subject Headings. In between would be a KOS for an organization, an online community, or a specific database such as AERS, FDA’s Adverse Events Reporting System database (fda.gov/cder/aers/default.htm). A broad scope of application (many departments in an organization, several online communities with related purposes, databases in the same general subject domain) improves the return on investment into the KOS and fosters semantic interoperability.

(2) **Determine the right function scope.** The more functions are served by a KOS the higher the return on investment. Therefore it is important to identify all possible applications that could profit from KOS support. This includes functions such as retrieval from any kind of database or through a Web search engine and structured output of results, project planning, insurance billing, natural language processing, reasoning in expert systems; see XXX for an extensive list. Understanding all these functions is crucial if one wants to maximize ROI. For each application state

- requirements and the spectrum of users’ level of knowledge.
- required characteristics of the KOS

(3) **Ensure adoption by supporting many views through an inclusive, flexible KOS database.** Different applications need different views for intrinsic reasons that are rooted in the nature of the application (for example, to optimize reasoning over large knowledge bases) or for historical reasons (a given user group may not want to change what they are used to, or a large body of material can not feasibly be re-indexed). The solution is a flexible comprehensive KOS database in which different views can coexist or from which different views can be easily extracted; that way each application’s or group’s view of the domain can be accommodated, and that removes a key barrier to engaging several groups that might otherwise go their own ways (at great expense), even though all groups may agree on 80% and disagree on just 20% of concepts, terms, and relationships. There is no need to force everybody into the same mold; many views can coexist within the same KOS database, where they can be related to each other to the extent possible, capturing commonalities and explicating differences. Through the KOS database, different groups can learn from each other; this may lead to improvements of each group’s view. In that sense, a KOS database

- records present usage by different groups (like a dictionary records usage);
- creates an overarching well-ordered structure for all the concepts from different views;
- through that structure, creates relationships between different views;
- through that structure, recommends improvements for each group’s view.

(4) **Build on existing KOS. Reuse, reuse, reuse.** There is enormous intellectual capital in existing KOS; use it! Identify the KOS presently used, explicitly or implicitly, in the organizational and functional scope. Find other KOS that could meet some or all of the requirements, perhaps with some modification, or that would be useful sources for developing a new KOS (see Box 1. Searching for KOS).

(5) **Use automation for efficiency.** Automation can be used to massage data from existing KOS for the purpose at hand. It can also be used to extract terms, concepts, relationships, and global structure from texts and document collections. While automation cannot entirely replace intellectual effort, a KOS constructed using automatic processing alone is better than no KOS at all.

(6) **Capitalize on collaborative creation and editing** with central expert support and some measure of control. Possibly have a cadre of KOS editors throughout the organization or throughout the world.
The process of KOS development

The overall process of KOS development

Sources of concepts, terms, relationships, definitions

Methods of data collection

Merging data from many sources
The overall process of KOS development

Diagram from DS 1974 copied in here. Need orig
Sources of concepts, terms, relationships, definitions

Reuse knowledge in existing Knowledge Organization Systems. Much intellectual capital was invested in their development.

But: Adapt content and structure to user requirements and background.

Most important source: search requests and other statements of user requirements.

Types of sources

(1) **Prearranged sources** (terms are already arranged according to some principle)

(2) **Open-ended sources** (terms are not ordered or terms must be inferred or derived)

Find machine-readable sources

Internal and external sources
Sources of concepts, terms, relationships, definitions

(1) **Prearranged sources** (terms are already arranged according to some principle)

(1.1) Descriptor lists, classification schemes, thesauri (universal classification schemes, such as LCC or UDC, and special classification schemes).

(1.2) Nomenclatures of single disciplines, esp. if approved by an international body.

(1.3) Treatises on the terminology of a subject field

(1.4) Encyclopedias, lexica, dictionaries, glossaries (universal or discipline-oriented; mono-, bi-, or multilingual).

(1.5) The tables of contents and indexes of conference proceedings, textbooks, handbooks, and course syllabi.

(1.6) Indexes of journals, abstracting journals, other publications, databases.

(1.7) Term-association lists produced by subjects in term association studies.

(1.8) Output from automatic classification programs based on term co-occurrence data or citations.
Sources of concepts, terms, relationships, definitions

(2) **Open-ended sources** (terms are not ordered or terms must be inferred or derived)

(2.1) Lists of search requests and interest profiles and other statements of user requirements obtained from search logs and user studies (individual interviews, focus groups).

(2.1a) Mooers’ method: Focus group, present documents, ask “Why would this be of interest?”

(2.2) Descriptions of R&D projects and other activities to be supported.

(2.3) Free indexing of a sample of documents, each by several experts (to get synonyms).

(2.4) Titles, abstracts, full text, reviews of books, journal articles, conference papers, Web sites, internal documents, etc.

(2.5) For more information on individual terms: Web searches
Methods of data collection

For prearranged sources

If machine-readable, include all information, can always delete later

If not machine-readable and highly relevant, scan or have transcribed

Otherwise go through and select

May need to reformat for input to thesaurus / KOS software; use Perl scripts or word processor macros
Methods of data collection

For open-ended sources

Extract terms and **phrases** automatically, using a large general phrase dictionary, syntactic analysis, or a system such as [http://www.nzdl.org/Kea/](http://www.nzdl.org/Kea/)

Possibly use frequency data for further selection.

Extract term relationship automatically (often a feature of text mining programs).

Extract terms manually, being on the look-out for term relationships that can be inferred from text.
Merging data from many sources

Merge terms

Need to consolidate term variants

Use broad-scope sources to get more information on terms collected

Assemble synonym sets / concepts

Use ST relationships from many sources
Source 1: elderly ST aged person,
Source 2: aged person ST senior citizen

Merge relationships

Need to consider that often the same conceptual relationship is expressed in different terms

“Afterburn” collection from specialized sources to fill gaps
Developing the conceptual structure

Facet analysis 1: Education

More facet examples:
   Yahoo Education (from Part 1)
   Job titles

Facet exercise (in pairs)

Principles for meaningful arrangement

Rules for selection of concepts as descriptors. Rules for selection of terms
Facet analysis

**Education** (starting with classes from DDC)

Conceptual analysis and synthesis in three steps:

**Step 1.** **Semantic factor compound concepts,** make a list of elemental concepts.

**Step 2.** **Arrange elemental concepts into facets.**
Arrange each facet in a well-structured hierarchy.

**Step 3.** **If needed, fit compound concepts into the framework of the hierarchy.**
## Concept list for conceptual analysis and synthesis
(from Dewey Decimal Classification)

**Note:** A broader class is given in ( ), if necessary to specify the meaning of a term.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>372.19</td>
<td>Curriculums of elementary schools</td>
</tr>
<tr>
<td>372.35043</td>
<td>Science in the elementary school curriculum</td>
</tr>
<tr>
<td>372.414</td>
<td>Methods of instruction for reading in elementary schools</td>
</tr>
<tr>
<td>372.72043</td>
<td>Arithmetic in the elementary school curriculum</td>
</tr>
<tr>
<td>373.19</td>
<td>Curriculums in secondary schools</td>
</tr>
<tr>
<td>373.243</td>
<td>Military schools (Secondary Education)</td>
</tr>
<tr>
<td>376</td>
<td>Education of women</td>
</tr>
<tr>
<td>376.63</td>
<td>Secondary education of women</td>
</tr>
<tr>
<td>378.19</td>
<td>Curriculum of colleges and universities</td>
</tr>
<tr>
<td>378.33</td>
<td>Fellowships (Higher Education)</td>
</tr>
<tr>
<td>371.911</td>
<td>Blind and partially sighted students</td>
</tr>
<tr>
<td>371.912</td>
<td>Deaf and hard-of-hearing students</td>
</tr>
<tr>
<td>371.95</td>
<td>Curriculums for gifted students</td>
</tr>
</tbody>
</table>
More facet examples

Job titles. Can you spot the facets?

Lawyer
Paralegal
Law office receptionist
Librarian
Library assistant
Library clerk
Physician
Physician’s assistant
Doctor’s office clerk
Ophthalmologist (eye doctor)
Ophthalmologic technician
Surgeon
Facet exercise (in pairs?)

Yahoo Health

Arrange the terms in front of you into meaningful groupings.

Use the blank strips to write a heading for each group.

Time: 30 minutes (leaving 10 minutes for discussion)

Note: Back page has a list of the terms
<table>
<thead>
<tr>
<th>Category</th>
<th>Yahoo Health</th>
<th>Men's Health</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative Medicine</td>
<td>Strips to cut for term sorting exercise</td>
<td>Mental Health</td>
</tr>
<tr>
<td>Business to Business</td>
<td></td>
<td>Midwifery</td>
</tr>
<tr>
<td>Chats and Forums</td>
<td></td>
<td>News and Media</td>
</tr>
<tr>
<td>Children's Health</td>
<td></td>
<td>Nursing</td>
</tr>
<tr>
<td>Conferences</td>
<td></td>
<td>Nutrition</td>
</tr>
<tr>
<td>Death and Dying</td>
<td></td>
<td>Organizations</td>
</tr>
<tr>
<td>Dentistry</td>
<td></td>
<td>Pet Health</td>
</tr>
<tr>
<td>Disabilities</td>
<td></td>
<td>Pharmacy</td>
</tr>
<tr>
<td>Diseases and Conditions</td>
<td></td>
<td>Procedures and Therapies</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td>Public Health and Safety</td>
</tr>
<tr>
<td>Emergency Services</td>
<td></td>
<td>Reference</td>
</tr>
<tr>
<td>Employment</td>
<td></td>
<td>Reproductive Health</td>
</tr>
<tr>
<td>Environmental Health</td>
<td></td>
<td>Senior Health</td>
</tr>
<tr>
<td>First Aid</td>
<td></td>
<td>Sexuality</td>
</tr>
<tr>
<td>Fitness</td>
<td></td>
<td>Shopping and Services</td>
</tr>
<tr>
<td>General Health</td>
<td></td>
<td>Teen Health</td>
</tr>
<tr>
<td>Health Administration</td>
<td></td>
<td>Traditional Medicine</td>
</tr>
<tr>
<td>Health Care</td>
<td></td>
<td>Travel Health and Medicine</td>
</tr>
<tr>
<td>Health Sciences</td>
<td></td>
<td>Web Directories</td>
</tr>
<tr>
<td>Hospitals and Medical Centers</td>
<td></td>
<td>Weight Issues</td>
</tr>
<tr>
<td>Institutes</td>
<td></td>
<td>Women's Health</td>
</tr>
<tr>
<td>Law</td>
<td></td>
<td>Workplace</td>
</tr>
<tr>
<td>Long Term Care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medicine</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Principles for meaningful arrangement

Sequence and two-dimensional graphical arrangements (concept maps) can convey important information about concept relationships.

Collocate closely related concepts.

Often a principle of arrangement intrinsic to the subject matter suggests itself. The following examples and guidelines are intended to sharpen “informed intuition”.
# Meaningful arrangement

## Example 1

<table>
<thead>
<tr>
<th>Art and Architecture</th>
<th>Suggested meaningful sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thesaurus</td>
<td>Alphabetical vs. meaningful sequence on same hierarchical level</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>size: photograph formats</th>
</tr>
</thead>
<tbody>
<tr>
<td>double whole plate</td>
</tr>
<tr>
<td>half plate</td>
</tr>
<tr>
<td>mammoth plate</td>
</tr>
<tr>
<td>ninth plate</td>
</tr>
<tr>
<td>quarter plate</td>
</tr>
<tr>
<td>sixteenth plate</td>
</tr>
<tr>
<td>sixth plate</td>
</tr>
<tr>
<td>whole plate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>size: photograph formats</th>
</tr>
</thead>
<tbody>
<tr>
<td>sixteenth plate</td>
</tr>
<tr>
<td>ninth plate</td>
</tr>
<tr>
<td>sixth plate</td>
</tr>
<tr>
<td>quarter plate</td>
</tr>
<tr>
<td>half plate</td>
</tr>
<tr>
<td>whole plate</td>
</tr>
<tr>
<td>double whole plate</td>
</tr>
<tr>
<td>mammoth plate</td>
</tr>
</tbody>
</table>
Meaningful arrangement

Example 2. **Body systems. Fuller version**

**XF** body system or organ

**XG** musculoskeletal system

**XH** skin system

**XJ** cardiovascular system

**XK** respiratory system

**XL** mouth, larynx, vocal organ

**XM** digestive system

**XN** urogenital system

**XP** urinary system

**XQ** reproductive system

**XR** blood, immune system

**XS** blood

**XT** immune system

**XU** endocrine system

**XV** sensory system

**XW** nervous system

**XX** nervous system structures and components

**XY** peripheral nervous system

**XZ** central nervous system
Meaningful arrangement

Example 3. Art genres

Trying to find a meaningful arrangement for a list of concepts often reveals a facet structure.

See the example in the tutorial notebook.

Graphical arrangement: Concept maps

See the examples in the tutorial notebook.
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
- computer 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

b. **Suggested meaningful sequence**

d. **Original alphabetical sequence**

Figure 3. *Example from the Art and Architecture Thesaurus*
Concept map PHD
Concept map instr design
Meaningful arrangement

Guidelines

“Natural” principles

(1) Chronological – e.g., historical events.
(2) Evolutionary – arrange entities in the order they evolved, e.g., biological species, ideas.
(3) Sequence of steps – e.g., production processes, research methods, sequence of logical steps
(4) Increasing extension
(5) Geographical – spatial proximity.

More conceptual principles

(6) Increasing complexity (integrative levels)
(7a) From abstract to concrete or vice versa
(7b) From general to specific
(7c) From universal to local
(8) Canonical – an order given by an authority, e.g., books of a holy scripture
(9) Consistency of comparable subdivisions that appear in two or more different places
(10) Importance for indexing and query formulation
Rules for selection of preferred terms from a group of synonyms

Include in the thesaurus / KOS any term that falls in scope.
A large lead-in vocabulary is good!

Then select preferred terms.

The preferred term should

• be the best to reflect the meaning of the concept;
• be recognized in the user community;
• be unambiguous;
• be simple and short in spelling.

These criteria may conflict

Frequency data and occurrence in authentic sources can help in the selection.
Rules for selection of concepts as descriptors

The following criteria are helpful:

- Usefulness for searching and other functions;

- Are there alternative solutions: use a combination of descriptors, use a broader descriptor, consolidate with another concept to form a broader concept;

- Logical structure: is the concept needed as a heading?

- Frequency in indexing.
Interoperability of thesauri/ontologies.
Crosswalks

Primary question:

- take a query formulated in vocabulary A,
- map the descriptors to vocabulary B,
- how good is the search in B as compared to using a query formulated in vocabulary B directly?

The answer determines searching compatibility.

Searching compatibility is directional, complex, and depends on the individual descriptors used.

<table>
<thead>
<tr>
<th>Vocabulary A</th>
<th>Vocabulary B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aircraft</td>
<td>Aircraft</td>
</tr>
<tr>
<td>Military aircraft</td>
<td>Airplane</td>
</tr>
<tr>
<td>Pest control</td>
<td>Helicopter</td>
</tr>
<tr>
<td>Pesticides</td>
<td>Aircraft AND Military</td>
</tr>
<tr>
<td></td>
<td>Pest control (no narrower terms)</td>
</tr>
</tbody>
</table>
Insert index language page here
Mapping through a Hub

Dewey | Hub | LCSH
---|---|---
387 Water, air, space transportation | Water transport | Shipping
386 Inland waterway & ferry transportation | Inland water transport | Inland water transport
387.5 Ocean transportation | Ocean transport | Merchant marine
386.8 Inland waterway tr. > Ports | Traffic station: Water transport | Harbors
| Traffic station: Ocean transport | Harren
Insert compat figures here
The structure of a KOS database

Thesaurus / KOS data are relational. **Relational database** is the most natural structure.

**Many types of relationships** – structure should not be restrictive. (See sample list in notebook.)

Examples of Synonymous-Term-type relationships
- ST Synonymous Term
- ET Equivalent Term
- SP Spelling Variant
- AB Abbreviation
- FT Full Term

Structure should allow for a **relationship to be the object of another relationship** (for example, a scope note explaining the relationship)

**Relationship strength**

Appendix 2. Relationship types presently recognized by TermMaster
Note: This list is extensible by simply updating a table in the program and recompiling.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Reciprocal</th>
<th>Group</th>
<th>Reference to</th>
</tr>
</thead>
<tbody>
<tr>
<td>FN</td>
<td>Full form Note (if full form of term &gt; 61 char)</td>
<td>-</td>
<td>SN</td>
<td>Text</td>
</tr>
<tr>
<td>SN</td>
<td>Scope Note</td>
<td>-</td>
<td>SN</td>
<td>Text</td>
</tr>
<tr>
<td>QN</td>
<td>Qualifier Note</td>
<td>-</td>
<td>SN</td>
<td>Text</td>
</tr>
<tr>
<td>HN</td>
<td>History Note</td>
<td>-</td>
<td>SN</td>
<td>Text</td>
</tr>
<tr>
<td>IN</td>
<td>Internal Note</td>
<td>-</td>
<td>SN</td>
<td>Text</td>
</tr>
<tr>
<td>AN</td>
<td>Action Note</td>
<td>-</td>
<td>SN</td>
<td>Text</td>
</tr>
<tr>
<td>SQ</td>
<td>Source (for additional subset record)</td>
<td>-</td>
<td>SN</td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>Detailed source</td>
<td>-</td>
<td>SN</td>
<td>Text</td>
</tr>
<tr>
<td>SI</td>
<td>Synonym, Internal</td>
<td>SI</td>
<td>ST</td>
<td>Term</td>
</tr>
<tr>
<td>SH</td>
<td>From non-hyphenated to hyphenated</td>
<td>SG</td>
<td>ST</td>
<td>Term</td>
</tr>
<tr>
<td>SG</td>
<td>From hyphenated to non-hyphenated</td>
<td>SH</td>
<td>ST</td>
<td>Term</td>
</tr>
<tr>
<td>SP</td>
<td>Spelling variant</td>
<td>SP</td>
<td>ST</td>
<td>Term</td>
</tr>
<tr>
<td>SB</td>
<td>Spelling British</td>
<td>SA</td>
<td>ST</td>
<td>Term</td>
</tr>
<tr>
<td>SA</td>
<td>Spelling American</td>
<td>SB</td>
<td>ST</td>
<td>Term</td>
</tr>
<tr>
<td>AB</td>
<td>Abbreviation</td>
<td>FT</td>
<td>ST</td>
<td>Term</td>
</tr>
<tr>
<td>FT</td>
<td>Full Term</td>
<td>AB</td>
<td>ST</td>
<td>Term</td>
</tr>
<tr>
<td>ST</td>
<td>Synonymous Term</td>
<td>ST</td>
<td>ST</td>
<td>Term</td>
</tr>
<tr>
<td>ET</td>
<td>Equivalent Term</td>
<td>ET</td>
<td>ST</td>
<td>Term</td>
</tr>
<tr>
<td>TR</td>
<td>Translation</td>
<td>TR</td>
<td>ST</td>
<td>Term</td>
</tr>
<tr>
<td>NA</td>
<td>Narrower of Facet</td>
<td>FA</td>
<td>NT</td>
<td>Term</td>
</tr>
<tr>
<td>NX</td>
<td>Narrower term of a broad category used in preliminary sorting</td>
<td>BX</td>
<td>NT</td>
<td>Term</td>
</tr>
<tr>
<td>NF</td>
<td>Narrower term - compound containing factor</td>
<td>BF</td>
<td>NT</td>
<td>Term</td>
</tr>
<tr>
<td>NM</td>
<td>Narrower Term - compound containing Modifier</td>
<td>BM</td>
<td>NT</td>
<td>Term</td>
</tr>
<tr>
<td>NC</td>
<td>Reciprocal of BC</td>
<td>BC</td>
<td>NT</td>
<td>Term</td>
</tr>
<tr>
<td>NT</td>
<td>Narrower Term</td>
<td>BT</td>
<td>NT</td>
<td>Term</td>
</tr>
<tr>
<td>NG</td>
<td>Narrower term - Generic</td>
<td>BG</td>
<td>NT</td>
<td>Term</td>
</tr>
<tr>
<td>NTT</td>
<td>Narrower term - Token</td>
<td>BTT</td>
<td>NT</td>
<td>Term</td>
</tr>
<tr>
<td>NPT</td>
<td>Narrower term - Partitive</td>
<td>BP</td>
<td>NT</td>
<td>Term</td>
</tr>
<tr>
<td>FA</td>
<td>Facet</td>
<td>NA</td>
<td>BT</td>
<td>Term</td>
</tr>
<tr>
<td>BX</td>
<td>Broader term for preliminary sorting</td>
<td>NX</td>
<td>BT</td>
<td>Term</td>
</tr>
<tr>
<td>BF</td>
<td>Broader term - Factor</td>
<td>NF</td>
<td>BT</td>
<td>Term</td>
</tr>
<tr>
<td>BM</td>
<td>Broader term - Modifier</td>
<td>NM</td>
<td>BT</td>
<td>Term</td>
</tr>
<tr>
<td>BC</td>
<td>Broader term that might have NT to be used in combination</td>
<td>NC</td>
<td>BT</td>
<td>Term</td>
</tr>
<tr>
<td>BT</td>
<td>Broader Term</td>
<td>NT</td>
<td>BT</td>
<td>Term</td>
</tr>
<tr>
<td>BG</td>
<td>Broader term - Generic</td>
<td>NG</td>
<td>BT</td>
<td>Term</td>
</tr>
<tr>
<td>BTT</td>
<td>Broader term - Type of token</td>
<td>NTT</td>
<td>BT</td>
<td>Term</td>
</tr>
<tr>
<td>BPT</td>
<td>Broader term - Partitive</td>
<td>NPT</td>
<td>BT</td>
<td>Term</td>
</tr>
<tr>
<td>RC</td>
<td>Related term for combination</td>
<td>RD</td>
<td>RT</td>
<td>Term</td>
</tr>
<tr>
<td></td>
<td>(pop-up menu showing terms to use)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RD</td>
<td>Inverse of RC</td>
<td>RC</td>
<td>RT</td>
<td>Term</td>
</tr>
<tr>
<td>RG</td>
<td>One-directional related term</td>
<td>RH</td>
<td>RT</td>
<td>Term</td>
</tr>
<tr>
<td>RH</td>
<td>Inverse of RG</td>
<td>RG</td>
<td>RT</td>
<td>Term</td>
</tr>
<tr>
<td>RN</td>
<td>Related term in scope note, generated by the program</td>
<td>RO</td>
<td>RT</td>
<td>Term</td>
</tr>
<tr>
<td>RO</td>
<td>Inverse of RN</td>
<td>RN</td>
<td>RT</td>
<td>Term</td>
</tr>
<tr>
<td>RT</td>
<td>Related Term</td>
<td>RT</td>
<td>RT</td>
<td>Term</td>
</tr>
<tr>
<td>EX</td>
<td>Excludes</td>
<td>EF</td>
<td>EX</td>
<td>Term</td>
</tr>
<tr>
<td>EF</td>
<td>Excluded From</td>
<td>EX</td>
<td>EX</td>
<td>Term</td>
</tr>
<tr>
<td>UN</td>
<td>Unspecified relationship</td>
<td>UN</td>
<td>RT</td>
<td>Term</td>
</tr>
<tr>
<td>HT</td>
<td>Homonymous Term</td>
<td>HB</td>
<td>HT</td>
<td>Term</td>
</tr>
<tr>
<td>HF</td>
<td>Homonym From</td>
<td>HT</td>
<td>HT</td>
<td>Term</td>
</tr>
<tr>
<td>ME</td>
<td>Meaning Equivalent</td>
<td>MF</td>
<td>HT</td>
<td>Term</td>
</tr>
<tr>
<td>MF</td>
<td>Meaning equivalent From</td>
<td>ME</td>
<td>HT</td>
<td>Term</td>
</tr>
<tr>
<td>BW</td>
<td>Broader Word</td>
<td>NW</td>
<td>RT</td>
<td>Term</td>
</tr>
<tr>
<td>NW</td>
<td>Narrower Word</td>
<td>BW</td>
<td>RT</td>
<td>Term</td>
</tr>
<tr>
<td>AF</td>
<td>Affects</td>
<td>AY</td>
<td>AF</td>
<td>Term</td>
</tr>
<tr>
<td>AY</td>
<td>Affected by</td>
<td>AF</td>
<td>AF</td>
<td>Term</td>
</tr>
<tr>
<td>PC</td>
<td>Precursor</td>
<td>PB</td>
<td>AF</td>
<td>Term</td>
</tr>
<tr>
<td>PB</td>
<td>Produced by</td>
<td>PC</td>
<td>AF</td>
<td>Term</td>
</tr>
<tr>
<td>RW</td>
<td>reacts with</td>
<td>RW</td>
<td>AF</td>
<td>Term</td>
</tr>
<tr>
<td>IB</td>
<td></td>
<td>IB</td>
<td>ID</td>
<td></td>
</tr>
</tbody>
</table>

```
// From a relationship to a term.
Internal symbol TH
//<
Relation
TI
// Inverse of //</n
Internal symbol TI
```
The structure of a KOS database

Three levels

Level 1: Link term variants to terms

AST  FT  aspartate aminotransferase
GOT  FT  glutamate oxaloacetate transaminase

(FT Full Term)

Level 2: Link terms to concepts

aspartate aminotransferase
ST  glutamate oxaloacetate transaminase

Level 3: Relate concepts to concepts

aspartate aminotransferase
BT  aminotransferases

Levels 1 and 2 are often confounded.
The structure of a KOS database

Two models

Concept-based model

Terms are mapped to concepts. This mapping expresses Synonymous Term relationships.

Concept relationships are expressed using concept identifiers.

Elegant, but in a multi-thesaurus / KOS database requires universal commitment to the term-concept mapping.

UMLS uses this model

<table>
<thead>
<tr>
<th>work</th>
<th>industrial relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>concept 1</td>
<td>RT concept 2</td>
</tr>
<tr>
<td>job</td>
<td>labor relations</td>
</tr>
</tbody>
</table>
The structure of a KOS database

Term-based model

All relationships are expressed as relationships between terms.

A concept relationship may be expressed in many ways, using different synonyms for each concept.

Requires extensive processing to discover all concept relationships starting from a given concept.

<table>
<thead>
<tr>
<th>job</th>
<th>ST work</th>
<th>job</th>
<th>RT industrial relations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ST labor relations</td>
<td></td>
<td>RT industrial relations</td>
</tr>
<tr>
<td>industrial relations</td>
<td>work</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RT industrial relations</td>
<td>job</td>
<td>RT labor relations</td>
</tr>
<tr>
<td></td>
<td>work</td>
<td></td>
<td>RT labor relations</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: Conceptual model for combining thesauri and ontologies
page from JODI
page from JODI
Standards 1
Standards 13
SKOS (Simplified Knowledge Organization System) example.
Sample records from AGROVOC

This example illustrates some of the core features of SKOS, a W3C standard
www.w3.org/2004/02/skos/.
Most records are very abbreviated.

The concepts of this example form the following polyhierarchy:

```
Processed products (#c_15472)
  . Processed animal products (#c_29107)
  .  . Milk products (#c_4830)
  .  .  . Cultured milk (#c_2023)
  .  .  .  . Acidophilus milk (#c_8602)
  .  .  .  . Milk protein (#c_4831)
  . Fermented products (#c_15734)
  .  . Cultured milk (#c_2023)
  .  .  . Acidophilus milk (#c_8602)
  . Protein products (#c_6256)
  .  . Animal protein (#c_439)
  .  .  . Milk protein (#c_4831)

Chemophysical properties (#c_1521)
  . Acidity (#c_8601)
```

SKOS File

```xml
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:rdfs="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:rdf="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
>

<skos:ConceptScheme rdf:about="http://www.fao.org/aims/aos/agrovoc"
  dc:title>AGROVOC</dc:title>
<dc:description>FAO Multilingual Thesaurus.</dc:description>
<dc:creator>
  <foaf:Organization>
    <foaf:name>FAO</foaf:name>
  </foaf:Organization>
</dc:creator>
<dc:rights>Free to all for non commercial use.</dc:rights>
<dcterms:issued>Tue Apr 29 09:49:31 CEST 2008</dcterms:issued>
<dcterms:modified>2008-04-14 10:17:36.0</dcterms:modified>
</skos:ConceptScheme>
```
<skos:Concept rdf:about="http://www.fao.org/aims/aos/agrovoc#c_3">
  <skos:prefLabel xml:lang="en">ABA</skos:prefLabel>
  <skos:prefLabel xml:lang="fr">ABA</skos:prefLabel>
  <skos:prefLabel xml:lang="es">ABA</skos:prefLabel>
  <skos:prefLabel xml:lang="ar">آبا</skos:prefLabel>
  <skos:prefLabel xml:lang="zh">ABA</skos:prefLabel>
  <skos:prefLabel xml:lang="pt">Aba</skos:prefLabel>
  <skos:prefLabel xml:lang="th">ABA</skos:prefLabel>
  <skos:prefLabel xml:lang="de">ABA</skos:prefLabel>
  <skos:prefLabel xml:lang="hu">Aba</skos:prefLabel>
  <skos:prefLabel xml:lang="pl">Aba</skos:prefLabel>
  <skos:prefLabel xml:lang="cs">kyselina abscisová</skos:prefLabel>
  <skos:prefLabel xml:lang="de">ABSCISINSAEURE</skos:prefLabel>
  <skos:prefLabel xml:lang="en">Abscisic acid</skos:prefLabel>
  <skos:prefLabel xml:lang="es">Ácido abscisico</skos:prefLabel>
  <skos:prefLabel xml:lang="fa">آبیسیس اسید</skos:prefLabel>
  <skos:prefLabel xml:lang="fr">Acide abscissique</skos:prefLabel>
  <skos:prefLabel xml:lang="hi">एबसिसिक अम्ल</skos:prefLabel>
  <skos:prefLabel xml:lang="pl">Kwas abscysynowy</skos:prefLabel>
  <skos:prefLabel xml:lang="sk">kyselina abscisová</skos:prefLabel>
  <skos:prefLabel xml:lang="th">กรดแบสิซิค</skos:prefLabel>
  <skos:prefLabel xml:lang="de">ABSCISINSAEURE</skos:prefLabel>
  <skos:prefLabel xml:lang="en">Abscisic acid</skos:prefLabel>
  <skos:prefLabel xml:lang="es">Ácido abscisico</skos:prefLabel>
  <skos:prefLabel xml:lang="fa">آبیسیس اسید</skos:prefLabel>
  <skos:prefLabel xml:lang="fr">Acide abscissique</skos:prefLabel>
  <skos:prefLabel xml:lang="hi">एबसिसिक अम्ल</skos:prefLabel>
  <skos:prefLabel xml:lang="pl">Kwas abscysynowy</skos:prefLabel>
  <skos:prefLabel xml:lang="sk">kyselina abscisová</skos:prefLabel>
  <skos:prefLabel xml:lang="th">กรดแบสิซิค</skos:prefLabel>
  <skos:prefLabel xml:lang="de">ABSCISINSAEURE</skos:prefLabel>
  <skos:prefLabel xml:lang="en">Abscisic acid</skos:prefLabel>
  <skos:prefLabel xml:lang="es">Ácido abscisico</skos:prefLabel>
  <skos:prefLabel xml:lang="fa">آبیسیس اسید</skos:prefLabel>
  <skos:prefLabel xml:lang="fr">Acide abscissique</skos:prefLabel>
  <skos:prefLabel xml:lang="hi">एबसिसिक अम्ल</skos:prefLabel>
  <skos:prefLabel xml:lang="pl">Kwas abscysynowy</skos:prefLabel>
  <skos:prefLabel xml:lang="sk">kyselina abscisová</skos:prefLabel>
  <skos:prefLabel xml:lang="th">กรดแบสิซิค</skos:prefLabel>
  <skos:changeNote rdf:parseType="Resource">
    <rdf:value>The last modification for this concept was for the term in CS</rdf:value>
    <dc:creator>
      <foaf:Organization>
        <foaf:name>FAO/GILW</foaf:name>
      </foaf:Organization>
    </dc:creator>
    <dc:date>2008-03-05 00:00:00.0</dc:date>
    <dc:language>
      <dcterms:RFC1766>
        <rdf:value>EN</rdf:value>
        <rdfs:label>English</rdfs:label>
      </dcterms:RFC1766>
    </dc:language>
  </skos:changeNote>
</skos:Concept>
In the following records, change notes and prefLabel and altLabel for most languages are omitted
For most records, everything but the English prefLabel and the concept relationships is omitted

```
<skos:Concept rdf:about="http://www.fao.org/aims/aos/agrovoc#c_439">
  <skos:prefLabel xml:lang="en">Animal protein</skos:prefLabel>
  <skos:prefLabel xml:lang="fr">Protéine animale</skos:prefLabel>
  <skos:prefLabel xml:lang="es">Proteínas de origen animal</skos:prefLabel>
  <skos:prefLabel xml:lang="de">TIERISCHES PROTEIN</skos:prefLabel>
  <skos:altLabel xml:lang="fr">Protéine d’origine animale</skos:altLabel>
  <skos:altLabel xml:lang="es">Proteínas de origen animal</skos:altLabel>
  <skos:altLabel xml:lang="de">TIERISCHES PROTEIN</skos:altLabel>
  <skos:scopeNote xml:lang="de">Proteinprodukte tierischer Herkunft; fuer die chemischen Verbindungen PROTEIN (6259) benutzen</skos:scopeNote>
</skos:Concept>

<skos:Concept rdf:about="http://www.fao.org/aims/aos/agrovoc#c_1521">
  <skos:prefLabel xml:lang="en">Chemico-physical properties</skos:prefLabel>
  <skos:scopeNote xml:lang="de">Fuer den Boden PHYSIKO-CHEMISCHE BODENEIGENSCHAFT (7182) benutzen</skos:scopeNote>
</skos:Concept>

<skos:Concept rdf:about="http://www.fao.org/aims/aos/agrovoc#c_2023">
  <skos:prefLabel xml:lang="en">Cultured milk</skos:prefLabel>
  <skos:prefLabel xml:lang="fr">Lait fermenté</skos:prefLabel>
  <skos:prefLabel xml:lang="es">Leche fermentada</skos:prefLabel>
  <skos:prefLabel xml:lang="de">SAUERMILCHPRODUKT</skos:prefLabel>
  <skos:altLabel xml:lang="en">Fermented milk</skos:altLabel>
  <skos:altLabel xml:lang="pl">Mleko ukwaszone</skos:altLabel>
  <skos:altLabel xml:lang="es">Bebidas lácticas</skos:altLabel>
  <skos:altLabel xml:lang="pl">Napoje mleczne fermentowane</skos:altLabel>
</skos:Concept>
```
<skos:altLabel xml:lang="de">SKYR</skos:altLabel>
<skos:altLabel xml:lang="en">Skyr</skos:altLabel>
<skos:altLabel xml:lang="fr">Skyr</skos:altLabel>
<skos:altLabel xml:lang="pl">Skyr</skos:altLabel>

<skos:Concept rdf:about="http://www.fao.org/aims/aos/agrovoc#c_4830">
  <skos:prefLabel xml:lang="en">Milk products</skos:prefLabel>
  ... 
  ... 
</skos:Concept>

<skos:Concept rdf:about="http://www.fao.org/aims/aos/agrovoc#c_4831">
  <skos:prefLabel xml:lang="en">Milk protein</skos:prefLabel>
</skos:Concept>

<skos:Concept rdf:about="http://www.fao.org/aims/aos/agrovoc#c_6256">
  <skos:prefLabel xml:lang="en">Protein products</skos:prefLabel>
  ... 
</skos:Concept>

<skos:Concept rdf:about="http://www.fao.org/aims/aos/agrovoc#c_8601">
  <skos:prefLabel xml:lang="en">Acidity</skos:prefLabel>
</skos:Concept>

<skos:Concept rdf:about="http://www.fao.org/aims/aos/agrovoc#c_8602">
  <skos:prefLabel xml:lang="en">Acidophilus milk</skos:prefLabel>
</skos:Concept>
<skos:Concept rdf:about="http://www.fao.org/aims/aos/agrovoc#c_15734">
  <skos:prefLabel xml:lang="en">Fermented products</skos:prefLabel>
</skos:Concept>

<skos:Concept rdf:about="http://www.fao.org/aims/aos/agrovoc#c_15742">
  <skos:prefLabel xml:lang="en">Processed products</skos:prefLabel>
</skos:Concept>

<skos:Concept rdf:about="http://www.fao.org/aims/aos/agrovoc#c_29107">
  <skos:prefLabel xml:lang="en">Processed animal products</skos:prefLabel>
</skos:Concept>

<!-- Concept Schemes and Top Concepts: -->
<!-- ================================= -->
<!-- Concept Schemes and Top Concepts: -->
<!-- ================================= -->
<skos:ConceptScheme rdf:about="http://www.fao.org/aims/asc#c_A">
  <rdfs:label>Agriculture</rdfs:label>
  <skos:hasTopConcept rdf:resource="http://www.fao.org/aims/aos/agrovoc#c_199"/>
  <skos:hasTopConcept rdf:resource="http://www.fao.org/aims/aos/agrovoc#c_15742"/>
  <skos:hasTopConcept rdf:resource="http://www.fao.org/aims/aos/agrovoc#c_1521"/>
</skos:ConceptScheme>
</rdf:RDF>
OWL (Web Ontology Language) example
A simple Pizza ontology

This example illustrates some of the core features of OWL, a W3C standard.
www.w3.org/TR/owl-ref/
It is fairly self-explanatory.
Tutorials
www.co-ode.org/resources/tutorials/protege-owl-tutorial.php,
www.cs.man.ac.uk/~horrocks/ISWC2003/Tutorial/

<!-- Preliminaries -->

<!DOCTYPE rdf:RDF [
<!ENTITY owl "http://www.w3.org/2002/07/owl#">
<!ENTITY owl11 "http://www.w3.org/2006/12/owl11#">
<!ENTITY xsd "http://www.w3.org/2001/XMLSchema#">
<!ENTITY owl11xml "http://www.w3.org/2006/12/owl11-xml#">
<!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#">
<!ENTITY pizza "http://www.pizza.com/ontologies/pizza.owl#">
<!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#">
]

<rdf:RDF
 xmlns="http://www.pizza.com/ontologies/pizza.owl#"
 xml:base="http://www.pizza.com/ontologies/pizza.owl"
 xmlns:owl11="http://www.w3.org/2006/12/owl11#"
 xmlns:pizza="http://www.pizza.com/ontologies/pizza.owl#"
 xmlns:owl11xml="http://www.w3.org/2006/12/owl11-xml#"
 xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
 xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:owl="http://www.w3.org/2002/07/owl#">
<owl:Ontology rdf:about="">
  <rdfs:comment>Protege tutorial for 775</rdfs:comment>
</owl:Ontology>

<!-- Object Properties -->

<!--Full URL of defined object. Given here as an example -->

<owl:ObjectProperty rdf:about="#hasIngredient">
  <rdf:type rdf:resource="#owl:TransitiveProperty"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#hasBase">
  <rdfs:subPropertyOf rdf:resource="#hasIngredient"/>
  <rdfs:domain rdf:resource="#Pizza"/>
  <rdfs:range rdf:resource="#PizzaBase"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="#hasTopping">
  <rdfs:subPropertyOf rdf:resource="#hasIngredient"/>
  <rdfs:domain rdf:resource="#Pizza"/>
  <rdfs:range rdf:resource="#PizzaTopping"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#isIngredientOf">
  <rdf:type rdf:resource="&owl;TransitiveProperty"/>
  <owl:inverseOf rdf:resource="#hasIngredient"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#isBaseOf">
  <rdfs:subPropertyOf rdf:resource="#isIngredientOf"/>
  <owl:inverseOf rdf:resource="#hasBase"/>
  <rdfs:domain rdf:resource="#PizzaBase"/>
  <rdfs:range rdf:resource="#Pizza"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#isToppingOf">
  <rdfs:subPropertyOf rdf:resource="#isIngredientOf"/>
  <owl:inverseOf rdf:resource="#hasTopping"/>
  <rdfs:domain rdf:resource="#PizzaTopping"/>
  <rdfs:range rdf:resource="#Pizza"/>
</owl:ObjectProperty>

<!-- Classes -->

<owl:Class rdf:about="&owl;Thing"/>  

<!-- Basic concepts: Pizza, PizzaBase, PizzaTopping -->

</owl:Class><owl:Class rdf:about="#Pizza">
  <rdfs:subClassOf rdf:resource="&owl;Thing"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasBase"/> <owl:someValuesFrom rdf:resource="#PizzaBase"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasTopping"/> <owl:someValuesFrom rdf:resource="#PizzaTopping"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <owl:disjointWith rdf:resource="#PizzaBase"/>
  <owl:disjointWith rdf:resource="#PizzaTopping"/>
</owl:Class>

<owl:Class rdf:about="#PizzaBase">
  <rdfs:subClassOf rdf:resource="&owl;Thing"/>
  <owl:disjointWith rdf:resource="#Pizza"/>
  <owl:disjointWith rdf:resource="#PizzaTopping"/>
</owl:Class>

<owl:Class rdf:about="#PizzaTopping">
  <rdfs:subClassOf rdf:resource="&owl;Thing"/>
  <owl:disjointWith rdf:resource="#Pizza"/>
  <owl:disjointWith rdf:resource="#PizzaBase"/>
</owl:Class>
<owl:Class rdf:about="#ThinAndCrispyBase">
    <rdfs:subClassOf rdf:resource="#PizzaBase"/>
    <owl:disjointWith rdf:resource="#DeepPanBase"/>
</owl:Class>

<owl:Class rdf:about="#DeepPanBase">
    <rdfs:subClassOf rdf:resource="#PizzaBase"/>
</owl:Class>

<owl:Class rdf:about="#PizzaTopping">
    <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    <owl:disjointWith rdf:resource="#Pizza"/>
    <owl:disjointWith rdf:resource="#PizzaBase"/>
</owl:Class>

<!-- PizzaToppings arranged in a hierarchy through subClassOf -->
<!-- Look at the first few, then skip down to Definition of specific types of Pizza -->

<owl:Class rdf:about="#CheeseTopping">
    <rdfs:subClassOf rdf:resource="#PizzaTopping"/>
    <owl:disjointWith rdf:resource="#SeafoodTopping"/>
</owl:Class>

<owl:Class rdf:about="#MozzarellaTopping">
    <rdfs:subClassOf rdf:resource="#CheeseTopping"/>
</owl:Class>

<owl:Class rdf:about="#ParmezanTopping">
    <rdfs:subClassOf rdf:resource="#CheeseTopping"/>
    <owl:disjointWith rdf:resource="#MozzarellaTopping"/>
</owl:Class>

<owl:Class rdf:about="#VegetableTopping">
    <rdfs:subClassOf rdf:resource="#PizzaTopping"/>
    <owl:disjointWith rdf:resource="#CheeseTopping"/>
    <owl:disjointWith rdf:resource="#SeafoodTopping"/>
</owl:Class>

<owl:Class rdf:about="#TomatoTopping">
    <rdfs:subClassOf rdf:resource="#VegetableTopping"/>
</owl:Class>

<owl:Class rdf:about="#PepperTopping">
    <rdfs:subClassOf rdf:resource="#VegetableTopping"/>
</owl:Class>

<owl:Class rdf:about="#GreenPepperTopping">
    <rdfs:subClassOf rdf:resource="#PepperTopping"/>
</owl:Class>

<owl:Class rdf:about="#RedPepperTopping">
    <rdfs:subClassOf rdf:resource="#PepperTopping"/>
    <owl:disjointWith rdf:resource="#GreenPepperTopping"/>
</owl:Class>
<owl:Class rdf:about="#JalapenoPepperTopping">
    <rdfs:subClassOf rdf:resource="#PepperTopping"/>
    <owl:disjointWith rdf:resource="#GreenPepperTopping"/>
    <owl:disjointWith rdf:resource="#RedPepperTopping"/>
</owl:Class>

<owl:Class rdf:about="#MushroomTopping">
    <rdfs:subClassOf rdf:resource="#VegetableTopping"/>
</owl:Class>

<owl:Class rdf:about="#OliveTopping">
    <rdfs:subClassOf rdf:resource="#VegetableTopping"/>
</owl:Class>

<owl:Class rdf:about="#OnionTopping">
    <rdfs:subClassOf rdf:resource="#VegetableTopping"/>
</owl:Class>

<owl:Class rdf:about="#MeatTopping">
    <rdfs:subClassOf rdf:resource="#PizzaTopping"/>
    <owl:disjointWith rdf:resource="#VegetableTopping"/>
    <owl:disjointWith rdf:resource="#CheeseTopping"/>
    <owl:disjointWith rdf:resource="#SeafoodTopping"/>
</owl:Class>

<owl:Class rdf:about="#HamTopping">
    <rdfs:subClassOf rdf:resource="#MeatTopping"/>
</owl:Class>

<owl:Class rdf:about="#PepperoniTopping">
    <rdfs:subClassOf rdf:resource="#MeatTopping"/>
    <owl:disjointWith rdf:resource="#HamTopping"/>
    <owl:disjointWith rdf:resource="#SalamiTopping"/>
    <owl:disjointWith rdf:resource="#SpicyBeefTopping"/>
</owl:Class>

<owl:Class rdf:about="#SalamiTopping">
    <rdfs:subClassOf rdf:resource="#MeatTopping"/>
    <owl:disjointWith rdf:resource="#HamTopping"/>
</owl:Class>

<owl:Class rdf:about="#SeafoodTopping">
    <rdfs:subClassOf rdf:resource="#PizzaTopping"/>
</owl:Class>

<owl:Class rdf:about="#AnchovyTopping">
    <rdfs:subClassOf rdf:resource="#SeafoodTopping"/>
    <owl:disjointWith rdf:resource="#PrawnTopping"/>
</owl:Class>
Definition of specific types of Pizza in terms of the toppings used

```xml
<owl:Class rdf:about="#NamedPizza">
    <rdfs:subClassOf rdf:resource="#Pizza"/>
</owl:Class>

<owl:Class rdf:about="#MargheritaPizza">
    <rdfs:comment>A pizza that has only Mozarella and Tomato toppings</rdfs:comment>
    <rdfs:subClassOf rdf:resource="#NamedPizza"/>
    <owl:Restriction>
        <owl:onProperty rdf:resource="#hasTopping"/> <owl:someValuesFrom rdf:resource="#MozzarellaTopping"/>
    </owl:Restriction>
    <owl:Restriction>
        <owl:onProperty rdf:resource="#hasTopping"/> <owl:someValuesFrom rdf:resource="#TomatoTopping"/>
    </owl:Restriction>
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasTopping"/> <owl:someValuesFrom rdf:resource="#PepperoniTopping"/>
        </owl:Restriction>
    </rdfs:subClassOf>
    <owl:disjointWith rdf:resource="#AmericanaPizza"/>
    <owl:disjointWith rdf:resource="#AmericanaHotPizza"/>
    <owl:disjointWith rdf:resource="#SohoPizza"/>
</owl:Class>

<owl:Class rdf:about="#AmericanaPizza">
    <rdfs:subClassOf rdf:resource="#NamedPizza"/>
    <owl:Restriction>
        <owl:onProperty rdf:resource="#hasTopping"/> <owl:someValuesFrom rdf:resource="#MozzarellaTopping"/>
    </owl:Restriction>
    <owl:Restriction>
        <owl:onProperty rdf:resource="#hasTopping"/> <owl:someValuesFrom rdf:resource="#TomatoTopping"/>
    </owl:Restriction>
    <owl:Restriction>
        <owl:onProperty rdf:resource="#hasTopping"/> <owl:someValuesFrom rdf:resource="#PepperoniTopping"/>
    </owl:Restriction>
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasTopping"/> <owl:someValuesFrom rdf:resource="#MozzarellaTopping"/>
        </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasTopping"/> <owl:someValuesFrom rdf:resource="#TomatoTopping"/>
        </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasTopping"/> <owl:someValuesFrom rdf:resource="#PepperoniTopping"/>
        </owl:Restriction>
    </rdfs:subClassOf>
    <owl:disjointWith rdf:resource="#AmericanaHotPizza"/>
    <owl:disjointWith rdf:resource="#SohoPizza"/>
</owl:Class>

<owl:Class rdf:about="#AmericanaHotPizza">
    <rdfs:subClassOf rdf:resource="#NamedPizza"/>
    <owl:Restriction>
        <owl:onProperty rdf:resource="#hasTopping"/> <owl:someValuesFrom rdf:resource="#MozzarellaTopping"/>
    </owl:Restriction>
    <owl:Restriction>
        <owl:onProperty rdf:resource="#hasTopping"/> <owl:someValuesFrom rdf:resource="#TomatoTopping"/>
    </owl:Restriction>
</owl:Class>
```
<owl:Restriction>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasTopping"/>
      <owl:someValuesFrom rdf:resource="#JalapenoPepperTopping"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Restriction>

<owl:Restriction>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasTopping"/>
      <owl:someValuesFrom rdf:resource="#PepperoniTopping"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Restriction>

<owl:Restriction>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasTopping"/>
      <owl:someValuesFrom rdf:resource="#CheeseTopping"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Restriction>

<owl:Restriction>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasTopping"/>
      <owl:someValuesFrom rdf:resource="#MozzarellaTopping"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Restriction>

<owl:Restriction>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasTopping"/>
      <owl:someValuesFrom rdf:resource="#ParmezanTopping"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Restriction>

<owl:Restriction>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasTopping"/>
      <owl:someValuesFrom rdf:resource="#TomatoTopping"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Restriction>

<owl:Restriction>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasTopping"/>
      <owl:someValuesFrom rdf:resource="#OliveTopping"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Restriction>

<owl:Class rdf:about="#CheesyPizza"/>

<owl:Class rdf:about="#SohoPizza"/>

<owl:Class rdf:about="#CheesyPizza"/>

<owl:Class rdf:about="#SohoPizza"/>
Appendix B. The Zthes Abstract Model in XML (www.loc.gov/z3950/agency/profiles/zthes-04.html)

Appendix B.1. The Zthes DTD for XML

This DTD was supplied by Thomas Place. It is put forward not as a "good" XML representation of thesaurus information (whatever that might be construed to mean) but as a pragmatically valuable alternative encoding of the Zthes abstract record. Real Zthes data sets have been exchanged in the form of XML documents conforming to this DTD.

<!-- Zthes DTD
    Based on Z39.50 Profile for Thesaurus Navigation, version 0.1 (20 Feb 1999)
    Version of DTD: 25 Feb 1999 -->

<!-- #PCDATA: parseable character data = text
    occurrence indicators (default: required, not repeatable):
    ?: zero or one occurrence (optional)
    *: zero or more occurrences (optional, repeatable)
    +: one or more occurrences (required, repeatable)
    |: choice, one or the other, but not both
-->

<!ENTITY % term "termId, termName, termQualifier?, termType?, termLanguage?”>
<!ENTITY % admin "termCreatedDate?, termCreatedBy?, termModifiedDate?, termModifiedBy?”>

<!ELEMENT Zthes (%term;, termNote?, %admin;, relation*)>

<!ELEMENT relation (relationType, sourceDb?, %term;)>

<!ELEMENT termId           (#PCDATA)>
<!ELEMENT termName         (#PCDATA)>
<!ELEMENT termQualifier    (#PCDATA)>
<!ELEMENT termType         (#PCDATA)>
<!ELEMENT termLanguage     (#PCDATA)>
<!ELEMENT termNote         (#PCDATA)>
<!ELEMENT termCreatedDate  (#PCDATA)>
<!ELEMENT termCreatedBy    (#PCDATA)>
<!ELEMENT termModifiedDate (#PCDATA)>
<!ELEMENT termModifiedBy   (#PCDATA)>
<!ELEMENT relationType     (#PCDATA)>
<!ELEMENT sourceDb         (#PCDATA)>

(This appendix should include a crosswalk with any pre-existing thesaurus DTDs if appropriate.)
Appendix B.2. Sample Zthes-in-XML Document

This document was supplied by Thomas Place.

<?XML version="1.0" ?>
<!DOCTYPE Zthes SYSTEM "zthes.dtd">
<Zthes>
  <termId>102067</termId>
  <termName>video art</termName>
  <termType>PT</termType>
  <termNote>
    Use for works of art that employ video technology, especially videotapes. For the study and practice of the art of producing such works, use "video."
  </termNote>
  <relation>
    <relationType>UF</relationType>
    <termId>102067/001</termId>
    <termName>art, video</termName>
    <termType>ND</termType>
  </relation>
  <relation>
    <relationType>BT</relationType>
    <termId>185191</termId>
    <termName>[time-based works]</termName>
    <termType>NL</termType>
  </relation>
  <relation>
    <relationType>RT</relationType>
    <termId>54153</termId>
    <termName>video</termName>
    <termType>PT</termType>
  </relation>
  <relation>
    <relationType>RT</relationType>
    <termId>253827</termId>
    <termName>video artists</termName>
    <termType>PT</termType>
  </relation>
</Zthes>
Dagobert Soergel dsoergel@umd.edu www.dsoergel.com

Elements of an XML thesaurus / KOS data specification

This proposed schema is parsimonious yet allows the recording of many types of data. It gives enough information to derive a full XML specification. It is more expressive than SKOS.

This spec assumes that data from each source are grouped, so that source attribution is not needed for each element; otherwise the structure would be much more complex. This works for a communications format but not for an internal database format.

The term itself is indicated in a relationship of type TERM. This allows for terms in multiple languages for the same concept and simplifies the schema since elements in term would be the same as in relationship target.

Addition of the scope element was inspired by the Topic Map Standard (see topicmap.com)

The scheme needs a method for indicating a relationship set defined elsewhere and used within the source or for defining a relationship set for the source.

Default is minOccurs="1" maxOccurs="1"

Source (minOccurs="0" maxOccurs="unbounded")
  Pointer to or definition of relationship set used
  Unit: Concept or term or group of terms (minOccurs="0" maxOccurs="unbounded")
    Unique identifier
    Hierarchy position (minOccurs="0" maxOccurs="unbounded")
      Hierarchical level
      Class number / notation
    Scope for which this concept/term holds (minOccurs="0" maxOccurs="unbounded")
  Relationship (minOccurs="0" maxOccurs="unbounded")
    Relationship type
    Relationship target
    /* See below for structure. */
  Relationship strength (minOccurs="0" maxOccurs="1")
  Audience level /* Of this relationship */ (minOccurs="0" maxOccurs="unbounded")
  Perspective /* Of this relationship */ (minOccurs="0" maxOccurs="unbounded")
  Scope for which this relationship holds (minOccurs="0" maxOccurs="unbounded")
  Relationship, added information (minOccurs="0" maxOccurs="unbounded")
    /* This could be a scope note explaining the relationship, an image illustrating the relationship,
      another term, etc. */
    Type of added information /* Relationship types might be reused here. */
    Relationship target
    Audience level /* Of this piece of info. */ (minOccurs="0" maxOccurs="unbounded")
    Perspective /* Of this piece of information */ (minOccurs="0" maxOccurs="unbounded")
Where relationship target has this structure (unifying term, text, images, multimedia document)

Relationship target
  Type
    /* Includes types of terms (descriptor, other preferred term, non-preferred term and types of texts and other
documents, may be an elaborate hierarchy. */
  Target value (a term or a document)
    Term
      Term variant (minOccurs="0" maxOccurs="unbounded")
        Type of variant
          /* Such as Preferred Spelling, other SPelling, ABBreviation, Full Term. */
        Term form (complete term or Stem plus suffix)
          Complete term
          Stem
          Suffix
    Document
      Language (zero to many, exactly one for terms)
      Audience level /* Of this relationship target */ (minOccurs="0" maxOccurs="unbounded")
      Perspective /* Of this relationship target */ (minOccurs="0" maxOccurs="unbounded")
      Scope for which this/term holds (minOccurs="0" maxOccurs="unbounded")
Thesaurus / KOS software and its evaluation

Different types of software

- Thesaurus management software specifically
- Concept mapping software
- Ontology editors
- Description-logic-based software
KOS software selection criteria

General criteria for evaluation of software
Customizable
**Distributed, support for collaboration**
Track sources & changes. Version control

Special functions of KOS management

A  **General system parameters**
Multiple KOS. Multiple languages
Relationship types supported

B  **Input and editing (batch and online)**
Preserve meaningful arrangement

C  **Output in various formats**
Nicely formatted hierarchical displays, concept maps. Web pages
Map detailed internal relationship types to less detailed external

D  **Processing of data**
Check or create reciprocal relationships
Create Description Logic (DL) formulas
Derive concept relationships from DL formulas
Create notations
A detailed list of criteria follows
Requirements for KOS Management Software.

KOS = Knowledge Organization System. Also thesaurus management software

Criteria for Evaluation

Outline

General criteria for description and evaluation of software

Special functions of thesaurus / KOS management

A General system parameters
B Input and editing (of input data files and online)
C Output in various formats
D Processing of data

General criteria for description and evaluation of software

Only a few points that are especially important in connection with thesaurus / KOS software are dealt with here.

Database management system used. Is it easy to produce tailor-made output. Performance.

Deployment and collaboration support

- Can be deployed in a framework such as Web services or SOAP
- Can be accessed by other programs that need to obtain information
- Supports collaborative creation and editing. distributed

Efficiency of storage

Version control. Does the program keep track of all changes

User interface

- Menus versus commands. Use of function keys, etc.
- Use of windows
  - Window positions fixed in program
  - Window positions on the screen can be specified by user
- Navigation possibilities (see editing)
- Program asks for verification before actually recording a change in the KOS database.
- Consistency of the user interface
- Help
**Case sensitivity.** Are upper and lower case treated the same or different in sorting and retrieval? If the same, is this true for all characters or are there exceptions (for example, in Index 4.1 sorting is different for upper and lower case umlauts).

Note: Case is often important to distinguish words, e.g. turkey and Turkey. If case insensitive, need turkey (bird), Turkey (country)

**User influence on how the program works.**

- The user can influence the program behavior through data input without changing the program itself.
  - The program reads parameter from a file (possible from the line(s) at the beginning of an input file), that can be modified by the user.
  - Program uses external files that can be changed by the user.
  - Program accepts specifications written by the user (e.g. specification of a record structure through giving data fields) (example: database system).
  - The user can change menus, error messages, help messages, etc.
- The program itself can be modified according to user wishes
  - Program change through the user himself or herself (source code available)
  - Program modification only through the producer
  - Effort needed for changing the program (this depends on the modularity of the program and the programming technique used. Example: in the program language C constants such as the maximum length of a term or the character used to mark a line as bold can be defined in a header file. To change these constants one needs only to change the header file and then compile the program anew, which could be done by a properly instructed non-programmer.)

For the rest, see the pdf file
Special functions of thesaurus / KOS management

Note.

For all parameters and functions of the program being evaluated the question arises how much the user can influence it. This criterion is always applicable and is explicitly mentioned only in special cases. For example, one should know whether the user can define term types, relationship types, etc. One should keep in mind, however, that many such values have a semantics which must be operated on by the program. For example, if the program has the ability to construct an overall hierarchal structure by binary NT relations, the parts of the program doing this function must use all NT-type relations, and only those. If the user defines a new relation that is a special case of NT then this can become complicated.

Whenever there are user choices, the system should provide default values so that the user who has no special requirements can use those defaults without further ado and need not concern herself with the choice of parameters and the methods for changing the parameters.
General system parameters

Types of KOS supported

Remark: The following types of vocabularies overlap considerably

1 Vocabularies used primarily for information retrieval
  1.1 Classifications and thesauri
    1.1.1 Thesauri without a well-structured classification
    1.1.2 Well-structured classification
    1.1.3 Concept map
  1.2 Topic map (relationally rich KOS)
  1.3 Indexes for books or journals
  1.4 Record filing scheme
  1.5 Data dictionary (in systems analysis and software development)

2 Full-fledged ontology with precisely defined relationship types so that the information can be used for reasoning

3 Nomenclatures and taxonomy (chemistry, biology, etc.)

4 Dictionaries or lexica, general or special
  4.1 Mono- or multi-lingual dictionaries
    4.1.1 Mono-lingual dictionaries
    4.1.2 Multi-lingual dictionaries
  4.2 Glossaries
  4.3 Lexica
  4.4 Picture dictionary

KOS database as a whole

1 Number of KOS in a KOS database
  1.1 One KOS per database
    1.1.1 One of several KOS being worked on can be specified when calling the program (but each KOS is stored in its own database)
  1.2 Multiple KOS integrated in one database
    1.2.1 Number of KOS that can be included
    1.2.2 Only KOS which are subsets of one unified KOS (micro-thesauri within one large KOS), or really different KOS?
    1.2.3 All KOS on an equal footing or one main KOS with connections to terms of other KOS
    1.2.4 Are there relationships between terms from different KOS? How are these relationships determined?
1.2.4.1 Derived from the structure of the database
1.2.4.2 Through reference to a "switching language"
1.2.4.3 Through direct bilateral relationships between pairs of KOS
1.3 Marking subsets in a single KOS (notations are the same across subsets)

2 Is there a starting database of terms and concepts that can be processed by the program?

3 Languages that can be processed: number of languages and list of languages. (This is relevant for functions that depend on the language such as normalization of plural forms to singular, decomposition of terms that include several roots - multi-word terms in English, composite words in German, spell checking, or use of a stop word list.)

3.1 All languages on an equal footing
3.2 One main language

4 Subjects that the program can work on: number and list (This is relevant for spell checking and possibly for operations that use certain structural properties of the terms in a special subject.)

5 Maximum number of terms

6 Stop word list

6.1 For data input (for example for the decomposition of terms that contain multiple roots, in English these are usually multi-word terms.)

6.2 Additional stop word list for KWIC or KWOC Index

6.3 Can the stop word list be changed by the user?

7 Does the program support hierarchical arrangement?

7.1 Maximum number of hierarchical levels

7.2 Does the program preserve sequencing on same level of the hierarchy (see below)
Other characteristics of the system as a whole

1 Code lists for various types of data (term types, relationship types, languages, etc. that are used for checking input and/or for presentation of menus. Can the user change these lists?)

Data that can be given for each term and for relations between terms

Note: This list is just a small subset of all the data that might be needed by varied applications.

1 Maximum term length (Recommended at least fifty, especially if there are many multi-word (or multi-root) terms and long names. Also important for input of source term lists that have long terms.)

1.1 Maximum defined by the system

1.2 Maximum can be defined by the user (within system limits) (This is needed if a KOS is produced for an ISAR system that has its own maximum term length.)

1.2.1 Is it possible to define a separate term length for each of multiple KOS integrated in a KOS database

2 Treatment of homonyms. How are the separate meanings of homonyms identified.

3 Language of the term. Maximal length of the language indication. Does the KOS use a standard list of language symbols (In a multilingual KOS databases indication of the language is necessary for the unique identification of a term.)

4 Sort form (if different from display form)

5 Part of speech for a term

6 The gender of a term

7 Other syntactic or morphological data

8 Language level (day-to-day language, discipline specific language, outdated, etc.)

9 Indication of whether this term may participate in relationships to other terms

10 Term types (See attached list for examples)

10.1 Term types predefined in the system: number and list

10.2 User definable term types: how many

10.3 Can a separate list of term types be defined for each KOS included in an integrated KOS database?

11 Perspective, a value that can be used for selecting terms into lists (Index 4.1)
Marker, another value that can be used to select terms into lists (Index 4.1)

Notation

13.1 Coarse notation (for example, for identifying broad subject groupings or facets)

13.2 Detailed notation (can at the same time fulfill the functions of a coarse notation)

13.3 External notation

13.4 Internal notation (for example, a notation expressing the hierarchical structure to be used by a retrieval program for inclusive searching)

13.5 For each kind of notation: maximum length (can the maximum length be specified by the user?)

13.6 Can the user specify whether a descriptor can have several or only one notation (MeSH, for example, has for each descriptor as many notations as the descriptor has places in the parley hierarchy.)

13.7 How much influence does the user have on the form of the notation

13.8 Support for the generation of notations

14 The sequence of the terms on the same level of a hierarchy can be stored (This can be implemented through notation)

15 Relationship types (See attachment for examples.) (At a minimum, KOS software should support the relationship types specified in KOS standards.)

15.1 Relationship types predefined in the system: number and list

15.2 Relationship types that can be defined by the user: number (But see note at the beginning.)

15.2.1 Can the user define/change the rules used by the system in processing relationship types?

15.3 Possibility of specifying many detailed relationship types in the database but map these to a few general relationship types in the user version

15.4 Can the relationship type names for the user version be freely chosen

16 Rules for relationship types (Rules serve for consistency checking but can also introduce unnecessary restrictions.) Examples for rules:

16.1 Synonym relationship always from descriptor to non-descriptor

16.2 Abbreviation relationship always from descriptor to non-descriptor

17 Data about relations

17.1 Strength of connection
17.2 Aspect used in establishing the relation. For hierarchical relationships: The characteristic of subdivision (However, it is preferable to create an own heading for each characteristic of subdivision to group all the narrower terms that correspond to that characteristic.)

17.3 Qualification through context (that is, the relation is valid only for a certain context, or in any case the connection strength is dependent on the context. Put differently, the relation is itself an object related to another object, such as a term.)

17.4 Scope note for a relation. Explains why the relation was introduced.

17.5 In what output formats should the relation appear (This does not refer to the relationship type, but to the specific relation between two terms)

18 Maximum number of relations of a given relationship type that can be given for a term (This may differ from one relationship type to another.) (For example, some systems allow only one BT; this is not good, since mono-hierarchy is too restrictive.)

19 Maximum number of relations for a term altogether

  19.1 In a single KOS
  19.2 In the integrated KOS database

20 Scope note and other text information

  20.1 How many types of text information (for example, is it possible to have internal notes)
  20.2 How many notes of each type of term
  20.3 Maximum text length
  20.4 Can descriptors inside a scope note be marked and treated specially?

21 For terms, notations and/or relations

  21.1 Status value (Such as included in present edition, kept for later decision, deleted. The deleted status is important so that decisions on the same term must not be made again when, for example, this term appears in a newly processed source; it is also needed in order to reconstruct the state of the KOS at the time of indexing a given document.)

  21.2 Source indication

  21.2.1 Maximum length of the source indication
21.2.2   . . Maximum number of sources of a term or relation

21.3   . Date indications (Dates for various events such as inclusion in the KOS database, inclusion in a given KOS, approval by an editor, deletion from the KOS, etc.)

21.4   . Frequency of use (in a system that indexes with weights: Frequency of use with weight 2, frequency of use with weight 1 or 2) (Keep in mind that one always must specify the frequency with a time span.)

21.5   . Indication of the editor/lexicographer and reviser

21.6   . Editing history (edited when and by whom, revised and approved when and by whom)

22   Data on the sources as such (Does the program allow for a directory of sources?)

23   Other kinds of data provided for in the program

24   Can the user define additional types of data? In what limits? (Since many data about a term can be given through relationships the possibility of defining additional relationship types is important.)

25   Data Structure

Level of logic supported (description logic, first order predicate calculus, etc.)
Data input and editing

Data input

1

Batch input

1.1

Batch input of other KOS

1.2

Batch input of KOS files that have been created with a word processor or otherwise. (In many cases this is the most efficient method of inputting data. This method also allows editors to work independently from the program wherever there is a computer.)

1.2.1

Format(s) of such input files

1.3

Command structure that allows for scheduling the input of several files in sequence (this is important because the input of a file may take a long time. With such a command one can input several files over night without intervention.)

2

Online data input (see also online editing)

2.1

Input of individual terms and data about them

2.1.1

Input of term and data about the term in one step

2.1.1.1

Online form for all data about a term. Details about this form (for example, are there fixed fields for relationship types or is the relationship type given through an explicit name, scope note as one continued text for a number of lines, scrolling if not all information fits on one screen)

2.1.1.2

Script: The system prompts for the various data for a term in a fixed sequence. Is the content and the sequence of these prompts defined by the system or definable by the user?

2.1.1.3

How does the system treat cross-terms that have not yet been entered as main terms

2.1.2

Input of terms and relations in separate steps

2.1.3

Are all data about a term shown on the screen once input is completed? Can they be modified at that point?

2.2

Input of whole lists, especially hierarchies, that have been composed on the screen under KOS program control (Functionally this is very similar to batch input of KOS files as discussed.)
Editing

For the selection of the terms to be edited in an editing session and for the format of display of the data to be edited see the criteria under Output

1 General functions in editing (Some of these are also important for input.)
1.1 Effort for different types of changes
1.2 Consistency check for changes made (see consistency check under D)
1.3 Is the user asked to verify the change?
1.4 Can changes be made with "hierarchical force"? (E.g., deleting a broad term and all its narrower terms.)
1.5 Does the system give a message if the user enters a term or relation that was considered earlier and either rejected for inclusion or deleted after it was once included?

2 Types of changes. For each type: How much effort

2.1 Changes for terms
2.1.1 Adding a term
2.1.1.1 Specifying of the position of the new term in the hierarchical sequence (the input of a BT relation alone is not sufficient if one wants to maintain a meaningful sequence of terms on the same level)
2.1.2 Deleting a Term
2.1.2.1 Are all relations deleted as well (or at least not output any more? Possibility differentiated by output formats for editing and output formats for the user version)
2.1.2.2 Is there a consistency check after a term was deleted? Especially the effects on the hierarchy need to be checked. It is problematic to delete a descriptor that has narrower descriptors that ought to be kept. Some systems do not allow deletion of a term that is linked to other terms through relations; the editor must first delete these relations.
2.1.3 Adding a term that was deleted earlier
2.1.3.1 Are the relations that were in the system while the term was still there also added automatically?
2.1.4 Change in term type (especially from descriptor to non-descriptor and vice-versa)
2.1.5 Replace one term through another
2.2 Notation changes
2.2.1. Are other affected notations automatically changed accordingly (important especially when a term is added at a given position)

2.3. Changes in relationships

2.3.1. Adding a relationship

2.3.2. Deleting a relationship

2.4. Global changes (for example, add EN to all terms in the KOS database if one wants to change from an English-only KOS database to a multilingual database)

3. Batch Editing

3.1. File of editing commands

3.2. The program produces a file for editing (as part of its output functions). This file can be edited and re-input (All data in the file for editing where given a temporary deleted status. For any data not contained in the edited file, that deleted status becomes permanent.)

3.2.1. Format of the file for editing (for example, Generic Word Processor format or a format that can be used by an outline processor)

3.2.2. See also online editing and output regarding the criteria for selection of terms and the display format

4. Online editing (most of the functions apply also to online input)

4.1. History functions

4.1.1. Is navigation history kept? Can the user retrace steps?

4.1.2. Complete transaction log for error recovery?

4.2. Manipulation of lists of terms that must undergo editing

4.2.1. Editing lists can be stored and recalled

4.2.1.1. During one session

4.2.1.2. From one session to the next

4.2.1.3. Editing lists named by the user or by the system

4.2.2. Navigation in the editing list

4.2.2.1. Screen by screen

4.2.2.2. Scrolling

4.2.3. Deleting elements from an editing list

4.3. Navigation in a batch of forms

4.4. Switching between editing lists and editing batch of forms
4.5 . . Editing data for an individual term

4.5.1 . . . Editing data about an individual term in a list

4.5.1.1 . . . Which data are displayed (see C)

4.5.1.2 . . . Which data can be edited (These editing changes can be changes to the database or they can be changes that influence further editing, such as marking a term as processed or moving a term to another list.)

4.5.1.3 . . . Can new terms be input while working on an editing list?

4.5.1.4 . . . Does system display available options (for example, when working on BT relationships, the system might display a list of the terms that would be legal and the user would select; see consistency checks)

4.5.2 . . Editing data for a term on an online form (most systems would always allow input of new terms in this context by having the user request an empty form)

4.5.2.1 . . . Screen format and editing options (for example, is it possible to do full screen editing as in a word processor using the general keys like arrows and delete, can text be copied from one place to another, from one form to another, can scope notes be edited as continuous text, mouse support.)

4.5.2.2 . . . Function for exchanging descriptor with one of the synonyms

4.5.3 . . Jumping to a cross-referenced term, editing it, and returning to the term previously worked on (possibly do this multiple steps)

5 Editing entire structure, especially a section of a hierarchy, without detailed data for each term. This is functionally equivalent to editing and re-input of an editing file in hierarchical format as discussed above, but may be more convenient.

5.1 . . Functions offered for editing (for the editing of hierarchies the functions of outline processing are especially useful)

Reports on inconsistencies (For example, relationships to a nonexistent term) in a form that facilitates the input.

1 Batch

2 Online

Reports on changes, especially if there is a procedure for the edition and final approval.
Output

Note: Output can be for human use, either printed or online, by KOS users or for editing, or for use by another system. Furthermore, many of the functions/criteria discussed here apply also to the selection of a group of terms for online editing. This includes the selection and sequencing of terms to be edited online, the data displayed on the screen, and the extent to which the user can control these parameters.

General criteria for all output functions

(One and the same KOS management program can have different values for different output formats.)

1 Domain of the output

1.1 An individual KOS (either the only KOS in the database or an individual KOS from an integrated database)

1.2 Terms that appear in multiple KOS

1.2.1 User can specify a list of KOS

1.2.2 Concordance

1.2.3 Comparison print: a printout that shows how the terms occurring in one or more source KOS are dealt with in a target KOS, highlighting especially terms missing from the target KOS

2 Selection of terms from the domain (Many of these criteria are important especially for editing.)

2.1 Scope in a hierarchy (identified by beginning and ending notation or all terms under a broad term)

2.2 Selection by relationship to another term or object

2.3 Selection by facet

2.4 Selection by hierarchical level

2.5 Scope in alphabetical sequence (identified by beginning and ending term)

2.6 Selection by status

2.7 Selection by markers or perspective

. . Selection by absence from a given KOS. (This is important for editing: If a new source is added to the KOS database, check terms absent from the KOS being worked on to see whether they should be included.)

2.8 Selecting terms that are not yet revised and approved
2.9. Select terms not included in the last printed or otherwise published version

2.10. Selection by language

2.10.1. Selection by string pattern contained (free text searching). How powerful are the possibilities for defining patterns (wild cards for characters, for strings, etc., phrase searching vs. just word searching, etc.)

2.11. Selection by internal term number (record number)

2.12. Selection by a boolean combination of the criteria

2.13. Selection of a small list by marking terms in a big list

3. Sequencing of the selected terms for presentation (this is important to achieve a meaningful sequence for editing)

3.1. Hierarchical sequence

3.1.1. Stored hierarchical sequence (usually implemented through notation)

3.1.1.1. If the domain includes several KOS: Can the editor select one KOS as a guide that will determine the hierarchical sequence?

3.1.2. Hierarchical sequence generated by the program based on hierarchical relationships (This usually implies alphabetical sequence of the children under the same parent.)

3.2. Alphabetical sequence

4. Determining the entry point for the list

5. Method for calling up a list (This may be different for the different selection criteria. For example: Index 4.1 the editor working on a term can position the cursor on the facet field and call up a list corresponding to the value; when the facet field for the term being worked on has the value "Person" then the list called up includes all terms from the facet "Person".)

6. Content and format of the output (for screen forms and for lists) (For each criterion: how much control does the user have?)

6.1. Data for each term

6.1.1. Suppressing relationships that are shown through arrangement (especially suppression of hierarchical relationships that are shown through the sequence and indication of the hierarchical level)
6.2 Differentiation of relationships types
6.3 Symbols for relationship types
6.4 Sequence of data and relationships for one term
6.5 Sequence of the cross terms within the same relationship type
6.6 Are cross terms shown with their notation
6.7 Are cross terms that have narrower terms identified (for example, by a plus before or after the notation or before or after the term) (This is important because the searcher or indexer should check to see whether one of the narrower terms is more suitable than the cross term.)
6.7.1 If yes, is this indication fixed by the system or selectable by the user? Is the symbol chosen (in the example plus) user selectable?
6.8 Orientation aids for the user (such as giving the first and last term on the page in an alphabetical list or the first and list notation on a page in a hierarchical list)

7 Number of languages presented in the output format
7.1 Monolingual KOS
7.2 Multilingual KOS
7.2.1 Parallel arrangement with a column for each language

8 Specification of the output format
8.1 Only predefined formats (The evaluation of a KOS management program should include detailed descriptions and sample pages of these redefined formats.)
8.2 Specification of the output format through the user
8.2.1 Specification online. Can the resulting specification be stored and recalled under a name?
8.2.2 Specification through a special specification file that can be produced with a word processor
8.2.3 How complex is this specification (this must be seen in relation to the number of formatting options offered)
8.2.4 How compact is the specification
8.2.5 Does the program come with predefined formats or specification files which the user can simply use as is or modified, which would be less work than creating these files from scratch. (Include in the evaluation detailed description and sample pages of these redefined specifications.)

9 Possibility to order several outputs at the same time (e.g., for overnight processing)
Printed KOS for public use

Note: Many of the format specifications listed here apply also to online displays, particularly Web displays.

1 Printing methods supported: especially laser printer support (for example, through output of a file in the format of a word processing or desktop publishing program), Photo Type Setting Support. File with general markup language

1.1 Formatting into pages, especially considering proportional fonts and different font sizes

1.2 Formatting into columns

Note: Formatting into pages or columns important for producing orientation aids for the user

2 Can the output file be edited before printing?

3 Output formats

3.1 Hierarchical lists of terms

3.1.1 Sequence of the hierarchy, see above

3.1.2 Specificity of the hierarchical list

3.1.2.1 Hierarchical outline

3.1.2.2 Hierarchical list of all terms

3.1.3 Degree of detail of the hierarchy

3.1.3.1 Quick hierarchical list

3.1.3.2 Annotated hierarchical list

3.1.4 Method for showing the hierarchical level

3.1.4.1 Showing the hierarchical level through indentation

3.1.4.1.1 Indentions with a special symbol (for example, a dot) for each level

3.1.4.1.2 With additional explicit indication of the hierarchical level

3.1.4.1.3 Indentation, type size, and normal/bold as a function of the hierarchical level

3.1.4.1.4 Maintaining the hierarchical context through repeating the hierarchical change at the beginning of each (left that is even)

3.1.4.2 Hierarchy without indentation with explicit indication of the hierarchical level, esp. for two column printouts
3.2 Graphical representation of conceptual relationships (concept maps, topic maps)
3.3 Alphabetical lists of terms
3.4 Alphabetical index
3.4.1 KWOC index
3.4.1.1 KWOC index in which the access words are normalized to singular form

Online search for navigation in the KOS using the Web or the program itself (also important for editing)

1 Web files
1.1 Generation of hyperlinks and anchors for jumping from an outline to a quick hierarchy to an annotated hierarchy and for following relationships
1.2 Explorer-type expandable hierarchy
1.3 Control over partitioning the KOS to get Web files of reasonable size
1.4 Capability for showing coordinated windows on the Web

Files for communicating KOS data to retrieval systems (such as DIALOG or search engines or intranet retrieval engines) or to other KOS management programs

1 Files compliant with a given standard, for example ZThes
2 Files that can be input into a database system for searching the KOS. If the database is Web-enabled, this can be combined with KOS Web files.

Change reports

1 Report of changes since a given date
2 Report of changes since the last printed or otherwise published edition

Statistical reports (Number of descriptors and entry terms, number of descriptors in each major class, number of descriptors on each hierarchical level, number of each type of relationship)
Processing of data (consistency checks, etc.) through the program

In general: how much support does the program offer in the processing and generation of data (for example, constructing a hierarchy from BT/NT relationships, generation of notations)? The other way around: To what extent is the program limited to managing the data input by the user.

Checking input data for formal correctness (in batch input or during online editing)

1. Checking the term length for main terms and cross terms
2. Checking the relationship symbols, term type symbols, language symbols, etc.
3. Checking for illegal terms in a hierarchy (A jump by more than one level down is illegal.)
4. Checking completeness (for example, checking whether a notation is given for a term when one is required)
5. Spell check

Consistency checks (during batch input and online editing)

1. General characteristics of consistency checks
   1.1. Is the check mandatory or user selectable
   1.2. Force of the check (maybe be different for different kinds of check)
       1.2.1. There is no way to input inconsistent data
       1.2.2. Merely a warning to the editor
2. Consistency checks for terms
   2.1. Check for form of term
       2.1.1. Check whether the term agrees with the rules of form established for the KOS (for example, preference for singular, preference for nouns over adjectives or verbs)
       2.1.2. Singular/plural check (whichever is preferred in the KOS) or conversion
       2.1.3. Spell check
       2.1.4. Capitalization check or conversion (Some terms must always be capitalized; can this be enforced, for example by having these terms in the KOS database?)
2.2. Duplication check for terms
2.2.1 . . Does duplication check consider singular and plural as the same? (In an integrated database both can appear.)

2.2.2 . . Does the duplication check consider variant spellings the same?

2.2.3 . . Does the duplication allow the editor to take care of homonyms

2.2.4 . . Can the program handle identical strings of characters that denote separate terms in different languages?

3 Consistency check for relationships

3.1 . Duplication checks for relationships

3.2 . Check for reciprocal relationships and creation of reciprocal relationships where needed

3.3 . Check for several relationships between the same ordered pair of terms (If this is not allowed, it should be checked, but only within an individual KOS.)

3.4 . Check for a relation of a term to itself

3.5 . Check for terms that are not preferred terms used as cross term in a concept relationships. Alternatively, replacing a term through the appropriate preferred term when producing output

3.6 . Check for ST-type relationship from descriptor to descriptor

3.7 . Check the consistency of hierarchical relationships

3.7.1 . . Check for hierarchical relationships that jump a level, for example, A NT B, B NT C, A NT C

3.7.2 . . Check for hierarchy cycles, for example, A NT B, B NT C, C NT A (Such cycles could throw the program for a loop in the generation of a complete hierarchical structure from hierarchical relationships.)

3.8 . Check for incomplete relationships, for example, semantic factoring with only one semantic factor

3.8.1 . . Check for terms that do not participate in any relationship (orphan terms)
3.9. More complex checks of the semantic consistency of a relationship (example for a rule: hierarchical relationships are allowed only between terms belonging to the same category (we do not say here whether this rule is good or bad). ST-type relationship only in some language if TRanslation relationship is used from one language to another (using ST-type relationships regardless of the languages involved might actually be better). Formal ontologies define properties of concepts such that only concepts that agree in these properties can be hierarchically related. For example, for a concept that represents a class of objects, each instance has identity, but for a concept that designates an amount this is not the case.

4. Check whether input data conform to the field or relationship rules (For example, some relationship may be valid only to places, so the cross term must be a place name.)

Support hierarchical inheritance

Support reasoning over the knowledge base

Support in the editing of terms

1. Normalization of terms to singular (while storing the original form)

Generation of notations

1. The editor can input some or all notations, but where notations are missing the program generates them through hierarchical extension.

2. Format of the notations generated

Support for the processing of relationships, for example

1. Support for the generation of relationships, for example

2. Extracting single words from a multi-word term and presenting them as candidates for semantic factors

3. Presenting candidates for semantic factors based on hierarchical inheritance from the broader terms

4. Generating hierarchical relationships from an input list in hierarchical format

Conversely, generation of a hierarchical sequence from binary hierarchical Relationships

5. In an integrated KOS database: use of synonym relationships in detecting the identity of conceptual relationships, for example KOS 1: A BT B; KOS 2: A BT C; any KOS: B ST C; conclusion: the two hierarchical relationships are the same conceptual relationships.
Appendix

KOS development process, with emphasis on multilingual KOS

Building a KOS, especially a multilingual KOS, takes a lot of effort. Some term relationships can be derived by statistical analysis of term occurrence in corpora, but this will not result in the kind of well-structured conceptual system described above. Developing such a structure requires intellectual effort.

A common method for KOS development in a single language is to work bottom-up: One collects a list of terms (words and phrases), preferably from search requests, but also from documents, free-term indexing, and other KOS. These terms are then sorted into increasingly fine-grained groups, until a group contains only synonyms or terms that, for purposes of the KOS, can be considered synonyms. In this process at least some homonyms will be detected; they must be disambiguated into several senses, each expressed by its own (possibly newly coined) term having one meaning and being grouped accordingly. A group of synonyms can be considered to represent a concept; usually a preferred term to designate the concept is selected, but some other concept identifier can be used. A first rough hierarchy of concepts emerges from this process.

Now perform conceptual analysis, especially facet analysis at various levels, resulting in a well-structured faceted hierarchy. Next write definitions (scope notes) (often results in rethinking the hierarchy) and introduce relationships between concepts that complement the hierarchy.

The development of a multilingual KOS is, naturally, an even more complex undertaking; the basic approaches are summarized in the table below. The ideal way to develop a multilingual KOS is to start from a pool of terms in all covered languages and carry out the process without regard to the language of the terms. This will bring together terms from different languages that have the same meaning into one group. This process gives all languages an equal chance to contribute concepts and concept relationships. It also forces a careful analysis of the meaning of each term in each language to determine the degree of equivalence, making it possible to develop the fine-grained structure of definitions that has the potential of providing powerful support to free-text cross-language retrieval.

Of course, this process requires a lexicographer knowledgeable in the subject matter of the KOS and fluent in all covered languages, not a very practical requirement. A more practical variation that still maintains the spirit of this approach is to start with two languages and develop the conceptual structure — a bi-lingual lexicographer is needed in any event. Definitions should be written in both languages. One would then work on a pool of terms in a third language and fit it into the structure, creating new concepts as necessary. This is not at all the same as translating the KOS into the third language. This requires a lexicographer fluent in one of the starting languages and the third language. Add other languages the same way.

The result of such a process is a conceptual system that brings the conceptual structures embedded in the different languages under one roof, so to speak.

The most common approach to the development of a multilingual KOS is to translate an existing monolingual KOS into one or more languages. But this approach is problematic: The original language and its vocabulary determine the conceptual structure, and one merely looks for equivalent terms in the second language with-out covering its terminological richness. In some multilingual KOS, only one term in the target languages is provided, making the KOS unsuitable for query term expansion in free-text searching.

In between is an approach in which one starts with a monolingual KOS as the center and fits terms from one or more other languages into the structure of this central KOS without changing the concepts or the hierarchy. EuroWordNet (Gillaranz 1997) takes an improved variation of this approach, working with the English WordNet as its central KOS. In EuroWordNet, separate and independent word nets are constructed in each language in parallel efforts, each identifying synonym sets in that language (A synset can be considered a concept). The synsets of each language are then mapped independently to WordNet synsets; no changes are made to WordNet. In addition to identity, this mapping allows for hyponym and hypernym relationships, thus indicating that the concept identified in the language being worked on is not included in WordNet, but giving at least the hierarchical location. EuroWordNet also uses a very weak variation of approach 5: The participants developed a “top ontology”, which presumably reflects and integrates perspectives from their individual cultures. In addition to being mapped to WordNet, the individual language synsets are also mapped to this top ontology.
Building a multilingual KOS

Requirements: Must cover all concepts of interest to the users in the various languages, at a minimum all domain concepts lexicalized in any of the participating languages. Must accommodate hierarchical structures suggested by different languages.

Approaches (by increasing complexity and quality)

(1) Start from monolingual KOS and translate. This approach does not capture concepts lexicalized only in another language and is biased to the conceptual structure underlying the starting language. May not produce all synonyms in the second language.

(2) Start from a monolingual KOS as the center. Collect terms from other languages and establish correspondences of these terms to the central KOS. Suffers from similar bias toward the starting language as (1), but may cover more synonyms in the other languages.

(3) Work with a central KOS as in (2), but after collecting terms from a second language first group them into synsets, that is, derive concepts each of which is represented by a set of terms, and then map each concept to the corresponding concept in the central KOS or indicate that the concept is new and give the nearest broader or narrower concept in the central KOS. Note that the central KOS remains unchanged.

(4) As (2), but add concepts not in the starting KOS. This mitigates bias, but the central KOS now becomes a moving target.

(5) Start from a pool of terms from all participating languages and organize them into a conceptual framework, establishing term correspondence in the process. This approach results in a true "conceptual interlingua" not biased to any one language, but offering a home to multiple conceptual perspectives. This approach requires most effort.
KOS development example

Audience/Demographic Characteristics
Raw term list

Terms collected from lists used in three NCADI databases, from the NCADI request form, and from Breaking New Ground for Youth At Risk, duplicates eliminated, in alphabetical order

A/D prevention professional
A/D treatment professionals
Administrator/Manager
Adults (25-59 years)
African Americans/Black
Asians and Pacific Islanders (Chinese, Japanese, Vietnamese, etc.)
Attorney
Biomedical researchers
Blacks
Caucasians
Children subjected to abuse and neglect
Children and youth who are economically disadvantaged
Children (pre-adolescents)
Children of alcoholic or other drug-abusing parents
Clergy
College students
Community organization leader
Community service groups
Correction officer
Criminal/juvenile justice
Disabled
EAP Practitioners
Educator/teacher/trainer (specify grade[s])
Elderly (60 + years)
Elementary youth (5-12)
Employees
Employer
General public
General public, personal concern
General public, concern for family/friend
Grantee
Handicapped/Disabled
Health care providers (physicians, nurses, Pas, NAs, pharmacists)
Health care professional
High-risk families
High-risk youth
High-risk families/youth (including COAs and ACOAs)
Hispanics/Latinos
Homeless or runaway youth
Homosexuals (males and females)
IV drug users
Jr. High Youth
Judge
Latchkey children
Legislator
Librarian/Information Specialist
Media representatives
Mental health professional
Native Americans (American Indians and Eskimos)
Other
Parents (specify age of child)
Parole/Probation officer
Patients
Police officer
Policy makers/Administrators
Preschool (age 4 and under)
Psychosocial researcher
Recreation/Sports Personnel
Reporter/writer
Researcher
School dropouts or those at risk of dropping out
School Administrator
Scientists and researchers
Single teenage mothers and their children
Social service professional
Sr. High Youth (16-18)
Student
Unemployed youth or those in danger of being unemployed
Unknown/anonymous
Women
Young Adults (18-25 years) (19-25 years)
Youth who use gateway drugs
Youth (adolescents)
Youth who are suicidal or physically or mentally disabled
Youth who are engaged in violent or delinquent acts
### Terms collected arranged in broad groupings

#### Age
- Preschool (age 4 and under)
- Elementary youth (5-12)
- ST Children (pre-adolescents)
- Youth (adolescents)
  - Jr. High Youth
  - Sr. High Youth (16-18)
- Young Adults (18-25 years) (19-25 years)
  - College students
- Adults (25-59 years)
- Elderly (60+ years)
- Student?

#### Gender
- Women
- Men

#### Sexual preference
- Homosexuals (males and females)

#### Racial/ethnic group
- African Americans/Black
- Asians and Pacific Islanders
- Caucasians
- Blacks
- Hispanics/Latinos
- Native Americans (Amer. Indians, Eskimos)

#### Group by ability/handicap
- Disabled
- Handicapped/Disabled

#### Groups at high risk of drug use
- Children subjected to abuse and neglect
- Children and youth economically disadvantaged
- Children of alcoholic or other drug-abusing parents
- High-risk families
- High-risk families/youth (including COAs and ACOAs)
- High-risk youth
- Homeless or runaway youth
- Latchkey children
- School dropouts or those at risk of dropping out
- Single teenage mothers and their children
- Unemployed youth or those in danger of being unemployed
- Youth who use gateway drugs
- Suicidal or physically or mentally disabled youth
- Youth who are engaged in violent or delinquent acts
- IV drug users

#### By profession or position
- A/D prevention professional
- A/D treatment professionals
- Administrator/Manager
- Attorney
- Clergy
- Community organization leader
- Community service groups
- Correction officer
- Criminal/juvenile justice
- EAP Practitioners
- Educator/teacher/trainer (specify grade[s])
- Health care providers (physicians, nurses, Pas, NAS, pharm.)
- Health care professional
- Judge
- Legislator
- Librarian/Information Specialist
- Media representatives
- Mental health professional
- Parole/Probation officer
- Police officer
- Policy makers/Administrators
- Recreation/Sports Personnel
- Reporter/writer
- Researcher
  - Biomedical researcher
  - Psychosocial researcher
- School Administrator
- Scientists and researchers
- Social service professional

#### By employer/employee relationship
- Employees
- Employer

#### Other groupings
- Patients
- Parents (specify age of child)
- General public
  - General public, concern for family/friend
  - General public, personal Concern
- Grantee
- Other
- Unknown/anonymous
One area conceptually refined

Groups at high risk of drug use

- Suicidal or physically or mentally disabled
- Persons from unstable or low-cohesion families
- Children of alcoholic or other drug-abusing parents
  - Grown up or still under age
- Children of single teen-age mothers
- Persons subjected to abuse or neglect
  - Now or in the past
  - Persons subjected to abuse and neglect by parents
  - Latchkey children
  - Persons subjected to abuse and neglect by their spouse
- Single teenage mothers
- School dropouts or those at risk of dropping out
- Unemployed or in danger of being unemployed
- Economically disadvantaged
- Homeless
  - Runaway youth
- Gateway drug users
- Persons engaged in violent or delinquent acts

The concept Youth at risk of drug use or any of its subordinate concepts (as specified by group and age range) can be produced by combination with Adolescent.

An observation on defining groups by combination: For any group defined by personal/demographic characteristics, there are several derivative groups, e.g.:

- Parents of members of the group
- Children of members of the group
- Spouses of members of the group
- Teachers of members of the group

The KOS needs to make provision for forming such combinations.
Examples of Thesauri
and other
Knowledge Organization Systems (KOS)

The paper tutorial notebook includes copies of sample pages from print thesauri / KOS, Web pages on thesauri / KOS, and of search results in various Web thesauri / KOS. The pdf file does not include these pages, but it does include the URLs of applicable Web pages.
Alcohol and Other Drug Thesaurus


Number of:

<table>
<thead>
<tr>
<th></th>
<th>2nd ed.</th>
<th>3rd. ed.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptors</td>
<td>10,315</td>
<td>11,323</td>
</tr>
<tr>
<td>Lead-in Terms</td>
<td>6,675</td>
<td>7,783</td>
</tr>
<tr>
<td>Total Terms</td>
<td>16,990</td>
<td>19,106</td>
</tr>
<tr>
<td>History Notes</td>
<td></td>
<td>2,900</td>
</tr>
<tr>
<td>Scope Notes (2.ed. incl. HN)</td>
<td>2,351</td>
<td>2,085</td>
</tr>
<tr>
<td>Total descriptor cross-references</td>
<td>35,108</td>
<td>39,720</td>
</tr>
</tbody>
</table>

Web: [http://etoh.niaaa.nih.gov/AODVol1/Aodthome.htm](http://etoh.niaaa.nih.gov/AODVol1/Aodthome.htm)

To order:
   CSR Inc
   1400 Eye St, NW, Ste 200
   Washington, DC 20005
   tel. (202) 842-7600
   US$ 100 plus shipping
Art and Architecture Thesaurus


US$ 375

Number of:

<table>
<thead>
<tr>
<th>Descriptor Type</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptors</td>
<td>24,500</td>
</tr>
<tr>
<td>Guide Terms</td>
<td>2,750</td>
</tr>
<tr>
<td>Synonyms</td>
<td>20,000</td>
</tr>
<tr>
<td><strong>Total Terms</strong></td>
<td><strong>47,000</strong></td>
</tr>
<tr>
<td>British Variants</td>
<td>2,000</td>
</tr>
<tr>
<td>Alternate Terms</td>
<td>16,000</td>
</tr>
<tr>
<td><em>(singular/plural)</em></td>
<td></td>
</tr>
<tr>
<td>Permutations</td>
<td>27,000</td>
</tr>
<tr>
<td><strong>Terms and Variants</strong></td>
<td><strong>92,000</strong></td>
</tr>
</tbody>
</table>

*Electronic Editions:*

See [http://www.getty.edu/research/tools/vocabulary/obtain.html](http://www.getty.edu/research/tools/vocabulary/obtain.html)

*On the Web*


Getty Vocabulary Program home page (copy included in this package)


*Review article*


A few sample pages from an expanded version of this article are included here.

1 Introduction: Thesauri in information retrieval

What is a thesaurus and what is its purpose? Describing the functions of a thesaurus in a nutshell will provide the background for a critical examination of the AAT. A thesaurus is a structured collection of concepts and terms for the purpose of improving the retrieval of information. A thesaurus should help the searcher to find good search terms, whether they be descriptors from a controlled vocabulary or the manifold terms needed for a comprehensive free-text search — all the various terms that are used in texts to express the search concept. Most thesauri establish a controlled vocabulary, a standardized terminology, in which each concept is represented by one term, a descriptor, that is used in indexing and can thus be used with confidence in searching; in such a system the thesaurus must support the indexer in identifying all descriptors that should be assigned to a document or other object in light of the questions that are likely to be asked. A good thesaurus provides, through its hierarchy augmented by associative relationships between concepts, a semantic road map for searchers and indexers and anybody else interested in an orderly grasp of a subject field.

A good thesaurus can be used for automatic search query expansion in two ways:

1. Synonym expansion, adding all the synonyms for a search term needed for free-text searching. For example,
   - **color proofs** add color separations
   - **barrel vaults** add cradle vaults, tunnel vaults, wagon vaults, wagonhead vaults
   - **bluish gray** add aqua gray, baby blue, blue black, blue gray, centroid color 191, light Payne's gray, pewter, powder blue, slate

2. Hierarchic expansion, adding all the narrower terms for a search term (also called inclusive searching). This is needed whether one searches with a controlled vocabulary or free-text, for example,
   - **humanities** add arts, linguistics, literature, philosophy, history, etc.
   - **gold** add electrum, chryselephantine sculpture
   - **barrel vaults** add annular vaults, half barrel vaults, rampant barrel vaults, spiral vaults
   - **saints** add hagiography, hagiographies
B  Associated concepts facet (1,018)
BM  Associated concepts (1018)

D  Physical attributes facet (890)
DC  Attributes and properties (353)
DE  Conditions and effects (46)
DG  Design elements (162)
DL  Color (329)

F  Styles and periods facet (3,382)
FL  Styles and Periods (3,382)

H  Agents facet (1,093)
HG  People (958)
HN  Organizations (135)

K  Activities facet (2,034)
KD  Disciplines (318)
KG  Functions (287)
KM  Events (177)
KQ  Physical activities (87)
KT  Processes and techniques (1,165)

M  Materials facet (2,869)
MT  Materials (2,869)

P/V  Objects facet (13,210)
PC  Object groupings and systems (202)
PE  Object genres (154)
PJ  Components (3,066)

R  Build Environment (1,943)
RD  Settlements and landscapes (241)
RG  Built complexes and districts (287)
RK  Single built works (1,185)
RM  Open spaces and site elements (230)

T  Furnishings and equipment (5,592)
TC  Furnishings (1,363)
TE  Costume (721)
TH  Tools and equipment (1,463)
TK  Weapons and ammunition (256)
TN  Measuring devices (315)
TQ  Containers (622)
TT  Sound devices (607)
TV  Recreational artifacts (183)
TX  Transportation vehicles (462)

V  Visual and verbal communication (1,853)
VC  Visual works (574)
VK Exchange media (169)
VW Information forms (1,110)

Numbers in parentheses give the number of descriptors to indicate emphasis.

Figure 1. **Top-level outline**
<table>
<thead>
<tr>
<th>Facet</th>
<th>Sample descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>physical attributes</td>
<td>quarter plate, opacity, vivid red</td>
</tr>
<tr>
<td>styles and periods</td>
<td>Rococo</td>
</tr>
<tr>
<td>agents</td>
<td>painters (artists), photographers</td>
</tr>
<tr>
<td>activities and processes</td>
<td>gilding, gelatin silver process, color photography, carving, deterioration</td>
</tr>
<tr>
<td>materials</td>
<td>color film, wood</td>
</tr>
<tr>
<td>objects</td>
<td>chairs, negatives</td>
</tr>
</tbody>
</table>

Figure 4. **Facets and sample descriptors**
VC1  <visual works>
VC2  <visual works by form>
VC34 <visual works by function>
VC70 <visual works by location or context>
VC75 <visual works by medium or technique>
VC283 photographs
VC284 <photographs by form>
VC285 negatives
VC289 <negatives by color>
VC290 black-and-white negatives
VC291 color negatives
VC292 <negatives by process>
VC295 gelatin silver negatives
VC299 positives
VC310 photographic prints
VC312 later prints
VC315 <photographic prints by color>
VC316 black and white prints (photographs)
VC317 color prints (photographs)
VC318 <photographic prints by process>
VC322 chromogenic color print
VC346 <photographs by form: color>
VC347 black-and-white photographs
VC348 color photographs
VC349 <photographs by form: format>
VC357 slides (photographs)
VC358 black-and-white slides
VC359 color slides
VC360 <photographs by function>
VC363 news photographs
VC364 <photographs by technique>
VC365 <photographs by picture-taking technique>
VC366 aerial photographs
VC381 <photographs by processing or presentation technique>
VC390 manipulated photographs
VC391 composite photographs
VC400 <photographs by subject type>
VC406 studio portraits

Figure 5. Example for minor facets and precombined descriptors
Photography

D Physical Attributes Facet, DC Attributes and Properties

DC111  <size: photograph formats>
DC116  quarter plate

D Physical Attributes Facet, DE Conditions and Effects

DE38  <conditions and effects: photography>
DE39  oxidative-reductive deterioration

H Agents Facet, HG People

HG299  photographers

K Activities Facet, KT Processes and Techniques

KT487  <photography and photographic processes and techniques>
KT503  photographic processes
KT526  gelatin silver process
KT567  <photographic techniques>
KT570  <picture-taking techniques>
KT571  chronophotography
KT598  <photographic processing and presentation techniques>
KT602  enlarging
KT616  reduction (photography)

M Materials Facet, MT Materials

MT1416  paper
MT1463  <paper by function>
MT1481  photographic paper
MT2364  photographic materials
MT2367  photographic film

P/V Objects Facet, TH Tools and Equipment

TH746  photographic equipment
TH747  <cameras and camera accessories>
TH788  <photographic processing equipment>
TH794  enlargers
        [no reducers]
P/V Objects Facet, VC Visual Works

VC283    photographs
VC284    <photographs by form>
VC285    negatives
VC292    <negatives by process>
VC295    gelatin silver negatives
VC364    <photographs by technique>
VC364    <photographs by picture-taking technique>
VC367    chronophotographs

Figure 6. Facet arrangement dispersing concepts from same subject area.
   a. Hierarchy excerpts concerning the subject Photography
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]

didactic art
commercial art
funerary art

a. Original alphabetical sequence

Meaningful arrangement

b. Suggested meaningful sequence

Example from the Art and Architecture Thesaurus
Medical Subject Headings

Medical Subject Headings, 2005 Supplement to Index Medicus
(black and white MeSH)
Alphabetical and categorized list of subject descriptors used to analyze the biomedical literature in NLM.
$66.00. GPO S/N: 917-033-00000-5

MeSH Files Available to Download: www.nlm.nih.gov/mesh/filelist.html

MeSH on the Web
www.nlm.nih.gov/mesh/MBrowser.html (more powerful)

Used in searching the bibliographic database Medline through PubMed

PubMed uses MeSH and UMLS for synonym expansion and the MeSH hierarchy for hierarchic expansion

For UMLS also Internet Grateful Med: http://igm.nlm.nih.gov/

Unified Medical Language System (UMLS)
## Structure of the UMLS Metathesaurus


<table>
<thead>
<tr>
<th>Strings</th>
<th>Terms</th>
<th>Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,593,730 (1,718,083 tokens)</td>
<td>1,338,650</td>
<td>730,155</td>
</tr>
</tbody>
</table>

| Substance Dependence          | Substance Dependence         | Substance Dependence |
| Substance dependence          | Substance dependence         | Substance dependence |
| Substance dependence          | Substance dependence         | Substance dependence |

| Addiction, chemical addiction | Addiction, chemical          | Addiction, chemical |
| Addiction, chemical addiction | Addiction, chemical          | Addiction, chemical |
| chemical addiction            | Addiction, chemical          | Addiction, chemical |
| chemical addictions           | Addiction, chemical          | Addiction, chemical |

| adolescent                    | adolescent                   | adolescent |
| adolescent                    | adolescent                   | adolescent |
| Teenager                      | Teenager                     | Teenager   |
| Teenagers                     | Teenager                     | Teenager   |
| teenager                      | teenager                     | teenager   |

| teen                           | teen                         | teen        |
| teens                          | teens                         | teens       |

| youth (young person)           | youth (young person)         | youth (young person) |
| youth                          | youth                         | youth         |
| youth <1>                      | youth <1>                    | youth <1>     |

| youth <2>                      | youth <2>                    | youth <2>     |
| youth (stage of life)          | youth (stage of life)        | youth (stage of life) |
UMLS semantic types

https://umlsks.nlm.nih.gov/KSS/00/Specialist/Semantic_Net/semtype.list.html
Last Modified: Monday, February 07, 2000, copied May 31, 2000

Entity
  Conceptual Entity
    Idea or Concept
    Functional Concept
    Body System
    Temporal Concept
    Qualitative Concept
    Quantitative Concept
    Spatial Concept
      Body Location or Region
      Body Space or Junction
      Geographic Area
    Molecular Sequence
      Amino Acid Sequence
      Carbohydrate Sequence
      Nucleotide Sequence
  Finding
    Laboratory or Test Result
    Sign or Symptom
  Organism Attribute
    Clinical Attribute
  Intellectual Product
    Classification
    Regulation or Law
  Language
  Occupation or Discipline
    Biomedical Occupation or Discipline
  Organization
    Health Care Related Organization
    Professional Society
    Self-help or Relief Organization
  Group Attribute
  Group
    Age Group
    Family Group
    Professional or Occupational Group
    Population Group
    Patient or Disabled Group
Physical Object
  Anatomical Structure
  Anatomical Abnormality
    Acquired Abnormality
    Congenital Abnormality
  Embryonic Structure
  Fully Formed Anatomical Structure
    Body Part, Organ, or Organ Component
    Cell
  Manufactured Object
    Clinical Drug
    Medical Device
    Research Device
  Organism
    Animal
      Invertebrate
      Vertebrate
      Amphibian
      Bird
      Fish
      Mammal
      Human
      Reptile
    Archaeon
    Bacterium
    Fungus
    Plant
      Alga
      Virus
    Rickettsia or Chlamydia
  Substance
    Body Substance
    Chemical
    Chemical Viewed Functionally
      Biologically Active Substance
      Neuroreactive Substance or Biogenic Amine
      Hormone
      Enzyme
      Vitamin
      Immunologic Factor
      Receptor
      Biomedical or Dental Material
      Pharmacologic Substance
Antibiotic Indicator, Reagent, or Diagnostic Aid Hazardous or Poisonous Substance Chemical Viewed Structurally Organic Chemical Amino Acid, Peptide, or Protein Carbohydrate Lipid Eicosanoid Steroid Nucleic Acid, Nucleoside, or Nucleotide Organophosphorus Compound Inorganic Chemical Element, Ion, or Isotope Food

Event Activity Behavior Social Behavior Individual Behavior Daily or Recreational Activity Occupational Activity Educational Activity Governmental or Regulatory Activity Health Care Activity Diagnostic Procedure Laboratory Procedure Therapeutic or Preventive Procedure Research Activity Molecular Biology Research Technique Machine Activity Phenomenon or Process Human-caused Phenomenon or Process Environmental Effect of Humans Injury or Poisoning Natural Phenomenon or Process Biologic Function Physiologic Function Cell Function Molecular Function Genetic Function Organ or Tissue Function Organism Function Mental Process Pathologic Function Cell or Molecular Dysfunction Disease or Syndrome Mental or Behavioral Dysfunction Neoplastic Process Experimental Model of Disease
UMLS semantic relations

Last Modified: Monday, February 07, 2000, copied May 31, 2000

associated_with

physically_related_to
branch_of
connected_to
consists_of
contains
ingredient_of
interconnects
part_of
tributary_of

spatially_related_to
adjacent_to
location_of
surrounds
traverses

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

概念ually_related_to
analyzes
assesses_effect_of
conceptual_part_of
evaluation_of
degree_of
assesses_effect_of
measurement_of
measures
diagnoses
property_of
derivative_of
developmental_form_of
method_of
issue_in

isa
Dewey Decimal Classification


Volume 2: Schedules 000 - 599, 1200 p.
Volume 3: Schedules 600 - 999, 1105 p.
Volume 4: Relative Index, 1207 p.

Cost: US$ 325, UK 220

World Wide Web:

Dewey Decimal Classification home page
www.oclc.org/oclc/fp/
www.oclc.org/dewey/products/webdewey/about.htm

Good display of top three levels
www.oclc.org/dewey/about/ddc_21_summaries.htm
www.tnrdlib.bc.ca/dewey.html
www.anthus.com/CyberDewey/CyberDewey.html

Examples of Internet Resources Classified by Dewey
www.oclc.org/dewey/worldwide/
http://link.bubl.ac.uk:80/linkbrowse
http://www.oclc.org/oclc/man/colloq/v-g

More complete listing to be on
www.clis.umd.edu/faculty/soergel/dlthestut
WordNet

WordNet Lexical Database. Version 3.0. Princeton University, Cognitive Science Laboratory, December 2006 (sample pages are from an older version)

WordNet is an online lexical database that is organized semantically rather than alphabetically.

<table>
<thead>
<tr>
<th></th>
<th>synsets (concepts)</th>
<th>word senses (terms, homonyms disambiguated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of:</td>
<td>nouns</td>
<td>117,798</td>
</tr>
<tr>
<td>(version 2.0)</td>
<td>verbs</td>
<td>11,529</td>
</tr>
<tr>
<td></td>
<td>adjectives</td>
<td>21,479</td>
</tr>
<tr>
<td></td>
<td>adverbs</td>
<td>4,481</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>155,287</td>
</tr>
</tbody>
</table>

More WordNet statistics on the back

Web home page: http://wordnet.princeton.edu/

Especially useful:

- http://wordnet.princeton.edu/obtain
- http://wordnet.princeton.edu/links
- http://www.cogsci.princeton.edu/~geo/reader (WNet as lexical aid: click on word in text)
- http://wordnet.princeton.edu/doc/man/ and the glossary link there
- www.illc.uva.nl/EuroWordNet/
- www.illc.uva.nl/EuroWordNet/docs/
  esp. D017, an overview of EuroWordNet

Best search: www.notredame.ac.jp/cgi-bin/wn.cgi
**WordNet Statistics** from http://wordnet.princeton.edu/man/wnstats.7W

**Number of words, synsets, and senses**

<table>
<thead>
<tr>
<th>POS</th>
<th>Unique Strings</th>
<th>Synsets</th>
<th>Total Word-Sense Pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>117,798</td>
<td>82,115</td>
<td>146,312</td>
</tr>
<tr>
<td>Verb</td>
<td>11,529</td>
<td>13,767</td>
<td>25,047</td>
</tr>
<tr>
<td>Adjective</td>
<td>21,479</td>
<td>18,156</td>
<td>30,002</td>
</tr>
<tr>
<td>Adverb</td>
<td>4,481</td>
<td>3,621</td>
<td>5,580</td>
</tr>
<tr>
<td>Totals</td>
<td>155,287</td>
<td>117,659</td>
<td>206,941</td>
</tr>
</tbody>
</table>

**Polysemy information**

<table>
<thead>
<tr>
<th>POS</th>
<th>Unique Strings</th>
<th>Mono- semous Words and Senses</th>
<th>Poly- semous Words</th>
<th>Poly- semous Senses</th>
<th>Avg no.of meanings, all words</th>
<th>Avg no.of meanings, polysemous words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>117,798</td>
<td>101,863</td>
<td>15,935</td>
<td>44,449</td>
<td>1.24</td>
<td>2.79</td>
</tr>
<tr>
<td>Verb</td>
<td>11,529</td>
<td>6,277</td>
<td>5,252</td>
<td>18,770</td>
<td>2.17</td>
<td>3.57</td>
</tr>
<tr>
<td>Adjective</td>
<td>21,479</td>
<td>16,503</td>
<td>4,976</td>
<td>14,399</td>
<td>1.40</td>
<td>2.71</td>
</tr>
<tr>
<td>Adverb</td>
<td>4,481</td>
<td>3,748</td>
<td>733</td>
<td>1,832</td>
<td>1.25</td>
<td>2.50</td>
</tr>
<tr>
<td>Totals</td>
<td>155,287</td>
<td>128,391</td>
<td>26,896</td>
<td>79,450</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
WordNet Lexical Database

The WordNet database is an encoding of relations between synonym groups representing nouns, verbs, adjectives, and adverbs. Read the manual page for further information.

**synonyms and immediate hypernyms**
Nouns Verbs Adjectives Adverbs

**antonyms**
Nouns Verbs Adjectives Adverbs

**familiarity and polysemy information**
Nouns Verbs Adjectives Adverbs

**recursive hypernyms**
Nouns Verbs

**immediate hyponyms**
Nouns Verbs

**hyponym tree**
Nouns Verbs

**coordinates**
Nouns Verbs

**meronyms**
Substance Part Member All Tree

**holonyms**
Part Member Substance All Tree

**entailment relations**

**verb sentence frames**

**Cause to relations**

**pertainyms**
Adjectives Adverbs

**adjective values / noun attributes**
Nouns Adjectives

**links between nous and verbs that are morphologically related**

From dduchier@csi.uottawa.ca with some adaptation (Web site no longer active)
D. Soergel  **Top level hierarchy of WordNet's main categories**
Arranged building on the structure from the WordNet literature. Categories in [] added.

<table>
<thead>
<tr>
<th>nouns</th>
<th>verbs</th>
<th>adjectives</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>thing, entity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>living thing, organism</td>
<td>verbs of bodily function</td>
<td></td>
</tr>
<tr>
<td>plant, flora</td>
<td></td>
<td></td>
</tr>
<tr>
<td>animal, fauna</td>
<td></td>
<td></td>
</tr>
<tr>
<td>person, human being and care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>non-living thing, object</td>
<td></td>
<td></td>
</tr>
<tr>
<td>natural object</td>
<td></td>
<td></td>
</tr>
<tr>
<td>body, corpus</td>
<td>contact verbs</td>
<td></td>
</tr>
<tr>
<td>artifact</td>
<td></td>
<td></td>
</tr>
<tr>
<td>substance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>food</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[other things or entities]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>group, collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>process, action, event</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>process</td>
<td>[process verbs]</td>
<td></td>
</tr>
<tr>
<td>act, action, activity</td>
<td>verbs of change</td>
<td></td>
</tr>
<tr>
<td>event, happening</td>
<td>creation verbs</td>
<td></td>
</tr>
<tr>
<td>natural phenomenon</td>
<td>motion verbs</td>
<td></td>
</tr>
<tr>
<td><strong>time and place</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>[social interaction and competition verbs]</td>
<td></td>
</tr>
<tr>
<td>place</td>
<td>verbs of social interaction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>competition verbs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>consumption verbs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>weather verbs</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>knowledge, communication, feeling</strong></td>
<td>[knowledge, communication, feeling verbs]</td>
<td>descriptive adjectives color adjectives</td>
</tr>
<tr>
<td>cognition, knowledge</td>
<td>cognition verbs</td>
<td></td>
</tr>
<tr>
<td>communication</td>
<td>perception verbs</td>
<td></td>
</tr>
<tr>
<td>feeling, emotion</td>
<td>verbs of communication</td>
<td></td>
</tr>
<tr>
<td>motive</td>
<td>emotion or psych verbs</td>
<td></td>
</tr>
<tr>
<td><strong>attributes and relations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>attribute, property</td>
<td>stative verbs</td>
<td>[stative and possession verbs]</td>
</tr>
<tr>
<td>state, condition</td>
<td>verbs of possession</td>
<td>descriptive adjectives color adjectives</td>
</tr>
<tr>
<td>shape</td>
<td></td>
<td></td>
</tr>
<tr>
<td>quantity, amount</td>
<td></td>
<td></td>
</tr>
<tr>
<td>possession</td>
<td></td>
<td></td>
</tr>
<tr>
<td>relation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>relational adjectives</td>
</tr>
<tr>
<td></td>
<td></td>
<td>reference-modifying adjectives</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(e.g., <em>former</em> president)</td>
</tr>
</tbody>
</table>
Cyc Ontology

Cyc Ontology. OpenCyc

The old Cyc Ontology is a subset of the CYC system, a multi-conceptual knowledge base and inference engine. It is produced by

Cycorp, 3721 Executive Center Dr., Austin, TX 78731

Now replaced by OpenCyc which is not accessible directly through the Web but must be downloaded

Number of: Concepts in the old Cyc Ontology guide (upper ontology) "the topmost few percent of the hierarchy in the Cyc® Knowledge Base.”

Concepts in the Cyc Knowledge Base

3000

For info look at www.cyc.com/ and www.opencyc.org/

Some Web pages printed out here

Some samples from the old CYC ontology are attached.
CYC ontology top level outline (43 classes)


Reformatted

<table>
<thead>
<tr>
<th>Fundamentals</th>
<th>Biology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top Level</td>
<td>Chemistry</td>
</tr>
<tr>
<td>Time and Dates</td>
<td>Physiology</td>
</tr>
<tr>
<td>Types of Predicates</td>
<td>General Medicine</td>
</tr>
<tr>
<td>Spatial Relations</td>
<td>Materials</td>
</tr>
<tr>
<td>Quantities</td>
<td>Waves</td>
</tr>
<tr>
<td>Mathematics</td>
<td>Devices</td>
</tr>
<tr>
<td>Contexts</td>
<td>Construction</td>
</tr>
<tr>
<td>Groups</td>
<td>Financial</td>
</tr>
<tr>
<td>&quot;Doing&quot;</td>
<td>Food</td>
</tr>
<tr>
<td>Transformations</td>
<td>Clothing</td>
</tr>
<tr>
<td>Changes Of State</td>
<td>Weather</td>
</tr>
<tr>
<td>Transfer Of Possession</td>
<td>Geography</td>
</tr>
<tr>
<td>Movement</td>
<td>Transportation</td>
</tr>
<tr>
<td>Parts of Objects</td>
<td>Information</td>
</tr>
<tr>
<td>Composition of Substances</td>
<td>Perception</td>
</tr>
<tr>
<td>Agents</td>
<td>Agreements</td>
</tr>
<tr>
<td>Organizations</td>
<td>Linguistic Terms</td>
</tr>
<tr>
<td>Actors</td>
<td>Documentation</td>
</tr>
<tr>
<td>Roles</td>
<td></td>
</tr>
<tr>
<td>Professions</td>
<td></td>
</tr>
<tr>
<td>Emotion</td>
<td></td>
</tr>
<tr>
<td>Propositional Attitudes</td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td></td>
</tr>
</tbody>
</table>

Supporting Documentation

The Syntax of CycL
The CYC® Functional Interface
Glossary of Common CYC® Terms
CYC Social Vocabulary Outline (created by DS from full file)

Some groupings, indicated by blank lines, introduced by DS (this outline would profit from better organization)

controls : <Agent> <Individual>
SocialBeing
affiliatedWith : <Agent> <Agent>
acquaintedWith : <IndividualAgent> <IndividualAgent>

Workplace
spectators : <Event> <Agent>
beneficiary : <Event> <Agent>

owns : <Agent> <SomethingExisting>
recipientOfService : <ServiceEvent> <Agent>
providerOfService : <ServiceEvent> <Agent>

socialParticipants : <SocialOccurrence> <Agent>
residesInDwelling : <Animal> <ShelterConstruction>
residesInRegion : <Animal> <GeographicalRegion>

HumanOccupationConstructResident
languageSpoken : <IntelligentAgent> <NatLanguage>
fieldsOfFormalEducation : <Person> <FieldOfStudy>
fieldsOfCompetence : <Person> <FieldOfStudy>
fieldsOfActivity : <Person> <FieldOfStudy>

representsAgentToAgent : <Agent> <Agent> <Agent>
socialClass : <Person> <SocialClass-Lifestyle>

competingAgents : <Competition> <Agent>
eventHonors : <SocialOccurrence> <Agent>
positiveVestedInterest : <Agent> <TemporalThing>
negativeVestedInterest : <Agent> <TemporalThing>

AdultFemalePerson
HumanInfant
HumanChild
HumanAdult

SportsEvent
EntertainmentPerformance
EntertainmentEvent

spouse : <Person> <Person>
acquaintances : <Person> <Agent>

SimpleContactAcquaintance
AcquaintanceAttribute
friends : <Animal> <Animal>
boss : <Person> <Person>

cohabitingFamilyMembers : <Animal> <Animal>
cohabitants : <Animal> <Animal>
likesAsFriend : <SentientAnimal> <SentientAnimal>
loves : <SentientAnimal> <Agent>

maritalStatus : <Person> <MaritalStatusOfPeople>

socialStatus : <Person> <SocialStatusAttributeType>

EducationLevelAttribute
schooling : <Person> <EducationalOrganization>

StudentStatusAttribute
educationLevel : <Person> <EducationLevelAttribute>

FieldOfStudy
ScientificFieldOfStudy
Religion

Title

CourtesyTitle
titleOfPerson-String : <Person> <CharacterString>

firstName : <Person> <HumanGivenNameString>
middleName : <Person> <HumanGivenNameString>
lastName : <Person> <HumanFamilyNameString>

ProperNameString
HumanNameString
HumanGivenNameString
HumanFamilyNameString

salutation : <Person> <CourtesyTitle>

nameOfAgent : <Agent> <ProperNameString>

ethnicity : <Person> <EthnicGroupType>
skinColor : <Person> <HumanSkinColor>

PersonalityAttribute
HumanCultureType
EthnicGroupType
Nationality
SocialOccurrence
SociabilityBasedAction
PublicEvent
SocialGathering
MeetingTakingPlace
Transaction
Party-Celebration
SocialRitual
Ritual

RudeAction
HostileSocialAction

TransferringPossession
GreetingSomeone
MeetingSomeone
VisitingSomeone

Competition
AthleticActivity
Bartering
MakingSomethingAvailable
AppropriatingSomething
ObtainingPermission

CommercialActivity
Advertising
Negotiating
BusinessRelationshipActivity

PhysicallyAttackingAnAgent
Battle
WagingWar
DisputeEvent
Trial
$controls : <#$Agent> <#$Individual>

($controls X Y) represents that assertion that agent X controls the object Y, in one of the following 2 senses: X can influence (prohibit, enable or constrain) the behavior of Y; or else X can at least influence (prohibit, enable or constrain) the behavior of other #$Agents in/concerning Y. For example, Fred may control his horse directly, forcing it to do things, or not do them; and he also could control the horse indirectly, by deciding who else has access to and use of that horse. Control of one agent over another agent is rarely total, of course, so this predicate is most likely to apply to a Y which is a non-living possession, and/or to apply in a very narrow context. X's control over Y is usually either actual (de facto) control or legal (de jure) control. It is usually #$cotemporal, meaning that some time slice of X controls the same temporal time slice of Y.

isa: #$BinaryPredicate #$CotemporalObjectsSlot
genlPreds: #$positiveVestedInterest #$cotemporal

some more specialized predicates: (1 addtl more specialized public predicate, 4 unpubl. ones)

#SocialBeing

The collection of beings whose existence is accepted by some social system. (Thus, the elements of #$SocialBeing will vary with social contexts.) Social beings are entities able to perform social roles in the system that recognizes them. #$SocialBeing includes elements of #$Organization (e.g., the #$QueensGuard) as well as the elements of #$LegalAgent (in that system), so, for example, in modern industrial social systems, the elements of #$LegalCorporation and #$Person are instances of #$SocialBeing.

isa: #$ExistingObjectType
genls: #$IntelligentAgent

some subsets: #$JudicialAgent #$MedicalCareProvider #$Family-SocialEntity #$LegalAgent #$Organization #$Court-Judicial #$MedicalCareProfessional #$MedicalCareOrganization #$GeopoliticalEntity #$SoleProprietorship #$Partnership #$LegalCorporation #$LegalGovernmentOrganization #$Person #$ManufacturingOrganization (plus 157 more public subsets, 1992 unpublished subsets)

#affiliatedWith : <#$Agent> <#$Agent>

...
The collection of places where people customarily work (not the employing organizations).
#$Workplace includes offices, restaurant buildings, construction sites, agricultural sites, the
#$SpaceNeedle, etc. Some places may be #$Workplaces only during a small part of their existence
(a piece of residential property while the house is being built, perhaps); some may almost always be
#$Workplaces (grocery store buildings, office buildings, smithies, hospitals, etc.).
isa: #$ContactLocationType
genls: #$HumanlyOccupiedSpatialObject #$PhysicalContactLocation
some subsets: (10 unpublished subsets)

#$AdultFemalePerson
The collection of all women; i.e., #$Persons who are adult and female
isa: #$ExistingObjectType
genls: #$HumanAdult #$FemalePerson

#$HumanInfant
The collection of #$Persons in the infant stage of life. Functionally, this ends when the infant learns
to walk (even just toddle) and/or talk (even a few words)... or, at latest, when the person's age
greatly exceeds that at which most people develop those skills. Generally, this means that it spans
the period from birth to about 12 - 18 months old. One of the subsets of this collection is
#$NewbornBaby.
isa: #$ExistingObjectType #$TemporalObjectType
genls: #$HumanChild
some subsets: (3 unpublished subsets)

#$HumanChild
The collection of all #$Persons in the childhood stage of life. Functionally, this ends when the child
begins to take responsibility for themselves, work, have children of their own,... or, at latest, when
the person's age greatly exceeds that at which most people reach those milestones. Generally, this
means that it spans the period from birth to teenage years. This is highly dependent on context, of
course; childhood in Shakespeare's culture ended around age 12.
isa: #$ExistingObjectType #$TemporalObjectType
genls: #$JuvenileAnimal
some subsets: (1 more public subset, 8 unpublished subsets)

#$HumanAdult
The collection of human beings old enough to participate as independent, mature members of
society. Since different societies have different age or maturity requirements for people to be
considered adults, different axioms in various society-specific microtheories express these
requirements. For most modern, Western, middle-class,... purposes, e.g., the current view is that
anyone over 18 is an adult. In many cultures, adulthood occurs when one reaches puberty.
Adulthood is #$contiguousAfter childhood; that is, a #$Person is a #$HumanChild for a while, and
then is a #$HumanAdult.
isa: #$ExistingObjectType #$TemporalObjectType
genls: #$AdultAnimal #$Person
some subsets: #$AdultFemalePerson (plus 16 unpublished subsets)
Additional schemes

**Bloom**  
Taxonomy of educational objectives 1956 (1 copy in the cataloging laboratory) (LB17.B55.1956), a summary at  
http://www.unesco.org/webworld/ramp/html/r8810e/r8810e0e.htm  
http://websites.ntl.com/~james.atherton/learning/bloomtax.htm,  
http://sweep.riv.csu.edu.au/td/bloom.html,  
http://faculty.washington.edu/~krummeguides/bloom.html

**SOC**  
Standard Occupational Classification 2000  
Bureau of Labor Statistics (BLS) + other agencies  
http://stats.bls.gov/soc/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/contstan.html

**ERIC**  
Education Resources Information Center Thesaurus. 13th ed.  
http://searcheric.org/
Yahoo

The Yahoo classification. Web pages www.yahoo.com
Assignment 14

Assigned: Nov. 11
Due: Nov. 18/Dec. 2

Yahoo classification

Time: 6 hours (first half - 3.5 hours; second half - 2.5 hours)

Materials needed/available

0 The Yahoo classification itself as available on the Yahoo Web site

Materials attached in print.

1 Figures 1a and b. Yahoo Home (first summary): top level classes (called categories in Yahoo) in the original Yahoo alphabetical arrangement and in a revised meaningful arrangement.

2 Figures 2a and b. First-level breakdown of Health in the original Yahoo alphabetical arrangement and in a revised meaningful arrangement.

3 Figure 3. Second summary (first two levels of the hierarchy).

4 Figure 4. Excerpts from the Yahoo classification 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. It does illustrate patterns that are repeated throughout the classification. 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.

5 Figure 5. Yahoo top level compared with LCC and Dewey top level.

6 Figure 6. Yahoo top level compared with the subdivisions under State and City.

7 Figure 7. Subdivisions of some Yahoo categories for comparison and analysis.

The Yahoo worksheet starts on the back of this page.

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.

You can prepare your worksheet answers on a word processor, which gives you the opportunity to copy and paste examples from Yahoo.

Deliverables

The filled-in worksheet (Examine the Outline for the analysis of subject access vocabularies, which is already filled in)

The filled-in document forms and query forms
Yahoo Worksheet

Name:

60 min. 1. General layout of the classification. Formal structure

Familiarize yourself with the general layout of the classification.
- Look at the top level on the Yahoo home page and click down to Education.
- Look at Figures 1 and 2 and skim Fig. 3, the Yahoo classification second summary.
- Skim through Figure 4, the excerpt from the Yahoo classification; note the many places in which education and transportation concepts appear.

Important principle: In Yahoo, as in the Library of Congress Classification, the meaning of a category (class) is always defined by its total context. Thus in

Education
  .  K-12
  .  .  Academic Competitions
  .  .  .  Debate
  .  .  .  .  Clubs, Teams, and Societies

The last category means

Clubs, Teams, and Societies that engage in Debates staged as Academic Competitions for students in the K-12 level of Education

If you click down to this level, this category will be shown with its full caption:

Home > Education > K-12 > Academic Competitions > Debate > Clubs, Teams, and Societies

What is the degree of precombination in Yahoo?

Search for categories in Yahoo

A search for one or more words in Yahoo returns both categories whose caption contains the word(s) and individual sites. So Yahoo functions as a kind of descriptor find index. But, as a rule, retrieval of categories 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. Try a search for vehicles; do you find ships or aircraft? To look at an exception, search for driver education.

Some good searches to try to explore the structure of the Yahoo classification

<table>
<thead>
<tr>
<th>CDs, Records, and Tapes</th>
<th>Mexican Americans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transportation</td>
<td>Bridge (a homonym)</td>
</tr>
<tr>
<td>Teaching and Learning Aids</td>
<td>Humor</td>
</tr>
</tbody>
</table>

You can restrict your search to a category.
Yahoo category search (descriptor find index): Words vs concepts

<table>
<thead>
<tr>
<th>Search 1: law education Canada (implied AND)</th>
<th>Search 2: law school Canada</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yahoo! Category Matches (1 - 4 of 4)</td>
<td>Yahoo! Category Matches (1 - 2 of 2)</td>
</tr>
<tr>
<td>Regional &gt; Countries &gt; Canada &gt; Provinces and Territories &gt; Saskatchewan &gt; Cities &gt; Saskatoon &gt; Education &gt; College and University &gt; University of Saskatchewan &gt; Departments and Programs &gt; College of Law [Note: College rather than School]</td>
<td>Regional &gt; Countries &gt; Canada &gt; Provinces and Territories &gt; Ontario &gt; Counties and Regions &gt; Frontenac &gt; Cities &gt; Kingston &gt; Education &gt; College and University &gt; Queen's University &gt; Departments and Programs &gt; School of Law</td>
</tr>
<tr>
<td>Regional &gt; Countries &gt; Canada &gt; Provinces and Territories &gt; Ontario &gt; Counties and Regions &gt; Frontenac &gt; Cities &gt; Kingston &gt; Education &gt; College and University &gt; Queen's University &gt; Departments and Programs &gt; School of Law</td>
<td>Regional &gt; Countries &gt; Canada &gt; Provinces and Territories &gt; Ontario &gt; Counties and Regions &gt; Frontenac &gt; Cities &gt; Kingston &gt; Education &gt; College and University &gt; Queen's University &gt; Departments and Programs &gt; School of Law</td>
</tr>
<tr>
<td>Regional &gt; Countries &gt; Canada &gt; Provinces and Territories &gt; British Columbia &gt; Counties and Regions &gt; Capital &gt; Cities &gt; Victoria &gt; Education &gt; College and University &gt; University of Victoria &gt; Departments and Programs &gt; Law [Note: Simply a department]</td>
<td>Regional &gt; Countries &gt; Canada &gt; Government &gt; Law &gt; Law Schools</td>
</tr>
<tr>
<td>Regional &gt; Countries &gt; Canada &gt; Provinces and Territories &gt; Alberta &gt; Counties and Districts &gt; Edmonton &gt; Cities &gt; Edmonton &gt; Education &gt; College and University &gt; University of Alberta &gt; Departments and Programs &gt; Law [Note: Simply a department]</td>
<td></td>
</tr>
</tbody>
</table>

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 Home > Government > Law > Law Schools; this category has under it categories that say College of Law) and vice versa (such as Home > Government > Law > Continuing Legal Education).

Note: Searching for United States finds no categories; you need to search for US.
Explore 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 thorough multiple paths down. How does Yahoo handle this problem? Probe the following example:

Click down to

- Home > Education > K-12
- Schools

Click on Christian@

Try Home > Education > K-12 > By Region > Countries > France > Cities

Probe some on your own

Summarize your observations. What does the @ mean?

A

Explore the 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

B
Example 1

Home > 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)
  ...

Example 2

Home > 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)
The remaining questions deal with content aspects of the Yahoo classification

30 min  2. **Begin to develop a meaningful arrangement of the categories one level below Education**

You can use the meaningful arrangement of the subcategories of *Health* (Figure 2b) as a general model. Just listing some facets with sample terms under them would be an acceptable answer.

C  Put your answer on a separate sheet

15 min  3. **Compare Yahoo with Dewey and Library of Congress Classification**

Look over Figure 5 and get a sense of how different topics are treated in these classifications.

Why is *literature* 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?

D

15 min  4. **Compare Yahoo Home, State subdivision and City subdivision**

Look over Figure 6 and briefly describe the differences you see between Yahoo Home and the *State* subdivision and the *State* subdivision and the *City* subdivision

E
5. **Examine some principles Yahoo uses when designing subdivisions**

Figure 7 gives a number of examples of category subdivisions. Where there are two groups, can you tell the difference between them?

Write your observations on any two of the examples or state a general principle.
60 min 6. Overall facet analysis of the Yahoo classification

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 core classification for Yahoo (see Organizing Information, p. 299 and Section 15.6, p. 322-323).

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 extended classification.

G
# Outline for the analysis of subject access vocabularies

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 [ ] [Ch. 12, Introduction]

1.3 **Type of file and search mechanism** for which originally designed

   - Shelving [ ] Card catalog [ ] 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. (Section16.2.2). 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 indexing documents with Yahoo

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 Home):

Home > Education > K-12 > Schools > Elementary Schools

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.

Instructions for formulating queries with Yahoo

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).

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.

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.

You can formulate each query in terms of elemental concepts and than use these to find categories (see Worksheet, Task 6).
<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arts &amp; Humanities</td>
<td>Literature, Photography ...</td>
</tr>
<tr>
<td>Business &amp; Economy</td>
<td>B2B, Finance, Shopping, Jobs ...</td>
</tr>
<tr>
<td>Computers &amp; Internet</td>
<td>Internet, WWW, Software, Games ...</td>
</tr>
<tr>
<td>Education</td>
<td>College and University, K-12 ...</td>
</tr>
<tr>
<td>Entertainment</td>
<td>Cool Links, Movies, Humor, Music ...</td>
</tr>
<tr>
<td>Government</td>
<td>Elections, Military, Law, Taxes ...</td>
</tr>
<tr>
<td>Health</td>
<td>Medicine, Diseases, Drugs, Fitness ...</td>
</tr>
<tr>
<td>News &amp; Media</td>
<td>Full Coverage, Newspapers, TV...</td>
</tr>
<tr>
<td>Recreation &amp; Sports</td>
<td>Sports, Travel, Autos, Outdoors ...</td>
</tr>
<tr>
<td>Reference</td>
<td>Libraries, Dictionaries, Quotations ...</td>
</tr>
<tr>
<td>Regional</td>
<td>Countries, Regions, US States ...</td>
</tr>
<tr>
<td>Science</td>
<td>Animals, Astronomy, Engineering ...</td>
</tr>
<tr>
<td>Social Science</td>
<td>Archaeology, Economics, Languages ...</td>
</tr>
<tr>
<td>Society &amp; Culture</td>
<td>People, Environment, Religion ...</td>
</tr>
</tbody>
</table>

Figure 1a. **Yahoo classification. Home** (first summary)
## Reference and General Interest

### Reference
- Libraries, Dictionaries, Quotations ...

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

## Subjects

### Science
- Animals, Astronomy, Engineering ...

### Health
- Medicine, Diseases, Drugs, Fitness ...

### Social Science
- Archaeology, Economics, Languages ...

### Society & Culture
- People, Environment, Religion ...

### Government
- Elections, Military, Law, Taxes ...

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

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

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

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

### Education
- College and University, K-12 ...

### Arts & Humanities
- Literature, Photography ...

### Regional
- Countries, Regions, US States ...

---

Figure 1b. Yahoo Classification. Home. Meaningful arrangement
Home > Health

Categories

<table>
<thead>
<tr>
<th>Alternative Medicine (480)</th>
<th>Men's Health (30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business to Business@</td>
<td>Mental Health (682)</td>
</tr>
<tr>
<td>Chats and Forums (52)</td>
<td>Midwifery (60)</td>
</tr>
<tr>
<td>Children's Health (153)</td>
<td>News and Media (201)</td>
</tr>
<tr>
<td>Conferences (19)</td>
<td>Nursing (431)</td>
</tr>
<tr>
<td>Death and Dying@</td>
<td>Nutrition (207)</td>
</tr>
<tr>
<td>Dentistry@</td>
<td>Organizations (21)</td>
</tr>
<tr>
<td>Disabilities@</td>
<td>Pet Health@</td>
</tr>
<tr>
<td>Diseases and Conditions (7392)</td>
<td>Pharmacy (1096)</td>
</tr>
<tr>
<td>Education (39)</td>
<td>Procedures and Therapies (292)</td>
</tr>
<tr>
<td>Emergency Services (236)</td>
<td>Public Health and Safety (740)</td>
</tr>
<tr>
<td>Employment (108)</td>
<td>Reference (93)</td>
</tr>
<tr>
<td>Environmental Health (194)</td>
<td>Reproductive Health (659)</td>
</tr>
<tr>
<td>First Aid (14)</td>
<td>Senior Health (79)</td>
</tr>
<tr>
<td>Fitness (166)</td>
<td>Sexuality@</td>
</tr>
<tr>
<td>General Health (81)</td>
<td>Shopping and Services@</td>
</tr>
<tr>
<td>Health Administration (65)</td>
<td>Teen Health (13)</td>
</tr>
<tr>
<td>Health Care (356)</td>
<td>Traditional Medicine (179)</td>
</tr>
<tr>
<td>Health Sciences (26)</td>
<td>Travel Health and Medicine (16)</td>
</tr>
<tr>
<td>Hospitals and Medical Centers (38)</td>
<td>Web Directories (50)</td>
</tr>
<tr>
<td>Institutes (34)</td>
<td>Weight Issues (77)</td>
</tr>
<tr>
<td>Law@</td>
<td>Women's Health (153)</td>
</tr>
<tr>
<td>Long Term Care (116)</td>
<td>Workplace (6)</td>
</tr>
<tr>
<td>Medicine (4955)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2a. **Yahoo classification. Health.**
### Categories

<table>
<thead>
<tr>
<th>Reference</th>
<th>Health by place</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference (93)</td>
<td>General Health (81)</td>
</tr>
<tr>
<td>Web Directories (50)</td>
<td>Public Health and Safety (740)</td>
</tr>
<tr>
<td>News and Media (201)</td>
<td>Environmental Health (194)</td>
</tr>
<tr>
<td>Chats and Forums (52)</td>
<td>Workplace (6)</td>
</tr>
<tr>
<td></td>
<td>Travel Health and Medicine (16)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Health Sciences Fields</th>
<th>Health by population group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health Sciences (26)</td>
<td>Children's Health (153)</td>
</tr>
<tr>
<td>Medicine (4955)</td>
<td>Teen Health (13)</td>
</tr>
<tr>
<td>Dentistry@</td>
<td>Women's Health (153)</td>
</tr>
<tr>
<td>Nursing (431)</td>
<td>Men's Health (30)</td>
</tr>
<tr>
<td>Midwifery (60)</td>
<td>Senior Health (79)</td>
</tr>
<tr>
<td>Pharmacy (1096)</td>
<td>Pet Health@</td>
</tr>
<tr>
<td>Traditional Medicine (179)</td>
<td></td>
</tr>
<tr>
<td>Alternative Medicine (480)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Individual health condition</th>
<th>Health Care (356)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diseases and Conditions (7392)</td>
<td>Emergency Services (236)</td>
</tr>
<tr>
<td>Disabilities@</td>
<td>First Aid (14)</td>
</tr>
<tr>
<td>Fitness (166)</td>
<td>Long Term Care (116)</td>
</tr>
<tr>
<td>Nutrition (207)</td>
<td></td>
</tr>
<tr>
<td>Weight Issues (77)</td>
<td></td>
</tr>
<tr>
<td>Reproductive Health (659)</td>
<td></td>
</tr>
<tr>
<td>Sexuality@</td>
<td></td>
</tr>
<tr>
<td>Death and Dying@</td>
<td></td>
</tr>
<tr>
<td>Mental Health (682)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Procedures and Therapies (292)</th>
<th>Health care organization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hospitals and Medical Centers (38)</td>
</tr>
<tr>
<td></td>
<td>Institutes (34)</td>
</tr>
<tr>
<td></td>
<td>Organizations (21)</td>
</tr>
<tr>
<td></td>
<td>Conferences (19)</td>
</tr>
<tr>
<td></td>
<td>Health Administration (65)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shopping and Services@</td>
</tr>
<tr>
<td></td>
<td>Business to Business@</td>
</tr>
<tr>
<td></td>
<td>Law@</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Education (39)</td>
</tr>
<tr>
<td></td>
<td>Employment (108)</td>
</tr>
</tbody>
</table>

---

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
  - 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)
- 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@
- 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)
<table>
<thead>
<tr>
<th>News &amp; Media</th>
<th>Recreation &amp; Sports</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN Full Coverage, Newspapers, TV</td>
<td>SN Sports, Travel, Autos, Outdoors</td>
<td>SN Libraries, Dictionaries, Quotations</td>
</tr>
<tr>
<td>. By Region (21280)</td>
<td>. Amusement and Theme Parks@</td>
<td>. Acronyms and Abbreviations (25)</td>
</tr>
<tr>
<td>. Columns and Columnists (276)</td>
<td>. Automotive (5657)</td>
<td>. Almanacs (13)</td>
</tr>
<tr>
<td>. Commercial Services@</td>
<td>. Aviation (810)</td>
<td>. Arts and Humanities@</td>
</tr>
<tr>
<td>. Content Ratings@</td>
<td>. Booksellers@</td>
<td>. Ask an Expert (556)</td>
</tr>
<tr>
<td>. Industry Information (821)</td>
<td>. Chats and Forums (7)</td>
<td>. Bibliographies (7)</td>
</tr>
<tr>
<td>. Internet Broadcasts (396)</td>
<td>. Cooking@</td>
<td>. Booksellers@</td>
</tr>
<tr>
<td>. Journals (33)</td>
<td>. Dance@</td>
<td>. Calendars (81)</td>
</tr>
<tr>
<td>. Magazines (3845)</td>
<td>. Employment (7)</td>
<td>. Codes (24)</td>
</tr>
<tr>
<td>. Newspapers (8447)</td>
<td>. Events (10)</td>
<td>. Dictionaries (153)</td>
</tr>
<tr>
<td>. Photojournalism@</td>
<td>. Fitness@</td>
<td>. Directories (3)</td>
</tr>
<tr>
<td>. Radio (9428)</td>
<td>. Gambling (315)</td>
<td>. Encyclopedia (23)</td>
</tr>
<tr>
<td>. Television (15866)</td>
<td>. Games (17999)</td>
<td>. English Language Usage@</td>
</tr>
<tr>
<td>. Web Directories (99)</td>
<td>. Hobbies (3044)</td>
<td>. Environment and Nature@</td>
</tr>
<tr>
<td>. Automotive@</td>
<td>. Home and Garden (760)</td>
<td>. Etiquette@</td>
</tr>
<tr>
<td>. Business (133)</td>
<td>. Magazines (63)</td>
<td>. FAQs (18)</td>
</tr>
<tr>
<td>. College and University (1003)</td>
<td>. Motorcycles@</td>
<td>. Finance and Investment@</td>
</tr>
<tr>
<td>. Computers and Internet@</td>
<td>. Outdoors (9884)</td>
<td>. Flags (26)</td>
</tr>
<tr>
<td>. Crime@</td>
<td>. Pets@</td>
<td>. General (13)</td>
</tr>
<tr>
<td>. Cultures and Groups (15)</td>
<td>. Sports (47097)</td>
<td>. Geographic Name Servers@</td>
</tr>
<tr>
<td>. Disabilities@</td>
<td>. Television@</td>
<td>. Health@</td>
</tr>
<tr>
<td>. Education@</td>
<td>. Toys (948)</td>
<td>. Journals@</td>
</tr>
<tr>
<td>. Entertainment@</td>
<td>. Travel (112681)</td>
<td>. Libraries (4803)</td>
</tr>
<tr>
<td>. Environment and Nature@</td>
<td></td>
<td>. Maps@</td>
</tr>
<tr>
<td>. Good News (9)</td>
<td></td>
<td>. Measurements and Units@</td>
</tr>
<tr>
<td>. Government@</td>
<td></td>
<td>. Music@</td>
</tr>
<tr>
<td>. Health@</td>
<td></td>
<td>. Parliamentary Procedure (13)</td>
</tr>
<tr>
<td>. History@</td>
<td></td>
<td>. Patents@</td>
</tr>
<tr>
<td>. Home and Garden@</td>
<td></td>
<td>. Phone Numbers and Addresses</td>
</tr>
<tr>
<td>. Humor (218)</td>
<td></td>
<td>(166)</td>
</tr>
<tr>
<td>. Law@</td>
<td></td>
<td>. Postal Information (21)</td>
</tr>
<tr>
<td>. Outdoors@</td>
<td></td>
<td>. Quotations (235)</td>
</tr>
<tr>
<td>. Personalized News (14)</td>
<td></td>
<td>. Research Papers@</td>
</tr>
<tr>
<td>. Philanthropy@</td>
<td></td>
<td>. Searching the Net@</td>
</tr>
<tr>
<td>. Politics@</td>
<td></td>
<td>. Standards (75)</td>
</tr>
<tr>
<td>. Real Estate@</td>
<td></td>
<td>. Statistics (27)</td>
</tr>
<tr>
<td>. Religion@</td>
<td></td>
<td>. Thesauri (22)</td>
</tr>
<tr>
<td>. Science@</td>
<td></td>
<td>. Time@</td>
</tr>
<tr>
<td>. Sexuality@</td>
<td></td>
<td>. Web Directories (16)</td>
</tr>
<tr>
<td>. Sports@</td>
<td></td>
<td>. World Population Counts@</td>
</tr>
<tr>
<td>. Technology (69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>. Traffic and Road Conditions@</td>
<td></td>
<td></td>
</tr>
<tr>
<td>. Transportation@</td>
<td></td>
<td></td>
</tr>
<tr>
<td>. Travel@</td>
<td></td>
<td></td>
</tr>
<tr>
<td>. Weather (1088)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>. Weird News (20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>. World (71)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>. ABC News</td>
<td></td>
<td></td>
</tr>
<tr>
<td>. BBC News</td>
<td></td>
<td></td>
</tr>
<tr>
<td>. CNN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>. Fox News</td>
<td></td>
<td></td>
</tr>
<tr>
<td>. MSNBC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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)
  - Books@
  - Chats and Forums (21)
  - 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 (354)
  - 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. 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
  • Education
    • Dance
      • Nonfiction (130)
        • Genres (2194)
          • Authors (9971)
            • Literature (16329)
              • Transcontinental Railroad
  • Humanities
    • History
      • Education
        • Art History@
Figure 4. Excerpts

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)

Shopping and Services (378101)
•Driving Schools (35)

Education (378909)
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 Medicine@
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@
Traffic 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**

- 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
  - Kazakhstan
  - Korea, South
  - Kyrgyzstan
  - Kuwait
  - Lebanon
  - Macau
  - Malaysia
  - Mongolia
  - Myanmar
  - Nepal
  - Oman
  - Pakistan
  - Philippines
  - Qatar
  - Russia
  - Saudi Arabia
  - Singapore
  - Sri Lanka
  - Taiwan
  - Thailand
  - United Arab Emirates
  - 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)
... 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)
  - College Bowl (13)
  - Companies
  - Distance Learning
  - GED
  - Institutes (203)
  - Literacy (66)
  - Media Education
  - Organizations (27)
  - Publications (2)
  - Special Education (4)

- Adult & Continuing Education
  - Career Specific Training
  - Apprentice
  - 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 Programming
  - Real Estate
  - Religious
  - Security and Law Enforcement
  - 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
  - Character Education
  - Bilingual (24)
    - English as a Second Language
    - Bilingual Education
    - Chats and Forums
    - College and University Departments (84)
    - Commercial Products
    - Conferences (1)
    - IELTS (5)
    - Language Schools (143)
    - Lessons & Tutorials Online
    - 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
  - 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
  - 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
  - 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)
Figure 4. Excerpts

- Literary (52)
- Newspapers (436)
- Business Schools (8)
- Medical Schools (2)
- News wires (8)
- Web Directories (2)
- Radio Stations (13)
- Countries (3)
  - Australia (1)
  - Belgium (1)
  - Cities (2)
  - Provinces (3)
  - Complete List (1)
- Brazil (1)
- Canada (1)
- Denmark (1)
- New Zealand (1)
- Singapore (1)
- South Africa (1)
- Sweden (1)
- Switzerland (1)
- United Kingdom (1)
- Vietnam (1)
- Eastern United States (168)
- Internet Broadcasts (110)
- Western United States (110)
- Organizations (96)
- Alumnae/I Associations (10)
- Community College (20)
- Graduate Education (10)
- Transfer Student (3)
- Policy Research Centers (10)
- Seminaries (11)
- Christian (1)
- Jewish (1)
- Student Life (25)
- Books (10)
- Classifieds (10)
- Organizations (14)
- Web Directories (14)
- College Entrance (10)
- Colleges and Universities (10)
- Graduate Programs (10)

**Instructional Technology** (327)
- Conferences (23)
- Online Teaching and Learning (10)
- 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 (1)
- Projects (20)
- Higher Education (4)
- Regional (4)
- Virtual Field Trips (1)
- Research Institutes (3)
- Teacher Resources (43)
- Companies (2)
- Email (2)
- English as a 2nd Language (1)
- Mathematics (1)
- Science (1)
- Social Studies (1)
- Telementoring (12)
- Papers (2)
- Web Directories (11)
- Web Directories (8)
- Telementoring (1)
- WebQuests (18)
- Collections (8)
- Science (1)
- Social Studies (1)
- School Technology Funding (1)
- Organizations (7)
- Programs (5)
- Software Reviews (1)
- Titles (64)
- College Preparatory (6)
- Math (1)
- Reading (23)
- Reader Rabbit (7)
- Web Directories (6)

**Journals** (33)
- Educational Theory (1)
- Instructional Technology (1)
- Music (1)

**K-12** (52293)
- By Region (27417)
  - Countries (7635)
  - Regions (9)
  - U.S. States (19773)

**Academic Competitions**
- Debate (1)
- Clubs, Teams, & Societies (25)
- Lincoln-Douglas (1)
- Institutes (2)
- Policy Debate (7)
- 1998-99 Debate Topic (3)
- 1999-2000 Debate Topic (1)
- History (1)
- Mathematics (1)
- International (6)
- Science (1)
- International (4)
- Mathematics (1)

- International (6)
- Projects and Ideas (20)
- Science Project Books (1)
- Web Directories (2)
- Teams (13)
- Debate (1)
- Lincoln-Douglas (1)
- Forensics (1)
- Drama (1)
- Lesson Plan (4)
- School Departments (10)
- Lesson Plans (17)
- Theater (1)
- Schools (35)
- Departments (5)
- Usenet (2)
- Conferences (10)

**Countries** (41)
- Australia (1)
- Belgium (1)
- Brazil (1)
- Canada (1)
- Chile (1)
- China (1)
- Columbia (1)
- Costa Rica (1)
- Cyprus (1)
- Denmark (1)
- Egypt (1)
- Estonia (1)
- Ethiopia (1)
- France (1)
- Germany (1)
- Ghana (1)
- Greece (1)
- Hong Kong (1)
- India (1)
- Ireland (1)
- Israel (1)
- Italy (1)
- Japan (1)
- Korea, South (1)
- Macau (1)
- Malaysia (1)
- Malta (1)
- Mexico (1)
- Netherlands (1)
- New Zealand (1)
- ...
Education

K-12

Curriculum Standards (65)
  By Region (52)
    Canadian Provinces (8)
    U. S. States (44)
  By Subject (13)
    Agriculture@
    Business@
    Dance@
    English Language Arts@
    Family and Consumer Sciences@
    Health@
    History@
    Languages@
    Library and Information Literacy@
    Mathematics@
    Music@
    Physical Education@
    Science@
    Social Studies@
    Distance Learning@
    Home Schooling@
    Christian@
    Environment and Nature@
      Programs (18)
    Gifted Youth (65)
    Schools (14)
    Guidance Counseling (17)
    K-12 Offices (7)
    Organizations (3)

Home Schooling@
  Christian@
    Companies@
      Publishers@
    Teaching & Learning Aids
      Curriculum (17)
      Used (3)
    Distance Learning (11)
    Organizations (10)
    Usenet (2)
  Companies@
    Administrative Software@
      Books@
      Christian Home Schooling (36)
    Publishers@
    Teaching & Learning Aids
      SN This is its own category under Christian Home Schooling
      Curriculum (17)
      Used (3)
      Publishers@
      Christian (9)

Teaching & Learning Aids
  SN This is a reference to the generic Teaching & Learning Aids (see above), not restricted to K-12, much less to Home Schooling
  Conferences (2)
  Curriculum (1)
  Distance Learning (8)
  Islamic@
  Magazines (10)
  Opposing Views (1)
  Organizations (37)
  Christian@
  Personal Experience (7)
  Unschooling (6)
  Web Directories (11)
  Usenet (2)

Humanities@
  Classics@
    Organizations (15)
    Student (15)
    Teacher Resources (2)
  History@
    Courses (15)
    Curriculum Standards (5)
    Fairs and Competitions (3)
  Social Studies@
    Archaeology@
      Companies@
      Presenters (4)
    Professional Development
    Teaching & Learning Aids
    Curriculum Standards (25)
  History@
    Lesson Plans (31)
    Literature@
    Courses (4)
  Reading@
    Children's Reading Lists@
    Young Adult@
      Companies@
    Professional Development
    Teaching & Learning Aids
    Lesson Plans (8)
    Literacy (37)
    Organizations (16)
    Organizations (5)
    Whole Language (3)
    Young Adult Reading Lists@
    Teacher Resources (40)

Classroom Projects (2)
  Lesson Plans (27)
  Philosophy@
  Institutes (7)

Issues (31)
  Bullying@
  Class Size@
  Religion in Public Schools (24)
  School Prayer (16)
  School Choice@
  School Phobia@
  Shyness@
  School Violence@
  Organizations (10)
  School Shootings (80)

Mathematics@
  Academic Competitions (22)
  Courses (1)
  Statistics@
  Curriculum Standards (38)
  Exercises@
    Algebra@
    Geometry@
    Online@
      Online Cryptarithmic & Alphametic Puzzle Solvers@
      Pre-Algebra@
      Organizations (6)
      Programs (26)
      Magnet Schools@
      Summer (7)
      School Departments (8)
      Teaching (32)
      Lesson Plans (12)
      Newspapers (142)
    Individual School Papers (130)

Organizations (202)
  Administrators@
  Alternative (7)
  Alumnae/I Associations@
    Reunions@
    Alumnae/I Associations@
    Business@
  Charter School@
  Computer Clubs@
  Home Schooling@
  International Schools@
  Lesbian, Gay, and Bisexual@
  Parent@
  Reading@
    Literacy@
    School Associations (24)
    School Board Associations (13)
    School Funding@
    Social Studies@
    Violent Crime Prevention@
    Parental Involvement@
    Physical Education@
SN Following various branches of the hierarchy

. . . . . . Culture Jammers@
. . . . . . . . Anarchism@
. . . . . . . . Magazines (7)
. . . . . . . . United Kingdom@
. . . . . . . . United States@
. . . . . . . . Political Theory@
. . . . . . . . History (11)
. . . . . . . . Spanish Civil War@
. . . . . . . . Libraries (6)
. . . . . . . . Organizations@
. . . . . . . . People (28)
. . . . . . . . Situationists@
. . . . . . . . Bakunin, Mikhail Aleksandrovich (1814-1876)@
. . . . . . . . Bookchin, Murray (1921- ) (5)
. . . . . . . . Chomsky, Noam@
. . . . . . . . Durruti, Buenaventura (1896-1936)@
. . . . . . . . Godwin, William (1756-1836) (3)
. . . . . . . . Goldman, Emma (1869-1940)@
. . . . . . . . Kropotkin, Peter (1842-1921) (4)
. . . . . . . . Makhnov, Nestor (1889-1934)@
. . . . . . . . Malatesta, Errico (1853-1932) (2)
. . . . . . . . Michel, Louise (1830-1905)@
. . . . . . . . Parsons, Lucy (1853-1942) (3)
. . . . . . . . Prominent Anarchists & Left-Libertarians -
. . . . . . . . . Proudhon, Pierre-Joseph (1809-1865) (2)
. . . . . . . . Stirner, Max (1806-1856)@
. . . . . . . . Political Opinion@
. . . . . . . . Web Directories (3)
. . . . . . . . Publishers@
. . . . . . . . Anti-Television@
. . . . . . . . Cacophonists@
. . . . . . . . Guerilla Art@
. . . . . . . . Billboard Liberation (3)
. . . . . . . . Culture Jammers@
. . . . . . . . Graffiti@
. . . . . . . . Anti-Graffiti (6)
. . . . . . . . Companies@
. . . . . . . . Exhibits (77)
. . . . . . . . Magazines (11)
. . . . . . . . Virtual Graffiti (4)
. . . . . . . . Writers and Crews (38)
. . . . . . . . Postering (6)
. . . . . . . . Youth Audience Issues (10)
. . . . . . . . Television Ratings@
. . . . . . . . Organizations (9)
. . . . . . . . News and Media (84)
. . . . . . . . Books@
. . . . . . . . Retail@
. . . . . . . . Publishers@
. . . . . . . . Academic@
. . . . . . . . College & Univ. Presses@
. . . . . . . . Textbooks@
. . . . . . . . Bookstores@
. . . . . . . . Home Schooling Materials (21)
. . . . . . . . Christian (9)
. . . . . . . . Language Education@
. . . . . . . . Special Education Books (9)
. . . . . . . . Supplementary Materials (91)
. . . . . . . . Forensics and Debate (5)
. . . . . . . . Teacher Aids (12)
. . . . . . . . Textbooks (88)
. . . . . . . . Vocational Training Books
. . . . . . . . . Textbooks (43)
. . . . . . . . Higher Education (53)
. . . . . . . . K-12 (17)
. . . . . . . . Titles (98)
. . . . . . . . Textbooks (8)
. . . . . . . . College & University
. . . . . . . . . Bookstores@
. . . . . . . . Science@
. . . . . . . . Used (19)
. . . . . . . . . . . Titles@
. . . . . . . . . . . . Administration (2)
. . . . . . . . . . . . College and University (26)
. . . . . . . . . . . . Admissions (10)
. . . . . . . . . . . . Athletic Recruiting@
. . . . . . . . . . . . Financial Aid (6)
. . . . . . . . . . . . University Life (4)
. . . . . . . . . . . . Distance Learning (3)
. . . . . . . . . . . . Educating Your Child@
• 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@
<table>
<thead>
<tr>
<th>Category</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Islamic</td>
<td>. . . Islamic@</td>
</tr>
<tr>
<td>Jewish</td>
<td>. . . Jewish@</td>
</tr>
<tr>
<td>Sikh</td>
<td>. . . Sikh@</td>
</tr>
<tr>
<td>Residence Hall Associations</td>
<td>. . . Residence Hall Associations (9)</td>
</tr>
<tr>
<td>Web Directories</td>
<td>. . . Web Directories (2)</td>
</tr>
<tr>
<td>Student Affairs</td>
<td>. . . Student Affairs (15)</td>
</tr>
<tr>
<td>Web Directories</td>
<td>. . . Web Directories (1)</td>
</tr>
<tr>
<td>Trade Associations</td>
<td>. . . Trade Associations (3)</td>
</tr>
<tr>
<td>Vocational Education</td>
<td>. . . Vocational Education@</td>
</tr>
<tr>
<td>Private Industry Councils</td>
<td>. . . Private Industry Councils@</td>
</tr>
<tr>
<td>Training &amp; Enterprise Councils</td>
<td>. . . Training &amp; Enterprise Councils@</td>
</tr>
<tr>
<td>Women</td>
<td>. . . Women@</td>
</tr>
<tr>
<td>Law</td>
<td>. . . Law@</td>
</tr>
<tr>
<td>Policy</td>
<td>. . . Policy (52)</td>
</tr>
<tr>
<td>College &amp; Univ. Dpts &amp; Programs</td>
<td>. . . College &amp; Univ. Dpts &amp; Programs (16)</td>
</tr>
<tr>
<td>News and Media</td>
<td>. . . News and Media (3)</td>
</tr>
<tr>
<td>Organizations</td>
<td>. . . Organizations (6)</td>
</tr>
<tr>
<td>Research Centers</td>
<td>. . . Research Centers (23)</td>
</tr>
<tr>
<td>Higher Education</td>
<td>. . . Higher Education (6)</td>
</tr>
<tr>
<td>Programs</td>
<td>. . . Programs (309)</td>
</tr>
<tr>
<td>Co-operative Programs</td>
<td>. . . Co-operative Programs (8)</td>
</tr>
<tr>
<td>K-12</td>
<td>. . . K-12@</td>
</tr>
<tr>
<td>Environment and Nature</td>
<td>. . . Environment and Nature@</td>
</tr>
<tr>
<td>National Guard Youth Challenge</td>
<td>. . . National Guard Youth Challenge Program@</td>
</tr>
<tr>
<td>STARBASE</td>
<td>. . . STARBASE (3)</td>
</tr>
<tr>
<td>Student Exchange</td>
<td>. . . Student Exchange (4)</td>
</tr>
<tr>
<td>Study Abroad</td>
<td>. . . Study Abroad (6)</td>
</tr>
<tr>
<td>Companies</td>
<td>. . . Companies (5)</td>
</tr>
<tr>
<td>Student Exchange</td>
<td>. . . Student Exchange@</td>
</tr>
<tr>
<td>Admissions</td>
<td>. . . Admissions@</td>
</tr>
<tr>
<td>Counseling</td>
<td>. . . Counseling@</td>
</tr>
<tr>
<td>Credential Evaluation</td>
<td>. . . Credential Evaluation (8)</td>
</tr>
<tr>
<td>Financial Aid</td>
<td>. . . Financial Aid (1)</td>
</tr>
<tr>
<td>Summer Programs</td>
<td>. . . Summer Programs (56)</td>
</tr>
<tr>
<td>Learning Disabilities</td>
<td>. . . Learning Disabilities (20)</td>
</tr>
<tr>
<td>Math</td>
<td>. . . Math@</td>
</tr>
<tr>
<td>Science</td>
<td>. . . Science (22)</td>
</tr>
<tr>
<td>Model Congress</td>
<td>. . . Model Congress@</td>
</tr>
<tr>
<td>Model United Nations</td>
<td>. . . Model United Nations@</td>
</tr>
<tr>
<td>College and University</td>
<td>. . . College and University (35)</td>
</tr>
<tr>
<td>Events</td>
<td>. . . Events (12)</td>
</tr>
<tr>
<td>College and University</td>
<td>. . . College and University (2)</td>
</tr>
<tr>
<td>High School</td>
<td>. . . High School (33)</td>
</tr>
<tr>
<td>Events</td>
<td>. . . Events (16)</td>
</tr>
<tr>
<td>National Writing Project</td>
<td>. . . National Writing Project (27)</td>
</tr>
<tr>
<td>Student Exchange</td>
<td>. . . Student Exchange (78)</td>
</tr>
<tr>
<td>Study Abroad</td>
<td>. . . Study Abroad (157)</td>
</tr>
<tr>
<td>Summer Programs</td>
<td>. . . Summer Programs (18)</td>
</tr>
<tr>
<td>K-12</td>
<td>. . . K-12@</td>
</tr>
<tr>
<td>Reform</td>
<td>. . . Reform (67)</td>
</tr>
<tr>
<td>Books</td>
<td>. . . Books@</td>
</tr>
<tr>
<td>Class Size</td>
<td>. . . Class Size (6)</td>
</tr>
<tr>
<td>Organizations</td>
<td>. . . Organizations (35)</td>
</tr>
<tr>
<td>School Choice</td>
<td>. . . School Choice (9)</td>
</tr>
<tr>
<td>Special Education</td>
<td>. . . Special Education (167)</td>
</tr>
<tr>
<td>Companies</td>
<td>. . . Companies (9)</td>
</tr>
<tr>
<td>Administrative Software</td>
<td>. . . Administrative Software@</td>
</tr>
<tr>
<td>Teaching &amp; Learning Aids</td>
<td>. . . Teaching &amp; Learning Aids@</td>
</tr>
<tr>
<td>Publishers</td>
<td>. . . Publishers@</td>
</tr>
<tr>
<td>Conductive Education</td>
<td>. . . Conductive Education (19)</td>
</tr>
<tr>
<td>Organizations</td>
<td>. . . Organizations (4)</td>
</tr>
<tr>
<td>Schools and Institutes</td>
<td>. . . Schools and Institutes (14)</td>
</tr>
<tr>
<td>Employment</td>
<td>. . . Employment@</td>
</tr>
<tr>
<td>Inclusion</td>
<td>. . . Inclusion (7)</td>
</tr>
<tr>
<td>Institutes</td>
<td>. . . Institutes (29)</td>
</tr>
<tr>
<td>College and Univ. Departments</td>
<td>. . . College and Univ. Departments (69)</td>
</tr>
<tr>
<td>Conductive Education</td>
<td>. . . Conductive Education (16)</td>
</tr>
<tr>
<td>Learning Disabilities</td>
<td>. . . Learning Disabilities (30)</td>
</tr>
<tr>
<td>Attention Deficit Disorder</td>
<td>. . . Attention Deficit Disorder@</td>
</tr>
<tr>
<td>Dyslexia</td>
<td>. . . Dyslexia@</td>
</tr>
<tr>
<td>Organizations</td>
<td>. . . Organizations (12)</td>
</tr>
<tr>
<td>Shopping and Services</td>
<td>. . . Shopping and Services@</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>. . . Dysphagia@</td>
</tr>
<tr>
<td>Dyspraxia</td>
<td>. . . Dyspraxia@</td>
</tr>
<tr>
<td>Gifted &amp; Learning Disabled</td>
<td>. . . Gifted &amp; Learning Disabled (21)</td>
</tr>
<tr>
<td>Institutes</td>
<td>. . . Institutes (29)</td>
</tr>
<tr>
<td>College &amp; Univ. Departments</td>
<td>. . . College &amp; Univ. Departments (69)</td>
</tr>
<tr>
<td>Conductive Education</td>
<td>. . . Conductive Education (16)</td>
</tr>
<tr>
<td>Organizations</td>
<td>. . . Organizations (20)</td>
</tr>
<tr>
<td>K-12</td>
<td>. . . K-12@</td>
</tr>
<tr>
<td>Summer Programs</td>
<td>. . . Summer Programs (3)</td>
</tr>
<tr>
<td>U.S. Boarding Schools</td>
<td>. . . U.S. Boarding Schools (9)</td>
</tr>
<tr>
<td>Organizations</td>
<td>. . . Organizations (18)</td>
</tr>
<tr>
<td>Conductive Education</td>
<td>. . . Conductive Education (16)</td>
</tr>
<tr>
<td>Schools</td>
<td>. . . Schools (69)</td>
</tr>
<tr>
<td>Adult &amp; Continuing Education</td>
<td>. . . Adult &amp; Continuing Education (16)</td>
</tr>
<tr>
<td>Colleges and Universities</td>
<td>. . . Colleges and Universities (3)</td>
</tr>
<tr>
<td>Conductive Education</td>
<td>. . . Conductive Education (16)</td>
</tr>
<tr>
<td>Early Education</td>
<td>. . . Early Education (16)</td>
</tr>
<tr>
<td>Schools for the Blind</td>
<td>. . . Schools for the Blind (46)</td>
</tr>
<tr>
<td>Schools for the Deaf</td>
<td>. . . Schools for the Deaf (46)</td>
</tr>
<tr>
<td>Learning Disabilities</td>
<td>. . . Learning Disabilities (30)</td>
</tr>
<tr>
<td>Schools for the Blind</td>
<td>. . . Schools for the Blind (46)</td>
</tr>
<tr>
<td>Schools for the Deaf</td>
<td>. . . Schools for the Deaf (46)</td>
</tr>
<tr>
<td>Web Directories</td>
<td>. . . Web Directories (1)</td>
</tr>
<tr>
<td>Usernet</td>
<td>. . . Usernet</td>
</tr>
<tr>
<td>Standards and Testing</td>
<td>. . . Standards and Testing (63)</td>
</tr>
<tr>
<td>ACT</td>
<td>. . . ACT (1)</td>
</tr>
<tr>
<td>GED</td>
<td>. . . GED (2)</td>
</tr>
<tr>
<td>GMAT Preparation Companies</td>
<td>. . . GMAT Preparation Companies@</td>
</tr>
<tr>
<td>GRE</td>
<td>. . . GRE (3)</td>
</tr>
<tr>
<td>IELTS</td>
<td>. . . IELTS@</td>
</tr>
<tr>
<td>K-12 Curriculum Standards</td>
<td>. . . K-12 Curriculum Standards@</td>
</tr>
<tr>
<td>By Region</td>
<td>. . . By Region (55)</td>
</tr>
<tr>
<td>By Subject</td>
<td>. . . By Subject (14)</td>
</tr>
<tr>
<td>Agriculture</td>
<td>. . . Agriculture@</td>
</tr>
<tr>
<td>Arts</td>
<td>. . . Arts@</td>
</tr>
<tr>
<td>Business</td>
<td>. . . Business@</td>
</tr>
<tr>
<td>Dance</td>
<td>. . . Dance@</td>
</tr>
<tr>
<td>English Language Arts</td>
<td>. . . English Language Arts@</td>
</tr>
<tr>
<td>Health</td>
<td>. . . Health@</td>
</tr>
<tr>
<td>History</td>
<td>. . . History@</td>
</tr>
<tr>
<td>Languages</td>
<td>. . . Languages@</td>
</tr>
<tr>
<td>Library &amp; Information</td>
<td>. . . Library &amp; Information (20)</td>
</tr>
<tr>
<td>Literacy</td>
<td>. . . Literacy (91)</td>
</tr>
<tr>
<td>Mathematics</td>
<td>. . . Mathematics@</td>
</tr>
<tr>
<td>Music</td>
<td>. . . Music@</td>
</tr>
<tr>
<td>Physical Education</td>
<td>. . . Physical Education (9)</td>
</tr>
<tr>
<td>Science</td>
<td>. . . Science@</td>
</tr>
<tr>
<td>Social Studies</td>
<td>. . . Social Studies (5)</td>
</tr>
<tr>
<td>LSAT</td>
<td>. . . LSAT (4)</td>
</tr>
<tr>
<td>MCAT</td>
<td>. . . MCAT (5)</td>
</tr>
<tr>
<td>Companies</td>
<td>. . . Companies@</td>
</tr>
<tr>
<td>Occupational Standards</td>
<td>. . . Occupational Standards (9)</td>
</tr>
<tr>
<td>Australian Qualific Framework</td>
<td>. . . Australian Qualific Framework (9)</td>
</tr>
<tr>
<td>U.K. Natl Vocational Qualific.</td>
<td>. . . U.K. Natl Vocational Qualific. (5)</td>
</tr>
<tr>
<td>U.S. National Skill Standards</td>
<td>. . . U.S. National Skill Standards (5)</td>
</tr>
<tr>
<td>SSAT</td>
<td>. . . SSAT (1)</td>
</tr>
<tr>
<td>Test Preparation Companies</td>
<td>. . . Test Preparation Companies@</td>
</tr>
<tr>
<td>Testing Companies</td>
<td>. . . Testing Companies@</td>
</tr>
<tr>
<td>Computer-Based Testing Serv.</td>
<td>. . . Computer-Based Testing Serv.</td>
</tr>
<tr>
<td>Software</td>
<td>. . . Software@</td>
</tr>
<tr>
<td>TOEFL</td>
<td>. . . TOEFL@</td>
</tr>
<tr>
<td>TOEIC</td>
<td>. . . TOEIC@</td>
</tr>
<tr>
<td>Voluntary National Testing</td>
<td>. . . Voluntary National Testing (4)</td>
</tr>
<tr>
<td>Statistics</td>
<td>. . . Statistics (6)</td>
</tr>
<tr>
<td>Teaching</td>
<td>. . . Teaching (91)</td>
</tr>
<tr>
<td>Arts</td>
<td>. . . Arts@</td>
</tr>
<tr>
<td>Curriculum</td>
<td>. . . Curriculum (9)</td>
</tr>
<tr>
<td>Companies</td>
<td>. . . Companies@</td>
</tr>
<tr>
<td>By Subject</td>
<td>. . . By Subject (9)</td>
</tr>
<tr>
<td>Art</td>
<td>. . . Art@</td>
</tr>
<tr>
<td>Computers and Technology</td>
<td>. . . Computers and Technology (9)</td>
</tr>
<tr>
<td>Health and Fitness</td>
<td>. . . Health and Fitness (9)</td>
</tr>
<tr>
<td>Languages</td>
<td>. . . Languages@</td>
</tr>
<tr>
<td>Life Skills</td>
<td>. . . Life Skills (9)</td>
</tr>
<tr>
<td>Mathematics</td>
<td>. . . Mathematics (9)</td>
</tr>
<tr>
<td>Reading and Writing</td>
<td>. . . Reading and Writing (9)</td>
</tr>
<tr>
<td>Science</td>
<td>. . . Science (9)</td>
</tr>
<tr>
<td>Social Studies</td>
<td>. . . Social Studies (9)</td>
</tr>
</tbody>
</table>
• **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 MOO's (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
    . . . 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@
              . 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@
News and Media
SN Full Coverage, Newspapers, TV
Industry Information
. •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@
. Bookstores@
. 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)
. Hearse (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

State Web Directory
. . Area Guides
. . Arts & Humanities
**Yahoo classification**

**Figure 4. Excerpts**

- 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@
  - 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)

**Countries**

- **Washington, D.C.**
  - Education
    - Adult, Career, & 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)
    - Distance Learning@
    - Environment and Nature@
    - Gifted Youth (1)
    - Home Schooling@
    - Mathematics@
    - Newspapers@
    - Organizations (20)
    - Physical Involvement@
    - 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)
  - **New York**
    - Cities
      - **New York**
        - Business and Shopping
        - Shopping and Services
        - •Automotive
        - •Driving Schools ***
        - •Education
        - Adult, Career, & Continuing Education (22)
        - 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)
        - Organizations (47)
        - Primary and Secondary (1)
        - Programs (7)
        - Shopping and Services@
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)
      •Education
    - German®
      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
Figures 5 and 6

Yahoo top level comparisons
<table>
<thead>
<tr>
<th>Dewey Decimal</th>
<th>Library of Congress</th>
<th>Yahoo</th>
</tr>
</thead>
<tbody>
<tr>
<td>000 Computers, information, &amp; general reference</td>
<td>A General works</td>
<td>Computers &amp; Internet Reference</td>
</tr>
<tr>
<td>000 Computers, Internet, &amp; systems</td>
<td>QA Math, incl. computer science</td>
<td>News &amp; Media</td>
</tr>
<tr>
<td>010 Bibliography</td>
<td>Z Bibliography and library science</td>
<td></td>
</tr>
<tr>
<td>020 Library and information sci.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>070 News media</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 Philosophy &amp; psychology</td>
<td>B-BJ Philosophy, Psychology.</td>
<td>A &amp; H &gt; Humanities &gt; Philosophy</td>
</tr>
<tr>
<td>200 Religion</td>
<td>BL-BX Religion</td>
<td>Society and Culture &gt; Religion and Spirituality</td>
</tr>
<tr>
<td>300 Social sciences</td>
<td>H Social sciences</td>
<td>Social Science</td>
</tr>
<tr>
<td>330 Econ, 380 Commerce</td>
<td>HB-HJ Economics</td>
<td>Society &amp; Culture</td>
</tr>
<tr>
<td>320 Pol. sci., 350 Pub. admin</td>
<td>J Political science</td>
<td>Business &amp; Economy</td>
</tr>
<tr>
<td>340 Law</td>
<td>K Law</td>
<td>Government</td>
</tr>
<tr>
<td>370 Education</td>
<td>L Education</td>
<td>Government &gt; Law</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Education</td>
</tr>
<tr>
<td>400 Language</td>
<td>P-PM Language</td>
<td>Social Science &gt; Linguistics and Human Languages</td>
</tr>
<tr>
<td>500 Natural sciences &amp; math.</td>
<td>Q Science</td>
<td>Science (parts of)</td>
</tr>
<tr>
<td>600 Technology (Applied sciences)</td>
<td>R Medicine</td>
<td>Health</td>
</tr>
<tr>
<td>610 Medical sciences and medicine</td>
<td>S Agriculture</td>
<td>Science &gt; Agriculture</td>
</tr>
<tr>
<td>630 Agriculture</td>
<td>T Technology</td>
<td>Science (parts of)</td>
</tr>
<tr>
<td>Most of 600</td>
<td></td>
<td></td>
</tr>
<tr>
<td>700 Arts &amp; recreation</td>
<td>M Music and books on music</td>
<td>Arts &amp; Humanities</td>
</tr>
<tr>
<td></td>
<td>N Fine arts</td>
<td>Entertainment</td>
</tr>
<tr>
<td></td>
<td>GV Recreation, leisure</td>
<td>Recreation &amp; Sports</td>
</tr>
<tr>
<td>800 Literature &amp; rhetoric</td>
<td>PN-PZ Literature</td>
<td>Arts &amp; Humanities &gt; Humanities &gt; Literature</td>
</tr>
<tr>
<td>900 Geography &amp; history</td>
<td>C-F History, G Geography</td>
<td>A &amp; H &gt; Humanities &gt; History Regional</td>
</tr>
<tr>
<td></td>
<td>U Military science</td>
<td>Government &gt; Military</td>
</tr>
<tr>
<td></td>
<td>V Naval science</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 5. 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.
<table>
<thead>
<tr>
<th>Yahoo Home</th>
<th>Subdivision of states</th>
<th>Subdivision of cities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regional</td>
<td><strong>Massachusetts Locations</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metropolitan Areas</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Counties and Regions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cities</td>
<td></td>
</tr>
<tr>
<td>(Yellow Pages)</td>
<td></td>
<td>Find Businesses</td>
</tr>
<tr>
<td>(Maps)</td>
<td></td>
<td>Yellow Pages, Maps</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Driving Dir</td>
</tr>
<tr>
<td></td>
<td></td>
<td>City Guides</td>
</tr>
<tr>
<td>(Classifieds)</td>
<td></td>
<td>Classifieds</td>
</tr>
<tr>
<td></td>
<td>Real Estate</td>
<td>Real Estate (category, general)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Local Real Estate (listings search)</td>
</tr>
<tr>
<td></td>
<td>Employment</td>
<td>Employment (category, general)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Local Jobs (job listings search)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Online Community</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computers &amp; Internet</td>
<td></td>
<td>Computers &amp; Internet</td>
</tr>
<tr>
<td>News &amp; Media</td>
<td></td>
<td>News &amp; Media</td>
</tr>
<tr>
<td>Entertainment</td>
<td></td>
<td>Entertainment &amp; Arts</td>
</tr>
<tr>
<td>Arts &amp; Humanities</td>
<td></td>
<td>(See Entertainment &amp; Arts below)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td>Education</td>
</tr>
<tr>
<td>Recreation &amp; Sports</td>
<td></td>
<td>Recreation &amp; Sports</td>
</tr>
<tr>
<td>Travel &amp; Transportation</td>
<td></td>
<td>Travel &amp; Transportation</td>
</tr>
<tr>
<td>Business &amp; Economy (Auctions) (Shopping)</td>
<td>Business &amp; Economy</td>
<td>Business &amp; Shopping</td>
</tr>
<tr>
<td>Government</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Society &amp; Culture</td>
<td>Community &amp; Culture</td>
<td>Community</td>
</tr>
<tr>
<td>Social Science</td>
<td>Social Science</td>
<td></td>
</tr>
<tr>
<td>Health</td>
<td>Health</td>
<td></td>
</tr>
<tr>
<td>Science</td>
<td>Science</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 6. Yahoo Home compared with State subdivision and City subdivision**

My arrangement (compare Figure 1b). The items in () in column 1 are not categories in Yahoo 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.
**Figure 7. Subdivisions of selected Yahoo categories for comparison and analysis**

*Arranged by ease of analysis.*

<table>
<thead>
<tr>
<th>Libraries</th>
<th>Categories (divided into two groups)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conferences (9)</td>
<td>Organizations (59)</td>
</tr>
<tr>
<td>Countries (27)</td>
<td>Professional Resources (112)</td>
</tr>
<tr>
<td>History (6)</td>
<td>Serials (6)</td>
</tr>
<tr>
<td>Librarians (29)</td>
<td>Web Directories (19)</td>
</tr>
<tr>
<td>Library and Information Science (306)</td>
<td></td>
</tr>
<tr>
<td>Academic Libraries (451)</td>
<td>Military Libraries@</td>
</tr>
<tr>
<td>Archives@</td>
<td>Music Libraries@</td>
</tr>
<tr>
<td>Arts Libraries@</td>
<td>National Libraries (33)</td>
</tr>
<tr>
<td>Business Libraries@</td>
<td>Native American Libraries@</td>
</tr>
<tr>
<td>Commercial Library Services@</td>
<td>Performing Arts Libraries@</td>
</tr>
<tr>
<td>Dance Libraries@</td>
<td>Philatelic Libraries@</td>
</tr>
<tr>
<td>Digital Libraries (74)</td>
<td>Physics Libraries@</td>
</tr>
<tr>
<td>Education Libraries@</td>
<td>Presidential Libraries@</td>
</tr>
<tr>
<td>Environmental Libraries@</td>
<td>Prison Libraries@</td>
</tr>
<tr>
<td>Government Documents@</td>
<td>Public Libraries (3494)</td>
</tr>
<tr>
<td>Health Libraries@</td>
<td>Religious Libraries@</td>
</tr>
<tr>
<td>Intellectual Property Libraries@</td>
<td>School Libraries (35)</td>
</tr>
<tr>
<td>Internet Filtering in Libraries@</td>
<td>Science Libraries@</td>
</tr>
<tr>
<td>Law Libraries@</td>
<td>Social Science Libraries@</td>
</tr>
<tr>
<td>Lesbian, Gay and Bisexual@</td>
<td>Special Collections (41)</td>
</tr>
<tr>
<td>Libraries for the Blind@</td>
<td>Sports Libraries@</td>
</tr>
<tr>
<td>Literary Libraries@</td>
<td>Theological Libraries@</td>
</tr>
<tr>
<td>Literature@</td>
<td>Transportation Libraries@</td>
</tr>
<tr>
<td>Map Libraries@</td>
<td>U.S. State Libraries (49)</td>
</tr>
<tr>
<td>Masonic Libraries@</td>
<td></td>
</tr>
<tr>
<td>Categories (divided into two groups)</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>--</td>
</tr>
<tr>
<td><strong>By Region (21222)</strong></td>
<td><strong>Magazines (3835)</strong></td>
</tr>
<tr>
<td><strong>Columns and Columnists (286)</strong></td>
<td><strong>Newspapers (8419)</strong></td>
</tr>
<tr>
<td><strong>Commercial Services@</strong></td>
<td><strong>Photojournalism@</strong></td>
</tr>
<tr>
<td><strong>Content Ratings@</strong></td>
<td><strong>Radio (9418)</strong></td>
</tr>
<tr>
<td><strong>Industry Information (824)</strong></td>
<td><strong>Television (15900)</strong></td>
</tr>
<tr>
<td><strong>Internet Broadcasts (400)</strong></td>
<td><strong>Web Directories (100)</strong></td>
</tr>
<tr>
<td><strong>Journals (33)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Arts and Humanities@</strong></td>
<td><strong>Law@</strong></td>
</tr>
<tr>
<td><strong>Automotive@</strong></td>
<td><strong>Outdoors@</strong></td>
</tr>
<tr>
<td><strong>Business (133)</strong></td>
<td><strong>Personalized News (14)</strong></td>
</tr>
<tr>
<td><strong>College and University (844)</strong></td>
<td><strong>Philanthropy@</strong></td>
</tr>
<tr>
<td><strong>Computers and Internet@</strong></td>
<td><strong>Politics@</strong></td>
</tr>
<tr>
<td><strong>Crime@</strong></td>
<td><strong>Real Estate@</strong></td>
</tr>
<tr>
<td><strong>Cultures and Groups (15)</strong></td>
<td><strong>Religion@</strong></td>
</tr>
<tr>
<td><strong>Disabilities@</strong></td>
<td><strong>Science@</strong></td>
</tr>
<tr>
<td><strong>Education@</strong></td>
<td><strong>Sexuality@</strong></td>
</tr>
<tr>
<td><strong>Entertainment@</strong></td>
<td><strong>Sports@</strong></td>
</tr>
<tr>
<td><strong>Environment and Nature@</strong></td>
<td><strong>Technology (69)</strong></td>
</tr>
<tr>
<td><strong>Good News (9)</strong></td>
<td><strong>Traffic and Road Conditions@</strong></td>
</tr>
<tr>
<td><strong>Government@</strong></td>
<td><strong>Transportation@</strong></td>
</tr>
<tr>
<td><strong>Health@</strong></td>
<td><strong>Travel@</strong></td>
</tr>
<tr>
<td><strong>History@</strong></td>
<td><strong>Weather (1087)</strong></td>
</tr>
<tr>
<td><strong>Home and Garden@</strong></td>
<td>** Weird News (20)**</td>
</tr>
<tr>
<td><strong>Humor (219)</strong></td>
<td><strong>World (72)</strong></td>
</tr>
</tbody>
</table>
### Categories (divided into two groups)

<table>
<thead>
<tr>
<th>Humor Categories</th>
<th>Humor Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archives (122)</td>
<td>Lists (65)</td>
</tr>
<tr>
<td>Chats and Forums (34)</td>
<td>News and Media@</td>
</tr>
<tr>
<td>Columns and Columnists (84)</td>
<td>Poetry@</td>
</tr>
<tr>
<td>Comedy@</td>
<td>Quizzes and Tests@</td>
</tr>
<tr>
<td>Comic Strips@</td>
<td>Quotations@</td>
</tr>
<tr>
<td>Companies@</td>
<td>Stories (21)</td>
</tr>
<tr>
<td>Jokes (371)</td>
<td>Web Directories (20)</td>
</tr>
<tr>
<td>Advertising (41)</td>
<td>Job Humor (90)</td>
</tr>
<tr>
<td>Advice (127)</td>
<td>Military (3)</td>
</tr>
<tr>
<td>Animals (176)</td>
<td>Movies and Film (62)</td>
</tr>
<tr>
<td>Begging (40)</td>
<td>Murphy's Laws (6)</td>
</tr>
<tr>
<td>Bitterness (38)</td>
<td>Music (93)</td>
</tr>
<tr>
<td>Bizarre (248)</td>
<td>Names (25)</td>
</tr>
<tr>
<td>Boredom (4)</td>
<td>Parenting (20)</td>
</tr>
<tr>
<td>Bubblewrap (4)</td>
<td>Parody (350)</td>
</tr>
<tr>
<td>Cars (40)</td>
<td>People (19)</td>
</tr>
<tr>
<td>Clean Humor (12)</td>
<td>Philosophy (18)</td>
</tr>
<tr>
<td>Codes (8)</td>
<td>Politics (16)</td>
</tr>
<tr>
<td>Computers and Internet (631)</td>
<td>Procrastination (13)</td>
</tr>
<tr>
<td>Cultures and Groups (177)</td>
<td>Rants (75)</td>
</tr>
<tr>
<td>Distorted Pictures (14)</td>
<td>Religion (216)</td>
</tr>
<tr>
<td>Drugs and Addictions (16)</td>
<td>Science (109)</td>
</tr>
<tr>
<td>Duct Tape (11)</td>
<td>Science Fiction and Fantasy (235)</td>
</tr>
<tr>
<td>Education (43)</td>
<td>Sex (130)</td>
</tr>
<tr>
<td>Fights (9)</td>
<td>Sports (11)</td>
</tr>
<tr>
<td>Food and Drink (139)</td>
<td>Stupidity (70)</td>
</tr>
<tr>
<td>Furniture (5)</td>
<td>Tasteless (370)</td>
</tr>
<tr>
<td>Gender Wars (37)</td>
<td>Traffic Cones (2)</td>
</tr>
<tr>
<td>Hair (27)</td>
<td>Useless Pages (123)</td>
</tr>
<tr>
<td>Health and Medicine (22)</td>
<td>Warning Labels (3)</td>
</tr>
<tr>
<td>Holidays and Observances (109)</td>
<td>Wedding and Marriage (13)</td>
</tr>
</tbody>
</table>
## Comics and Animation

### Categories (not divided into groups)

<table>
<thead>
<tr>
<th>Animation (1821)</th>
<th>Editorial Cartoons (110)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artists@</td>
<td>Graphic Novels (15)</td>
</tr>
<tr>
<td>Characters (163)</td>
<td>Magazines (6)</td>
</tr>
<tr>
<td>Chats and Forums (18)</td>
<td>Organizations (22)</td>
</tr>
<tr>
<td>Comic Books (1781)</td>
<td>People (6)</td>
</tr>
<tr>
<td>Comic Strips (1023)</td>
<td>Shopping and Services@</td>
</tr>
<tr>
<td>Conventions (22)</td>
<td>Small Press Comics (15)</td>
</tr>
<tr>
<td>Cultures and Groups (21)</td>
<td>Web Directories (8)</td>
</tr>
<tr>
<td>Categories (divided into two groups)</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Alternative Dispute Resolution (38)</td>
<td>Law Enforcement@</td>
</tr>
<tr>
<td>Attorneys General@</td>
<td>Law Schools (296)</td>
</tr>
<tr>
<td>Booksellers@</td>
<td>Lawyer Jokes@</td>
</tr>
<tr>
<td>Cases (287)</td>
<td>Legal Ethics (4)</td>
</tr>
<tr>
<td>Continuing Legal Education (24)</td>
<td>Legal Research (98)</td>
</tr>
<tr>
<td>Countries (43)</td>
<td>News and Media (49)</td>
</tr>
<tr>
<td>District Attorneys@</td>
<td>Organizations (340)</td>
</tr>
<tr>
<td>Employment Resources (48)</td>
<td>Self-Help (15)</td>
</tr>
<tr>
<td>Events (3)</td>
<td>Software Companies@</td>
</tr>
<tr>
<td>Firms and Services@</td>
<td>U.S. Judiciary and Supreme Court@</td>
</tr>
<tr>
<td>History (27)</td>
<td>U.S. States (50)</td>
</tr>
<tr>
<td>Journals (101)</td>
<td>Web Directories (40)</td>
</tr>
<tr>
<td>Jury Duty (15)</td>
<td></td>
</tr>
</tbody>
</table>

|行政 (9)                              | 移民和归化 (106) |
|商业 (65)                              | 原住民 (106) |
|宪法 (175)                            | 知识产权 (167) |
|消费者 (25)                           | 国际 (48) |
|刑事司法 (110)                       | 遗产和遗嘱 (10) |
|残疾人 (29)                          | 隐私 (35) |
|特定狗种 (29)                        | 财产 (70) |
|高龄 (10)                            | 性取向 (9) |
|就业 (24)                            | 税 (54) |
|娱乐 (9)                             | 技术 (104) |
|环境 (72)                            | 贸易 (104) |
|遗产和遗嘱 (10)                     | 妇女资源 (9) |
|健康 (12)                            | Usenet (9) |
**Home > Government > Politics**

**Categories** (not divided into groups, except for the fairly standard By Region)

<table>
<thead>
<tr>
<th>By Region (7647)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. Politics@</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Activism Resources (47)</th>
<th>Parties (134)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chats and Forums (33)</td>
<td>Political Consulting@</td>
</tr>
<tr>
<td>Civic Participation@</td>
<td>Political Issues (46)</td>
</tr>
<tr>
<td>Elections (3151)</td>
<td>Political Opinion (218)</td>
</tr>
<tr>
<td>General Information (3)</td>
<td>Political Science@</td>
</tr>
<tr>
<td>Humor@</td>
<td>Political Theory@</td>
</tr>
<tr>
<td>News and Media (24)</td>
<td>Regional Conflicts (33)</td>
</tr>
<tr>
<td>Organizations (78)</td>
<td>Web Directories (8)</td>
</tr>
</tbody>
</table>
Assignment 14. Yahoo classification

Answer sheet

What is the degree of precombination in Yahoo? Very high

A Every Yahoo category (precombined descriptor) has a *home place* which is chosen from its many broader terms. But it appears also at other places where it logically belongs; at these other places the category is marked with @. Clicking on a category marked with @ takes you to a different place in the hierarchy. You could say that the @ represents a cross-reference.

The linkages are by no means complete. For example, under
(1) Home > Education > K-12 > Schools > Elementary Schools

There is no reference to
(2) Home > Society and Culture > Religion and Spirituality > Faiths and Practices > Christianity > Denominations and Sects > Anglican > Education > K-12 > Elementary Schools

However, sites are double-indexed. For example,

Christ Episcopal School

Is found under (2) and also under

Home > Education > K-12 > Schools > Elementary Schools > By Region > U.S. States > Maryland > Complete List

B Under Canada there are first cities and provinces (narrower by autonomous subdivision) and then precombined descriptors, such as

... > Canada > Arts and Humanities

which are narrower by combination

Similarly, under Transportation there are categories narrower by autonomous subdivision, such as Auto-Free Transportation, Aviation, and Highways and Road, and narrower categories formed by combination, such as ... > Transportation > History

C Meaningful arrangement under Education see attached.

D DDC -LCC - Yahoo

There are quite a few similarities, even at the top level. Differences in emphasis are based on literary warrant in the (book) literature (at the time DDC and LCC were devised) and the Web and/or difference the interests of users of the library vs. a Web search engine.

E Compare Yahoo Home, State subdivision and City subdivision

Some categories are unique to state and city subdivision: Employment and Real Estate. Market Place, Yellow Pages, Maps are unique to city subdivision. State subdivision replicates most of the categories under Home, but city subdivision combines some categories (less specific) and omits others altogether (less exhaustive).
Examine some principles Yahoo uses when designing subdivisions

In Libraries, you first have a group of categories that combine a general concept with Libraries. The categories in the next group are all formed by adding a subject component. The same principle is applied in News and Media and in Humor, but not in Comics and Animation, which has categories of both kinds mixed together. In Law the principle is applied again. In Politics there are no categories that are formed by combining with a subject, even though there could be plenty (economic policy, energy policy, etc.)

Some facets discernible in the Yahoo classification (in no particular order)

This list has just a few facets with a few sample concepts in each. A bit more than you were expected to do, but only the tip of the iceberg of a real analysis

<table>
<thead>
<tr>
<th>Form of document</th>
<th>Grade level</th>
<th>Other elemental concepts repeated throughout</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web Directories</td>
<td>Early Childhood Ed.</td>
<td>Software</td>
</tr>
<tr>
<td>Chats and Forums</td>
<td>K - 12</td>
<td>History</td>
</tr>
<tr>
<td>Bibliographies</td>
<td>College and University</td>
<td>Standards</td>
</tr>
<tr>
<td>Software</td>
<td></td>
<td>Tests</td>
</tr>
<tr>
<td>Videos</td>
<td></td>
<td>Shopping ans Services</td>
</tr>
<tr>
<td>Flash cards</td>
<td></td>
<td>Accident</td>
</tr>
<tr>
<td>Games</td>
<td></td>
<td>Companies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Theory</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Method</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Employment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Models</td>
</tr>
</tbody>
</table>

Subject
- Business
- Education
- Health
- Law

Languages
- Chinese
- Japanese
- Greek

Places

Religions
- Christian
- Hindu
- Islamic
Yahoo classification. Education

Home >
Education

Categories

Browse by Region (170)
By Culture or Group (398)
By Subject (11)

<table>
<thead>
<tr>
<th>Academic Competitions (79)</th>
<th>Higher Education (16638)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult and Continuing Education (325)</td>
<td>Instructional Technology (334)</td>
</tr>
<tr>
<td>Bibliographies (4)</td>
<td>Journals (36)</td>
</tr>
<tr>
<td>Bilingual (24)</td>
<td>K-12 (53910)</td>
</tr>
<tr>
<td>Career and Vocational (236)</td>
<td>Literacy (12)</td>
</tr>
<tr>
<td>Chats and Forums (40)</td>
<td>News and Media (83)</td>
</tr>
<tr>
<td>Companies@</td>
<td>Organizations (3008)</td>
</tr>
<tr>
<td>Conferences (52)</td>
<td>Policy (52)</td>
</tr>
<tr>
<td>Correctional@</td>
<td>Programs (322)</td>
</tr>
<tr>
<td>Disabilities@</td>
<td>Reform (70)</td>
</tr>
<tr>
<td>Distance Learning (476)</td>
<td>Special Education (168)</td>
</tr>
<tr>
<td>Early Childhood Education (90)</td>
<td>Standards and Testing (63)</td>
</tr>
<tr>
<td>Employment (143)</td>
<td>Statistics (6)</td>
</tr>
<tr>
<td>Equity (27)</td>
<td>Teaching (63)</td>
</tr>
<tr>
<td>Financial Aid (395)</td>
<td>Theory and Methods (659)</td>
</tr>
<tr>
<td>Government Agencies (77)</td>
<td>Web Directories (47)</td>
</tr>
<tr>
<td>Graduation (53)</td>
<td></td>
</tr>
</tbody>
</table>


### Yahoo classification. Education. Meaningful arrangement.

**Home >**

**Education**

**Categories**

- Browse by Region (170)
  - By Culture or Group (398)
  - By Subject (11)

<table>
<thead>
<tr>
<th>Information sources</th>
<th>Educational methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliographies (4)</td>
<td>Theory and Methods (659)</td>
</tr>
<tr>
<td>Web Directories (47)</td>
<td>Teaching (63)</td>
</tr>
<tr>
<td>News and Media (83)</td>
<td>Instructional Technology (334)</td>
</tr>
<tr>
<td>Chats and Forums (40)</td>
<td>Distance Learning (476)</td>
</tr>
<tr>
<td>Conferences (52)</td>
<td>Standards and Testing (63)</td>
</tr>
<tr>
<td>Journals (36)</td>
<td>Academic Competitions (79)</td>
</tr>
<tr>
<td>Statistics (6)</td>
<td>Graduation (53)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Education by level</th>
<th>Political and economic aspects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Childhood Education (90)</td>
<td>Policy (52)</td>
</tr>
<tr>
<td>K-12 (53910)</td>
<td>Reform (70)</td>
</tr>
<tr>
<td>Higher Education (16638)</td>
<td>Equity (27)</td>
</tr>
<tr>
<td>Adult and Continuing Education (325)</td>
<td>Financial Aid (395)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Special students and subjects</th>
<th>Organizational aspects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Special Education (168)</td>
<td>Government Agencies (77)</td>
</tr>
<tr>
<td>Disabilities@</td>
<td>Organizations (3008)</td>
</tr>
<tr>
<td>Literacy (12)</td>
<td>Companies@</td>
</tr>
<tr>
<td>Bilingual (24)</td>
<td>Programs (322)</td>
</tr>
<tr>
<td>Career and Vocational (236)</td>
<td></td>
</tr>
<tr>
<td>Correctional@</td>
<td></td>
</tr>
</tbody>
</table>