160

Knowl. Org. 35(2008)No.2/No.3
M. L. Zeng. Knowledge Organization Systems (KOS)

Knowledge Organization Systems (KOS)[†]

Marcia Lei Zeng

Kent State University, School of Library and Information Science,
Kent, OH, USA 44242-0001, <mzeng@kent.edu>

Marcia Lei Zeng has been involved in the development, teaching, and research of knowledge organization systems (KOS) for over 20 years. She has served on standards committees and working groups for IFLA, Special Libraries Association (SLA), American Society for Information Science and Technology (ASIST), and US National Information Standards Organization (NISO). She is a member of the Advisory Group for NISO Z39.19-2005 for monolingual controlled vocabularies. Her services include chairs of the SLA Technical Standards Committee, ASIST Standards Committee, IFLA Classification and Indexing Section, and IFLA Functional Requirements for Subject Authority Records (FRSAR) Working Group.

ABSTRACT: Knowledge organization systems (KOS) can be described based on their structures (from flat to multidimensional) and main functions. The latter include eliminating ambiguity, controlling synonyms or equivalents, establishing explicit semantic relationships such as hierarchical and associative relationships, and presenting both relationships and properties of concepts in the knowledge models. Examples of KOS include lists, authority files, gazetteers, synonym rings, taxonomies and classification schemes, thesauri, and ontologies. These systems model the underlying semantic structure of a domain and provide semantics, navigation, and translation through labels, definitions, typing, relationships, and properties for concepts.

The term knowledge organization systems (KOS) is intended to encompass all types of schemes for organizing information and promoting knowledge management, such as classification schemes, gazetteers, lexical databases, taxonomies, thesauri, and ontologies (Hodge 2000). These systems model the underlying semantic structure of a domain and provide semantics, navigation, and translation through labels, definitions, typing, relationships, and properties for concepts (Hill et al. 2002, Koch and Tudhope 2004). Embodied as (Web) services, they facilitate resource discovery and retrieval by acting as semantic road maps, thereby making possible a common orientation for indexers and future users, either human or machine (Koch and Tudhope 2003, 2004).

1. Overview of types of knowledge organization systems

Figure 1 shows the types of KOS, arranged according to the complexity of their structures and major functions. It visualizes the understanding of the author based on: 1) the *Taxonomy of Knowledge Organization Sources/Systems* (2000) originated by Hodge (2000) and adopted by the Networked Knowledge Organization Systems/Services (NKOS) group; 2) NISO Z39.19-2005 *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* issued by the National Information Standards Organization (NISO 2005) in the U.S.; and 3) a recent JISC (Joint Information Systems Committee) state-of-the-art review and report, *Terminology Services and Technology,* prepared by Tudhope, Koch, and Heery (2006).

The class of KOS can be explained according to four major groups, from simpler to more complicated
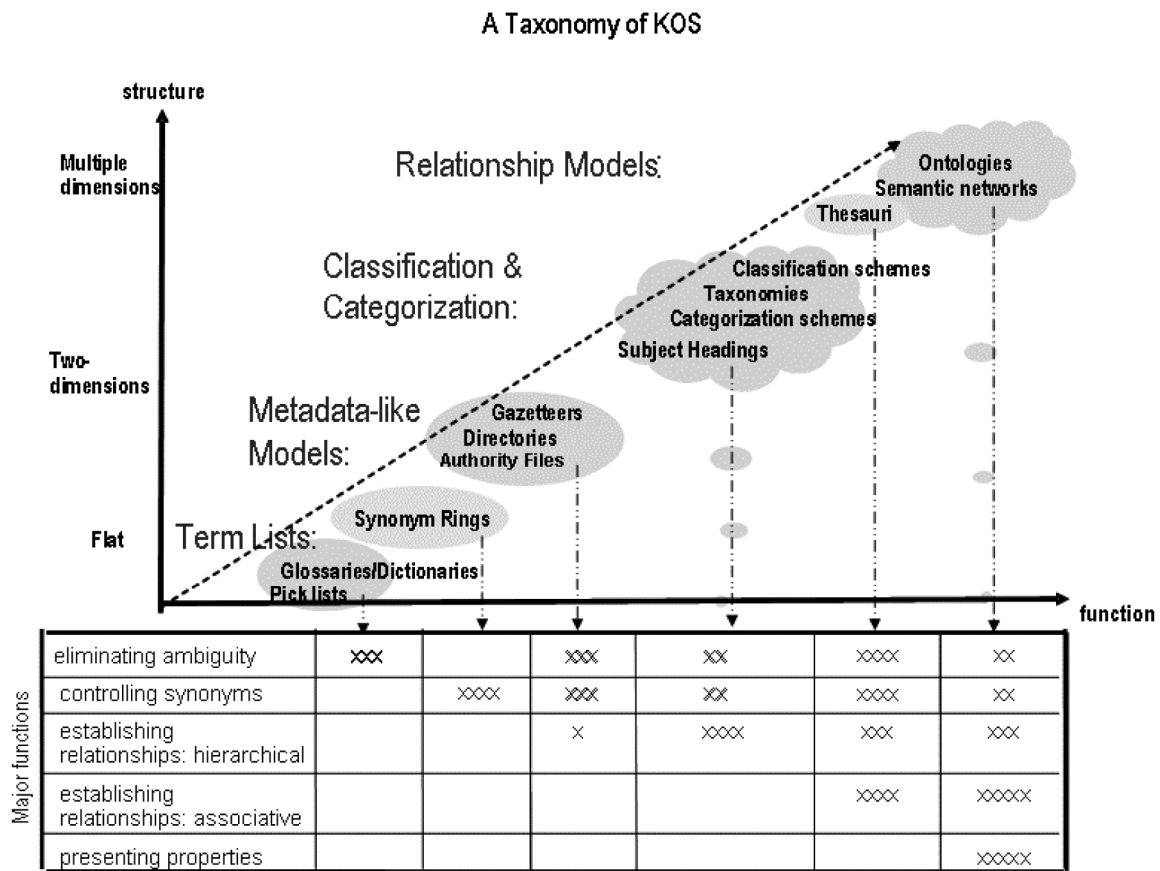
A Taxonomy of KOS

Major functions						
eliminating ambiguity	xxx		xxx	xx	xxxx	xx
controlling synonyms		xxxx	xxx	xx	xxxx	xx
establishing relationships: hierarchical			x	xxxx	xxx	xxx
establishing relationships: associative					xxxx	xxxxx
presenting properties						xxxxx

Figure 1. *An overview of the structures and functions of KOS*

structures: term lists, metadata-like models, classification and categorization, and relationship models (Hodge 2000, NISO 2005, Hill et al. 2002). Please note that these groupings are not mutually exclusive:

– Term Lists
  - Lists (pick lists): limited sets of terms in some sequential order.
  - Dictionaries: alphabetical lists of terms and their definitions that provide variant senses for each term, where applicable.
  - Glossaries: alphabetical lists of terms, usually with definitions.
  - Synonym Rings: sets of terms that are considered equivalent for the purpose of retrieval.

– Metadata-like Models
  - Authority Files: lists of terms that are used to control the variant names for an entity or the domain value for a particular field.
  - Directories: lists of names and their associated contact information.

  - Gazetteers: geospatial dictionaries of named and typed places.

– Classification and Categorization
  - Subject Headings: schemes that provide a set of controlled terms to represent the subjects of items in a collection and sets of rules for combining terms into compound headings.
  - Categorization Schemes: loosely formed grouping schemes.
  - Taxonomies: divisions of items into ordered groups or categories based on particular characteristics.
  - Classification Schemes: hierarchical and faceted arrangements of numerical or alphabetical notations to represent broad topics.

– Relationship Models
  - Thesauri: sets of terms representing concepts and the hierarchical, equivalence, and associative relationships among them. Thesaurus structures of this type are based on NISO Z39.19-2005 and ISO 2788 -1986 standards. Another type of the-

saurus (e.g., *Roget's Thesaurus*) represents only the equivalence (synonymy) of terms, with the addition of classification categories.

- Semantic Networks: sets of terms representing concepts, modeled as the nodes in a network of variable relationship types.
- Ontologies: specific concept models representing complex relationships between objects, including the rules and axioms that are missing in semantic networks.

2. Structures and characteristics of common KOS

Intending to fulfill fundamental functions, different types of KOS have been structured and implemented. These functions are: eliminating ambiguity, controlling synonyms, establishing relationships (hierarchical and associative), and presenting properties. The rest of this paper will introduce different types of KOS based on these functions. It is important to note that some of the structures enable a system to fulfill multiple functions.

2.1 Structures that focus on eliminating ambiguity

Ambiguity occurs in natural language when a word or phrase (a homograph or polyseme) has more than one meaning. Figure 2 provides an example and shows how a single word may be used to represent multiple and very different concepts. Without appropriate controls, these terms will result in poor precision in information retrieval.

There are different ways to eliminate ambiguity. Adding a qualifier to the term Mercury, e.g. "Mercury (automobile)", is one of the major methods used by almost every type of KOS, especially lists of subject headings and thesauri.
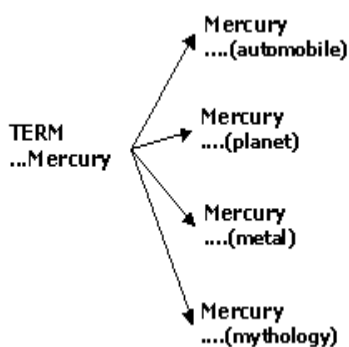

Figure 2. *Ambiguity caused by homographs and polysemes. Source: NISO 2005, 13*

Another approach to making a term's meaning clear is providing a context for the term. For example, for any of the following terms, the meaning is not clear:

Flying Horse, King Fisher, Royal Challenge

After seeing other terms listed in the cluster, the meanings of the terms in the whole group become clearer:

Heineken, Budweiser, Miller-Lite, Bud-Light

Now a heading is added to the group and a list is made, and the ambiguity is eliminated:

Drinks:
  Bud-Light
  Budweiser
  Flying Horse
  Hayward's 2000
  Heineken
  King Fisher
  Miller-Lite
  Royal Challenge
  Taj Mahal

This was a real situation the author encountered at an Indian restaurant in Columbus, Ohio. This kind of list is, in fact, a KOS structure that focuses on the function of eliminating ambiguity. A list (also called a "pick list") is a limited set of terms arranged in a simple alphabetical list or in some other logically evident way, such as chronological, numerical, etc. (NISO 2005). Lists are used to describe aspects of content objects or entities that have a limited number of possibilities. The defining characteristics of a pick list are that the terms:

– are all members of the same set or class of items (e.g., content type, language),
– are not overlapping in meaning, and
– are equal in terms of specificity or granularity (e.g., the geographic areas listed in Figure 3 do not mix continents with country or state names.)

Lists can be used effectively for both browsing and searching. In browsing, items are directly accessed when the list of terms is reviewed and one term is selected as in Figure 4.

Knowl. Org. 35(2008)No.2/No.3
M. L. Zeng. Knowledge Organization Systems (KOS)

163

Content Type	Geographic Area	Language	Target Audience
Book	Africa	Arabic	Parents
Brochure	Asia	Chinese	Students
Journal Article	Australia	English	Teacher
Report	Europe	French	
White Paper	North America	German	
	South America	Russian	
		Spanish	

Figure 3. *Examples of lists*

Figure 4. *Screenshot of the* Human Genome Resources *browsing page provided by the National Center for Biotechnology Information, National Library of Medicine.*
Source: http://www.ncbi.nlm.nih.gov/genome/guide/human/resources.shtml

Figure 5. *Screenshot of Google's advanced search. Source: http://www.google.com*

164

Knowl. Org. 35(2008)No.2/No.3
M. L. Zeng. Knowledge Organization Systems (KOS)

In searching, a list may be used to access content in a single term search, or the terms from the list may be used to limit a retrieved set by another attribute of interest for the user (one or more terms in the search). An example is Google's advanced search as shown in Figure 5. Several pick lists are provided for users to limit a retrieved set by choosing additional attributes such as language, format, time, location, and so on.

Lists are simple to implement, use, and maintain. They are frequently used to display small sets of terms that are used for narrowly defined purposes, such as a Web pull-down list or a list of menu choices.

2.2 Structures that focus on controlling synonyms or equivalents

In information retrieval, another major problem that affects search effectiveness is caused by the uncontrolled synonyms or equivalents, i.e., a concept is represented by two or more synonymous or words or phrases that can be considered as near synonymous (see Figure 6). This means that desired content may be scattered around an information space or database because it can be described by different but equivalent terminology. This is a common problem that results in poor recall during information retrieval.
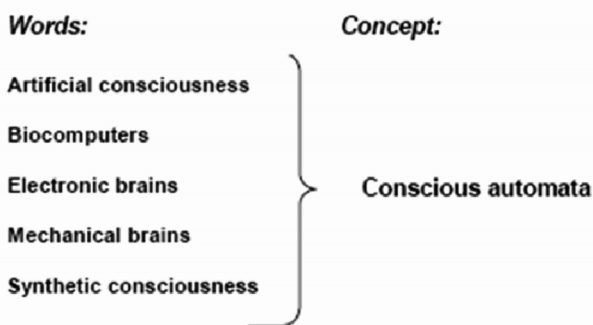

Figure 6. *Information scatter caused by synonyms. Source: NISO 2005, 13*

True synonyms include common and technical names, changes in usage of terms over time, terms from different languages, acronyms, and variant spellings. The most common problems, however, are the near synonyms whose meanings are generally regarded as different, but which are treated as equivalents for the purposes of a controlled vocabulary. The first situation includes overlapping concepts (such as *medicine* and *drugs*, *forest* and *woods*, *arid* and *dry*, etc.) Another situation may include antonyms or represent points on a continuum. For example:

sea water / salt water [variant terms] meteors / meteorites / meteoroids [points on a continuum] smoothness / roughness [antonyms] (NISO 2005, 45)

Information or content that is provided to a user should not be spread across a system with multiple access points, but should be gathered together at one point. Each distinct concept should refer to a unique linguistic form.

Libraries and information services have a history of creating authority files to establish forms of names (for persons, places, meetings, and organizations), titles, and subjects used in bibliographic records. An authority record is the record of authority decisions, all or some of which may be used in a system display. Basically, it is the process of reaching a consensus on the name(s) of an entity, making cross references from variant names, keeping track of those decisions, and displaying those decisions in information systems. A typical authority record using MARC format is illustrated in Figure 7.

```
000    nz n
001    435303
003    OCoLC
005    20021209141403.0
008    021209nneanz||babn n ana d
040    OCoLC  $b eng  $c OCoLC  $f fast
053  0 HF5548.32  $b HF5548.33
150    Electronic commerce
450    Cybercommerce
450    E-business
450    E-commerce
450    eBusiness
450    eCommerce
450    Internet commerce
450    Online commerce
550    Commerce
550    Information superhighway
688    LC usage 76; WC usage 468 (1999)
750  0 Electronic commerce  $0 (DLC)sh 96008434
```

Figure 7. An authority record for "**electronic commerce**" in the FAST Authority File. Source: FAST: Faceted Application of Subject Terminology. *http://fast.oclc.org/*

The authoritative term is recorded in field 150. Therefore, according to this record, *Electronic commerce* is the preferred term (or the established heading) while other terms recorded in field 450 (*Cybercommerce, E-business, E-commerce, eBusiness, eCommerce, Internet commerce*, and *Online commerce*) are treated as non-preferred terms, even though those headings have been used in documents as well.

Several authority files are well known. *The Union List of Artist Names (ULAN)* is a structured vocabulary containing more than 293,000 names with biographical and bibliographic information about artists and architects, including a wealth of variant names, pseudonyms, and language variants. *The Getty Thesaurus of Geographic Names (TGN)* is a structured, world-coverage vocabulary of over 1.1 million names, including vernacular and historical names, coordinates, place types, and descriptive notes, focusing on places important for the study of art and architecture. The *Library of Congress (LC) Authorities* has expanded to become the *Anglo-American Authority File (AAAF)* since 1994, holding several million name authority records for personal, corporate, meeting, and geographic names. The LC Cataloging Policy and Support Office announced recently that the number of subject authority records had reached 300,000 by the end of February 2007, making it by far the largest subject authority file in the world (PCC 2007). *FAST (Faceted Application of Subject Terminology)* adapted the *Library of Congress Subject Headings (LCSH)* with a simplified syntax. It retains the very rich vocabulary of *LCSH* while making the schema easier to understand, control, apply, and use. The headings have been built into *FAST* authority records. As of the end of March 2007, the *FAST* project had completed authority records for topicals, personal names (as subjects), corporate names (as subjects), geographics, periods, titles, events, and forms (*FAST* 2007).

Gazetteers can be regarded as a special kind of authority file. A gazetteer is a spatial dictionary of named and typed places. Originally (in the simplest case), a gazetteer is only the "index" in an atlas, providing the basic set of information (name, type, location) in this spatial dictionary. *The Getty Thesaurus of Geographic Names (TGN)* is also a gazetteer although constructed in a thesaurus format. With the development of digital libraries, digital gazetteers now have extended to become a service where relationships between places are represented inherently

Figure 8. *A record from the Alexandria Digital Library, reported in a standard format.*
Source: ADL Gazetteer http://middleware.alexandria.ucsb.edu/client/gaz/adl/index.jsp

166

Knowl. Org. 35(2008)No.2/No.3
M. L. Zeng. Knowledge Organization Systems (KOS)

through geospatial representations as well as through explicitly stated relationships such as "IsPartOf"; the schemes are extendable to the representation of events (e.g., hurricanes) and named time periods where the geospatial representations become time ranges. Digital gazetteers merge information about a place from multiple sources. A well-known digital gazetteer is the Alexandria Digital Library (ADL) project of the University of California at Santa Barbara (ADL Gazetteer Development [2002]). As a specialized type of KOS, it maps place names and types of places to map-based locations and thus integrates word-based georeferencing to map-based georeferencing. A standard report of an ADL record is displayed in Figure 8. Another output format of an ADL record uses XML (not shown here).

Name authority files, gazetteers, lists of subject headings, and thesauri must all compensate for the problems caused by synonymy by ensuring that each concept is represented by a single preferred term. The lists of subject headings and thesauri usually provide other synonyms and variants as non-preferred terms with USE references to the preferred term. The vocabulary control for the same set of terms shown in an authority record using MARC format (Figure 9) would be displayed in a thesaurus

with USE-UF references (Figure 10), where a preferred term is used for (UF) the non-preferred terms, while each non-preferred term becomes an entry term pointing to (i.e., USE) the preferred term.

Synonym rings, however, are an exception to the above rule. This different approach for controlling synonyms or equivalents should be given close attentions as well. While a synonym ring is considered a type of controlled vocabulary and has been written into the NISO Z39.50 standard, it plays a somewhat different role from other types of KOS. Unlike other KOS which are used during the indexing process, synonym rings are used only during retrieval. A synonym ring, therefore, is a set of terms that are considered equivalent for the purposes of retrieval (NISO 2005, 18). When a concept is described by multiple synonymous or quasi-synonymous terms, a synonym ring ensures that a set of documents will be retrieved as long as any one of the terms is used in a search. For example, a search for the activities of *astronauts* should be able to retrieve a set of documents that are indexed under *astronauts* as well as under *cosmonaut, taikonaut, spationaut*, and *spaceman*, while there is no requirement for picking one of them as the "preferred" term in searching. Rings

```
... ...
150   World War, 1939-1945
450   European War, 1939-1945
450   Second World War, 1939-1945
450   World War 2, 1939-1945
450   World War II, 1939-1945
450   World War Two, 1939-1945
```

Figure 9. *An established heading and its equivalent terms displayed in an authority record encoded with MARC format. Source: FAST:* Faceted Application of Subject Terminology. *http://fast.oclc.org/*

```
World War, 1939-1945
UF    European War, 1939-1945
UF    Second World War, 1939-1945
UF    World War 2, 1939-1945
UF    World War II, 1939-1945
UF    World War Two, 1939-1945

European War, 1939-1945
USE   World War, 1939-1945

Second World War, 1939-1945
USE   World War, 1939-1945

World War 2, 1939-1945
USE   World War, 1939-1945

World War II, 1939-1945
USE   World War, 1939-1945

World War Two, 1939-1945
USE   World War, 1939-1945
```

Figure 10. *The set of terms in Figure 9 displayed in a thesaurus. Source: Created by the author based on Figure 9*

can include all kinds of synonyms: true synonyms, misspellings, predecessors, abbreviations, near synonyms, etc. Sometimes the rings also contain terms that are more general or specific than other terms on the ring. For example, users may look for information regarding *cholesterol* with any of the following terms: *cholesterol, blood cholesterol, serum cholesterol, good cholesterol, bad cholesterol*, and *LDL*. An excellent example from another domain (Figure 11) is provided by Bedford (2006).

Synonym rings usually occur as sets of flat lists. Creating synonym rings involves going through word stocks and deciding what terms should be considered interchangeable when searching. Terms that are considered to form a synonym ring can be stored as a unit in a search system. A search using any term in the ring will retrieve all documents tagged as des-

ignated. Because users can be confused by results that do not actually include their keywords, interface design and an understanding of user goals become the keys for proper balance. A search interface may provide a clue about what terms are considered synonyms. In the following example, (Figure 12), after the term *silicon* is entered into the search box, a message will inform the searcher: *Your search was submitted as "SILICON" or "SI".*

Synonym rings are used to expand queries for content objects, especially in systems where the underlying content objects are left in their unstructured natural language format. Synonym rings are often used in conjunction with search engines and provide a minimal amount of control of the diversity of the language found in the texts of the underlying documents. Another important characteristic is that,
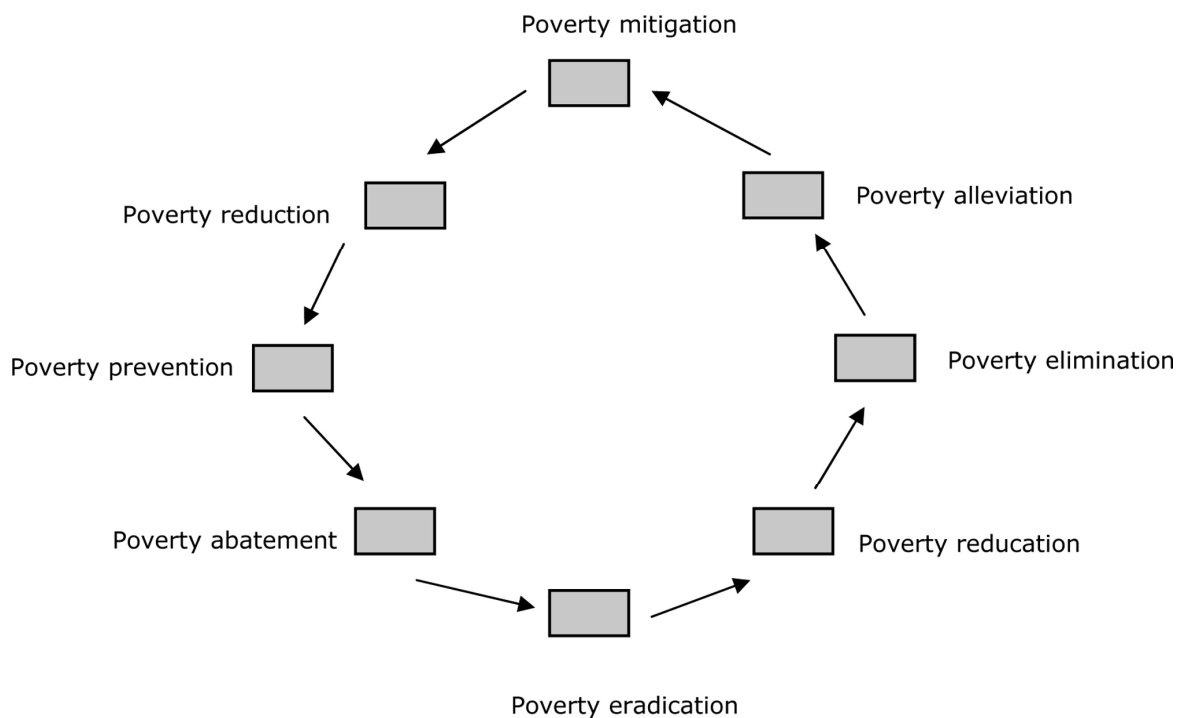

Figure 11. *An example of terms considered to form a synonym ring. Source: Bedford 2006, modified August 7, 2007*

Figure 12. *A search interface showing the submitted synonyms after a search term is entered. Source: Leise et. al. 2003*

168

Knowl. Org. 35(2008)No.2/No.3
M. L. Zeng. Knowledge Organization Systems (KOS)

unlike other KOS which require large investment up front and usually take a long time to build, synonym rings can be built on-demand, according to user needs, in a timely fashion. Search logs of any time period are one of the best sources for building effective synonym rings. Other sources are dictionaries, authority files, and lexical databases.

To increase effectiveness (including recall and precision), a system needs to implement a one-to-one principle, i.e., each term has only one meaning and only one term may be used to represent a given concept or entity in a search. Information or content that is provided to a user should not be spread across the system under multiple access points, but should be gathered together at one entry point. KOS types introduced in the above two sections fulfill these basic functions.

2.3 Structures that focus on making explicit semantic relationships

2.3.1 Hierarchical relationships

The use of hierarchical relationships is the primary feature that distinguishes a taxonomy or a thesaurus from other simpler forms such as lists and synonym rings. Hierarchical relationships are based on degrees or levels of superordination and subordination (NISO 2005, Iyer 1995). Classes at the same level of division are described as coordinate. Equal classes may be grouped together into higher level classes which are superordinate to the original classes. A class may be divided into a number of subclasses, where each subclass is a subset of the original class. This process may be repeated and the subclasses divided into a lower level of subclasses. Classes at the same level of division share a set of common properties inherited from the parent class. In the following example, levels of classes are indicated through indentation.

```
superordinate classes (e.g., parents)
    coordinate classes (e.g., siblings)
        subordinate classes (e.g., children)
        subordinate classes
    coordinate classes
    coordinate classes
        subordinate classes
```

When represented by terms, every subordinate term should refer to the same basic kind of concept as its superordinate term; that is, both the broader and the narrower term should represent a thing, an action, a property, etc. For example:

– anatomy (a discipline) and central nervous system (a body part that can be an object of study of that discipline) represent different kinds of concepts; therefore, these terms cannot be related hierarchically;
– central nervous system and brain both represent body parts; these terms can therefore be related hierarchically (NISO 2005, 47).

Hierarchical relationships cover three logically different and mutually exclusive conditions: generic relationships, instance relationships, and whole-part relationships.

1. The *generic relationship* identifies the link between a class and its members or species. This type of relationship is often called "IsA" and is specified as "KindOf." A simple way to apply the test for validity described above is to formulate the statement "[narrower term] is a [broader term]." For example, a *boot sector virus* is a kind of computer virus (*Viruses (computer)*).

    Viruses (computer)
        Boot sector viruses
        Companion viruses
        Email viruses
        Logic bombs
        Time bombs
        Macro viruses
        Sentinels
        WB Microworm
        Cross-site scripting virus

2. The *instance relationship* identifies the link between a general category of things or events, expressed by a common noun, and an individual instance of that category, often a proper name. This type of relationship is also known as an "IsA" relationship and expressed as "InstanceOf." For example, *Mydoom* and *ILOVEYOU* are two instances of computer worms (*Worms (computer)*), expressed by proper names.

    Worms (computer)
        Mydoom
        ILOVEYOU

3. The *whole-part relationship* covers situations in which one concept is inherently included in an-

Knowl. Org. 35(2008)No.2/No.3
M. L. Zeng. Knowledge Organization Systems (KOS)

169

other, regardless of context, so that the terms can be organized into logical hierarchies, with the whole treated as a broader term. This relationship can be applied to several types of terms such as geographical names and hierarchical organizational structures. The relationship is still known as an "IsA" and is usually specified as "part of." In the following example, parts are indicated through indentation. In a personal computer there is a *motherboard* or system board with slots for expansion cards and holding parts such as *Central processing unit (CPU)* and *Random Access Memory (RAM)*.

    Motherboard
        Central processing unit (CPU)
        Computer fan
        Random Access Memory (RAM)
        Basic Input-Output System (BIOS)
        Buses

In addition, some concepts belong, on logical grounds, to more than one category. They are then said to possess polyhierarchical relationships. For instance, *pianos* would be a subordinate term of both *stringed instruments* and *percussion instruments* (NISO 2005, 50).

A taxonomy is a type of KOS which consists of preferred terms, all of which are connected in a hierarchy or polyhierarchy. The original use of the term taxonomy has its roots in the work of Carolus Linnaeus, who grouped biological species according to shared physical characteristics. These groupings have since been revised with the advancement in science (Cain 1959). Today, the term taxonomy is applied in a broader and more general sense and now may refer to the classification of things, as well as to the principles underlying such a classification. In building classificatory structures people partition areas of knowledge into groups or classes, and further partition each group into smaller sets, continuing this process of successive division until the scheme is as specific as required.

The process of classifying suggests not only the scientific aspects of the scientific taxonomy, but also its cognitive aspects. It is generally believed that basic-level categories exist in abstraction (Rosch 1978). Categories can be organized into a hierarchy from the most general to the most specific. However, the level that is most cognitively basic is "in the middle" of the hierarchy: a category which is a family of events, objects, patterns, emotions, spatial relation-

ships, or social relationships that are cognitively basic. Examples would include "dog," "chair," "ball," and "cup." This is the level first named and understood by children: the level at which subjects are fastest identified as category members, and the highest level at which a single mental image can reflect the entire category. It is at this level that most of our knowledge is organized.

In constructing taxonomies, both scientific aspects of categorization and cognitive aspects of categorization need to be taken into account. A related and important principle of constructing any KOS is selecting and testing under the assumption of three warrants:

– the natural language used to describe content objects (*literary warrant*),
– the language of users (*user warrant*), and
– the needs and priorities of the organization (*organizational warrant*) (NISO 2005, 16).

The Tree of Life web project (http://www.tolweb.org/tree/) gives a very good example of using a classificatory structure to represent knowledge. In the following screenshot (Figure 13), a tree diagram provides an overview of the phylogenetic relationships among subgroups, which allows a visitor to move up the branches of the tree of life all the way to leaf pages.

Figure 14 shows a different display, also by the Tree of Life Web Project, in which the information is presented in a way most users can immediately understand based on the "most popular groups" of life.

In libraries and information services, there is already a long history of using classifications. They have established hierarchical or faceted structures and used numeric or alphabetic notations to represent broad topics. Famous universal classification schemes include the *Dewey Decimal Classification* (*DDC*), the *Universal Decimal Classification* (*UDC*), and the *Library of Congress Classification* (*LCC*). Many specialized classification schemes have also been developed and widely used in different subject domains, such as the *NLM Classification* of the National Library of Medicine.

Nowadays, the taxonomy approach is being applied to many domains and disciplines. With or without notations, these structures have fully employed classificatory principles and hierarchical relationships to represent the knowledge of a domain. Some KOS are attempting to provide a high level taxonomic organization from which many efforts may benefit.
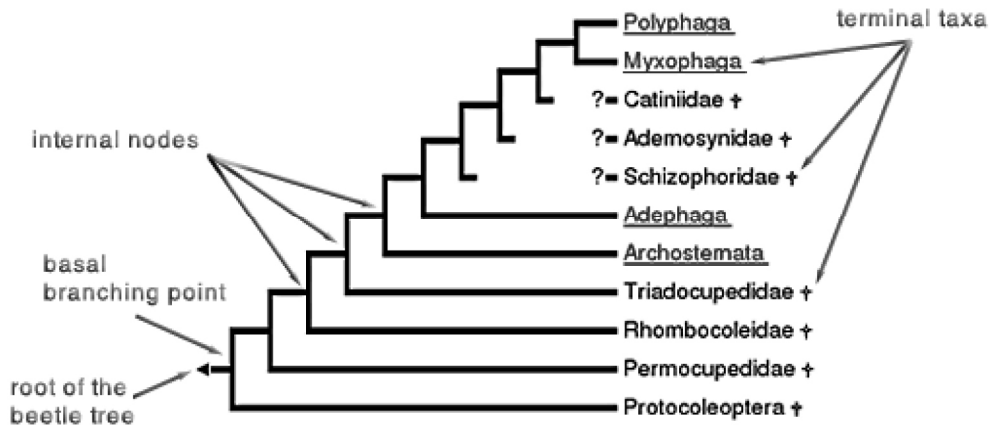
Figure 13. *The tree diagram on the beetle (Coleoptera) page showing the relationships between the major beetle subgroups. Source: http://tolweb.org/tree/home.pages/structure.html ©Tree of Life Web Project.*

Figure 14. *Tree of Life project's "Popular Groups display." Source: http://tolweb.org/tree/home.pages/popular.html ©Tree of Life Web Project.*

Figure 15. *A screenshot of the Open Directory Project's main categories. Source: http://dmoz.org/*

The *UNSPSC* (*The United Nations Standard Products and Services Code®*) offers a global electronic coding convention that intends to arrange the entire universe of products and services into over ten thousand hierarchical categories according to a five-level umbrella structure and numbering system, in order to facilitate and standardize spending analysis, finding and purchasing, and product awareness and discovery in the global marketplace (*UNSPSC* 2001).

The terms taxonomy, classification, and categorization have been used interchangeably by different disciplines and professions. An "unofficially" classified group of products is called categorization schemes which consist of loosely formed grouping schemes. The Open Directory Project's scheme is a good example of a comprehensive human-edited directory of the Web (Figure 15). It is constructed and maintained by a vast, global community of volunteer editors.

2.3.2 Associative Relationships

Hierarchical relationships are probably the most commonly recognized relationships in KOS. Beyond them are associative relationships, which cover relations between terms that are neither equivalent nor hierarchical, yet the terms are semantically or conceptually associated and co-occurring so that the link between them should be made explicit in the controlled vocabulary. The grounds for explicit links between such terms are that additional terms may be suggested for use in indexing or retrieval (NISO 2005).

In general, associative relation links are established among the terms belonging to different hierarchies (Figure 16). Most commonly considered associative relationships fall into these categories (Lancaster 1986; NISO 2005; Aitchison 2000):

Relationships	Examples
Cause/Effect	Accident/Injury
Process/Agent	Velocity Measurement/Speedometer
Action/Product	Writing/Publication
Action/Patient	Teaching/Student
Concept or Thing/Properties	Steel Alloy/Corrosion Resistance
Thing or Action/Counter-Agent	Pest/Pesticide
Raw Material/Product	Grapes/Wine
Action/Property	Communication/Communication Skills
Antonyms	Single People/Married People

Figure 16. *Examples of associative relationships*

Associative relations can also be established among sibling terms with overlapping meanings, such as *ships* and *boats*, where each of the terms can be precisely defined (so they do not form an equivalence set), yet they are sometimes used loosely and almost interchangeably (NISO 2005, 52-53).

By definition, "[a] thesaurus is a controlled vocabulary arranged in a known order and structured so that the various relationships among terms are displayed clearly and identified by standardized relationship indicators (NISO 2005, 18)." Here "various relationships" include the hierarchical relationships and associative relationships we have discussed so far.

Thesauri are the most typical form of controlled vocabulary developed for use in indexing and searching applications because they provide the richest structure and cross-reference environment. Thesauri are helpful to both indexers and searchers who need to discover the most appropriate and specific terms for their purposes.

Figure 17 shows an example from the *Thesaurus for Liquid Crystal Research and Applications*. The left side box gives an extracted hierarchical structure which is exactly like a taxonomy. It is two-dimensional, allowing a user to explore the terms through hierarchies. The hierarchical relationships are presented as narrower terms (NT) in the thesaurus entry on the right side box. The thesaurus also introduces another dimension by establishing networks among terms beyond hierarchies (see RT terms in Figure 17).

The entry for the term LIQUID CRYSTAL PHASES which shows the equivalent relationship (used-for terms (UF)), hierarchical relationship (narrower terms (NT)), as well as associative relationship (related terms (RT)) provides a clear picture about the individual term. A term's meaning is usually made clear through a scope note (SN). In thesauri, relationship indicators are usually employed reciprocally. A strong structure builds a strong network.

More and more Internet search engines tend to adopt the idea of displaying and suggesting related topics in the search results display as well. Searching "global warming" in both Yahoo! and Google will
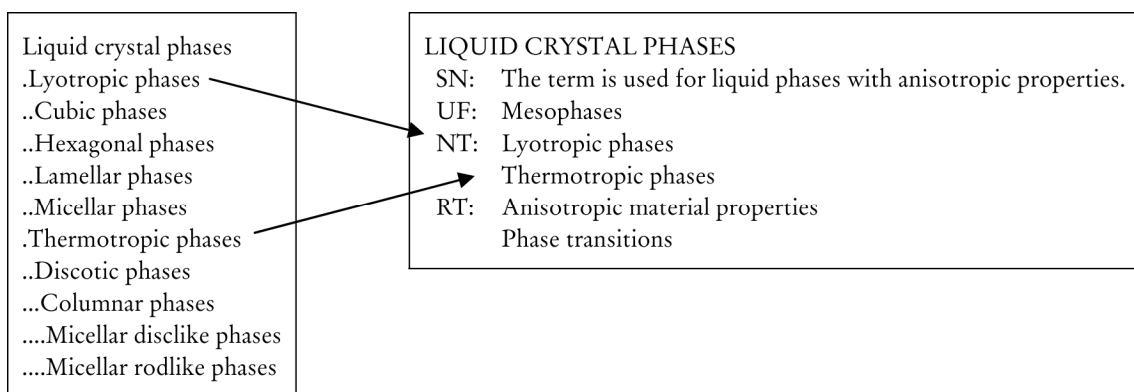
Figure 17. *An example of exploring one term's multiple dimensions. Source:* Thesaurus for Liquid Crystal Research and Applications

Also try: causes of global warming, global warming articles, effects of global warming, global warming pictures, global warming solutions, greenhouse effect global warming, global warming newspaper articles, al gore global warming, definition of global warming, global warming hoax

Figure 18. *"Also try" terms suggested by Yahoo! for the "global warming" search. Source: http://www.yahoo.com/*

Knowl. Org. 35(2008)No.2/No.3
M. L. Zeng. Knowledge Organization Systems (KOS)

173

Figure 19. *"Searches related to" suggested by Google for the "global warming" search.*
Source: http://www.google.com/

obtain a set of related topics (both hierarchical and associative) that one may further explore. Yahoo! provides an expendable list of terms under its "Also try" label on the top of the screen after a search term is entered (Figure 18).

Although most of the terms suggested by the two search engines contained the same terms as the query ("global warming"), Google did return links to the movie "an inconvenient truth" and a narrower term "greenhouse effect"; neither of these results contained the words used in the query (Figure 19).

When talking about thesauri, it is necessary to discuss lists of subject headings. Nowadays the lists of subject headings are presented similarly to thesauri and even the labels of relationships (NT, BT, RT) may be the same. A list of subject headings is a set of controlled terms to represent the subjects of items in a collection. They can be extensive, covering a broad range of subjects, e.g. the *Library of Congress Subject Headings* (*LCSH*). Typically, their structure is generally shallow and has a limited hierarchy. They also tend to be pre-coordinated, with rules for how subject headings can be joined to provide more specific concepts. *Medical Subject Headings* (*MeSH*) is an other widely used list of subject headings. Because of its comprehensive tree structure, it has a stronger structure than most subject headings lists. Sometimes it is regarded as a thesaurus even though it has restricted rules for pre-coordinating sub-headings in applications.

Within a thesaurus, faceted structures can be employed to overcome the problems of traditional systematic classification structures in which the central process is choosing the characteristics to divide knowledge by as well as the order in which to use them. Together, the chosen characteristics and sequence determine the structure of a classification scheme. In other words, those characteristics and se-

quences that are not chosen and reflect different views and needs may be ignored, although some modern classification schemes also have employed limited facets. A thesaurus' post-coordinating nature already helps to reduce such problems. Moreover, a faceted approach employed in a thesaurus provides the most flexible structure to represent the many aspects of a knowledge domain. For example, the narrower terms for 'flowers', as shown in this entry, (Figure 20), are grouped according to two criteria: by plant type or by flowering season (NISO 2005, 61).

Figure 20. *Displaying narrower terms with node labels.*
Source: NISO 2005, 61

Here two node labels are used to group both sets of narrower terms in categories. Although displayed in the hierarchies, they are not to be used in indexing or searching, therefore they are distinguished from terms by placing them in square brackets.

Figure 21. *A screenshot of an AAT online full record display. Source: Art and Architecture Thesaurus Online, http://www.getty.edu/research/conducting_research/vocabularies/aat/*

In the *Art and Architecture Thesaurus*, the whole thesaurus is constructed on seven facets and many sub-facets are used inside of classes at different levels (Figure 21). It provides the most flexible structure to represent the many aspects of knowledge in the domain of art and architecture.

While semantic networks establish relationships like taxonomies and thesauri would, they also define types of entities and relationships more specifically. Semantic networks organize sets of terms representing concepts, modeled as the nodes in a network of variable relationship types.

The *Unified Medical Language System* (*UMLS*) has specified 135 semantic types and 54 relationships (*UMLS* 2004a). The top level types are *Entities* (including *Physical Object* and *Conceptual Entity*) and *Events* (including *Activity* and *Phenomenon or Process*). The primary link in the network is the "*isa*" relationship link. This establishes the hierarchy of types within the network and is used for deciding on the most specific semantic type available for assignment to a concept in the *Metathesaurus* of the *UMLS*

(*UMLS* 2004b). In addition, a set of non-hierarchical relations between the types has been identified. These are grouped into five major categories as the "*associated_with*" relationships, which are actually relations themselves. These include: "physically related to," "spatially related to," "temporally related to," "functionally related to," and "conceptually related to" (UMLS 2004a). Figure 22 is a portion of the *UMLS* Semantic Network presented in a graph.

A visualized interface, *Visual Concept Explorer*, which is developed by Lin and Aluker (2004) and built on the vocabularies provided by *UMLS*, may be helpful to explain the advantages of specifying types of concepts in an effort to understand complicated topics. Figure 23 recorded a search process conducted recently by the author. When the term "herbs" (a non-*MeSH* term) was entered, the system first provided a list of *MeSH* headings that were found in the top 200 documents in PubMed that matched this term. After selecting a particular *MeSH* heading from this list, the interface provided an interactive concept map with lines and nodes as well as
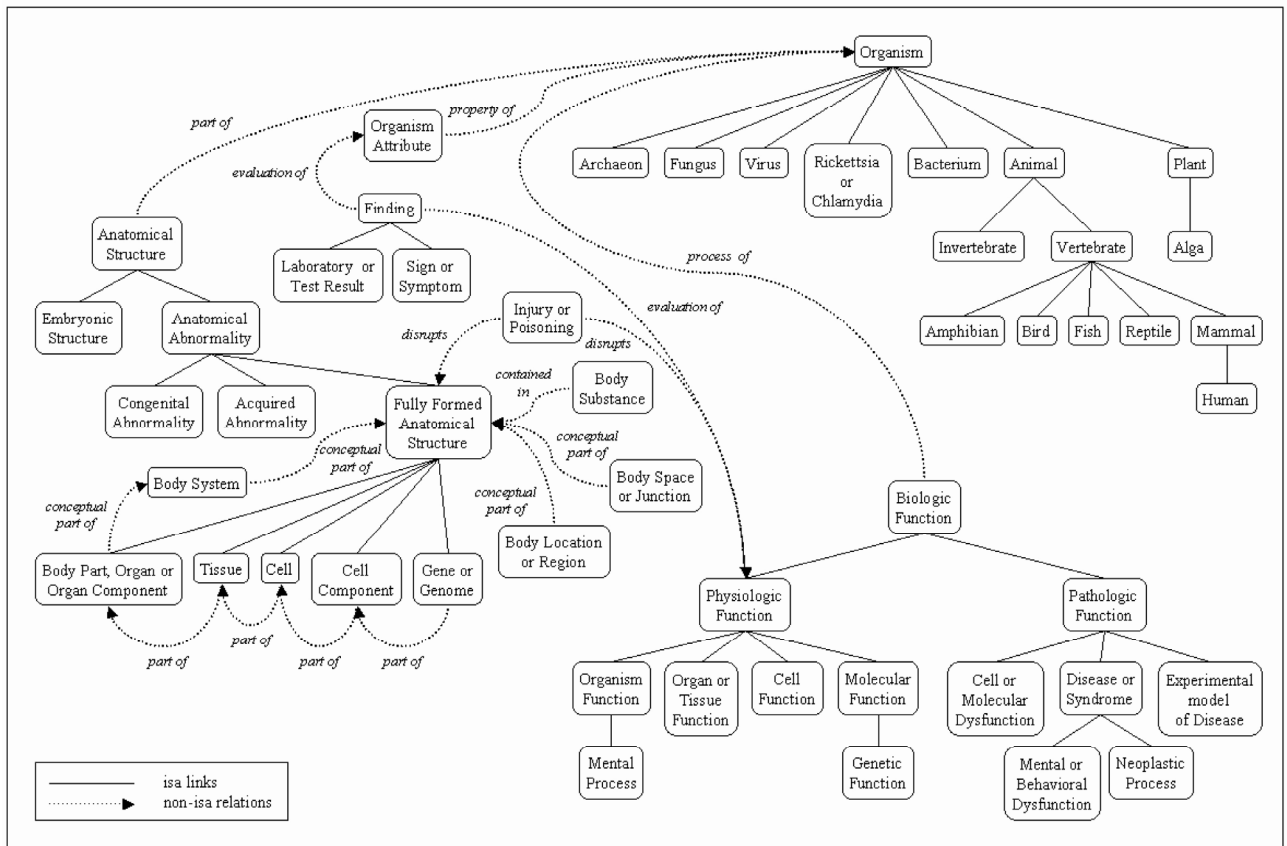
Figure 22.   *A portion of the UMLS® Semantic Network of the National Library of Medicine.*
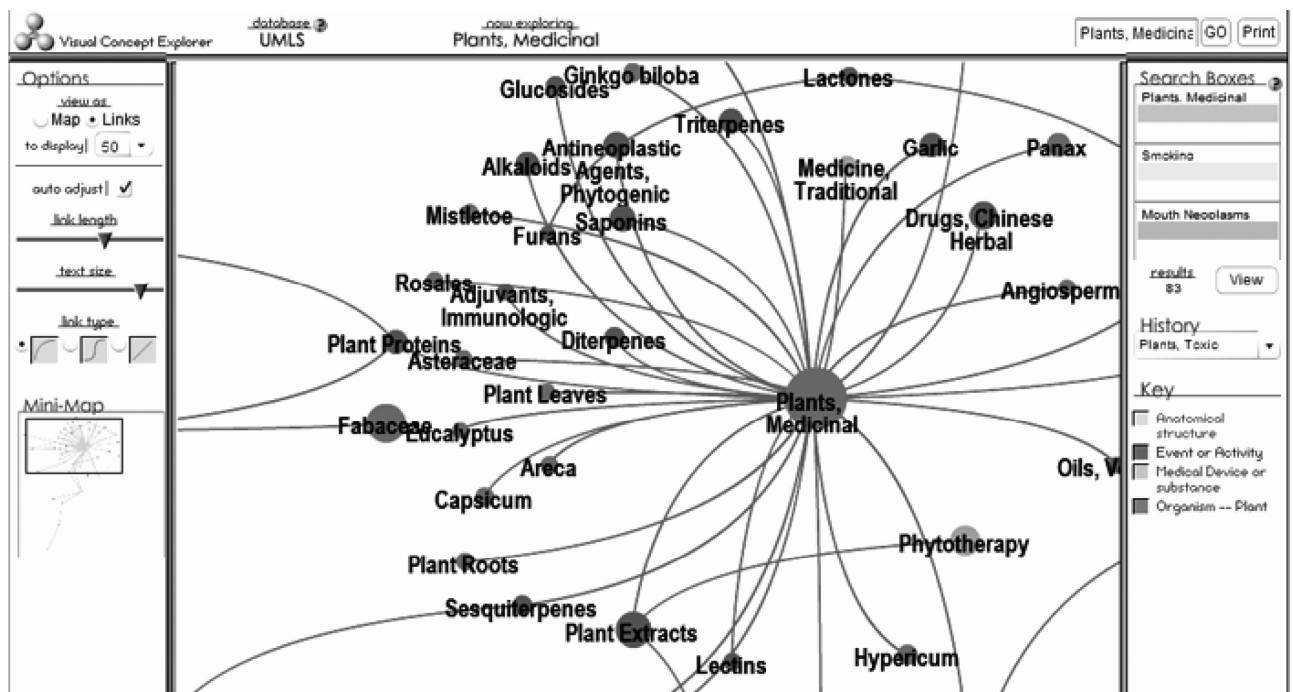*Source: http://www.nlm.nih.gov/research/umls/META3_Figure_3.html*

Figure 23. *A screenshot of the Visual Concept Explorer. Source: http://cluster.cis.drexel.edu/vce/*

176

Knowl. Org. 35(2008)No.2/No.3
M. L. Zeng. Knowledge Organization Systems (KOS)

a geo map with zones. *Visual Concept Explorer* uses different colors to represent different types of concepts. According to the sequence appearing under "Key" in Figure 23, they are: light green for *anatomic structure*, magenta for *event or activity*, dark green for *medical device or substance*, orange for *organism*, and brown for *phenomenon or process*. Thus in this search, "Plants, Medicinal" was marked with an orange circle (for *organism)*; "Smoking" was marked in magenta (for *event or activity)*; and "Mouth Neoplasm" was marked in brown (for *phenomenon or process)*. By exploring different maps and right-mouse-clicking a particular *MeSH* term to load into one of the three search boxes located on the upper-right corner of the screen, the number of hits responding to the query (in this case, 83) was reported under the search boxes. A further click of "view" would bring a visitor to the PubMed search results. With a better understanding of the types of concepts one is looking for, it is much easier to navigate amid the terms, modify search strategies by adding or changing particular types of concepts, and monitor the changes of search results according to the changed concept types.

Note that both examples in Figures 22 and 23 use concept maps to present information and semantic relations. A concept map is a visual representation of concepts and their relationships. Figure 22 demonstrates a typical concept map that consists of nodes (points/vertices) that represent concepts and links (arcs/edges) that represent the relations between concepts. The links can be labeled and denote direction with an arrow symbol (non-, uni- or bi-directional) that describes the direction of the relationship. Concept maps can be used to represent any type of KOS structures, containing simple or complicated relationships.

FACET (*Faceted Access to Cultural Heritage Terminology*), a terminology service prototype, has been developed at the Hypermedia Research Unit, University of Glamorgan (UK). The project has explored the potential of semantic expansion in search and browsing based on faceted thesaurus relationships (Tudhope 2006). All terms in the query expansion interface are from the *Art and Architecture Thesaurus*. Here different types of concepts are again marked with different colors (indicated according to the order under "Legend" in Figure 24). They are: blue for properties, teal for time, purple for agents, red for processes, gold for materials, and green for objects). Figure 24 shows the whole steps used in making this example by the author: (1) find a term

in the thesaurus, (2) add terms one by one to the query boxes, (3) run query, and (4) view matching items. Colors are displayed for all of the terms appearing in the term selection and view boxes (left side), the query term boxes (right side), and the results display box (at the bottom).

2.4 Structures that present both semantic relationships and properties

The KOS class has been extended since the introduction of the term ontology to knowledge acquisition, representation, and organization fields by communities other than philosophy and library and information science. The definition of ontology is still being debated and the use of this term has been varied, particularly during the beginning years when the term entered into the main stream of the World Wide Web. A widely accepted explanation is that ontology is a formal, explicit specification of a shared conceptualization. It is a specification of a representational vocabulary for a shared domain of discourse— definitions of classes, relations, functions, and other objects (Gruber 1993, Studer et al. 1998). At implementation level, many ontologies published on the Web not only represent complex relationships between objects, but also include the rules and axioms.

Ontology embraces the classificatory structure used by taxonomies and thesauri. Its unique feature is the presentation of properties for each class within the classificatory structure. With a full taxonomy and exhaustive properties, an ontology functions as both a conceptual vocabulary and a working template which allows for storing, searching, and reasoning that is based on instances and rules. A project reported by Wielinga et al. (2001) built an ontology prototype based on the existing *Art and Architecture Thesaurus* and Visual Resource Association's (VRA) Core Categories metadata element set version 3.0. The purpose was to create a knowledge-rich description of art objects using Protégé-2000 software (http://protege.stanford.edu/). The ontology contained a taxonomy of furniture and a template showing the properties of class "furniture". This template includes the 17 VRA Core metadata elements and eight additional elements defined by the project.

The Foundational Model of Anatomy (*FMA*) ontology is another excellent example of a domain ontology that represents a coherent body of explicit declarative knowledge about human anatomy. Using the Protégé ontology editor, anatomical classes ranging from macroscopic to molecular levels are organized

hierarchies. According to project documentation (*FMA* [2006]), the *FMA* consists of 75,000 anatomical classes, 130,000 unique terms, over 205,000 frames, and 174 unique slots showing different types of relations, attributes and attributed relationships. There are over 44,000 English synonyms, of a class' preferred name, as well as more than 15,000 non-English equivalents. The relationship network of the *FMA* contains more than 2.5 million relationship occurrences. Over one million of these occur in classes, of which 450,000 relate classes directly to other classes. This symbolic modeling of the structure of the human body is in a format that is understandable

by humans and is also navigable and interpretable by machine-based systems. In the following figure, the concept "ear" is presented in a hierarchy on the left side. The properties of "ear" and the facts (instances) are given in detail on the right side (Figure 25).

Properties in a knowledge model are represented with "slots" in an ontology editor such as the one used in the above example. Slot attributes and slot relationships of a class or instance collectively define the frame. Every slot is given a name that identifies the relationship. In Protégé, slots are attached to frames in two distinct ways: a) "own slots" and their values describe the relationships and attributes that
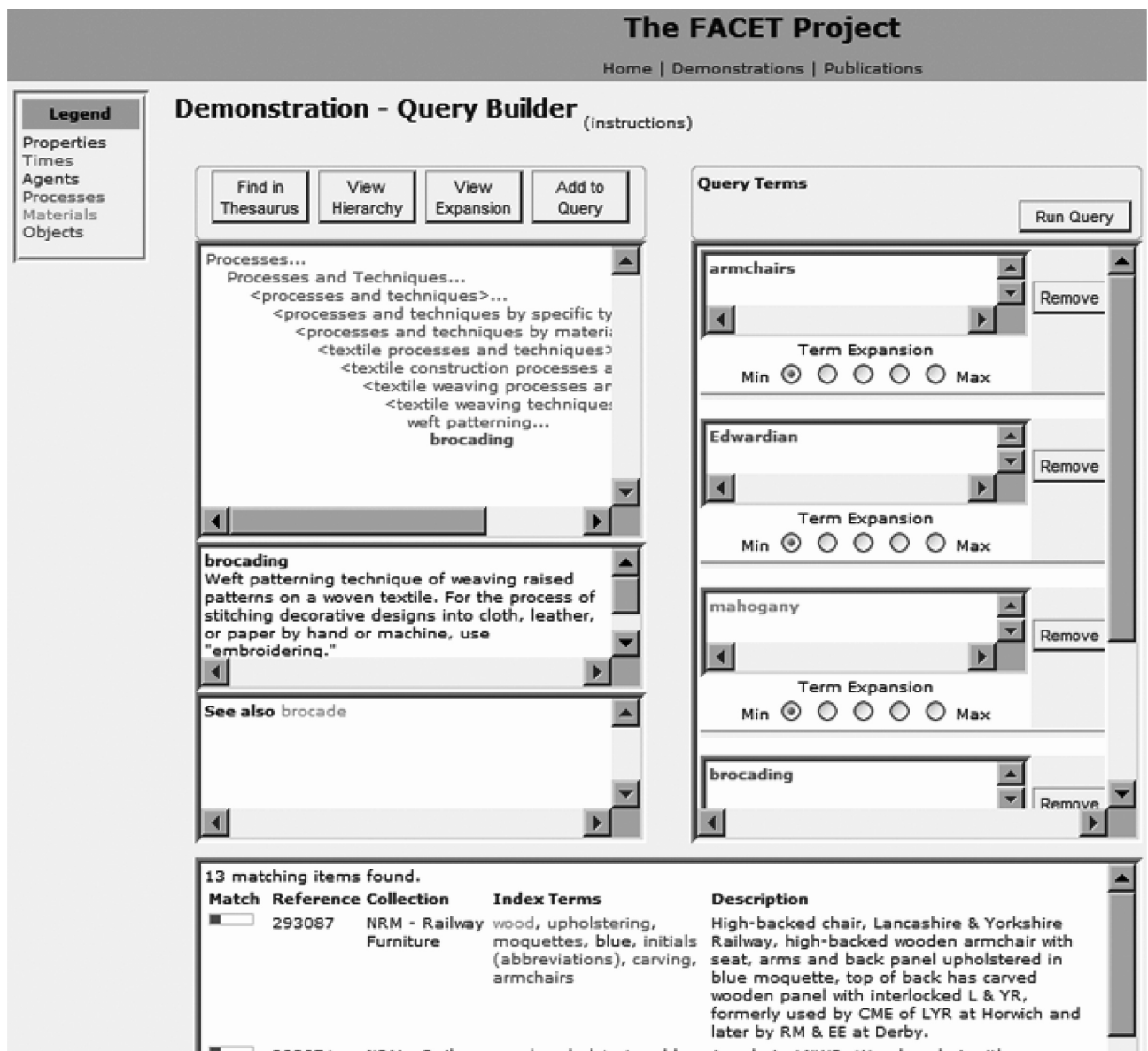
Figure 24. *A screenshot illustrates the thesaurus-based semantic query expansion in a prototype Web application. Source:*
http://www.comp.glam.ac.uk/~FACET/webdemo/default.htm
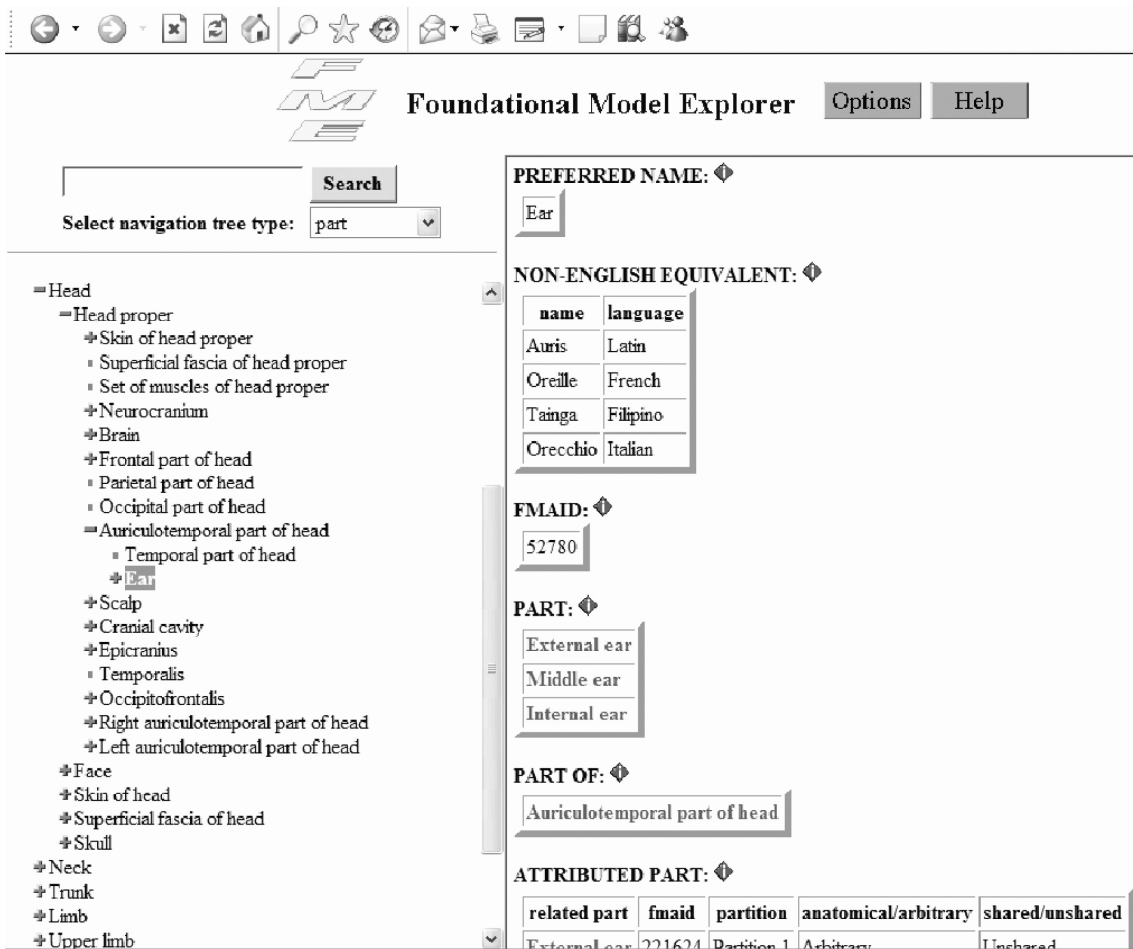
Figure 25. *Browsing terms through FME Foundational Model Explorer. Source: http://fme.biostr*
           .washington.edu:8089/FME/index.html, Foundational Model of Anatomy (FMA) On-
           tology, Structural Informatics Group, University of Washington.

Figure 26. *A snapshot of a subset of slots attached to the class Heart (from the classes-tab), in Pro-*
           tégé editor. Source: FME 2007, http://sig.biostr.washington.edu/projects/fm/FAQs.html,
           Foundational Model of Anatomy (FMA) Ontology, Structural Informatics Group,
           University of Washington

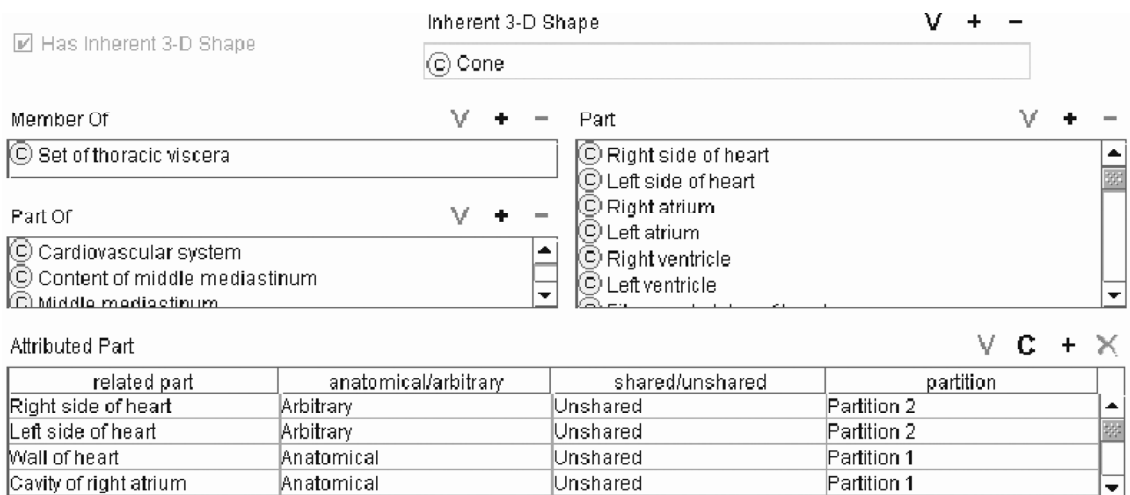pertain to the frame on which they are attached; and, b) "template slots" represent the attributes/relationships (and possibly values) that will be propagated to all of their instance frames. Only frames that represent classes have template slots, as illustrated in Figure 26, where a subset of slots attached to the class *Heart* include "member of," "part of," "part," "inherent 3-d shape," etc.

One of the fundamental characteristics of ontologies is their function for recording instances, such as a gene product, which follow the rules of logical reasoning. An example of this kind is the *Gene Ontology* (GO) which describes genes and gene products. According to the Gene Ontology Consortium (1999, 2000), the GO project has developed three structured controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components, and molecular functions that are species-independent. A gene product might be associated with, or located in, one or more cellular component; it is active in one or more biological process(es) during which it may perform one or more molecular function(s). An annotation of

gene products entails linking associations between the ontologies and the genes/gene products in the collaborating databases. The ontologies are structured so that they can be queried at different levels. For example, one can use GO to find all the gene products in the mouse genome that are involved in signal transduction, or one can zoom in on all the receptor tyrosine kinases. The structure also allows annotators to assign properties to genes or gene products at varying levels—depth dependent—based on knowledge about that entity.

An interesting statement in a GO document is that although the ontologies are structured similarly to regular hierarchies, they differ in that a "child", or more specialized term, can have many "parents", or less specialized terms. Every GO term must obey the true path rule: if the "child" term describes the gene product, then all its "parent" terms must also apply to that gene product (Gene Ontology Consortium 1999). The following three screenshots show the results after searching "chronological cell aging". In addition to the synonyms, definitions, belonging ontologies, and other basic information,
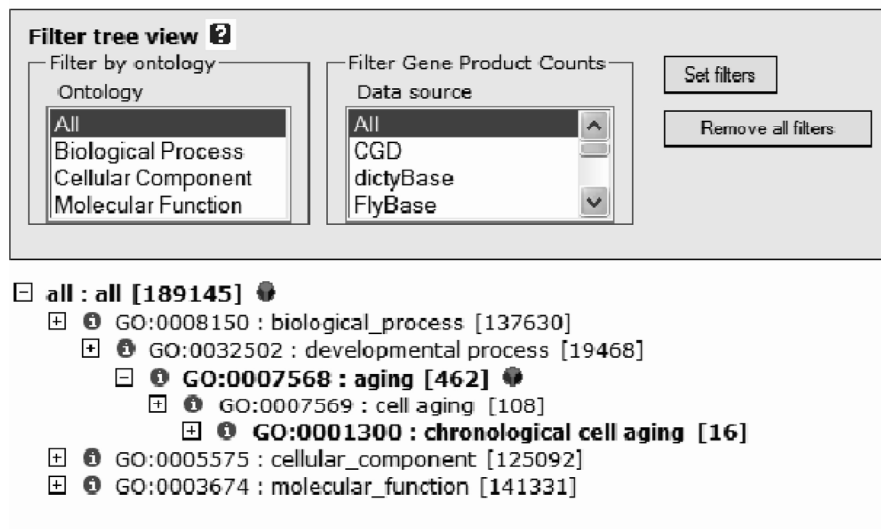

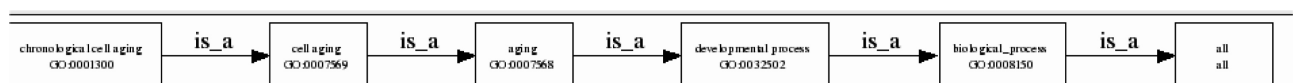Figure 27. *The* Tree Browser *view. Source: http://www.geneontology.org/GO.doc.shtml, Gene Ontology Consortium*

Figure 28. *The* Graph *view of relations. Source: http://www.geneontology.org/GO.doc.shtml, Gene Ontology Consortium*

180

Knowl. Org. 35(2008)No.2/No.3
M. L. Zeng. Knowledge Organization Systems (KOS)

Term Name	Total Gene Products	Percent of all
molecular_function ; GO:0003674	141331	74.7
biological_process ; GO:0008150	137630	72.7
cellular_component ; GO:0005575	125092	66.1
cell aging ; GO:0007569	108	0.05
obsolete_cellular_component ; obsolete_cellular_component	0	0
all ; all	0	0
obsolete_biological_process ; obsolete_biological_process	0	0
obsolete_molecular_function ; obsolete_molecular_function	0	0
All all	189145	100.0 %

Figure 29.   *View* Total Gene Products and Percent of all *(also accompanied by a pie graph, not showing here). Source: http://www.geneontology.org/GO.doc.shtml, Gene Ontology Consortium*

there are options of viewing the item through a tree browser (Figure 27), a concept map (Figure 28), and related gene products and the percentage of all (Figure 29).

3. Conclusion

Various types of KOS have been discussed in this article, with examples of KOS instances. When looking at the structures, one can see simple flat structures such as pick lists and synonym rings, two-dimensional structures such as those employing hierarchies, and multiple-dimensional structures which build networks based on various semantic types and semantic relationship types. Employing the underlying principles of KOS, one can understand those structures that focus on fulfilling primary functions: eliminating ambiguity, controlling synonyms or equivalents, establishing explicit semantic relationships such as hierarchical and associative relationships, and presenting both relationships and properties of concepts in the knowledge models. The more complex structures usually carry most or all of the functions.

With the research and development of the new generation Web, represented by the Semantic Web and Web 2.0 movement, all knowledge organization systems have one common concern: in the networked environment, KOS must become machine-understandable, not just machine-readable. This article does not address the enabling technologies such as the encoding standards XML, SKOS (Simple Knowledge Organization System), and OWL Web Ontology Language that will allow this to occur;

however, very soon they will be embedded with all the KOS products. Another significant trend is that KOS is not used in isolation. Various structures have been integrated into web-based services. They are used not only for organizing, indexing, cataloging, and searching, but also in learning, knowledge modeling, reasoning, and many other environments. The KOS in the networked environment do inherit most of the structures that the world has witnessed for at least a hundred years, yet networked knowledge organization systems/services/ structures are not simply a repetition of the past. They are forming new semantic structures that will function with a greater impact far more extensive then imagined.

References

ADL Gazetteer development. [2002]. Alexandria Digital Library Project. University of California, Santa Barbara. Last updated 4 June 2004. Available at http://www.alexandria.ucsb.edu/gazetteer.

ADL Gazetteer server client. [2002]. Alexandria Digital Library Project. University of California, Santa Barbara. Available at http://middleware.alexandria.ucsb.edu/client/gaz/adl/index.jsp.

Aitchison, Jean. 2000. *Thesaurus construction and use: a practical manual*. 4th ed. London: Fitzroy Dearborn.

Bedford, Denise. 2006. Ontologies, taxonomies and search. Presentation at the Special Libraries Association Annual Conference, Baltimore, Maryland, June 2006. Available at http://units.sla.org/division/dsoc/Conference%20Archive/D_Bedford_OntologiesSLA2006.ppt.

Cain, Arthur James. [1959]. *Function and taxonomic importance*. London. Systematics Association.

FACET - Faceted Access to Cultural Heritage Terminology. [2006]. The Hypermedia Research Unit, University of Glamorgan, UK. Available at http://www.comp.glam.ac.uk/~FACET/webdemo/.

FAST: Faceted Application of Subject Terminology. [2007]. OCLC Online Computer Library Center. Available at http://www.oclc.org/research/projects/fast/.

FMA. [2006]. *Foundational Model of Anatomy Ontology*. School of Medicine, University of Washington. Available at http://sig.biostr.washington.edu/projects/fm/AboutFM.html.

The Gene Ontology (GO). 1999. The Gene Ontology Consortium. Available at http://www.geneontology.org/.

The Gene Ontology Consortium. 1999. An Introduction to the Gene Ontology. Last modified 22 January 2007. Available at http://www.geneontology.org/GO.doc.shtml.

The Gene Ontology Consortium. 2000. Gene Ontology: tool for the unification of biology. *Nature Genet* 25, 25-29.

Getty Vocabulary Program. 1988. *Art & Architecture Thesaurus (AAT)*. Los Angeles: J. Paul Getty Trust, Vocabulary Program. Available at http://www.getty.edu/research/conducting_research/vocabularies/aat/.

Getty Vocabulary Program. 2000. *The Getty Thesaurus of Geographic Names* (TGN). Los Angeles: J. Paul Getty Trust, Vocabulary Program. Available at http://www.getty.edu/research/tools/vocabulary/tgn/.

Getty Vocabulary Program. 2000. *The Union List of Artist Names* (ULAN). Los Angeles: J. Paul Getty Trust, Vocabulary Program. Available at http://www.getty.edu/research/tools/vocabulary/ulan/.

Gruber, Tom R. 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5:2, 199-220. Available at http://ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html.

Hill, Linda, Buchel, Olha, Janee, Greg, and Zeng, Marcia L. 2002. Integration of knowledge organization systems into digital library architectures: In Mai, Jens-Erik, et al., ed. *Advances of classification research* vol. 13, Proceedings of the 13th ASIST SIG/CR Workshop, 17 November 2002, Philadelphia PA, 62-68.

Hodge, Gail. 2000. *Systems of knowledge organization for digital libraries: beyond traditional authority files*. Washington, DC: Council on Library and Information Resources. CLIR Pub91. Available at http://www.clir.org/pubs/reports/pub91/pub91.pdf.

Human Genome Resources. [2003]. Bethesda, MD: National Center for Biotechnology Information, U.S. National Library of Medicine. Available at http://www.ncbi.nlm.nih.gov/genome/guide/human/resources.shtml.

ISO. 1986. *ISO 2788:1986 Documentation—Guidelines for the establishment and development of monolingual thesauri.* International Organization for Standardization (ISO) Technical Committee (TC) 46.

Iyer, Hemelata. 1995. *Classificatory structure: concepts, relations and representation*. Frankfurt am Main: Indeks Verlag.

Koch, Traugott and Tudhope, Douglas. 2003. New applications of knowledge organization systems: call for papers.

Koch, Traugott and Tudhope, Douglas. 2004. User-centred approaches to Networked Knowledge Organization Systems/Services (NKOS): Background. Available at http://www2.db.dk/nkos-workshop/#Background.

Lancaster, F.W. 1986. *Vocabulary control for information retrieval*. 2nd ed. Arlington, Virginia: Information Resources Press.

Leise, Fred, Fast, Karl and Steckel, Mike. 2003. Synonym rings and authority files. Boxes and Arrows. Available at http://www.boxesandarrows.com/view/synonym_rings_and_authority_files.

Library of Congress Authorities. The Library of Congress. Available at http://authorities.loc.gov/.

Lin, Xia and Aluker, Serge. 2004. *Visual Concept Explorer*. (Software) Available at http://cluster.cis.drexel.edu/vce.

NISO. 2005. *ANSI/NISO Z39.19-2005 Guidelines for the construction, format, and management of monolingual controlled vocabularies.* Bethesda, Md.: NISO Press. Available at http://www.niso.org/standards/standard_detail.cfm?std_id=814.

PCC. 2007. Program for Cooperative Cataloging (PCC) news sent to the listserv PCCLIST@LISTSERV.LOC.GOV with the subject: 300,000 Subject Authorities. Wednesday, 14 March 2007 10:43 AM.

Rosch, Eleanor. 1978. Principles of categorization. In: Rosch, Eleanor and Lloyd, Barbara B., eds: *Cognition and categorization*. Hillsdale, New Jersey: Lawrence Erlbaum, 27-48.

182

Knowl. Org. 35(2008)No.2/No.3
M. L. Zeng. Knowledge Organization Systems (KOS)

Studer, Rudi, Benjamins, V. Richard and Fensel, Dieter. 1998. Knowledge engineering: principles and methods, *Data and Knowledge Engineering* 25, 161-197. Available at http://www.ubka.uni-karlsruhe.de/cgi-bin/psgunzip/1997/wiwi/33/33.pdf.

Taxonomy of Knowledge Organization Sources/Systems. 2000. Available at http://nkos.slis.kent.edu/KOS_taxonomy.htm.

Thesaurus for liquid crystal research and applications. 1993. Compiled by Zumer, Maja. Kent, Ohio: Kent State University.

Tree of Life project. 2005. *Tree of Life web project*. Website hosted by The University of Arizona College of Agriculture and Life Sciences and The University of Arizona Library. Available at http://tolweb.org/tree/home.pages/popular.html.

Tudhope, Douglas. 2006. Towards terminology services, reflections from the FACET project. Presentation given at OCLC Distinguished Seminar Series, Dublin, Ohio, April 2006. Available at http://www.oclc.org/research/dss/.

Tudhope, Douglas, Koch, Traugott and Heery, Rachel. 2006. *Terminology services and technology. JISC state of the art review.* Bath, UK: UKLON.

Available at http://www.ukoln.ac.uk/terminology/JISC-review2006.html.

UMLS. 2004a. Section 3. Semantic Network. In: UMLS® Knowledge Sources documentation. Bethesda, MD: U.S. National Library of Medicine. Last updated: 12 January 2007. Available at http://www.nlm.nih.gov/research/umls/meta3.html#s3_0.

UMLS. 2004b. Section 2. Metathesaurus. In: UMLS® Knowledge Sources documentation. Bethesda, MD: U.S. National Library of Medicine. Last updated: 02 July 2007. Available at http://www.nlm.nih.gov/research/umls/meta2.html.

UNSPSC. 2001. Using the UNSPSC. United Nations Standard Products and Services Code. White Paper. Granada Research. September 1998, updated October 2001, p. 13. Available at http://www.unspsc.org/AdminFolder/Documents/UNSPSC_White_Paper.doc.

Wielinga, B. J., Schreiber, A. Th., Wielemaker, J. and Sandberg, J. A. C. 2001. From thesaurus to ontology. In *International Conference on Knowledge Capture, Proceedings of the 1st international conference on knowledge capture*, 22-23 October 2001, Victoria, British Columbia, Canada, 194–201.