Knowledge Organization Systems. Overview

Dagobert Soergel
504 North Quaker Lane
Alexandria, VA 22304
Home: (703) 823-2840 Office: (716) 645-1474 Mobile: 703-585-2840
Fax home: (703) 823-6427 Fax office (716) 645-3775
dsoergel@.umd.edu www.dsoergel.com

Knowledge Organization Systems. Overview

KEYWORDS

Knowledge Organization Systems
KOS
Knowledge Organization System functions
ontology
metadata schema
taxonomy
classification
Web directory structure
filing plan
thesaurus
dictionary
folksonomy
authority file
controlled vocabulary
gazetteer
faceted classification
subject headings

ABSTRACT

This article provides an overview of the many different types of Knowledge Organization Systems (KOS) – ontology, metadata schema, taxonomy, classification, Web directory structure, filing plan, thesaurus, dictionary, folksonomy, etc., explicates commonalities and differences, and provides a guide through the terminological wilderness. Section 1 introduces the many functions of KOS, the many benefits that can be derived from a KOS; this can be used as a rationale for investing into KOS development. Section 2 discusses the intrinsic and extrinsic characteristics of KOS, resulting in a core metadata schema that allows for the precise description and analysis of a KOS and its features; at the same time this section provides a rudimentary introduction into the structure of KOS. Section 3 provides a typology of KOS and gives for each KOS type a short definition and a pointer to a an example, sometimes augmented by a brief excerpt that illustrates KOS structure.

INTRODUCTION

Knowledge is the lifeblood of modern society, but without organization knowledge is dead. We could even say that without organization knowledge is not knowledge at all. Knowledge must be organized in order to be used, be it by people or by machines. If the Semantic Web is ever going to take off, Knowledge Organization will be the key. This will require

- the assembly and presentation of knowledge in different formats;
- the reuse of the vast capital of knowledge that resides in existing systems.

Both must be based on a solid understanding of the many different types of Knowledge Organization Systems (KOS); this article aims to give such an understanding.

Knowledge Organization Systems (KOS) cover a wide range of systems serving a wide range of purposes. They are known under names such as ontology, metadata schema, taxonomy, classification, Web directory structure, filing plan, thesaurus, dictionary, folksonomy, and more (see Table 0). Many of these terms are used in different contexts and by different communities to mean different things, and meanings overlap; to really understand an individual KOS's purpose and structure, one needs to carefully and precisely describe it. Section 2 presents a schema that outlines the categories of information needed in the precise description of a KOS – a core metadata schema for KOS. Section 3 then surveys the space of many types of KOS, applying the schema to the elucidation and prototypical description of different types of KOS – points in the multidimensional design space around which individual KOS congregate forming clouds. However, keep in mind that assigning any individual KOS to a type gives incomplete information – a full metadata profile is required.

KOS are used by people to find information and make sense of it; KOS must support people in their quest for meaning, they must present meaningful structures of concepts . KOS are also used by computer programs to reason about data; KOS must represent formal knowledge about concepts. Any meaningful discussion of KOS must start with understanding the full range of KOS functions so that KOS can be examined from the perspective of how well they serve which function (Section 1). Only by considering all possible applications can we reap the full benefit from the considerable investment that goes into developing a good KOS.

Knowledge Organization system vs. knowledge base. Prototypically, a KOS provides a framework or schema for storing and organizing data, information, knowledge about the world and about thoughts – about what is or is alleged to be, what could be or should be – for understanding, retrieval or discovery, for reasoning, and the many other purposes outlined in Section 2. So a KOS deals with concepts, categories, classes, relationships between them, and terms or other designations for these concepts and relationships. Prototypically, data, information, knowledge about the world and thought – empirically determined "facts", hypotheses, prescriptive statements – are not part of the KOS but are stored in a knowledge base or database, expressed in terms of the KOS. But things are not that simple. Empirical relationships are often useful for organizing and discovering information. For example:

- Facts of the form

  Oryza sativa (rice) <*hasPest*> Orselia oryzae (Rice Gall Midge)

  can be used to compile a list of organisms that is then used in formulating a query in a search for literature on all kind of rice pests.

- Facts of the form
  Thailand <*isPartOf*> Southeast Asia
  can be used to find all tours in Southeast Asia in December 2010 or to compute aggregate statistics on rice production in Southeast Asia.
- Prescriptive assertions of the form ($^{++}$ indicates *should*)
  (Oryza sativa, Thailand) <*isPlantedIn*>$^{++}$ June
  can be used to compile a planting schedule for a Thai farmer

Thus many KOS contain a large amount of empirical and theoretical assertions; there is no clear boundary between a KOS and a knowledge base, or between a dictionary and an encyclopedia. The information in many KOS serves a dual function: it is useful in and of itself – the KOS is an information source in its own right – and it is useful for organizing and searching for other information – the KOS is a guide to other information.

All KOS arise from the same conceptual basis, the representation of reality and thought through an entity-relationship schema. See Soergel 2004 and Soergel 2008 for an elaboration of this principle with examples.

Table 0. The many types of KOS

A	**KOS by generic function**
A1	**conceptual data schema, conceptual schema, metadata schema, data model, implementation-driven ontology** (defining entity types and relationship types)
A1.1	. entity-relationship schema
A1.2	. table schema for a relational database
A1.3	. frame structure schema
A1.4	. object class schema
A1.5	. record format
A1.6	. data dictionary
A1.7	. data definition for software
A1.7.1	. . property list
A2	**List of entity values**
A2.1	. authority list, authority file (terms or individual entities)
A2.1.1	. . controlled vocabulary
A2.1.2	. . code list
A2.1.3	. . pick list
A2.2	. index language
A2.3	. list of variables
A2.3.1	. . coding scheme
A2.4	. systematic arrangement or structure for browsing
A2.4.1	. . filing plan, file directory structure
A2.4.2	. . table of contents of a book, law, ...
A3	**knowledge base**
B	**KOS by content and structure**
B1	**KOS focusing on concepts and their words, terms, designations, and signs**
B1.1	. KOS focusing primarily on terms
B1.1.1	. . plain term list, keyword list, word list
B1.1.2	. . KOS focusing on providing synonyms
B1.1.2,1	. . . List of synonym rings (synonym sets)
B1.1.2,2	. . . General language thesaurus
B1.1	. . controlled vocabulary
B1.1.3	. . glossary
B1.1.4	. . (lexicographical) dictionary, general dictionary
B1.1.4,1	. . . spell check dictionary
B1.1.5	. . lexical database, NLP lexicon
B1.1.5,1	. . . database of verb frames
B1.1.5,2	. . . word segment list
B1.2	. KOS focusing primarily on concepts
B1.2.0	. . relationship-based KOS
B1.2.1	. . thesaurus
B1.2.2	. . KOS alphabetically arranged by terms (terms also important)
B1.2.2,1	. . . subject dictionary
B1.2.2,2	. . . subject heading list
B1.2.3	. . KOS arranged in a hierarchical format classification scheme, taxonomy, categorization scheme
B1.2.3,1	. . . subject category list

B1.2.3,2	. . . typology
B1.2.3,3	. . . subject directory structure, shelf classification
B1.2.3,4	. . . Taxonomy
B1.2.4	. . KOS arranged in a 2-D or 3-D graph
B1.2.4,1	. . . concept map, mind map
B1.2.5	. . semantic network, RDF graph, topic map
B1.2.6	. . KOS by further structural properties
B1.2.6,1	. . . enumerative versus synthetic KOS
B1.2.6,1,1	. . . . enumerative KOS, enumerative classification scheme
B1.2.6,1,2	. . . . synthetic KOS, synthetic classification scheme
B1.2.6,2	. . . faceted classification
B2	**KOS focusing on other types entities (individuals)**
	Examples
B2.1	. onomasticon
B2.2	. personal and corporate name authority file
B2.2.1	. . personal (often author) name authority file
B2.2.2	. . corporate name authority file
B2.3	. title list, title authority file (e.g., for serials)
B2.4	. biographical reference work, biographical dictionary
B2.5	. a reference work of myths arranged by underlying theme
B2.6	. gazetteer
B2.7	. event gazetteer
B2.8	. list of languages
B3	**ontology**
B3.1	. ontology by degree of formality
B3.1.1	. . lightweight ontology
B3.1.2	. . formal ontology
B3.2	. ontology by conceptual level and scope
B3.2.1	. . foundational ontology, upper-level ontology
B3.2.2	. . core ontology, reference ontology
B3.2.3	. . domain-specific ontology, domain ontology

C KOS by origin and editorial control

C1	editorially controlled KOS
C2	KOS generated automatically by text mining etc.
C3	**folksonomy**

1 The many functions of Knowledge Organization Systems

KOS are often constructed for a specific purpose but they can be useful in many ways and support many functions. To understand the full breadth of these functions is essential for planning and using KOS for greatest benefit and return on investment. Often adding a little more information to a KOS will expand its usefulness tremendously.

Table 1 gives a comprehensive list of these functions, considering support for information processing by both people and machines.

Table 1. KOS as conceptual infrastructure.
The Many Functions of Knowledge Organization Systems

Support thinking, sense-making, integration of knowledge, and the discovery of new knowledge and of gaps by people and computer programs. Support design and decision-making.

- Support individuals in thinking and sense-making by providing a conceptual structure that allows for organizing and integrating data and for the identification of gaps in the individual's knowledge.
- Support design of any kind of product, process, or service including the design of strategies and plans and software and database design, by providing design patterns. Specifically
  Provide 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.
  Assist 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.
- Support the integration of knowledge within fields and across fields, including functions such as
  - preparation of knowledge for application;
  - discovery of new relationships and generalizations and prediction of not-yet-observed phenomena (such as the prediction of not-yet-observed elements from the periodic table in chemistry);
  - formulation of new hypotheses;
  - discovery of gaps in general knowledge;
  - mapping out new research directions.
- Support reasoning and inference by people or programs (as in Artificial Intelligence (AI), especially knowledge-based systems), possibly across data from many sources.

.

Provide a semantic road map for an individual or to promote shared understanding. Improve communication generally. Support collaboration.

- Provide a semantic road map to individual fields and the relationships among fields.
- Map out a concept space, provide classification/typology, put concepts and terms into context, relate concepts to terms, and give definitions, thus provide orientation and serve as a reference tool, and possibly promote consensus on concepts and terms and the move towards a common language for a field.
- Provide shared understanding for collaboration, especially computer-supported cooperative work.
- Support agent-to-agent communication.

Support learning and assimilating information. Support e-learning. Support writing.

- Support learning through learner-appropriate conceptual frameworks to help the learner ask the right questions and to present information in a structured way.
- Support the (collaborative) development of instructional materials through conceptual frameworks.
- Support organization and discovery of all kinds of learning objects (see info. retrieval below).
- Assist readers in understanding text by giving the meaning and conceptual context of terms.
- Assist writers in producing understandable text by suggesting good terms and spell check.
- Support foreign language learning.

Support good research and practice. Support e-science.

- Assist researchers and practitioners with problem clarification.
- Provide consistent definition of variables for data collection and **compilation of statistics** to make research results comparable and cumulative (related to information analysis).

Provide classification for action and for social and political purposes; for example,

- Classification of diseases for diagnosis, treatment, and giving reason of death.
- Classification of medical procedures and concepts for insurance billing and patient records.
- Classification of commodities for customs.

Support discovery / information retrieval and analysis and filtering / routing Organizing and keeping track of goods and services for commerce (especially e-commerce).

- Provide a tool for searching and discovery, especially knowledge-based support for end-users, including
    - assistance to users in understanding the information need and formulating a good query, for example guided facet analysis and hierarchy browsing (for free-text or for controlled-vocabulary search);
    - synonym and hierarchic expansion to find terms for free-text and folksonomy search;
    - knowledge-base for improved retrieval, ranking, clustering, and "more like this" algorithms.
- Provide a tool for human indexing and support for automated indexing (see NLP below).
- Facilitate the integration of or unified access to multiple databases (syntactic and semantic interoperability).
- Support document processing after retrieval.
- Applicable to many types of objects: assertions, documents, software, job seekers and jobs, ...

Support meaningful, well-structured display of information (in information retrieval and learning).

Support processing of information after retrieval.

Provide the conceptual basis for knowledge-based systems. Artificial Intelligence (AI).

- Support automated reasoning
- Support Semantic Web applications. (Includes many functions listed elsewhere.)

Support data modeling. Data element definition, data element dictionary. May be enterprise-wide

- Organization of data in substantive databases.
- Organization of documents, providing document structure (document templates).
- Organization of metadata.
- Sharing/communicating/reusing data/knowledge across software applications and systems, for example communicating calendar/time data among scheduling software agents or process description data among different process management software packages or using a model of human anatomy for reasoning in many applications

Support syntactic & semantic interoperability and preservation of meaning across time.

- Support all functions across multiple systems, languages, cultures, historical periods.
- Support searching across multiple databases and combining and integrating the results as well as outright database integration.
- Support the analysis, comparison, and integration of other KOS.

Provide information on terms and concepts and other entities to readers, writers, and translators to assist readers in understanding text and with conceptualizing a topic and with finding the proper term. (mono-, bi-, or multilingual dictionary and thesaurus for human use). For example,

- spelling;
- term origin;
- meanings, definitions;
- usage: audience, prevalence of use;
- translation(s) into one or more other languages, with information assisting in selecting the best translation for the intended meaning.

Support natural language processing (NLP), serve as dictionary/lexical knowledge base for functions such as

- part-of-speech tagging;
- noun-phrase, verb-phrase, and full sentence parsing;
- word sense disambiguation;
- named entity identification and information extraction;
- question answering;
- summarization;
- sentiment detection;
- determining reading level or audience;
- machine translation;
- text generation;
- spell check and grammar check.

2 A metadata schema for the description of Knowledge Organization Systems

This section provides the elements necessary to understand the nature of KOS, laying the foundation for the KOS typology presented in Section 3. The design space for KOS is very large, so it is important to understand the characteristics that can be used to describe and analyze KOS. One can distinguish roughly between *intrinsic characteristics*, the dimensions that span that design space and that enter into the definition of KOS types, and *extrinsic characteristics*, such as size or publisher, that do not concern the intrinsic nature of a KOS. There are named KOS types that correspond loosely to points in the design space, but individual KOS can be anywhere. While an individual KOS can be associated with the nearest ideal type, to really know what is in a KOS requires a metadata profile in terms of the characteristics given in Table 2b.

The main content of this section resides in the tables. This short text provides a guide. The section starts with an introductory example, the *Gene Ontology (GO)* (Gene Ontology Consortium 2000, www.geneontology.org/index.shtml) (Table 2a), for the sole purpose of illustrating and preparing for the KOS metadata schema presented in Table 2b. The example gives some sample data and a metadata record describing the GO. Table 2b, *A metadata schema for the description of Knowledge Organization Systems*, is the core of the section. Further tables elaborate on specific aspects of the metadata schema – entity types a KOS may deal with, the ever more important facet principle, the types of information a KOS may provide, and some principles for arranging this information.

Table 2a. Introductory example: Gene ontology (GO)
Hierarchy excerpt, sample record, and metadata record

Gene Ontology Hierarchy Excerpt. Reformatted for ease of reading

biological_process
. i biological regulation
. . i regulation of biological process
. . . i negative regulation of biological process
. . . . i negative regulation of cellular process
. . . . . i negative regulation of cell activation
. . . . ▸i *negative regulation of cell differentiation*
. . . . i negative regulation of circadian rhythm
. . . . i negative regulation of growth
. . . i regulation of cellular process
. . . . i negative regulation of cellular process
. . . . . i negative regulation of cell cycle
. . . . ▸i *negative regulation of cell differentiation*
. . . . i regulation of cell differentiation
. . . . ▸i *negative regulation of cell differentiation*
. . . . . i positive regulation of cell differentiation
. i cellular process
. . i cellular developmental process
. . . i cell differentiation
. . . . i cardiac cell differentiation
. . . . p cell development
. . . . i cell differentiation in spinal cord
. . . . p cell fate commitment
. . . . i embryo sac central cell differentiation
. . . . r regulation of cell differentiation
. . . . ▸i *negative regulation of cell differentiation*
. r regulation of biological process
. . i negative regulation of biological process
. . . i negative regulation of cellular process
. . . ▸i *negative regulation of cell differentiation*
cellular_component
. i cell
. . p cell part
. i cell part
. . i cellular bud
. . . p cellular bud membrane]
. . . p cellular bud neck
. . . p cellular bud tip
. . i external encapsulating structure
molecular_function
. i antioxidant activity
. i enzyme regulator activity
. . i enzyme activator activity
. . . i protease activator activity

Legend

i	<*is_a*>	negative regulation of cell differentiation <*is_a*> regulation of cell differentiation
p	<*part_of*>	cellular bud membrane <*part_of*> cellular bud
r	<*regulates*>	regulation of cell differentiation <*regulates*> cell differentiation

Notes: ▸ denotes multiple occurrences of negat*ive regulation of cell differentiation* (polyhierarchy)
AMIGO displays as follows: GO:0045596 : negative regulation of cell differentiation [552 gene products]

Gene Ontology Descriptor Record. Slightly reformatted

negative regulation of cell differentiation	
Accession	GO:0045596
Ontology	biological process
Synonyms exact	down regulation of cell differentiation down-regulation of cell differentiation downregulation of cell differentiation
Synonyms narrower	inhibition of cell differentiation
Definition	Any process that stops, prevents or reduces the frequency, rate or extent of cell differentiation. [source: GOC:go_curators]
Comments	
Subset	[Subset of GO, for example descriptors especially relevant to a given organism]
Term lineage	An except from the tree that shows all the places the descriptor occurs with the chain up to the root.

A metadata record for the Gene Ontology

Intrinsic characteristic (basis for defining types) *Extrinsic characteristic*	
Formal data	**Gene ontology (GO)** Gene Ontology Consortium First established 1998 Updated continuously Information: www.geneontology.org/index.shtml Search: http://amigo.geneontology.org/cgi-bin/amigo/go.cgi Downloads: www.geneontology.org/GO.downloads.shtml
Type	Hierarchical thesaurus or slightly formal ontology
Entities on which the KOS is focused	Concepts, specifically processes (molecular function and biological process) and things (cellular components)
Domain	Biology, specifically genetics
Language(s)	English
Description *Summary evaluation*	A deep hierarchy of highly precombined descriptors (controlled vocabulary), divided into molecular function, biological process, and cellular component. The hierarchy is based on three types of relationships: *is_a*, *part_of*, and *regulates* (divided into *positively_regulates* and *negatively_regulates*). Gives for each descriptor synonyms and a definition. Accessible through multiple browsers listed on the site.
Purpose *Intended and possible uses* (see Table 1)	Intended and used for indexing (annotating) gene products (proteins and other molecules produced by a gene, also the gene itself) with respect to the molecular function(s) and biological process(es) in which they participate and the cell components in which they occur or which are affected by the processes. Used as a common language by many genome databases that store data about the biology of one or more species with emphasis on the role of their various genes.
Audience	Scientists
Size	26505 descriptors, 98.4% with definitions. as of January 2, 2009 No information on the number of synonyms
Information given. Structural characteristics	• The GO has a highly elaborated structure with high specificity of concepts and somewhat differentiated relationships. • It gives synonyms (called synonym exact); no differentiation between synonyms and spelling variants. It also gives non-descriptors related in other ways (for example narrower terms not used as descriptors, called synonym narrower) • There are three types of conceptual relationships between descriptors: *is_a*, *part_of*, and *regulates* (which is further divided into *positively_regulates* and *negatively_regulates*). • The conceptual structure is partially built on the facet principle. The three

	top-level concepts – molecular function , biological process, and cellular components – express aspects to be considered in indexing gene products. Many more facets – such as *anatomical part* (including type of cell) or *type of process or function* (such as adhesion) – are implied but not made explicit. • The descriptors are highly precombined. Compound concept are not explicitly expressed in terms of elemental concepts, but to a large degree the constituent concepts can be seen from the linguistic constituents of the terms. • Hierarchy is an essential feature of the GO. The hierarchy is based on the conceptual relationships, and the relationship on which subordination in the hierarchy is based is indicated in the hierarchy display (This is the only place where these relationships are indicated). The hierarchy is very deep (11 levels or more). It is strongly polyhierarchical. Due to the high degree of precombination, descriptors appear in many places in the hierarchy. Arrangement at each level of the hierarchy is alphabetical by descriptor rather than based on some meaningful principle. Even descriptors related to the superordinate descriptor in different ways (*is_a*, *part_of*) are intermingled in one alphabetical sequence. There is no notation.. • GO is entirely enumerative; the indexer cannot build new precombined descriptors on the fly. However, the indexer can request that a new descriptor be added, and GO is constantly updated • All descriptors (98.4%) have definitions, some have comments. • GO is divided into subsets, primarily based on what descriptors are especially relevant to a given organism; subset membership thus carries important information. The subsets a descriptor belongs to (if any) are given. • GO gives the source for definitions.
Degree of formality	Low to medium formality

Access, display, arrangement	Access and display information is for the AMIGO browser. External browsers may have other ways of access The hierarchical display gives • the descriptor • the nature of the relationship to the descriptor one level up in the hierarchy (*is_a*, *part_of*, or *regulates*) • the GO Accession Number, • the number of gene products indexed by the descriptor The record for a descriptor gives • the descriptor • the GO Accession Number • exact synonyms and spelling variants (no differentiation) • non-descriptors related in other ways (for example narrower terms not used as descriptors) • Definition • Comments • Subsets of GO One can search the GO using words or phrases; the search returns all descriptors where either the descriptor or a synonym (as broadly defined in the GO) contains the search word or phrase.
Formalism for internal representation. **Expressiveness**	OBO format flat file OWL
Availability: Media and formats	Downloads available in many different formats.
Construction method (overall or for specific pieces of information)	Not known
Origin, degree of control, editorial control	GO is maintained be the GO consortium through a group of GO curators working for consortium members using SourceForge as collaboration platform. Anybody can submit suggestions.
Scope of deployment and maintenance	GO is centrally (if collaboratively) maintained. Users can use it at the central location or download it and integrate into their local system.
Degree and scope of consensus	GO establishes a consensus view within the GO consortium and possibly beyond that in the field of genomics
Scope of application	GO is intended for wide (public) application
Relationship to other KOS	For a list of mappings to other KOS, see www.geneontology.org/GO.indices.shtml GO is part of the Open Biomedical Ontologies www.obofoundry.org
Additional information	

Table 2b. A metadata schema for the description of Knowledge Organization Systems

Intrinsic characteristic (basis for defining types) *Extrinsic characteristic*	
Formal data	• Title • Publisher • Original publication date • Revision cycle • Last updated • Informational URL • Online/Download URL
Type	From Table 0
Entities on which the KOS is focused	(concepts, terms, places, ...; see Table 2c for examples)
Domain	• Broad subject area, discipline, sub-discipline, interdisciplinary field, other context For a more precise description, give a list of subfields/topics and for each the completeness of coverage and depth of coverage (the specificity of the concepts included)
Language(s)	Also completeness of coverage of the terminology from each language
Description *Summary evaluation*	• Description is a (usually brief) text explaining the major features that are detailed below. • Summary evaluation deals with the KOS' adequacy for a given purpose based on the more detailed analysis as outlined below
Purpose *Intended and possible uses* (see Table 1)	• Purpose(s) for which the KOS was designed • Purposes for which the KOS can be used (for example, automated reasoning)
Audience	
Size	Number of • Total Terms • Top Terms • Descriptors • Non-descriptors • Total relationships (possibly differentiated by type) • Definitions
Information given. Structural characteristics	• **Overall characterization of structure** . Degree to which structure is elaborated. Weakly vs. strongly interconnected through relationships . Granularity. How specific are the concepts included? How differentiated are the relationships? • **Synonyms** (note if differentiated by type, e.g., spelling variant or acronym)

	For evaluation: How complete is the coverage of synonyms (if applicable, differentiated by language) Are all synonym relationships recognized or does the KOS contain terms that are synonymous or quasi-synonymous without indicating the relationship? • **Conceptual relationships types** For description: Types and degree of differentiation of conceptual relationships included: For evaluation: Completeness and correctness of conceptual relationships included . Hierarchical relationships . Associative relationships . Rich set of semantic relationships (towards a full-fledged ontology) • **Is the conceptual structure built on the facet principle** (see Table 2d) • **Degree of precombination** included or intended. Elemental concepts vs. compound concepts assigned in indexing/ annotating/ coding are compound concepts expressed in terms of elemental (primitive) concepts (for example, in description logic formulas)? • **Hierarchical structure** . What types of relationships considered in the hierarchy (generic, part-whole, topic inclusion, etc.)? Are they differentiated? . Depth of hierarchy (hierarchical levels) . Monohierarchy vs. polyhierarchy . . Monohierarchy (tree) . . Polyhierarchy (DAG, directed acyclical graph), degree of polyhierarchy . Meaningful arrangement given (see Display below) . Hierarchy expressed by notation (class/concept/term numbers)? (Indicate scheme or None) • **Extensibility. Enumerative vs. synthetic classification** (see Table 2d for explanation) . Not extensible, all descriptors enumerated (enumerative classification, especially if degree of precombination included is high) . Extensible, additional descriptors can be built using rules (synthetic classification, especially if the degree of precombination intended is high) • **Notes (**indicate if different types such as **definition**) For evaluation, give the quality of definitions, explications, scope notes (correctness, detail, clarity) • Source of each piece of information, mappings to other KOS (For additional examples of information that might be given in a KOS see Table 2e.)
Degree of formality	• **Very informal**, relationship types with loosely defined semantics. Intended interpretations of all terms are represented in a semiformal or informal language. • **Semiforma**l, relationship types with well-defined semantics. Intended interpretations are defined by sentences in a logical language for some terms, informally for others. • **Very formal**, axioms. Intended interpretations of all terms are defined by sentences in a logical language
Access, display,	• Only alphabetical

arrangement	• Concept relationships expressed through arrangement . Concept relationships expressed through linear arrangement . Concept relationships expressed through 2-D or 3-D arrangement (concept maps) (See Table 2f for elaboration.)
Formalism for internal representation. **Expressiveness**	• **Formalism, conceptual level**, e.g. . No special formalism . Description logic . Frames, records . Semantic network . Conceptual graphs • **Formalism, syntactic level** . Specific standard followed, e.g., . . Unified Modeling Language (UML) . . RDF . . . OWL (OWL Lite, OWL DL, OWL Full) . . Topic Map . . Other (e.g., OBO) . Tagging format followed (e.g., XML) • **Expressiveness** . Minimal expressiveness required to represent the KOS . Expressiveness of the formalism actually used to represent the KOS
Availability: Media and formats	
Construction method (overall or for specific pieces of information)	• **Construction agent** . Human editor . Automatically derived (term associations, clustering, etc.) . Combination • **Construction approach** (top-down, bottom-up, middle, merging…)
Origin, degree of control, editorial control	• Sources used in constructing the KOS • Relative contributions of . editors (lexicographers, taxonomists, ontologists), . users (folksonomies), . automated methods extracting terms, concepts, and relationships from text, user term/tag assignments, and other sources • Governance
Scope of deployment and maintenance	• Stand-alone, centrally maintained under one authority (but can be collaborative) • Federated, different components maintained by different authorities
Degree and scope of consensus	• KOS for a person or organization . Personal KOS . Organizational KOS • KOS for wide use
Scope of application	
Relationship to other KOS	(subset, mapped, ...)

Additional information	

See also Zeng 2008

Table 2c. Sample types of entities different KOS focus on

Concepts, words, terms, designations, signs	**Entities other than concepts (examples)**
. words, terms, designations, signs as focus . . Designations, signs (as focus) . . . Acronyms . . . Signs, symbols, codes . . Words, terms (as focus, as in a dictionary) . Concepts (as focus, as in a classification) Could list important types, such as . . Abstract concepts . . Chemical substances . . Biological taxa . . Diseases . . Commodities . Concepts by degree of combination . . Predominantly atomic concepts . . Includes many compound concepts	. People . Organizations . Documents (authority list of titles) . Myths . Places . Events . Languages . Religions

Table 2d. The facet principle and extensibility

The facet principle is at the core of understanding the conceptual structure of KOS and assumes increasing practical importance as more and more Web sites use faceted navigation.

A facet groups concepts that fill the same role; this basic idea can be put in many different ways: a facet groups

- concepts that fall under the same aspect or feature in the definition of more complex concepts;
- concepts that describe the same feature of a product or service;
- concepts that combine in similar patterns with other concepts.

For example, characterizing an ability of a person involves two aspects, *area of ability* and *degree of ability*, as shown in the box

A Area of ability combines with **B Degree of ability**

A1	psychomotor ability	B1	low degree of ability, disabled
A2	senses	B2	average degree of ability
A2.1	. vision	B3	above average degree of ability
A2.1.1	. . night vision	B3.1	. very high degree of ability
A2.2	. hearing		
A3	intelligence		
A4	artistic ability		

Examples	A2.1B1	visually impaired
	A2.1.1B3.1	excellent night vision
	A2.2B1	hearing impaired
	A3B1	mentally handicapped
	A3B3	intellectually gifted

For another example, consider *route of drug administration*; three aspects must be specified:
the scope of drug action (local/topical or systemic, as in local vs. general anesthesia);
the body site (mouth, arm, ...);
the method of application (ingestion, injection, rubbing on, etc.).
A combination of three concepts, one from each facet, describes a route of administration as discussed in a textbook or applied to a patient.

In the Gene Ontology example, we might have facets such as
regulation (positive or negative)
type of process or function (development, division, aging, metabolism, adhesion)
body system or anatomical part (including type of cell) involved
chemical substance involved

A KOS can give just elemental concepts arranged into facets; such a KOS is called a *faceted classification* or, more generally, a faceted KOS. Objects (documents, items for sale, gene products) can then indexed by multiple elemental descriptors, one or more from each applicable facet. Or the indexer combines elemental concepts on the fly to form one precombined descriptor that is then assigned; such systems are sometimes called synthetic, they are completely open and extensible. Note that in the ability classification example elemental concepts must be combined into precombined descriptors to avoid confusion; a person may be blind (vision : low degree) and individually gifted (intelligence : very high degree).

A KOS can use the facet principle but provide ready-made combinations of concepts (precombined descriptors, less aptly called pre-coordinate descriptors). A KOS enumerating (a usually large number of) precombined descriptors, whether explicitly faceted or not, is called enumerative. If the indexer is limited to the precombined descriptors that are enumerated, the system is not extensible. A KOS may enumerate some precombined descriptors and allow for building still more compound descriptors; this is the case for the Library of Congress Subject Headings, the Library of Congress Classification, and Dewey Decimal Classification, all systems with some facet features. Synthesis allows for a much wider range of subjects to be described than is practical to enumerate explicitly.

Faceted KOS are now increasingly used to help users formulate their queries and navigate Web sites. In this context, the notion of facet is often broadened to simply mean metadata field; for example, on a wine store site, the facets might be grape variety, color, sweetness, country of origin, and price.

For a recent discussion of the facet principle see Gnoli 2008.

Table 2e. Types of information given in a KOS. A few examples

- *Entry term, icon, concept* (or a group of terms or concepts with common characteristics).
- *Spelling variants* (other character strings in the same language).
- *Pronunciations* (with dialect/regional variations and frequency information), in a phonetic alphabet or as digitized sound (for educational and voice interface applications).
- *Word root* and derivation from the root.
- Part of speech, inflection rules, and other *syntactic information*.
- *Terminological information*: Other terms and icons with the same or similar meaning in the same language and in other (sub)languages/(sub)cultures/environments.
- *Definition and/or how-to description* (for functional concepts).
- *Usage notes*, usage examples and quotations, familiarity and frequency. Explanation of subtle differences in meaning between related terms. Hyperlinks to texts in which the term occurs.
- *Disambiguation rules*. Rules on how to determine the proper meaning of a homonym.
- *Detailed conceptual relationships* (broader terms / hypernyms, narrower terms / hyponyms, parts / meronyms, whole / holonyms) and pointers to the concept's place in overall classificatory structures and conceptual maps. Display of the structural relationships among subordinate concepts.
- *Rules on combination with other concepts to form expressions*. For verbs: *case frame*.
- *(Sub)language/(sub)culture/population group and audience level* as a tag in every slot.

Table 2f. Characteristics of KOS presentation. Access and display

Consider for each format access/retrieval by concepts versus access/retrieval by terms

Access can be provided through arrangement in a printed/displayed document or through a computerized search system.

1 **Format of printed document** (some of this also applies to Web display)

1.1 Overall format

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?

1.2 Display of conceptual relationships
- through arrangement
- through cross-references
- through descriptor-find index

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

1.3 Display of terminological relationships. Format of alphabetical index

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

2.1 Overall format. Available windows and their relationships

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.

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!)?

.

3 The landscape of Knowledge Organization Systems

There are many types of KOS as shown in Table 0. They are created and used in different communities for different purposes. Yet they overlap greatly in the types of information they provide, including information regarding the overall conceptual structure. A comprehensive multi-functional KOS database, including, for example, linguistic information for terms as found in a general dictionary and information on concepts in both informal and formal presentation would often be much more efficient. However, standards for different types of KOS have been developed separately by different communities, for example,

- Lexicography and dictionaries, for example ISO 12620;
- Semantic Web, OWL;
- Publishing industry: Topic Maps, ISO 13250 and XTM 1.0;
- Library and information services: Thesaurus standards ANSI/NISO Z39.19-2005, BS8723, MARC Format for authority data.

KOS development and maintenance software tends to follow standards, so no program to support a multi-functional comprehensive KOS exists and data exchange across different types of KOS is difficult.

Many KOS are designed for a particular function, yet the information they provide can be used for other functions as well. For example, an information retrieval thesaurus designed as a controlled vocabulary can also be used for query term expansion for free-text and folksonomy searching. Similarly, based on commonality of information, a KOS of a less-developed type, such as a common information retrieval thesaurus, can be used as a starting point for developing a more elaborate KOS, such as a full-fledged ontology with a richer set of relationships, at considerable savings.

As was mentioned in the beginning of Section 2, KOS can occupy many points in a multidimensional design space and defy easy division into types. To make matters worse, terms, especially *ontology* and *taxonomy*, are used loosely; many KOS that would be better described as *thesauri* are called *ontology* because that term is more fashionable. The bulk of this section gives brief characterizations of the usage of the many terms. The KOS types are arranged primarily by the type of entity a KOS focuses on and secondarily by the amount of information and structure provided, but even this simple classification is quite imperfect. For example, while a general dictionary focuses on terms for concepts, most dictionaries include place names and some include names of famous people.

There are two primary facets for defining KOS:

A The generic function of a KOS, regardless of content and
B Kos content and structure, defined by formality and by the entity type covered.

In B the main division is between KOS that focus on concepts and their words, terms, designations, and signs and those that focus on other types of entities, strictly speaking individuals but pragmatically including organisms, chemical substances, and other types of concepts that have a very large number of values. A third facet, C *origin and editorial control*, is needed to distinguish folksonomies. So overall KOS types can be roughly arranged as shown in Table 3a.

Table 3a. Broad overall arrangement of KOS

A	**KOS by generic function**
A1	. Conceptual data schema, conceptual schema, metadata schema, data model
A2	. List of entity values
B	**KOS by content and structure**
B1	. **KOS focusing on concepts and their words, terms, designations, and signs**
B1.1	. . KOS focusing primarily on terms
B1.2	. . KOS focusing primarily on concepts
B2	. **KOS focusing on other types entities (including individuals)**
B3	. **Formal ontology**
C	**KOS by origin and editorial control**
C1	. editorially controlled KOS
C2	. KOS generated automatically by text mining etc.
C3	. **folksonomy**

Table 3b. Annotated list of KOS types

A KOS by generic function

A1 **Conceptual data schema, conceptual schema, metadata schema, data model**
(also called implementation-driven ontology, thus belong to B3 Ontology)
A conceptual data schema specifies the structure of databases or documents; it specifies what kinds of information can be recorded in a database or knowledge base by defining entity types and relationship types with which statements can be constructed. If one thinks of a database as a set of statements one can think of the conceptual data schema as a set of statement templates. Perhaps the most natural way to represent this is an entity-relationship schema which gives a set of entity types and a set of relationship types through which entity types can be connected into statements. For example, with

entity types Taxon and ChemicalSubstance
relationship types Taxon <*hasPest*> Taxon and
[Taxon, Taxon]) <*usePesticide*> ChemicalSubstance
we can write the statements
Oryza sativa (rice) <*hasPest*> Orselia oryzae (Rice Gall Midge)
[Oryza sativa (rice), Orselia oryzae (Rice Gall Midge)] <*usePesticide*> Benfuracab

Examples: CIDOC Conceptual Reference Model (CRM) (http://cidoc.ics.forth.gr, http://en.wikipedia.org/wiki/CIDOC_Conceptual_Reference_Model);
the Unified Medical Language System (UMLS) Semantic Network (http://semanticnetwork.nlm.nih.gov/).
Possible forms of representation are listed below. This also includes **metadata registries** (MDR)

A1.1 **Entity-relationship schema**. See example under A1

A1.2 **Table schema for a relational database**

A1.3 **Frame structure schema**.
Example
A pesticide frame might have slots for

Substance (the pesticide)	*When applied*
Pest fought	*Dosage*
Crop or livestock	*How applied* (*spray*, *work into soil*, . . .)

In a hierarchy of frames one can define inheritance.

A1.4 **Object class schema**.
A frame system for defining the structure of objects in object-oriented programming or object-oriented databases.

A1.5 **Record format**.
A simple frame for organizing data about an entity.
Examples: The MARC format for bibliographic records, the format for an address book entry.

A1.6 **Data dictionary**.
A list of data items and information about them (especially data field definitions) – spanning a single database (giving all its tables and for each table the field/column definitions) or the entire computing environment of an organization. Used to assure uniformity of data (starting with simple things such as dates or person names). Can be organized alphabetically or systematically. For pointers to sources see http://en.wikipedia.org/wiki/Data_dictionary.

A1.7 **Data definition for software**
Many of the KOS types liste above can be used directly for this purpose, especially data dictionaries. Data definition may be embedded in a computer program or used from an external source. May include structural information and specific entity values.

A1.7.1 **Property list**
Data definition for computer programs or an entire software environment in a specific format. See Wikipedia.

A2 **List of entity <u>values</u>**
These KOS list entity values, such as the value Oryza sativa of the entity type Taxon. These entity values can be used in E-R statements, as slot fillers in frames, or in metadata fields. In description logic, this is called the TBox (T is for Terminology).

A2.1 **Authority list, authority file (terms or individual entities)**
A list of terms or names or other designations that are used to control the variant names for an entity or the values for a particular metadata field. Examples include terms for concepts (thus a controlled vocabulary is an authority list) and names for countries, individuals, and organizations. May include non-preferred designations linked to the preferred versions (in which case we have a thesaurus). Authority lists are often presented alphabetically or organized by a shallow classification scheme to support simple navigation.
In order to fulfil their functions, authority lists must often deal with information about the world, such as a hierarchy of organizations (for example, the departments and agencies of the US Federal Government), other relationships between organizations (for example, the complex network of corporate ownership), and changes in organizations, such as changes in name over time, or the change of an organization's nature, or divisions and mergers. Similar considerations apply to political subdivisions. See A3 Knowledge base below
Example: Library of Congress Authorities (names, titles, subjects) http://authorities.loc.gov/
For more examples dealing with specific entity types, see B1 and B2.

A2.1.1 **Controlled vocabulary**
An authority list of terms, self-standing or embedded in a more complex KOS, that have been enumerated explicitly, possibly by a registration authority, to be used in indexing and searching, often associated with a particular metadata field. The intent is that all authorized terms are unambiguously defined (homonyms disambiguated, usually by parenthetical qualifiers). May also contain non-preferred synonyms. In some contexts, controlled vocabularies are also used for writing, especially writing for audiences with limited literacy level; for example, many dictionaries published by Longman use a base vocabulary of 2,000 words for writing definitions.

A2.1.2 **Code list**
A list of entities associated with codes. Examples: ISO codes for countries (ISO 3166), languages (ISO 639-1 and ISO 639-2),and currencies (ISO 4217); a directory of periodical titles with ISSN and/or CODEN (such as Ulrich's). Strictly speaking, any classification that gives notations (class numbers) would fall under here, but that is not the customary use of the term *code list*.

Examples from	ISO 3166 Country Codes			License Plate Country Prefix	Telephone Country Code
China, Peoples Republic	CN	CHN	156	CN	86
France	FR	FRA	250	F	33
United States	US	USA	840	USA	1

A2.1.3 **Pick list** (also called menu)
A list of values for a data field from which an indexer or searcher can select ("pick"). Particularly appropriate if the list of values is small.
Example: The drop-down menu for selecting a country or a state/province within a country.

A2.2 **Index language**
The language used for indexing and searching in an information retrieval system. This can be a controlled vocabulary, possibly with a more or less complex syntax, or it can be natural language: the set of words or tags used in social tagging of a collection (folksonomy) or the set of all the words of one or more languages, or the full natural language including syntax (assuming powerful natural language processing capabilities for searching). In that sense, natural language is a KOS. When the elements in an index language are controlled, the index language is an authority list, possibly with a syntax for connecting elements. In book or periodical indexes the index language is often not or only minimally controlled.

A2.3 **List of variables**
A list of variables used in data collection and/or data analysis in an empirical study or statistical data collection or postulated in a theoretical work. The variables can be abstract concepts, in which case they might be arranged in a hierarchy, which might be called a typology. But variables might also be countries or organizations or from other entity types.
Example: Life Event Scale http://www.mentalhealthguide.co.uk/lifeevent.htm,
see http://etoh.niaaa.nih.gov/aodvol1/aodhnlh.htm#LK for a better structured list as part of the AOD Thesaurus

A2.3.1 **Coding scheme**
A list of variables and their values for coding collected data for analysis. For recording in a database or data analysis program, the variables are often expressed as codes, making the coding scheme a code list (A2.1.2).

A2.4 **Systematic arrangement or structure for browsing**
Most such structures deal with subjects, see *B1.2.3,2 Directory structure, shelf classification*. But there are also directories based on other entities, such as organizational units, such as the tree of US federal agencies at www.usa.gov/Agencies/federal.shtml.

A2.4.1 **Filing plan, file directory structure**
A list of terms, names, or codes used in arranging any type of paper or digital records. The elements can be concepts or, for example, names of organizational units arranged in the organizational hierarchy.

A2.4.2 **Table of contents of a book, law, conference**, ...

A3 **Knowledge base**
A collection of statements or assertions formulated in terms of the conceptual data schema and the entity values. In description logic, this is called the ABox. See the Introduction for a brief discussion.

B KOS by content and structure

Content refers primarily to entity types covered but on more detailed levels also to the relationship types used to relate entity values. Strictly speaking, there is another facet, level of formality. To keep things simple, the KOS types are arranged first by entity type they focus on (B1 and B2). Under B1 and B2 are listed KOS of a more informal nature. Formal ontologies are listed as B3 because of their practical separation, but they also belong under B1.2, KOS focusing primarily on concepts.

B1 **KOS focusing on concepts and their words, terms, designations, and signs**
Listed here are KOS of limited formality; these KOS have also been called "linguistic ontologies" or "terminological ontologies".
This subdivision covers most of the systems that are normally considered KOS. Put differently, it captures the core meaning of KOS.
A2.1.1 Controlled vocabulary and A2.2 Index language could be listed here since they deal primarily with concepts and terms, but their definition is based on their function, so they are more properly listed in Facet A.

B1.1 **KOS focusing primarily on terms**
KOS that are intended primarily for access by term or for finding terms, or that give information only on terms. Most of these KOS also deal with concepts: glossaries and dictionaries give definitions; any KOS that provides synonyms at least implicitly stipulates concepts.

B1.1.1 **Plain term list, keyword list, word list**

B1.1.2 **KOS focusing on providing synonyms**
Note: Information retrieval thesauri also provide synonyms, but they do much more.

B1.1.2,1 **List of synonym rings (synonym sets)**
A list of synonym sets where no preferred terms are selected. Used for synonym expansion of query terms in free-text and folksonomy searching. Perhaps the term *synonym ring* is used with the understanding that the synonyms in the set come from different systems, as in the second example.
Examples:
person, individual, someone, somebody, mortal, soul -- (a human being) (WordNet)
foaf:person *gjxdm*:Person *niem*:Person *sumo*:Human *cyc*:Individual
(http://en.wikipedia.org/wiki/Synonym_ring)

B1.1.2,2 **General language thesaurus**
A work that provides synonym sets primarily for the purpose of assisting writers in finding the right term, but clearly also useful for synonym expansion of query terms. May be arranged alphabetically or hierarchically.
Examples: Roget's Thesaurus of the English Language, Duden 08. Das Synonymwörterbuch

B1.1.3 **Glossary**
A list of terms with definitions. Many glossaries are restricted to terms from a specific subject field or those used in a particular work, such as a textbook, a brochure for health consumers, or a legal or policy document, and the terms are defined as used within that specific environment without giving variant meanings. While glossaries focus on access by term, they clearly establish and provide information about concepts. While usually arranged alphabetically, they can be arranged hierarchically with an alphabetical index. Glossaries do not themselves introduce new terms, but they often contain terms /concepts new to a domain that have been introduced in the work with which the glossary is associated.
Examples:
EPA Terms of the Environment (www.epa.gov/OCEPATERMS/)
Pesticide: Substance or mixture intended for preventing, destroying, repelling, or mitigating any pest. Also, any substance or mixture intended for use as a plant regulator, defoliant, or desiccant.

Credit And Banking Glossary (www.cardreport.com/glossary.html)
Default Failure to meet the terms of a credit agreement.

B1.1.4 **(Lexicographical) dictionary, general dictionary**
A dictionary is a listing of words and phrases giving information such as spelling(s), morphology and part of speech, senses (possibly senses across disciplines or other contexts), definitions, usage, origin, sometimes synonyms and, in the case of bi- or multilingual dictionaries, equivalents in other languages; most often arranged alphabetically. Usually there is no explicit hierarchical structure or attempt to group terms by concept. The term *dictionary* is also used more broadly, as in *Dictionary of Theories* or *Chamber's Biographical Dictionary*.

B1.1.4,1 **Spell check dictionary**
A dictionary used for spell checking in a word processor or other computer application. Usually includes proper names that are used with some frequency. Contains only spelling(s) and possibly morphological rules. Optimized for speedy look-up. A general spell check dictionary comes with most computers; domain-specific spell-check dictionaries (such as for chemistry, medicine, or law) are available.

B1.1.5 **Lexical database, NLP lexicon**
Machine-readable dictionary with the information needed by natural language processing (NLP) applications, such as part of speech, morphology, frequent co-occurrence with other words or phrases, disambiguation rules, case frames.
Examples: WordNet (http://wordnet.princeton.edu/), which is useful for both people and programs; Thamus Bilingual Dictionaries (http://catalog.elra.info/);
the UMLS SPECIALIST Lexicon
(http://lexsrv3.nlm.nih.gov/SPECIALIST/Projects/lexicon/2008/release/LEX/DOCS/techrpt.pdf);
Sample entries:

base=	anaesthetic	*base=*	give
spelling_variant=	anesthetic	*entry=*	E0029785
entry=	E0008770	*cat=*	verb
cat=	adj	*variants=*	irreg|gives|gave|given|giving|
variants=	inv [forms no comparative or superlative]	*intran*	[The bridge gave without warning]
position=	attrib(3) [appears after color adjectives]	*intran;*	part(out) [The disk drive gave out]
		tran=	np [He gave money]
		tran=	np;part(back) [He gave money back]
		ditran=	np,np|short shrift| [He gave strikers short shrift]
		ditran=	np,pphr(for,np) [he gave money for charity]
		nominalization=	gift|noun|E0029737

B1.1.5,1 **Database of frames for lexical units (especially verbs)**
A lexical resource based on frame semantics that documents the range of semantic and syntactic combinatory possibilities (valences) of word senses (a word with multiple senses has multiple frames). Similar information as SUMO (B3.2.1)

Example: FrameNet http://framenet.icsi.berkeley.edu/
Sample entry (simplified)

Intentionally_create

Definition:	The Creator creates a new entity, the Created_entity, possibly out of Components.
Core slots:	Created_entity (Semantic Type Artifact)
	Creator (Semantic Type Sentient)
Non-core slots:	Co_participant
	Components
	Instrument
	Purpose
	Place
	Time
	. . .

B1.1.5 ,2 **word segment list**
A list of word segments (morphemes) that occur in a language.

B1.2 **KOS focusing primarily on concepts**
These KOS, while of course dealing with terms, focus on establishing concepts, often for purposes of indexing and searching, and many provide access by concept through their structure. Such a KOS may introduce new terms.

B1.2.0 **Relationship-based KOS**
A term sometimes used to include thesauri and classifications (and semantic networks and ontologies) as KOS that are characterized by their linking concepts through relationships. They vary in granularity of relationships and degree of formality of definition. There tends to be a practical trade off between expressivity (e.g. number of relationship types) and both interoperability (via common agreement on meaning and use of the relationship types) and overhead in design.

B1.2.1 **Thesaurus**

Thesaurus, along with taxonomy and ontology, is a “global” term in the sense that it is used to refer to many different types of KOS. Most would agree that a thesaurus is a structure that

(1) manages the complexities of terminology by indicating multiple meanings for homonyms and by relating terms that are identical or similar in meaning (synonyms and quasi-synonyms) and

(2) provides conceptual relationships, ideally through an embedded classification/ontology.

While there are standards for thesauri, many KOS commonly understood to be thesauri do not adhere to all the specifications in a thesaurus standard. Most were developed for a specific subject domain.

When thesauri were first introduced, there purpose was controlling the terms used for indexing; this is still true for many thesauri today. These thesauri specify descriptors authorized for indexing and searching; these descriptors form a controlled vocabulary (authority list, index language). This controlled vocabulary is surrounded by a lead-in vocabulary consisting of non-descriptors that refer to the descriptor or descriptor combination to be used. However, any thesaurus can be used to support free-text or folksonomy searching

(1 through alerting the user to synonyms, and to broader, narrower, and related terms that might be useful for expanding a query term or through automatic query term expansion and

(2) through alerting a user if a term has multiple meanings and giving hints on how to amend the query to focus on the meaning the user has in mind..

A thesaurus designed only to support searching may be called a *search thesaurus*. A search thesaurus need not specify authorized descriptors.

Most thesauri use a limited set of relationships:

USE/UF [Used For] on the terminological level and

BT/NT [Broader Term/Narrower Term] and RT/RT [Related Term] at the conceptual level.

Some thesauri use more differentiated relationships; for example, the hierarchical relationship can be specialized into Generic (subclass/superclass), Instance (class/instance) and partitive (whole-part) relationships. Thesauri with differentiated relationships are just as likely to be called ontologies.

In most thesauri alphabetical arrangement of terms is predominant, but a hierarchical arrangement, often pieced together from BT/NT relationships, may be added as an afterthought. There are thesauri with a developed hierarchy as advocated by Soergel 1969,1974; an early examples is Thesaurofacet: A thesaurus & faceted classification for engineering & related subjects (Aitchison 1970).

Some well-known thesauri are

ERIC Thesaurus. Alphabetical (http://eric.ed.gov)

AGROVOC. Alphabetical, multilingual (www.fao.org/aims/ag_intro.htm)

AAT – Art and Architecture Thesaurus. Mostly well-structured, partially facet-based hierarchy, main part alphabetical (www.getty.edu/research/conducting_research/vocabularies/aat)

AOD Thesaurus – Alcohol and Other Drug Thesaurus. Entirely hierarchical (http://etoh.niaaa.nih.gov/aodvol1/aodthome.htm)

Sample entries from the ERIC Thesaurus (abridged and reformatted)

Critical Thinking
CA Learning and Perception
BT Cognitive Processes;
NT Evaluative Thinking;
RT . . .

Evaluative Thinking
SN Process of determining or judging the appropriateness, efficacy, or value of something with respect to specified objectives or standards
UF Judgmental Processes
CA Learning and Perception
BT Critical Thinking
NT Value Judgment
RT Data Interpretation
Decision Making
Evaluation
Thinking Skills

Judgmental Processes
USE Evaluative Thinking

SN Scope Note
UF Used For
USE Use
CA Category (one of 41 broad categories)
BT Broader Term
NT Narrower Term
RT Related Term

Excerpt from the AOD Thesaurus (abridged)

EF route of administration

SN The way in which a substance reaches its site of action in the body. . . .
ST *medication route*
ST *route of drug application*
ST *route of drug entry*
ST *route of exposure*
BT +EE12 pharmacokinetics
RT +AA2 AOD use
RT EE12.2e drug absorption
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

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) for 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**.
ST *topical application*

EF2.2.4. . local drug administration
SN The introduction of a substance into a localized area of the skin or other tissue, as through injection.
NT EF4.6.8 intracutaneous injection

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

EF4 route of administration by method or body site

. . .

EF6 drug administration by self vs. others

. . .

B1.2.2 **KOS alphabetically arranged by terms (terms also important)**
Many thesauri fall under here.

B1.2.2,1 **Subject dictionary**
A dictionary of terms in a subject domain. A monolingual subject dictionary tends to have definitions that are more complete than those found in a medium-sized general dictionary. A multilingual subject dictionary may have just terms and equivalences across languages

B1.2.2,2 **Subject heading list**
Prototypically, an alphabetical list of subject terms created for a subject catalog on cards or in book form. Since in these media it is not practical to combine descriptors in searching, subject headings have a high degree of precombination. They typically have a set of main headings, many of which are precombined in themselves and from which more compound headings can be formed by appending one or more subheadings from an approved list, making a subject heading list a synthetic KOS. Most subject heading lists consist of a controlled vocabulary used as the index language and a lead–in vocabulary that contains many non-descriptor terms with a USE (formerly SEE) reference to the appropriate descriptor. Among the controlled headings there are conceptual relationships, originally only one catch-all SEE ALSO, but many lists now use BT (Broader Term), NT (Narrower Term), RT (Related Term). Thus, a subject heading list meets the definition of thesaurus. What distinguishes a subject heading list from most systems that are known as thesauri is the high degree of precombination and the use of subheadings.

Examples: RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié) (http://rameau.bnf.fr/informations/rameauenbref.htm)
Medical Subject Headings (MeSH)(http://www.nlm.nih.gov/mesh/).
Library of Congress Subject Headings (LCSH) (http://authorities.loc.gov). Excerpts:

Coal
  BT Fossil fuels
  NT Anthracite coal
  NT Coal reserves
  Coal – Combustion
  Coal – Transportation
    Coal – Transportation – Law and legislation

Coal-burning power plants
  USE Coal-fired power plant

Coal-fired power plant
  UF Coal-burning power plants
    Power plants, coal-fired
  BT Fossil fuel power plants
  NT Geothermal-coal hybrid power plants

B1.2.3 **KOS arranged in a hierarchical format**:
Classification scheme, taxonomy, categorization scheme
A classification is a structure that organizes concepts into a hierarchy, possibly in a scheme of facets. Some thesauri, such as the ThesauroFacet, AAT, MeSH, and the AOD Thesaurus incorporate classification schemes. See the more specific KOS types for examples.

OpenGalen XXX

B1.2.3,1 **Subject category list**
A list of normally fewer than 100 subject categories, often arranged in a two-level hierarchy. Used to assign one or more subject categories to documents and/or to group descriptors from a thesaurus into broad subject categories.
Example: AGRIS SUBJECT CATEGORIES www.fao.org/scripts/agris/c-categ.htm
ERIC Thesaurus Subject Categories
http://eric.ed.gov/ERICWebPortal/Home.portal?_nfpb=true&_pageLabel=Thesaurus&_nfls=false
For more examples see www.dit.unitn.it/~knowdive/Lectures/Dagobert/SubjectDomains.pdf

B1.2.3,.2 **Typology**
The term typology is used for small classifications, often in the context of research studies. Many typologies consist of fewer than 12 types and are flat (no hierarchy)
Example: Pew Research Center Political Typology (http://people-press.org/report/?pageid=944, http://people-press.org/report/?pageid=949
See A2.3 for another example

B1.2.3,3 **Subject directory structure, shelf classification**
A classifications of subjects for the purpose of arranging books or other documents on shelves or in file drawers or Web pages in a subject directory, a kind of A2.4 *Systematic arrangement or structure for browsing*. The idea is to group items on the same subject together for ease of comparison and to collocate similar subjects for ease of browsing.
Examples
DMOZ directory (www.dmoz.org)
Dewey Decimal Classification (www.oclc.org/dewey)
Library of Congress Classification (www.loc.gov/catdir/cpso/lcco)
Excerpt from the Library of Congress classification

HE	Transportation and Communications	
HE199	. Freight (General)	
HE199.5.A-Z	. . Special Commodities, A-Z	
HE199.5.B35	. . . Banana	
HE199.5.C6	. . . Coal	[Note that HE2321.C6 is a narrower class]
	. Railways	
	. . Railway administration	
	. . . Traffic	
	. . . . Freight	
HE2321.A-Z	. . . . . Special commodities, A-Z	
HE2321.B35	. . . . . . Berries	[Note inconsistency]
HE2321.C6	. . . . . . Coal	

B1.2.3,4 **Taxonomy** The term taxonomy was originally used for the classification of living organisms but is now used very loosely by different communities to refer to any kind of classification or hierarchically structured thesaurus, including particularly Web subject directories and enterprise taxonomies used for intranets and other corporate information systems . Examples: Organism taxonomy (a comprehensive source: http://tolweb.org/tree/phylogeny.html) A Taxonomy for Learning, Teaching, and Assessing. A revision of Bloom's taxonomy of educational objectives. (A faceted classification of educational objectives with extensive definitions) (Anderson 2001, www.uwsp.edu/education/lwilson/curric/newtaxonomy.htm) LexisNexis Subject & Industry Taxonomies. A hierarchically structured thesaurus (http://www.taxonomywarehouse.com/vocabdetails_include.asp?vVocID=485)
B1.2.4 **KOS arranged in a 2-D or 3-D graph**
B1.2.4,1 **Concept map, mind map** A concept map is a graphical representation of concepts (nodes) and relationships ((usually labeled) edges) between them. The relationships can be conceptual (such as hierarchical) or empirical (such as cause and effect, influence); thus a concept map can represent a knowledge base (or a piece of one). Concept maps are used primarily in education. An entity-relationship graph is very similar to a concept map. A concept map displays a semantic network; the emphasis is on the visual display. Many sample concept maps at http://cmex.ihmc.us/cmex/Map%20of%20Maps.html See also the Atlas of Science Literacy www.project2061.org/publications/atlas/default.htm XXX Example of concept map still to come
B1.2.5 **Semantic network, RDF graph, topic map** A semantic network is an abstract graph where the nodes typically represent concepts (but can also represent other entities) and the (usually labeled) edges represent relationships. Many types of KOS and many databases (especially relational databases) can be represented as semantic networks – semantic networks are more a representation formalism than a type of KOS. This representation is associated with processing techniques such as spreading activation and hierarchical inheritance. RDF and the Topic Map standard are formats for encoding semantic networks. RDF and the Topic Map standard can be seen as formalisms for defining relational database schemas. Do not confuse topic maps (semantic networks encoded following a standard) and concept maps (visual representations). The Wikipedia article gives a good introduction with pointers (http://en.wikipedia.org/wiki/Semantic_network)
B1.2.6 **KOS by further structural properties** (See Table 2d)
B1.2.6,1 **Enumerative versus synthetic KOS**
B1.2.6,1,1 **Enumerative KOS, enumerative classification scheme**
B1.2.6,1,2 **Synthetic KOS, synthetic classification scheme**
B1.2.6,2 **Faceted classification**

B2 **KOS focusing on other types entities (individuals)**
These are often A2.1 *authority lists, authority files*. They could be arranged alphabetically or systematically. They could give codes and thus function as A2.1.2 *code lists*. Below are a few examples.

B2.1 **Onomasticon**
A list or dictionary of proper nouns, particularly names of persons or places.

B2.2 **Personal and corporate name authority file**
Examples: Virtual International Authority File (VIAF) (http://www.oclc.org/research/projects/viaf)
Library of Congress Name Authority File (http://authorities.loc.gov)
From two entries
Mann, Thomas, 1875-1955
LC Control Number: n 2002022302
400 Mann, Paul Thomas, 1875-1955
400 Thomas, Paul, 1875-1955
400 Mani, T'omas, 1875-1955 [Georgian]
670 b. 6-6-1875 in Lübeck as Paul Thomas Mann; used pseud. Paul Thomas as co-editor of student newspaper in 1893;
Thomas-Mann-Archiv
See: Eidgenössische Technische Hochschule Zürich. Thomas-Mann-Archiv

B2.2.1 **Personal (often author) name authority file**
Example: Personennamendatei (PND) (www.d-nb.de/standardisierung/normdateien/pnd.htm)
Union List of Artist Names (ULAN)
(www.getty.edu/research/conducting_research/vocabularies/ulan/)

B2.2.2 **Corporate name authority file**
Example: Gemeinsame Körperschaftsdatei (GKD)(www.d-nb.de/standardisierung/normdateien/gkd.htm)

B2.3 **Title list, title authority file (e.g., for serials**)
Example: Cultural Objects Name Authority (CONA)
www.getty.edu/research/conducting_research/vocabularies/contribute.html#cona

B2.4 **Biographical reference work, biographical dictionary**

B2.5 **A reference work of myths arranged by underlying theme**

B2.6 **Gazetteer**
A gazetteer is a list of place names. Traditional gazetteers have been published as books or they appear as indexes to atlases. Each entry may also be identified by feature type, such as river, city, or school. Geospatially referenced gazetteers provide coordinates for locating the place on the earth's surface and various types of 'footprint', such as centroid, bounding box, etc. Note that the term "gazetteer" has several other meanings, including an announcement publication such as a patent or legal gazetteer.

Examples:

Getty Thesaurus of Geographic Names
(www.getty.edu/research/conducting_research/vocabularies/tgn/)

U.S. Geographic Names Information System (http://geonames.usgs.gov/)

U.S. Gazetteer (www.census.gov/cgi-bin/gazetteer)

Sample entry
Alexandria, VA (city)
Population (1990): 111183
Location: 38.82100 N, 77.08617 W
Zip Code(s): 22041 22206 22301 22302 22304 22305 22310 22311 22312 22314
Browse Tiger Map of area.
Lookup 1990 Census STF1A, STF3A tables.

B2.7 **Event gazetteer**

B2.8 **List of languages**
Example: Ethnologue: Languages of the World. www.ethnologue.com/.

B3 **Ontology**
An ontology is a system of concepts and relationships for the (more or less) formal expression of statements in a domain to support question-answering and inference and for defining and labeling of subjects or topics in the domain to support metadata assignment to many kinds of resources. Often ontologies are seen or even defined as shared or consensual conceptualizations that are reusable; also, much discussion centers on universal schemes. While consensual ontologies of broad scope have an important role, in many contexts highly domain-specific or task-specific ontologies, even individual and highly idiosyncratic ontologies, will best support the work of users or computer programs.

Ontology is used in many different meanings, so reader beware! The Oxford English Dictionary defines *ontology* as "The science or study of being; that department of metaphysics which relates to the being or essence of things, or to being in the abstract." Part of such a study is a classification of "things that are" into basic types, often starting with *living vs non-living entities*. Thus the term *ontology* assumed the additional meaning of a shallow classification of basic categories. Such classifications or ontologies are needed in linguistics, for example, to formulate rules of the subjects or objects a verb can take, and in data element definition. As such rules became more and more refined, the classification supporting them needed to be more specific, so eventually *ontology* was used to designate any classification, particularly in the communities of linguistics, AI, and software engineering. Indeed, once these communities increased their awareness that there is not only a problem of classification but also of terminology, "ontologies" included lead-in vocabularies as well and became full-fledged hierarchical thesauri as well. Today the term *ontology*, because it is fashionable, is often used loosely for any knowledge organization system, particularly if it is represented using Semantic Web standards, such as RDF. Many KOS labeled *ontology* would be more properly called classification or thesaurus; as mentioned above, the term *taxonomy* is used in a similarly imprecise way.

However, as intended for AI modeling and inferencing purposes, ontologies tend to have precise and formal definitions of relationships, giving classes (concepts) and instances of those classes, being objects in the domain. Classes have attributes so that complex objects in the domain can be described. Relationships include *is-a* (for class hierarchies), *instance*, *partitive* and domain-specific relationships that are more differentiated than in a traditional thesaurus. For examples of entity types and relationships in Medicine see the UMLS Semantic Network (www.nlm.nih.gov/research/umls/META3_current_semantic_types.html, www.nlm.nih.gov/research/umls/META3_current_relations.html)

Even what is most obvious to the human mind needs to be specified with excruciating explicitness for use by computer programs.

An ontology can be very broad, giving only entity types and relationship types to serve as a A1 Conceptual data schema or it can be more specific and give a A2 List of entity values; some ontologies do both.

On ontology in general see Gruber 2008, Horrocks 2008, Aussenac-Gilles and Soergel 2005

B3.1 **Ontology by degree of formality**

B3.1.1 **Lightweight ontology**
Any KOS, often with the understanding that there are at least *is-a* relationships (a *is-a* hierarchy).

B3.1.2 **Formal ontology**
Ontologies can be associated with formally defined axioms and rules for processing and combining relationships and are intended for use with logical reasoning systems. Consequently, they are suited to applications with well-defined objects and operations.
Formal ontology requires a clear semantics for the language and clear motivations for the adopted distinctions as well as strict rules about how to specify terms and relationships.
This is obtained by relying on ontological analysis (in the philosophical sense) and by using formal logic, usually Description Logic (DL) up to subsets of Higher Order Logic (HOL), where the meaning of the terms is guaranteed by formal semantics.

See example under B3.2.1 Foundational ontology

B3.2 **Ontology by conceptual level and scope**

B3.2.1 **Foundational ontology, upper-level ontology, upper ontology**
These are the most general formal ontologies. They deal with very general and basic terms like entity, event, process, spatial and temporal location, part-of, quality-of, participation, and the like.
The purpose of these ontologies is to characterize entities and relations that are common in all domains and to provide a consistent and unifying view.
On foundational ontologies in general see Obrst et al. 2006

Examples: Major upper ontologies
BFO – Basic Formal Ontology www.ifomis.org/bfo
DOLCE – Descriptive Ontology for Linguistic and Cognitive Engineering
www.loa-cnr.it/DOLCE.html, Masolo et al. 2003
GUM – Generalized Upper Model
www.fb10.uni-bremen.de/anglistik/langpro/webspace/jb/gum/index.htm
OpenCyc. www.opencyc.org/
SUMO – Suggested Upper Merged Ontology www.ontologyportal.org, Niles and Pease 2001
Excerpts from SUMO (compare FrameNet, B1.1.5,1)

SUMO class (*is-a*) hierarchy excerpts	**Example definitions and axioms**
entity	Notation:

. **physical**
. . **object**
. . . self connected object
. . . . substance
. . . . corpuscular object
. . . . . organic object
. . . . . artifact
. . . . content bearing object
. . . . . symbolic string
. . . agent
. . . . sentient agent
. . . . geopolitical area
. . . . organism
. . . . group
. . **process**
. . . dual object process
. . . intentional process
. . . . making
. . . . . constructing
. . . . . manufacture
. . . . social interaction
. . . . . change of possession
. . . . . communication
. . . . . cooperation
. . . motion
. . . internal change
. . . shape change
. **abstrac**t
. . **quantity**
. . **attribute**
. . . internal attribute
. . . . physical state
. . . . perceptual attribute
. . . relational attribute
. . . . social role
. . set or class
. . **relation**
. . proposition
. . . field of study

Artifact a class (entity type) or a constant entity value
?ARTIFACT a variable ranging over the values of the entity type Artifact.
The example includes the formal logical notation only for some axioms

sentient agent (SentientAgent)
An Agent that has rights but may or may not have responsibilities and the ability to reason. If the latter are present, then the Agent is also an instance of CognitiveAgent.

Axioms
If ?AGENT is an instance of sentient agent,
then ?AGENT is capable to do perception in role experiencer.

```
(=>
    (instance ?AGENT SentientAgent)
    (capability Perception experiencer ?AGENT))
```

If ?ATTR is an instance of PsychologicalAttribute and ?ATTR is an attribute of ?AGENT,
then ?AGENT is an instance of SentientAgent.

```
(=>
    (and
        (instance ?ATTR PsychologicalAttribute)
        (attribute ?AGENT ?ATTR))
    (instance ?AGENT SentientAgent))
```

?AGENT is an instance of SentientAgent and Living is an attribute of ?AGENT **if and only if** there exists ConsciousnessAttribute ?ATTR so that ?ATTR is an attribute of ?AGENT.

making (Making)
The subclass of Creation in which an individual Artifact or a type of Artifact is made.

Axioms
Some ?ARTIFACT t is an instance of Artifact if and only if there exists some Making ?MAKING so that ?ARTIFACT is a result of ?MAKING.

social interaction (SocialInteraction)
The subclass of IntentionalProcess that involves interactions between CognitiveAgents

Axioms
If ?INTERACTION is an instance of SocialInteraction,
then there exist ?AGENT1, AGENT2 so that ?AGENT1 is an agent of ?INTERACTION and ?AGENT2 is an agent of ?INTERACTION and ?AGENT1 is not equal to ?AGENT2.

```
(=>
    (instance ?INTERACTION SocialInteraction)
    (exists
        (?AGENT1 ?AGENT2)
        (and
            (agent ?INTERACTION ?AGENT1)
            (agent ?INTERACTION ?AGENT2)
            (not
                (equal ?AGENT1 ?AGENT2)))))
```

B3.2.2 **Core ontology, reference ontology** A formal ontology that furnishes the organization of top-level (general) concepts used in (or across) a community or application area. The purpose is to facilitate reliable exchange of information within a community or area. Core ontologies seek to act as unifying frameworks for a general domain, sometimes bridging different domain ontologies. Example: The CIDOC Conceptual Reference Model (CIDOC CRM) is a widely used example from the cultural heritage domain. http://cidoc.ics.forth.gr
B3.2.3 **Domain-specific ontology, domain ontology** Detailed domain ontologies (which can be thesauri or classification schemes or enriched versions) these are (more or less) formal ontologies that focus on an application area (i.e., enterprise modeling, anatomy, astrophysics, etc.). The purpose is to provide a basic, stable and unambiguous description of concepts, entities and relations used in such a domain. Examples Gene Ontology (GO) www.geneontology.org/index.shtml, Gene Ontology Consortium 2000 For many more examples, see http://bioportal.bioontology.org/, http://protege.stanford.edu/download/ontologies.html
C KOS by origin and editorial control This criterion classifies KOS not by there intrinsic nature but how they come about and how they are controlled. Most KOS fall under C1, some under C2, but neither of these classes are named or specifically discussed. This facet is needed to place folksonomy.
C1 Editorially controlled KOS
C2 KOS generated automatically by text mining etc.
C3 Folksonomy A folksonomy is simple the list or collection of all the "tags" assigned on a social tagging site. These tags include subject terms (so a folksonomy can be considered a B1.1.1 term list) but also anything else useful to the tagger – to-do terms, dates, places, people's names, etc. There is no control over what is included in a folksonomy, so term variations abound. There is also no order. Some order can be generated automatically by clustering or other techniques resulting in "tag clouds", or users can contribute some order by submitting tag clouds or hierarchy segments of their own.

The following sources deal with KOS types and their definitions. Some have been used to some extent in constructing Table 3b, but as pieces of text where evaluated, mingled, and modified, keeping track would have been too cumbersome, so we give here global attribution. The sources are Tudhope 2006 (used primarily in B1 and B2), Borgo 2004 (used heavily for B3), Gruninger et al. 2008, Hodge 2000 July, Hodge 2000 April, Zeng 2008, Wright 2008,Uschold 2006, Taxonomy Glossary, Ontology Online, Taxonomy Warehouse.

Major organizations dealing with KOS

ISKO – International Society for Knowledge Organization www.isko.org
ASIST SIG/CR – American Society for Information Science Special Interest Group Classification Research www.asis.org/SIG/cr.html
NKOS – Networked Knowledge Organization Systems and Services http://nkos.slis.kent.edu
Classification Society http://thames.cs.rhul.ac.uk/~fionn/classification-society/
IFCS – International Federation of Classification Societies www.classification-society.org
ALA/ALCTS/CCS American Library Association / Association for Library Collections & Technical Services / CCS Cataloging and Classification Section http://www.ala.org/ala/mgrps/divs/alcts/mgrps/ccs/index.cfm
ONTOLOG Open, International, Virtual Community of Practice on Ontology, Ontological Engineering and Semantic Technology http://ontolog.cim3.net/
W3C Semantic Web Activity http://www.w3.org/2001/sw/

Bibliography

Aitchison, Jean. "The Thesaurofacet: A Multipurpose Retrieval Language Tool." *Journal of Documentation* 26, no. 3 (September 1970): 187-203.

Anderson, Lorin W., ed., David R. Krathwohl, ed., Peter W. Airasian, ed., Kathleen A. Cruikshank, ed., Richard E. Mayer, ed., Paul R. Pintrich, ed., James Raths, ed. and Merlin C. Wittrock, ed. *A Taxonomy for Learning, Teaching, and Assessing A Revision of Bloom's Taxonomy of Educational Objectives*. Addison Wesley Longman, 2001.

Aussenac-Gilles, Nathalie and Dagobert Soergel. "Text Analysis for Ontology and Terminology Engineering." *Applied Ontology* 1, no. 1 (2005): 35-46.

Borgo, Stefano. "Classifying (Medical) Ontologies. Presentation Slides." *FOIS 2004 (International Conference on Formal Ontologies in Information Systems)* (2004).

Gene Ontology Consortium. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics* 25 (1 May 2000): 25-29.

Gnoli, Claudio, ed. "Axiomathes Special Issue on Facet Analysis." *Axiomathes* 18, no. 2 (June 2008).

Gruber, Tom. "Ontology." In *Encyclopedia of Database Systems*, edited by ed. Liu Ling and Tamer Özsu, ed. Springer, 2008.

Gruninger, Michael, Olivier Bodenreider, Frank Olken, Leo Obrst and Peter Yim. "Ontology Summit 2007 Ontology, Taxonomy, Folksonomy: Understanding the Distinctions." *Applied Ontology* 3, no. 3 (2008): 191-200.

Hodge, Gail. *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files.* Washington, DC: Council on Library and Information Resources (CLIR), 2000-2004.

Hodge, Gail and NKOS. *Taxonomy of Knowledge Organization Sources/Systems (1) Draft June 7, 2000 (Revised July 31, 2000).* , 2000-2031.

Horrocks, Ian. "Ontologies and the Semantic Web." *Communications of the ACM* 51, no. 12 (2008-2012): 58-67.

Masolo, Claudio, Stefano Borgo, Aldo Gangemi, Nicola Guarino, Alessandro Oltramari and Luc Schneider. *The WonderWeb Library of Foundational Ontologies. WonderWeb Deliverable D17.* Padova, Italy: ISTC-CNR, 2003-2029.

Niles, Ian and Adam Pease. "Towards a Standard Upper Ontology." *2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)* (2001): 2-9.

Obrst, Leo, Patrick Cassidy, Steve Ray, Barry Smith, Dagobert Soergel, Matthew West and Peter Yim. “The 2006 Upper Ontology Summit Joint Communiqué.” *Applied Ontology* 1, no. 2 (2006): 203-11.

“Ontology Online.”.

Soergel, Dagobert. “Digital Libraries and Knowledge Organization Systems.” In *Semantic Digital Libraries*, edited by Sebastian Kruk, ed. Springer, 2008.

Soergel, Dagobert. *Indexing Languages and Thesauri: Construction and Maintenance*. New York, NY: Wiley, 1974.

Soergel, Dagobert. “Information Organization.” In *Berkshire Encyclopedia of Human-Computer Interaction*, edited by David Bainbridge, ed. Great Barrington, MA:: Berkshire, July 2004.

Soergel, Dagobert. *Klassifikationssysteme Und Thesauri. Eine Anleitung Zur Herstellung Von Klassifikationssystemen Und Thesauri Im Bereich Der Dokumentation*. Frankfurt/M.: Dt. Gesellschaft fur Dokumentation, 1969.

Taxonomy Glossary. .

“Taxonomy Warehouse.”.

Tudhope, Douglas, Traugott Koch and Rachel Heery. *Terminology Services and Technology. JISC State of the Art Review*. UKOLN, University of Bath, 2006-2015.

Uschold, Michael. “Ontologies Ontologies Everywhere. But Who Knows What to Think? *Protégé Users Conference Keynote Address*.” *International Protégé Users Conference. 9* (24 July 2006).

Wright, Sue Ellen. “Typology for KRRs - Knowledge Representation Resources.” *NKOS-CENDI Workshop* (September 2008).

Zeng, Marcia Lei. “Metadata for Terminology / KOS Resources.” *The 7th European NKOS Workshop at the 12th ECDL Conference, Aarhus, Denmark* (19 September 2008).