



Published on *technical foundations* (<http://technicalfoundations.ukoln.ac.uk>)

---

# Knowledge Organisation Systems

23 November 2011 - 4:53pm — Koraljka Golub

Knowledge organization systems (KOS) intended for information discovery such as classification schemes, thesauri and subject heading systems have traditionally been used in libraries and indexing and abstracting services, some since the 19th century. KOS model the underlying semantic structure of a domain, provide a semantic road map of individual fields and the relationships among and across fields, and relate concepts to terms. Today KOS are a necessity in that they inform and promote discovery, use and re-use of information. They serve as the prerequisites for enhanced semantic interoperability.

[Content reviewed by Gordon Dunsire, Aida Slavic, Douglas Tudhope, and by the JISC]

## Background:

The term "knowledge organisation systems" (KOS) is used in practice to denote systems, tools, and services developed to organize knowledge and to present the organized interpretation of knowledge structures, including automated categorization or knowledge mining software. Nowadays there is an increased interest in the type of KOS developed for information discovery and this text will focus on this particular aspect. However, one should note that the term "knowledge organization systems" would also include at least three other kinds of knowledge organization systems:

- knowledge systems that are created to support the function of presentation of knowledge (philosophical systems of knowledge)
- the application of knowledge (lexicographical and linguistic tools)
- scientific and administrative functions (e.g. classifications of professions etc.)

There are two terms related to KOS: *controlled vocabularies* and *indexing languages*. Both of these terms belong to KOS but have a different scope and, should there be need, they can be used with more precision. The term "controlled vocabularies" can be used to denote any controlled set of terms or controlled list of terms used in document description i.e. in descriptive metadata. Indexing languages are a specific kind of controlled vocabularies representing formalized languages designed and used to describe the subject content of documents for the information retrieval purposes.

There are two main types of indexing languages: alphabetical (using natural language terms, thus requiring terminology control such as thesauri, descriptor systems, and subject headings) and classifications (using symbols, operating with concepts and not being concerned with natural language). The main characteristics of indexing languages are that they are concerned primarily with the subject

content of documents and that they contain rules for applications and, in some cases, syntax rules for pre-combination of terms in the process of indexing.

Apart from indexing languages such as subject headings, thesauri, and descriptor systems, which are applied for indexing the content of documents, KOS types also include the following, which may have other purposes as described above (cf. [Hjørland 2008](#) [1] and [NKOS 2008](#) [2]):

- lists (a simple group of terms used for example in web site pick lists)
- synonym rings (a list of synonyms or near-synonyms used interchangeably for retrieval)
- authority files (e.g. names for countries, individuals, and organizations)
- glossaries (usually a subject-specific list of terms with definitions)
- dictionaries (a more general-subject list of terms and their definitions)
- encyclopædias
- gazetteers (dictionaries of place names)
- taxonomies (similar to classification schemes, but the term is more often used in knowledge management systems to indicate any grouping of objects based on a particular characteristic)
- lexical databases or semantic networks (with more defined relationships between terms, used in natural language applications: a major example being WordNet)
- ontologies (with even more defined relationships between terms as well as the rules and axioms, often applied in data mining and knowledge management)
- search-engine directories of web pages
- folksonomies

The general purpose of KOS is to provide a means for organizing information ([ANSI/NISO Z39.19](#) [3]), through:

- translation of the natural language of authors, indexers, and users into a vocabulary that can be used for indexing and retrieval
- ensuring consistency through uniformity in term format and in the assignment of terms
- indicating semantic relationships among terms
- supporting browsing by providing consistent and clear hierarchies in a navigation system
- supporting retrieval

KOS play a crucial role in resource retrieval and discovery. They improve the effectiveness of retrieval by helping to handle the sheer mass of information and they provide knowledge-based support for end users who access information without the help of an intermediary. In comparison to free-text searching, there are many advantages to searching by KOS terms, such as the following:

- the most relevant search terms are selected, and relevant search terms which are not explicitly mentioned in a document may be added
- search terms are controlled, i.e. disambiguated, so that there is no confusion between terms that look the same but have different meanings
- search terms can come from semantically structured vocabularies – hence documents can be found by searching for synonyms, narrower, broader, and even related terms that may not be present in the document itself (semantic query expansion)

A well-structured KOS can be used as the knowledge base for an interface that can assist users with search topic clarification (e.g. through browsing well-structured hierarchies and guided facet analysis) and with finding good search terms (through query term mapping and query term expansion: synonyms and hierarchical inclusion).

Additional functions of KOS are to ([Soergel 2003](#) [4]):

- help improve communication, support learning and assimilating information (e.g. through providing conceptual frameworks to help the learner ask the right questions, assist readers in understanding text by giving the meaning of terms, assist writers in producing understandable text by suggesting good terms, and support foreign language learning)
- provide the conceptual basis for the design of good research and implementation (e.g. assist researchers and practitioners with problem clarification)
- provide classification for action, classification for social and political purposes (e.g. classification of diseases for diagnosis)
- facilitate unified access to multiple databases
- serve as a source for data element definition and provide a conceptual basis for knowledge-based systems
- do all this across multiple languages

The different KOS overlap and vary in their purpose, structure, functionality, field of application as well as in other characteristics. For example, the oldest knowledge organization systems (classification schemes, thesauri and subject heading systems) are related to knowledge at three levels:

- purpose and/or function: communication and/or access to recorded knowledge
- knowledge is in their source (they are built on the basis of scientific, educational and/or cultural consensus)
- consequence of their use: they may facilitate the learning and contribute to the creation of new knowledge (e.g. serendipitous knowledge discovery)

As such, they differ from the other KOS. They also differ from one another: for example, classification schemes are better suited for browsing (library shelves or digital directories) than thesauri and subject headings are. Thesauri and subject headings are generally more suited for search-box information searching.

Current Usage:

With their wide range of characteristics, KOS are used in a variety of applications. Their most prominent use is for improved information retrieval through searching, disambiguation, query expansion and reformulation, or browsing. Different KOS serve different functions, which is why more than one KOS should ideally be used in information retrieval applications. For example, classification schemes generally serve to group together topically related documents into classes and are thus better suited to subject browsing than other KOS; thesauri are used to denote a number of detailed topics and are thus better suited for searching (although examples of KOS which aim to integrate both functions exist). When considering adopting a particular KOS from a type of KOS, the subject indexing policy for the collection at hand needs to be considered: for example, the bigger the collection, the more depth the classification hierarchy should contain, and more detailed topics should be listed in a thesaurus; quality and maintenance (e.g. home-grown KOS on the Web often lack principles from international standards on design and development of KOS), etc.

Other uses include aiding in the general understanding of a subject area, providing "semantic maps" by showing inter-relationships between concepts, and helping to provide definitions of terms. KOS can help improve automated classification and indexing, semantic reasoning, text mining, and information extraction. Topical crawlers or harvesters can utilize KOS to define topics using the high-quality terms for those topics. KOS can also provide support for social tagging, and consequently improve information retrieval and knowledge organization in Social Web applications.

Today KOS are used in a variety of contexts:

- in libraries: for shelf arrangement, information retrieval (both searching and browsing), and collection management (acquisition, circulation statistics, weeding)
- in museums and archives: for collection display, objects indexing and retrieval, and collection management
- in bibliographies, for subject information navigation
- in bibliographic databases (including repositories and subject gateways), for information retrieval
- in information services, for selective dissemination of information
- in journal articles (e.g. "keywords" or "index terms" in the abstract)
- in metadata (e.g. recommended as part of the Dublin Core element "subject")
- as a source for building various knowledge domain maps (ontologies) and other KOS
- in data mining
- in knowledge management

Examples of using KOS for improving the performance of automated subject indexing and classification also exist, and so do KOS as a feed for topical crawlers, as well as KOS as a source for social tagging (currently these are largely experimental but show considerable potential).

Current Issues:

## Interoperability

The fact that classification schemes use a system of notation to represent the hierarchical structure of concepts, where each concept is represented by a notation rather than a natural language term, provides the potential for interoperable search and browsing access to multilingual databases when the databases use the same classification schemes. However, if the KOS used in the databases differ in structure, domain, language, or granularity, the KOS will need to be transformed, mapped, or merged. Moreover, multilingual KOS mapping is complex because it involves translation of concepts, not terms, and there is often significant variation between languages. Different cultural perspectives also need to be integrated (e.g. the concept space of education in one country can be rather different to that in a neighbouring country). On the one hand, communities develop KOS specific to their concepts, terminology, and needs; on the other hand searchers want to use a single search to find resources in databases serving different domains and accessed by different KOS, across which there may be no consensus regarding concepts, terminology, and knowledge organisation.

Apart from semantic interoperability, there also needs to be interoperability with applications: KOS should work with search engines, Content Management Systems, Web publishing software, etc. In order to do this they need to be made available in existing formats and protocols for data exchange, such as [SKOS](#) [5] for representation of KOS in RDF in a simple way, and URIs for unique identification of the KOS, its concepts and terms. SKOS and URIs will allow KOS become [Linked Data](#) [6]. While early adopters exist, there is a long way to go before the potential of these approaches is fully explored and implemented in practice.

## Alternatives to manual KOS-based subject indexing and classification

Although it is very unlikely that there will be approaches that would entirely replace creating quality subject metadata by humans, there are two major attempts in current research and practice aimed at adding to subject metadata created by trained subject metadata specialists: social tagging using KOS as a basis, and automated or semi-automated means. Both approaches warrant further research:

1. Social tagging involves adapting KOS for end user tagging: it needs to be determined which modifications are most likely to make KOS more useful in this context. The changes may include more definitions, better displays and algorithms providing good automated suggestions. Motivation of end users for tagging also needs to be explored further, etc.

2. Although the vendors of today's research and commercial software sellers emphasise the high potential of automated tools for subject metadata generation, real evidence of their success is so far lacking. Software tools may be useful but only in very constrained subject domains; they are unlikely to improve with research because it is essentially "hard" artificial intelligence. The difference between reported high performance results and the reality is in part due to restricting the evaluation of these tools to comparison against existing or ad hoc metadata that serves as the gold standard in laboratory-like conditions which has inherent subjectivity problems in two areas: the correct interpretation of a document's subject matter; any evaluation of the tools is carried out in the context of a laboratory-like environment rather than a real operational system where the most commonly used measures are precision and recall. Although this issue has been discussed widely in the literature, mainstream research has not paid much attention, and published results are widely acknowledged nonetheless. However, existing human-assigned metadata cannot be used as a gold standard. For example, the classes assigned by algorithms, rather than by humans, might be wrong; alternatively, they might be right but mistakenly omitted during human indexing. Subject metadata creation involves determining subject terms or classes under which a document should be found: this goes beyond simply capturing what the document is about to what the document could be used for; algorithms might find such terms, given a good training set, but human indexers who are not well trained might miss them.

## Improvements to KOS

There are a number of areas in which existing KOS could be improved. One approach is to simplify complex KOS that are intended for use in the first instance by librarians and trained end users in a paper environment, for the benefit of non-specialists and for use on the Web. This should also include hierarchy browsing at different levels, hyperlinks for relationships, searching for compounds containing any combination of elemental concepts, adjustments for social tagging applications, etc. Replacing complex built-in concepts, which are present in some KOS, with a structure based on facets, would allow greater flexibility in building new specific concepts at the time of searching as required by the end-user and at the same time reduce the size of the KOS.

Another approach is to enrich one KOS with the benefits of other types of KOS. For example, enriching typical thesauri with hierarchical structure would enable their use both for searching and for browsing. Moreover, empowering end users in searching collections of ever increasing magnitudes, with performance far exceeding plain free-text searching, and developing systems that not only find but also process information, requires far more powerful and complex KOS: thus enriching thesauri with the characteristics of ontologies would be highly beneficial in such applications.

The slow maintenance and updating of some KOS is an issue for end-users who cannot find new concepts and terms or who cannot find out how to use them because of outdated structures, hierarchies and similar. A major reason why updating has been slow is that it would require re-indexing and re-classification of existing collections, which implies expensive re-shelving in libraries; changing the structure would also cause problems for end-users as they would have to learn the new structures when browsing either online or in a physical collection.

KOS do not simply represent the information, but also construct that information. For example, while existing classification schemes are intended to be universal, they are actually culturally specific (e.g. the [Chinese Library Classification](#) [7], BBK in the former Soviet Union). In the [Dewey Decimal](#)

[Classification](#) [8], the most widespread classification system in the world, regional variants had to be introduced as a compromise. In KOS there persists a historical bias on the basis of gender, sexuality, race, age, ability, ethnicity, language and religion, which limits the representation of diversity and effective library service for diverse populations. Now used globally and in interoperable systems, the KOS should be restructured in order to address these issues in a modern context: this once again implies re-classification and re-indexing efforts which are expensive in themselves, and getting the end users to re-learn the KOS they have been used to.

Further Reading:

[Selected KOS General Resources](#) [9]

[JISC Terminology Services Review](#) [10]

[ISO 25964 – the international standard for thesauri and interoperability with other vocabularies](#) [11]

Related:

[Metadata](#) [12]

[Digital Preservation](#) [13]

[Repositories](#) [14]

Other Resources:

[Selected KOS Further Resources](#) [15]

[Metadata and Terminology Registries](#) [16]

Notes:

Reviewed by: Aida Slavic Gordon Dunsire BN

The [Innovation Support Centre](#) was funded during 2010-13 by the [JISC](#), managed by [UKOLN](#) and hosted by the [University of Bath](#).

---

**Source URL:** <http://technicalfoundations.ukoln.ac.uk/subject/knowledge-organisation-systems>

### Links:

[1] [http://www.iva.dk/bh/lifeboat\\_ko/CONCEPTS/knowledge\\_organization\\_systems.htm](http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/knowledge_organization_systems.htm)

[2] [http://nkos.slis.kent.edu/KOS\\_taxonomy.htm](http://nkos.slis.kent.edu/KOS_taxonomy.htm)

[3] [http://www.niso.org/kst/reports/standards?](http://www.niso.org/kst/reports/standards?step=2&gid=&project_key=7cc9b583cb5a62e8c15d3099e0bb46bbae9cf38a)

[step=2&gid=&project\\_key=7cc9b583cb5a62e8c15d3099e0bb46bbae9cf38a](http://www.niso.org/kst/reports/standards?step=2&gid=&project_key=7cc9b583cb5a62e8c15d3099e0bb46bbae9cf38a)

[4] <http://www.ecdl2003.org/ecdl.tutorials.html#tutorial4>

[5] <http://www.w3.org/2004/02/skos/intro>

[6] <http://linkeddata.org/>

[7] [http://en.wikipedia.org/wiki/Chinese\\_Library\\_Classification](http://en.wikipedia.org/wiki/Chinese_Library_Classification)

[8] <http://www.oclc.org/dewey/about/default.htm>

[9] [http://www.ukoln.ac.uk/metadata/kos/selected\\_general\\_resources.html](http://www.ukoln.ac.uk/metadata/kos/selected_general_resources.html)

[10] [http://www.jisc.ac.uk/Terminology\\_Services\\_and\\_Technology\\_Review\\_Sep\\_06](http://www.jisc.ac.uk/Terminology_Services_and_Technology_Review_Sep_06)

[11] <http://www.niso.org/schemas/iso25964/>

[12] <http://technicalfoundations.ukoln.ac.uk/subject/metadata>

[13] <http://technicalfoundations.ukoln.ac.uk/subject/digital-preservation>

[14] <http://technicalfoundations.ukoln.ac.uk/subject/repositories>

[15] [http://www.ukoln.ac.uk/metadata/kos/selected\\_further\\_resources.html](http://www.ukoln.ac.uk/metadata/kos/selected_further_resources.html)

[16] <http://www.ukoln.ac.uk/metadata/kos/registries.html>