DESIDOC Journal of Library & Information Technology, Vol. 33, No. 4, July 2013, pp. 306-313

An Overview of Classification Technologies

K.S. Raghavan* and M. Krishnamurthy**

**Centre for Knowledge Analytics & Ontological Engineering (KanOE)*
PES Institute of Technology, Bangalore-560 059
E-mail: ksragav@hotmail.com

***DRTC, Indian Statistical Institute, Bangalore-560 059*
E-mail: krish@isibang.ac.in

ABSTRACT

Knowledge organisation is a sub-discipline of Information Studies and has its roots in Philosophy. Being application-oriented, it is an area of major interest to librarians, webpage designers, information architects, and semantic web community. The paper examines the scope of 'knowledge organisation' and its various facets. The different approaches to knowledge organisation are examined and the requirements in the context of digital environment are highlighted. An overview of the major trends and approaches is provided.

Keywords: Knowledge organisation, library classification systems, classification technologies

1. KNOWLEDGE ORGANISATION SYSTEMS

Knowledge organisation is a sub-discipline of Information Studies. It has its roots in Philosophy; however, for philosophers, knowledge organisation was primarily an exercise in the logical mapping of different branches of knowledge. As an application-oriented discipline, classification has been a major area of interest to librarians at least since the 19th century; librarians were confronted with two major tasks:

- Organising artifacts in the form of printed books, pamphlets, and similar knowledge resources on the shelves of a library in a logical sequence to support browsing; and
- Designing and building catalogues/databases to facilitate search and retrieval of library resources.

Modern library classification can be said to have begun with the publication of the first edition of Dewey Decimal Classification in 1876 under the title, A Classification and Subject Index for Cataloguing and Arranging the Books and Pamphlets of a Library. Henry E. Bliss used the term 'organisation of knowledge' as early as in 1933 in his book, 'Organisation of knowledge in libraries'. The term 'Knowledge Organisation' (KO) is today used to designate a field of study devoted to and encompassing document description, classification and indexing performed not only in libraries, but also in archives, databases and similar information environments and institutions. These activities are carried out by human indexers – members of the LIS community or subject specialists – as also, and increasingly so in recent years, by computers based on algorithms. Smiraglia[1] says: "Knowledge organisation is devoted to the conceptual order of knowledge. In the broadest sense KO is the arena in which the heuristics of ordering knowledge are studied. Specifically, KO is the research community devoted to classification and ontology, thesauri and controlled vocabulary, epistemology and warrant, and applied systems for all of the preceding (often, especially in North America, resource description is also considered to be a part of KO)". Based on an analysis of the domain of KO, Smiraglia suggests: "KO shows remarkable coherence as a research domain over time, which is witnessed … by commonality of terminology. The extension of the domain is consistently represented as including theoretical foundations, such as classification and ontology, and epistemology, which lies at the heart of both. The intension is represented by development and testing of applications". Evidently, there is both a theoretical dimension to KO concerned with development of principles to serve as the foundations of the domain and a practical dimension concerned with the development of tools and technologies for

Received 20 May 2013, online published 22 July 2013

KO. Hjorland[2] mentions of several approaches and research traditions in KO, viz., the facet-analytical approach, the information retrieval tradition, user-oriented and cognitive views, bibliometric approaches and the domain analytic approach. The principal goal in all practical applications of KO is and has always been to enhance the probability of a user seeking information resources to meet an information need finding the 'relevant' resources/ information objects–be it in a physical library, a digital repository or on the Web. The KO systems seek to support this at the macro-level by building taxonomy of classes that will facilitate browsing information objects themselves or their metadata records. At the micro-level KO systems support information retrieval by seeking to categorise information objects usually based on their 'aboutness' and or some other characteristic(s) to support matching with a query specifying an 'information need' to judge 'relevance'. The aim is to help identify objects from within a collection/ corpus relevant to an information need by means of a user-initiated query. In institutional repository (IR) the purpose of classification is to categorise information objects in such a way that we achieve acceptable levels of Recall and Precision.

2. TYPES OF KNOWLEDGE ORGANISATION SYSTEMS

A wide range of tools could be categorised as Knowledge organisation systems (KOS). Hodge[3] identified some common characteristics in KOS:

- The KOS imposes a particular view of the world on a collection and the items in it.
- Same entity can be characterised in different ways, depending on the KOS that is used.
- There must be sufficient commonality between the concept expressed in a KOS and the real-world object to which that concept refers that a knowledgeable person could apply the system with reasonable reliability. Likewise, a person seeking relevant material by using a KOS must be able to connect his or her concept with its representation in the system.

Hodge[3] also provided a taxonomy which grouped KOS into three broad categories: 'Term lists', which emphasise lists of terms often with definitions; 'Classifications and categories', which emphasise the creation of subject sets; and 'Relationship lists', which emphasise the connections between terms and concepts.

Term Lists:

- Authority files
- Glossaries
- Dictionaries
- Gazetteers

Classifications and Categories:

- Subject headings
- Classification schemes
- Taxonomies (According to Hodge are the last tree terms often used interchangeably)
- Categorisation schemes.

Relationship Lists:

- Thesauri
- Semantic networks
- Ontologies

Linda Hill[4], *et al*. modified this list as:

- Classification and Categorisation
  - Categorisation schemes
  - Classification schemes
  - Lists of subject headings
  - Taxonomies
- Metadata-like models
  - Directories
  - Gazetteers: Geo-spatial dictionaries of places
- Relationship Models
  - Ontologies (Concept Spaces): Specific concept models representing complex relationships between objects, including the rules and axioms missing from semantic networks.
  - Semantic networks: Sets of terms representing concepts, modeled as the nodes in a network of variable relationship types.
  - Thesauri: Sets of terms representing concepts and the hierarchical, equivalence, and associative relationships among them.
- Term Lists
  - Authority files: Lists of terms that are used to control the variant names for an entity or the domain value for a particular field.
  - 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.

In a broad sense, KOS includes even such systems as proprietary coding schemes used by some electronic health record systems, road classifications used by highway departments, classifications used by sports organisations, etc.

This paper is limited in scope to issues related to determining and representing the 'aboutness' of a resource whose principal constituent is text and classing it on the basis of its 'aboutness' into one or more classes. The paper also does not examine classification and clustering based on citation practices. It also does not, in any great detail, examine developments related to theoretical foundations of knowledge organisation. A review paper by Dahlberg published in 'Advances in Librarianship' in 1978, starts with the section heading 'Ranganathan started it'. She quotes a 'knowledgeable person' as remarking: 'No developments since Ranganathan'! There has been a revival of interest in the theoretical foundations of knowledge organisation in recent years and a great deal of work is being done in this area.

3. KNOWLEDGE ORGANISATION APPROACHES

Classification is and has always been 'purpose-oriented'; When applied to information objects its purposes are mainly two-fold: At the macro-level it is a taxonomy or hierarchy of 'classes' to support browsing metadata records of information objects; or browsing information objects themselves. Its practical applications include: Library classification schemes for arranging books on library shelves; Yellow Pages; Hierarchically-structured Web directories – Yahoo!, Open Directory project (ODP). Both manual and machine-based systems, tools and approaches have been employed in performing KO tasks. These include applying:

- Library classification systems; semi-automatic indexing systems;
- Keyword indexing-based on H.P. Luhn's experiments widely adopted by bibliographic databases, online retrieval systems and Web search engines; The techniques coming under this broad category are being continuously refined to improve the quality of indexing/metadata extraction using developments in text processing and a variety of statistical techniques;
- Pre-publication metadata (based on e.g., Dublin Core or some other schemes) by the author/publisher to enable service providers to more accurately identify the class(es) to which a resource belongs.

Whichever approach is adopted for classing and grouping information objects, its adequacy and value is largely measured by retrieval performance. The primary objective of the KOS is to help identify objects from within a collection/corpus relevant to an information need defined by means of a user-initiated query. A document is relevant if it contains information of value as seen by the user. Any evaluation campaign has a set of criteria that generally fall into one of two categories:

- Effectiveness (does the system do what it was designed to do?); and
- Efficiency (how fast, reliable and economical is it?)

The purpose of KO is to categorise information objects in such a way that we achieve acceptable levels of Recall, Precision and response time and, in today's context, another major requirement is to be able to rank the information objects in an order of decreasing relevance to the information need.

3.1 Traditional KOS

There are two main types of traditional KOS:

(a) Alphabetical (verbal) systems, and

(b) Classificatory systems.

The principal difference between the two is that the latter use notations in addition to verbal expressions to support arrangement of information objects on shelves or their metadata records in a catalogue. There are also significant differences in terms of how subjects and their inter-relationships are displayed in the two. Right from the times of Cutter and Dewey the relative merits and demerits of the two have been debated. Even the first Aslib-Cranfield study sought to compare the two. It is now fairly clear that the two are complimentary; classificatory schemes require alphabetical indexes (Dewey's 'Relativ Index' is perhaps the best example of what an alphabetical index can do to supplement and compliment classificatory structures; Ranganathan's 'Chain Indexing' is based on 'Relativ Index' and achieves the same purpose when applied to indexing a classified catalogue). Alphabetical schemes can be significantly enhanced by classificatory structures (the hierarchical display of the vocabulary is now fairly used by many alphabetical thesauri such as 'MeSH'. All traditional KOS are artificial languages and differ from natural languages with respect to their vocabulary, semantics and grammar. The vocabulary of KOS is normalised and controlled, and employs terms with well-defined semantics; in contrast it is not uncommon to find synonyms and homographs in natural languages. Just as natural languages classify words into categories (parts of speech), KOS also categorise terms in their vocabulary. Ranganathan's schema of fundamental categories (preceded by Kaiser's categories) carries this to such an extent that it became the model for analytico-synthetic KOS. Following the Ranganathan's approach several schemes of categories and associated syntactic rules were developed including those by B.C. Vickery, G. Bhattacharyya, Derek Austin, Jason Farradane (although Farradane came up with a schema of categories of syntactic relations (and not categories of concepts, his approach is analytico-synthetic in nature), and others. The British Classification Research Group (CRG), while adopting Ranganathan's technique of Facet Analysis rejected his schema of categories. To them the nature and number of categories was more an empirical issue and was largely a function of the discipline. While most KOS employ some synthetic devices, those that do not extensively employ syntactic categories came to be referred to as enumerative KOS. In terms of relationships expressed, most traditional KOS are restricted to expressing 'Equivalence' (Synonymy), 'Hierarchical', and 'Associative Relations (Lateral Relations)'. Traditional KOS continue to be widely used especially in libraries, national bibliographies and even in many structured bibliographic databases. What is however important to note is that the processes of building and applying traditional KOS have largely remained manual and will continue to be so. In other words these are labour intensive and do not scale well.

Major traditional library classification schemes such as UDC, DDC and LCC continue to be widely used in libraries and some bibliographies and are being regularly revised. However, they have not been widely accepted by the digital world of the Web–Most finding aids to Web resources, resources in digital repositories, indexes to open access resources (Directory of Open Access Journals (DOAJ), Open J-Gate, etc.) depend on keyword indexing. The application of universal general classification schemes such as UDC, DDC and LCC appears to be largely restricted to shelf arrangement of documents on library shelves (or entries in national bibliographies). In so far as controlled vocabularies (verbal indexing languages) are concerned the idea of a universal thesaurus was probably first suggested by Soergel in the early 1970s. Following trends can be seen:

- There is a clear move among designers of lists of subject headings to move towards making these more thesaurus-like and adopt the schema of relations widely employed by thesauri ('Equivalence', 'Hierarchical' and 'Associative Relations');
- There is also a move among some thesauri to evolve into faceted 'concept systems';
- Controlled vocabularies in the form of thesauri are trying to transform themselves into ontologies with a view to be able to not only express more adequately the relations between concepts (than is possible using the BT, NT, and RT types used in conventional thesauri), but also support machine manipulation based on rules of logical reasoning.
- There are also attempts to integrate vocabularies in a field; perhaps the best example of this is the UMLS, which is an effort to integrate major vocabularies in the field of medicine; In more recent times efforts along this direction have taken the form of building 'crosswalks' among vocabularies and metadata schema. An issue that has come to occupy importance in recent years is that of Interoperability. Interoperability is defined as the ability of two or more systems or components to exchange information and to use the information that has been exchanged. It has been recognised that this ability is basic to the effective functioning of information networks, and even more to the development of the semantic web (SW). Two levels of interoperability have been recognised as applicable to thesauri:
  – Presenting data in a standard way to enable import and use in other systems;
  – Mapping between the terms/concepts of one thesaurus and those of another to support their complementary use.

3.2 International Standards

ISO 25964 is the new international standard for thesauri; part 2 of this was published as recently as in March 2013 and deals with interoperability between thesauri and other types of vocabularies, including classification schemes, taxonomies, subject heading schemes, name authority lists, ontologies, terminologies, and synonym rings, more particularly with the principles and practice of mapping between them[5]. These mappings need to be prepared with care as recommended in ISO 25964 part 2 so as to ensure that semantic web inferences based on them do not lead to misleading conclusions to the Web surfer. A major feature of these guidelines is that they include structural models for mapping, guidelines on mapping types, and for handling pre-coordination (which occurs especially in classification schemes, taxonomies and subject heading schemes).

4. CHANGING CONTEXT

The emergence of digital resources and resources whose principal constituent is not merely text has led to a re-examination of the tasks to be accomplished by KO. The contexts in which KO tasks have to be performed have also grown and expanded (e.g. knowledge management in the corporate environment, e-governance, e-commerce, etc.). Some KO tasks in addition to the more conventional ones such as organising documents on library shelves in the present day context are:

- Classification of e-mails into classes, (e.g. spam and non-spam)
- Detecting a document's encoding (ASCII, Unicode UTF-8 etc.) and categorising them on that basis
- Identifying the language of a document
- Ranking retrieved documents (could be based on a document classifier)
- Classifying webpages–personal home page, etc. (Genre)
- Extracting metadata from textual e-resources.

Even by the middle of the 20th century, following the beginning of information explosion search was on for automated systems capable of indexing large volumes of information. H.P. Luhn's Keyword Indexing could be seen as an effort in this direction. The simplicity of its approach and its ability to handle and process very large volumes of textual information made the system acceptable in a wide range of information environments, particularly large bibliographic databases, indexing and abstracting services. Several different sub-species of keyword indexing came to be developed and today keyword indexes are widely accepted despite their limitations in effective information retrieval. The advent of World Wide Web and the digital world have thrown up many newFtable challenges. Information retrieval today is characterised by:

- Scale at which it has to operate;
- Range of material/resources from which to recall needed information–scholarly material, e-mails, balance sheets, hospital patient records, crime records, images, multimedia, etc.
- Operational space of IR–Institutional, domain-specific or genre-specific search

As such today's knowledge organisation involves indexing issues that support:

- Building systems that work effectively and efficiently at this scale; and
- Handling the extremely broad range of document types

The impact has also been on traditional libraries as pointed out by Hjorland[6]:

- There is a great deal of dependence on centralised agencies; many libraries rely on classification codes supplied, e.g., Library of Congress, rather than doing original classification in so far as their book collections are concerned;
- Many library directors expect that, in the future, large scanning projects (such as that which is being conducted by Google) may enable full text searches to be carried out of all available content. For this reason, they may consider it a waste of resources to classify or index books;
- Libraries have also come to rely on user tagging and may perhaps expect that this will somehow act as a substitute for professional indexing and classification; and
- Users mostly find the books they need using tools other than the library online public access catalog (OPAC); Even the 2010 OCLC survey of Users' Perception of Libraries confirms this finding[7].

All this is not to suggest that traditional KOS have become irrelevant in the present-day context. It is only to suggest that they have limitations in terms of the space/information environments within which they can operate and be effectively employed; we need to look for more scalable KOS to complement traditional KOS for the digital environment. This is not something entirely new and the need has been recognised for quite sometime. let us re-state our problem and requirements and illustrate these by means of a couple of examples. The amount of data available online has grown and we need to be able to search large-sized collections in the order of billions to trillions of words. There is a need of more flexible matching (with query) facilities; e.g., to be able to search for texts in which the word 'INFORMATION' is near the word 'RETRIEVAL' and to be able to define 'nearness' according to our requirements. In view of the volume of data, it is important to have a ranked output (Decreasing degree of relevance to the query). In other words, a KOS that is capable of informing which of the retrieved documents are more relevant. Some examples for the same are:

- The size of 'Shakespeare's Collected Works' is about one million words of text. Going through the entire text to retrieve relevant paragraphs could be effective, but it is time consuming;
- 'Reuters Corpus', vol. 1 released on 3 November 2000 contains about 810,000 English Language news stories (1996-08-20 to 1997-08-19) requiring about 2.5 GB for storage of the uncompressed files;
- 'Reuters Corpus', vol. 2 - a multilingual corpus released on 31 March 2005 and distributed on one CD contains nearly half million Reuters news stories in 13 languages (Dutch, French, German, Chinese, Japanese, Russian, Portuguese, Spanish, Latin American Spanish, Italian, Danish, Norwegian, and Swedish) (covering the period 1996-08-20 to 1997-08-19)

The task of organising knowledge contained in such resources to facilitate their effective retrieval is gigantic if it were to be carried out manually. The approach to classification has to be necessarily machine-based. The most elementary approach, also employed by many of the Web search engines, would be to consider every word (every Keyword) in a document as defining a class to which the document belongs and build a huge term-document matrix as below:

Table 1. Document-term matrix

Document / Term	D_1	D_2	D_3	D_n
T_1	1	0	0	0
T_2	0	0	0	1
T_3	1	0	0	1
T_n	0	0	1	0

While this is not a very effective or efficient method of classification, it proved adequate for creating document classes (subsets of documents in a corpus) matching an information need expressed using Boolean operators as the Boolean model merely views a document as a set of words; A document class matching an information need is formed using Boolean operators – AND, OR, NOT. However, the limitations of such an approach to classification were realised quite early and refinements followed. Let us consider a realistic situation of a corpus of 100000 documents requiring classification to support retrieval. If each document, on an average, is 1000 words long and there are 100000 unique words in the corpus, the term-document matrix will have 10 billion '0s' and '1s'; the '1s' will probably be less than 10 % of this with over 90 % of the cells in the matrix being '0s'; A more efficient approach is to record only the '1s' which is what an inverted file or back of the book index (with terms arranged alphabetically) does, a classificatory approach extensively used in today's database management systems. This still left unsolved the problem of 'dodging' irrelevant documents, a problem frequently faced while using Web search engines. The problems could be traced to:

- Issues arising out of the very broad definition of Keyword (Any word that is not in the stop list is a Keyword);
- The severe limitations of the Boolean 'AND' in adequately expressing the relations between concepts/terms (both in a documentary resource and in a query)

Some new approaches explored were aimed at addressing the two issues identified above.

Analytico-synthetic classifications did provide an appropriate and useful approach to the problems related to more clearly defining the relations between concepts in a given context. Jason Farradane's Relational Indexing is probably the most expressive of all the relational indexing systems in terms of its ability to explicitly and clearly specify relations between concepts. However, recognising that application of traditional KOS in the digital environment is an expensive proposition, the focus in research has been on further refining and extending Boolean 'AND' and 'OR', and on improving the quality of keyword identification and extraction. Proximity operators and Field-specific searches have been introduced to further refine the 'AND' operator and Truncation for string searching. These to a certain extent helped in restricting/widening (depending on the need) the class of documents to be retrieved. There were also several experiments aimed at implementing processes that helped refine the definition of keywords:

- To restrict the words to be inverted while indexing full texts or texts of abstracts, etc.;
- To make machines understand which phrases to be accepted as classes (Third World, Developing Countries, etc.);
- To be able to attach weights to keywords extracted from documents to support ranking of retrieved documents in contrast with Boolean logic that resulted in a binary classification of documents (those that are relevant and those that are not).

While the extended Boolean operators have been in use by major commercial information providers / databases, many did not support ranking of retrieved documents based on term weight. Indexing requirements since the arrival of the Web have changed in view of the volume and variety of data to be handled in Web indexing. Web is multilingual. An essential step in full-text indexing is the process of tokenisation; and tokenisation is language-specific.

和尚

The above sequence of Chinese characters if read as one word mean a monk and as a sequence of two words mean 'and'& 'still'. The number of index terms that need to be inverted is the principal issue in any language text. It has been shown that with good processing technologies for stemming and case folding, it is possible to reduce the number of words to be indexed. In a language that is morphologically richer the reduction could be substantial. However, parts of speech taggers, stemming algorithms, etc are yet to be developed fully for many of the languages of the world. There are also cases of tokens, which have specific meaning in certain domains (e.g., C++, IR 8, B52, etc.).

4.1 Text Classification

Given a set of pre-defined classes, text classification is essentially determining the classes to which a document belongs, Basically the following two kinds of technologies are in use:

(a) Supervised learning in which classes are distinguished by word patterns:

An example can be that the documents in the class China tend to have high values on dimensions like Chinese, Beijing, Shanghai and Mao Tse Tung whereas documents in the class India tend to have high values for New Delhi, Gandhi and Mumbai. Enhancing the effectiveness of classifiers has been the main focus of research and a range of new techniques (support vector machines, neural networks, etc.) has been developed.

(b) Clustering:

Algorithms that group a set of documents into subsets or clusters that are coherent; Documents in a cluster should be as similar to one another as possible. The assumption in class formation based on clustering is that documents in the same cluster behave similarly in terms of relevance to information needs. Clusty (www.clusty.com) employs Clustering to support retrieval.

In what has been presented in the foregoing paragraphs, we have considered a document as a set or sequence of terms. Many textual documents, in addition, also have a structure; for example, they carry metadata; names of authors, title, table of contents, date, abstract, format, etc; building indexes using these parameters (or zones) is referred to as parametric and zone indexes. The approach is essentially to build indexes by treating each field/zone as either a part of the term (dictionary) or posting. Weights could be attached to extracted terms based on the field in which the term occurs.

5. CONCLUSIONS

In the foregoing paragraphs, mechanisms that have the aim of arriving at a better representation of a document's aboutness have been looked. Enhancing the effectiveness of classifiers has been the main focus of research and a range of new techniques (Support Vector Machines, Neural Networks, etc) has been developed.. There are also technologies that help enhance queries; some global methods are discussed in the literature; a mention of these is appropriate here primarily because they make use of classificatory tools. Some of the widely employed technologies are:

- Suggesting related queries (by search engines);
- Relevance feedback; marking the relevant documents in the initial set and using this to reformulate the query; and
- Using lexical tools for query expansion/reformulation (e.g. thesaurus or WordNet)

There has also been Research on genre identification that could enhance IR effectiveness in the context of Web. Typical examples of genre include:

- Homepages
- Catalogue
- Article/paper
- News

Another important area that has been given considerable attention in recent years is to explore how existing KOS could be employed and made us of in the web environment, Faceted navigation has been recognised as an effective way forward by information architects and web designers. There have also been efforts at porting KOS to Web (Semantic Web?). The SKOS has been designed to provide a low cost migration path. It also provides a conceptual modeling language for developing and sharing new KOS. It can be used on its own, or in combination with more formal languages like the web ontology language (OWL). The role of SKOS is to bring the worlds of library classification and web technology together. Automatic classification based on supervised methods appears to perform better. However, availability of large training corpora for different domains, different languages, and different classificatory tasks is a major issue. For example, in the area of WSD (Word Sense Disambiguation), it has been estimated that to achieve reasonable levels of accuracy one needs about 3.2 million sense-tagged words. The human effort for constructing such a training corpus has been estimated to be 27 man-years! A significant portion of the work is in experimental stage and several approaches are being experimented with. When tools and applications are designed to work focused, on a specific domain of interest, the results appear to be better. Figure 1 presents an overview of the various approaches to text classification vis-à-vis the cost and degree of automation that can be achieved.

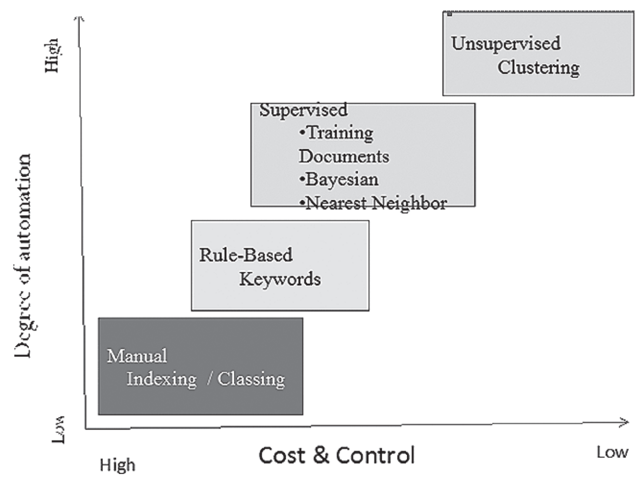

Figure 1. Approaches to text classification.

One thing that is certain is that automation will continue to make significant advances. While more and more intelligent search engines are being developed, it is important to clearly identify which of the tasks in KO are amenable for automation and which are not. The success of attempts to automatically class documents will undoubtedly depend upon developments in text-processing technologies, establishing links between lexical tools and controlled vocabularies. That significant enhancements could be made to information retrieval by linking thesauri with lexical tools has been demonstrated and explained in a few recent papers[8-11].

REFERENCES

1. Smiraglia, Richard P. Universes, dimensions, domains, intensions and extensions: Knowledge organisation for the 21st century. *In* Categories, contexts and relations in knowledge organisation, edited by A. Neelameghan and K.S. Raghavan. Ergon Verlag, Wurzburg, 2012.

2. Hjørland, B. Facet analysis: The logical approach to knowledge organization. *Inf. Process. Manag.* , 2012. http://dx.doi.org/10.1016/j.ipm.2012.10.001/.
3. Hodge, G. Systems of knowledge organisation for digital libraries: Beyond traditional authority files. The Council on Library and Information Resources, Washington DC, 2000. http://www.clir.org/pubs/reports/pub91/contents.html/.
4. Hill, L.L.; *et al*. Integration of knowledge organisation systems into digital library architectures: Position paper. *In* Advances in Classification Research, edited by J.E. Mai; *et al*. Proceedings of the 13th ASIST SIG/CR Workshop on Reconceptutalizing Classification Research. Philadelphia, PA, 2002, pp. 62-68. http://alexandria.sdc.ucsb.edu/~lhill/paper_drafts/KOSpaper7-2-final.doc
5. http://www.oclc.org/reports/2010perceptions/2010perceptions_all.pdf/.
6. Birger Hjørland. Is classification necessary after Google? *J. Doc.*, 2012, **68**(3), 299-317.
7. http://www.oclc.org/reports/2010perceptions/2010perceptions_all.pdf/.
8. Neelameghan, A. & Raghavan, K.S. A knowledge organizing system for humanistic disciplines with enhanced capabilities: Case studies. *Information Studies*, 2009, **15**(2), 75-94.
9. Neelameghan, A. & Raghavan, K.S. Relationships in multi-lingual thesauri in humanities: Issues and case study. *In* Libraries and information systems: From Alexandrian heritage to social networking, Essays in honour of Prof. S. Parthasarathy edited by K.S. Raghavan & K.N. Prasad. Ess Ess Publications, New Delhi, 2009, 318-40.
10. Raghavan, K.S. & Neelameghan, A. A bilingual information retrieval thesaurus: Design and value addition with online lexical tools. *In* ICADL-2008: Digital Libraries: Universal and Ubiquitous Access to Information, edited by G. Buchanan, M. Masoodian & S.J. Cunningham. LNCS 5362. Springer Verlag, Berlin, 2008, pp. 408-09.
11. Raghavan, K.S. & Neelameghan, A. Design and development of a bilingual thesaurus for classical Tamil studies: Experiences and issues. *In* Culture and Identity in knowledge organization, edited by Clement Arsenault & Joseph Tennis. Proceedings of the Tenth ISKO International Conference, 5-8 August 2008, Montreal, Ergon Verlag, Wurzburg, Germany, 2008, pp. 70-76.