Classificatory ontologies

A.R.D. Prasad
Devika P. Madalli
Documentation Research and Training Centre, Indian Statistical Institute

ABSTRACT: Digital Libraries and Digital Repositories are data-intensive with large numbers of full-text resources accessible online. Activities in the area of Semantic Web development recognize the significant part played by metadata and knowledge organization systems such as classification systems and thesauri in capturing and communicating 'meaning'. We now have Web ontology standards, such as Simple Knowledge Organization Systems (SKOS), a common data model for sharing and linking knowledge organization systems via the Semantic Web. Standards such as SKOS are also meant to be used as a vehicle for deployment of knowledge organization systems that were not born digital (or XML/RDF) such as thesauri and bibliographic classifications. This paper attempts to present an application of the faceted classification scheme as enunciated by Ranganathan in developing ontologies. It further explores the issues in modelling the faceted scheme of Ranganathan using SKOS.

KEYWORDS: classification, digital libraries, ontology

1. Introduction

The phenomenal presence of Digital Libraries (DLs) on the Web highlights the importance of structuring data. The size and complexity of the information stored in Digital Libraries (DLs) is steadily increasing. As opposed to the loosely-organized information sources available on the Web, DLs provide metadata-supported, well-structured data. In the context of the envisioned Semantic Web (Berners-Lee *et al.,* 2001), metadata becomes a cornerstone. The Semantic Web (SW) cannot realize its potential unless intelligent automatic procedures help to analyze and transform current natural language knowledge into its Semantic representation. Metadata, a key idea of the Semantic Web, has already been in practice for centuries in libraries through the use of different knowledge organization systems.

In DLs, metadata have many roles and functions, such as semantic and structural information representation, organization and retrieval. Metadata are usually embedded in the form of item descriptions, and the way metadata schemas are deployed depends on the DL's domain. The intrinsic complexity of heterogeneous content requires highly complicated and re-mixed metadata schemes rather than just the use of monolithic sets of elements. The fact is clearly emphasized by current approaches to metadata, such as the Warwick Framework, which assumes that "complex metadata needs are best met by a multiplicity of separate but functionally focused metadata schemes, relatively orthogonal and independently maintained by communities of expertise and practice, which can be mixed and matched as needed, rather than by just one comprehensive but monolithic set of elements. The metadata associated with a given object would be separated into a series of 'packages' that would be marked as using these various metadata schemes" (Baker & Lynch, 1998).

In DLs, a full range of information is included, spanning from descriptive and administrative metadata (technical metadata, rights management and digital provenance) to structural metadata. Subject metadata are represented by several knowledge organization systems (KOS), e.g. classification schemes, thesauri, taxonomies or ontologies. Such topical tools strengthen the legacy of the so-called "subject approach to information", maintaining this approach alive also in the digital scenario, which has been a significantly important issue for librarians since Cutter's time. The present paper tries to explore the synergy of DLs and KOS and the possibilities of embedding the advantages of faceted system in DLs.

2. Digital Libraries and KOS

The term "knowledge organization systems" is intended to encompass all types of schemes for organizing information and promoting knowledge management. Knowledge organization systems include classification schemes that organize materials at a general level (such as books on a shelf), subject headings that provide more detailed access, and authority files that control variant versions of key information (such as geographic names and personal names) (Hodge, 2000). They also include less traditional schemes, such as semantic networks and ontologies. Knowledge organization systems as mechanisms for organizing information find applications in every library, museum, and archive.

Digital libraries often tend to adopt one or more KOSs for organizing their collections and resources. Just as in a physical library, the KOS in a digital library provides an overview of the collection's content and supports retrieval. The scheme may be a traditional KOS, relevant to the scope of the material and the expected audience for the digital library (such as the Dewey Decimal System or the INSPEC Thesaurus), or a commercially developed scheme such as the Yahoo or Excite categories, or a locally developed scheme for a corporate intranet.

KOSs enhance the digital library in a number of ways. They can be used to connect a digital library resource to a related resource. The related information may reside within the KOS itself, or the KOS may be used as an intermediary file to retrieve the key needed to access information in another resource. According to Soergel (2008) in a DL, knowledge organization comes into play in several closely inter-related ways:

a. Organization of information in substantive databases;
b. Organization of information within documents;
c. Organization of information about documents and databases (metadata);
d. Organization of information about any type of subject treated in documents (needed to support finding documents and other digital objects);
e. Information about concepts, terms and their relationships; organization of ontological and lexical information.

The inter-related approach highlights the fact that users to DL can be varied, depending on the intended use of a given resource, but recognizes that the subject or topical approach is one of the most common in any domain. Therefore, a KOS can make digital library materials accessible to disparate communities. This may be realized by (Hodge, *op.cit.*):

a. providing alternate subject access,
b. adding access by different modes,
c. providing multilingual access, and
d. using the KOS to support free text searching.

In DLs, KOS are presented in the form of metadata schemas. Metadata standards bring into focus the fact that there is a need to describe the content of Internet resources using some "meaningful descriptors" (Smith & Schirling, 2006). Metadata standards are implemented in DLs to provide the necessary pointers to information by breaking up the content into bits and pieces. Metadata standards are particularly applicable to digital libraries as tools for finding information.

The function of metadata has a parallel with that of the card catalogue in traditional print libraries. The logical extension of such data description methods was carried forward by bibliographic databases. Data were structured into fields and subfields that followed standards in the element set. The element sets were formalized into standards such as MARC21 (Library of Congress, 2006) and UNIMARC (IFLA, 2004), widely adopted by libraries in their OPACs. Also, several standards emerged for the data elements themselves, including ISO date format, AACR2 rules for data extraction, and so forth.

For the purpose of this discussion, we would like to broadly classify metadata into two kinds: bibliographic and non-bibliographic metadata. Bibliographic metadata are based on traditional library catalogue data. Non-bibliographic metadata could be descriptions of any entity, such as a person, organization, service, market products, etc. Traditionally, machine-readable catalogues based on standards such as MARC21 include authority files for personal names, organizations and institutions, geographical places, etc. Most of these authority files are used for vocabulary control or standardization of terminology used within systems.

In a digital library environment, though, such authority files can be transformed and enriched by using metadata. In this paper only bibliographic metadata will be analysed. Metadata are much akin to bibliographic data as they are "structured data" describing the characteristics of a resource. Metadata are about knowledge, meaning the ability to turn information and data into effective action (Haynes, 2004). This implies that through metadata, more "contextual" information may be attributed to the resources in order to eventually help in their discovery by end users on the Web.

Metadata empowers search engines and end users in search and retrieval by providing them with essential clues (Prasad & Madalli, 2008). For example, searching "Johnson" would produce the results where "Johnson" is a player, "Johnson" is a company or "Johnson" the name of a software package. But if a mechanism is provided to state in the query the following equation:

> Look for "Johnson" where "Johnson" is the author

then only those resources where Johnson appears as an author would be retrieved.

This kind of mechanism was perfected in bibliographic databases on CD and online information systems. In Web parlance, metadata elements are used in the place of bibliographic elements. But the above option would fail again if we were looking for "Johnson" – an electrical geyser company. Obviously, it requires a different kind of metadata other than just bibliographic. Thus, each requirement is context-specific. Dublin Core, the widely used standard for metadata, describes Web pages with 15 structured elements that can be used optionally; each element can be repeated as well. Elements can be further qualified by element refinements and data standards (Dublin Core Metadata Initiative, 2006). The design and content of Dublin Core being very simple, the need was felt by different communities to extensively add on elements or change them to suit the description of different datasets. However, Dublin core is used as a minimum common denominator among and between digital libraries.

3. Some KOS implementations in DLs (related works)

In the digital library community, the importance of different KOSs has been recognized quite early. Perhaps the reason can be attributed to the contributions of library KOSs and their different application in library OPACs and indexing systems. DESIRE - *Development of a European Service for Information on Research and Education* (DESIRE, 1997) tried to implement automatic classification in networked environment. The RUBRIC Project (RUBRIC, 2007) came up with an XML file to provide a controlled vocabulary of ASRC (Australian Standard Research Classification) subject terms for use with the DSpace digital library suite. For achieving this goal, DSpace Ontology Add On is used. The ASRC subject terms XML file is really a list of the Research Fields, Courses and Disciplines (RFCD) classification. These classifications are used extensively by Australian institutional repositories and are known by the more generic term of ASRC subject headings. Another project, ERCIM (European Consortium for Informatics and Mathematics) Technical Reference Digital Library (ETRDL) (Baldacci *et al.*, 1998) has incorporated the ACM Computing Classification descriptors and AMS Mathematics Subject classification descriptors to be chosen while inputting the "subject" field.

The Alexandria Digital Library Project (ADL) from the University of California at Santa Barbara (Hill *et al*., 2002) used thesauri and authority lists to improve navigation, such as the Getty Thesaurus of Geographic Names and the Gazetteer Content Standard, by providing relations between both. With all these implementations, developers of popular DL packages also started incorporating features for embedding KOSs.

The Greenstone Digital library offers a unique way of knowledge organization by providing for hierarchical phrase browsing (Witten, 2003). This kind of browsing is performed through an interactive interface which allows access to a phrase hierarchy that has been extracted automatically from the full text of a document collection. It is designed to resemble a paper-based subject index or thesaurus. The user enters an initial word into a search box, and a list of phrases containing this word is shown. These phrases are simple at the first instance; each can be further expanded into a list of more complete phrases that contain the original one. This allows hierarchical access to the lexical content of a document collection. Ultimately, the user reaches a leaf of the hierarchy, which takes him/her straight to the unique document containing that phrase. Apart from this, semantic metadata that are structured in hierarchies, such as library classifications, can be presented as trees where the nodes open to reveal the data beneath. In this case, the user is provided an auxiliary file giving labels for intermediate nodes of the hierarchy (e.g., subject headings corresponding to each classification number) (Witten, *op.cit.*).

4. Significance of Faceted Classification Schemes in DLs

Most search engines deployed in DLs (e.g. Lucene in Dspace) use post-coordinate indexing and depend on ranking algorithms to display search results in decreasing order of relevance. However, it should be noted that emphasis in recall is the main reason for noise/false drops in the search. In contrast, pre-coordinated indexing system results are better in precision often at the cost of recall. It is generally accepted that precision and recall are inversely proportional. With vast quantities of information on the Web, perhaps it should be rethought whether the end user should be given the choice of searching either pre- or post-coordinated indexes. Thus far we hardly have computational models of pre-coordinated indexing systems in deployed system, though there have been experimental models of it (Prasad, 1995).

4.1. Library Classification and Indexing

Theoretically, library classification systems such as Dewey Decimal Classification (DDC), Universal Decimal Classification (UDC), and Colon Classification (CC) can serve to implement pre-coordinate indexing and models could be developed, e.g., on the lines of chain indexing. Although classification schedules are traditionally used to assign class numbers to documents for the purposes of shelf arrangement, in the context of bibliographic databases and the Web there is no requirement for a linear arrangement. One of the major advantages of forgoing class numbers is that documents can be arranged by different criteria, e.g. by subject, author or title, to facilitate browsing. Class numbers impose a mono-hierarchical system where, as with thesauri or classauri (Bhattacharyya, 1981), they are not constrained by such a requirement and therefore Web documents can be browsed in a user-defined order.

Computational models used to unscramble class numbers represented in DDC, UDC or CC into subject terms have been demonstrated (Prasad, *op.cit.*) while it is rather more complex to build systems that can automatically assign class numbers to documents. In a library environment, where it is important to physically arrange books by using class numbers, there is no purpose of class numbers in indexing. If the whole process is viewed as defined in terms of Ranganathan's three planes of work – idea plane, verbal plane and notational plane – unscrambling class number (notational plane) (Ranganathan, 1954a) into subject keywords (verbal plane) is similar to that of "one step forward, two steps back". While it is justified to convert class numbers into subject strings in the case of already worked out class numbers, it is not a justifiable approach to construct class numbers and translate them back to subject key terms. In the latter case, one can move from idea plane to verbal plane without bothering about notational plane altogether. However, developing computational models/ontologies of these systems is a very complex procedure. While linear arrangement was the focus of library classification system, subject indexing system was deployed to add context to concepts in content through post and pre-coordinate indexing methods (Ranganathan, 1954b).

Context Free indexing: At the time of indexing, context is not taken into consideration, though users are free to add context during search by using Boolean "and" or "not". Boolean "or" may facilitate multiple contexts, but again it may result in providing wrong context if care is not taken by nesting the search expression, whereas in the case of pre-coordination such ambiguities are avoided.

Context Sensitive Indexing: In post-coordinate indexing, though more terms could be added by using Boolean "and", which would add context, there is no specific order for the terms (lack of syntax). This was foreseen by the authors of PRECIS (Austin, 1974) and POPSI (Bhattacharyya, *op.cit.*). In fact, PRECIS and POPSI are considered as subject indexing languages with emphasis on structured context. POPSI, in particular, is a derivative of the Analytico-Synthetic Classification system propounded by S.R. Ranganathan (Ranganathan, 1954a). An advantage of the analytico-synthetic approach in subject indexing as found in POPSI is that associative relations between documents can be visualized.

5. Faceted Classification and SKOS

The vision of Semantic Web is that Web technology could be used to better organize the vast amounts of unstructured (i.e. human-readable only) information on the Web. To actually apply these technologies over large bodies of information requires the construction of detailed "maps" of particular domains of knowledge, in addition to the accurate description (i.e. annotation or cataloguing) of information resources on a large scale, much of which cannot be done automatically. The accumulated experience and best practices of library and information sciences for the organization of information and knowledge are obviously complementary, and they are applicable to the grand vision, since they include the many existing knowledge organization systems already developed and in use such as classification schemes, subject headings and thesauri. Representing such concept schemes in a machine-processible language and structure will help to realize the idea of concept-based instead of text-based searching, which is the main motto of Semantic Web.

Simple Knowledge Organization System (SKOS) (Miles & Bechhofer, 2009), which is presently a candidate for W3C Recommendation, is a step towards achieving this motto. The primary goal of SKOS is to provide a simple machine-understandable representation framework for KOSs with enough flexibility and extensibility to cope with the variations found in KOS idioms, and fully capable of supporting KOS publication and use within a decentralised, distributed information environment such as the Semantic Web.

SKOS is intended for use in three important information retrieval applications: Vocabulary development, Indexing and Searching. It is ultimately designed to support interoperation of these three key components. SKOS representation of classification schemes will add semantics to the visual-based searching techniques such as hierarchical browsing, faceted search, etc. A classification scheme is a taxonomy with a notation for each entity which denote its position in the hierarchy. Prominent examples are DDC, CC and UDC.

5.1. SKOS implementations

Some SKOS use cases are discussed in detail by Isaac *et al.* (2007). SKOS is used in the *Bio-Zen Ontology Framework* (http://neuroscientific.net/index.php?id=43) for the representation of many existing life sciences vocabularies, taxonomies and ontologies coming from the *Open Biomedical Ontologies* (OBO) collection. To represent such vocabulary elements as well as other types of information, the existing SKOS model has been integrated into a single OWL ontology, together with the DOLCE foundational ontology and the Dublin Core metadata model.

In the context of FAO Agricultural Information Management Standards (AIMS) project (http://www.fao.org/aims) Agrovoc has been converted into SKOS (ftp://ftp.fao.org/gi/gil/gilws/aims/kos/agrovoc_formats/skos/2006) and is being mapped to two other vocabularies: the Chinese Agricultural Thesaurus (CAT) and the National Agricultural Library thesaurus (NAL). This mapping uses links inspired by the SKOS mapping vocabulary. A Sesame RDF Web repository containing the SKOS version of the GTAA thesaurus (Common Thesaurus for Audiovisual Archives) is used in the project, CHOICE@CATCH for the automated ranking of candidate terms for description of radio and TV programs. SKOS is used to describe all lexical qualities in BIRNLex, a lexicon of concepts covering clinical neuroimaging research developed by the Biomedical Informatics Research Network (BIRN).

5.2. Need for faceted KOS

SKOS is proved to be good for thesauri. SKOS can clearly represent properties of concepts, and relationships among or between the concepts — both hierarchical and associative. Hierarchical relations can be displayed using skos:hasTopConcept, skos:broader, skos:narrower, etc., while associative relations can be represented using skos:related. But in an enumerative classification system, SKOS fails to provide the order of main classes and other topics. This issue could be tackled to an extent by treating the main classes as independent schemes and with use of skos:OrderedCollection to include main classes and skos:memberList to show the members in an array by preserving their order. Another problem is the absence of a vocabulary element to denote alternative classification notations or cross-listed notations. The skos:notation element does not record how a classification notation is built, which restricts the further manipulation of the class numbers in a machine-assisted environment. Although SKOS provides skos:hasTopConcept for classes in the main schedule, no provision is available to represent auxiliary tables without semantic loss (Zeng, Fan 6 Lin, 2008).

5.3. Colon Classification (CC) modelling in SKOS

In its development towards being a W3C standard, SKOS has been trying to add new elements for better representation of KOS. The present specification, as directly demonstrated above, does not support the requirements for representing faceted classification schemes. Unlike enumerative schemes, the structure of faceted schemes mainly consists of isolates distributed over fundamental categories and its rounds and levels. In addition, analytico-synthetic structures provide for the combination of the various basic subjects, isolates, etc. to represent more complex subjects. An approach to dealing with such kind of complex structures in SKOS could be by grouping concepts using vocabulary elements such as skos:Collection, skos:OrderedCollection, skos:member, skos:memberList, etc.

Consider an example of subject approach to a collection that we try to facetize. We attempt representing the structure of Colon Classification (CC) (Ranganathan, 1960) in SKOS. Consider a competency question where a user requires a report on "Eliminating drug use among urban youth in Northern India in the year 2008". CC analysis of the complex query is presented below (a detailed discussion of facet analysis is beyond the scope of this paper):

> **Title:** "A report on eliminating drug use among urban youth in Northern India in the year 2008".
> Facets:
>
> *Basic subject:* Sociology
> *Personality:* Youth (by age)
> Urban (by residence)
> *Energy:* Drug habits (round1)
> Treatment (round2)
> *Space Isolate*: North India
> *Time Isolate:* Year 2008
> *Common Isolate:* Report

From the above analysis we can figure out that for representing it into SKOS we would need proper labelling for *Basic subject*, *Personality* (two isolates superimposed), two rounds of *Energy*, and labels for *Space*, *Time* and other *Common isolates.*

5.4. Faceted structure in SKOS

Let us now examine how SKOS can be adopted for such a faceted structure. For representing the basic subject, *Sociology*, there can be two approaches i.e., it can be treated as an independent scheme by using skos:ConceptScheme or by treating sociology as a skos:Concept for representing the occurrences *urban* and *youth* of the facet *Personality*. For one instance, skos:OrderedCollection can be used, but again the superimposition is not possible as is required to represent concept behind *urban youth*. Similarly, for *Energy isolates* the concepts of rounds cannot be conveyed.

Likewise, in *Space* and *Time* there is again the problem of representing levels. To represent common isolates, independent schemes can be made for them and can be borrowed anytime using skos:inScheme property but with semantic loss.

While treating the basic subject as an independent scheme, the fundamental categories (PMEST) can be considered as skos:hasTopConcept of underlying skos:Collection, but in another approach where the basic subject is a skos:Concept, there should be an extension like skos:hasTopCollection.

6. Conclusion

As discussed in the section above, faceted structures point to concepts by the topical scope as well as by the context albeit being rather complex systems. Use of these provisions makes the SKOS syntax of faceted structures very lengthy and complex. There are other issues, such as skos:Collection not being semantically equivalent to "facets". Therefore, there is a high need to develop SKOS extensions in order to solve representation problems for complex structures such as those of faceted classifications. Alistair Miles reported that "…he decided to drop the skos:Facet class, and the skos:inFacet and skos:facetMember properties in SKOS-Core 1.0 after there was some contention as to whether this had been modelled in the right way" (Miles, 2005).

One of the reasons for advocating the analytico-synthetic classification approach is its ability to seamlessly accommodate (hospitality) new subjects/topics. As subjects are continuously evolving and new terminology keeps pouring in, it is essential to adopt a system that can accommodate emerging new terms. There have been experiments in this direction of automatic updating of classification schedules, which could be applied to building Classauri (Madalli, 2001). Systems incorporating combination of results of knowledge representation systems using AI techniques (Panigrahi & Prasad 2007), Natural Language Parser to generate subject strings based on existing library classification systems in a hybrid model may be more capable of attempting complex context sensitive retrieval. The pre-requisites for such systems are (Prasad, *op.cit.*):

a. Natural Language Parser to generate subject strings based on existing library classification systems.
b. A knowledge base for representation to represent the schedules/classauri.
c. A rule-base for constructing pre-coordinated subject strings.

Although, at present, most of the schemas for ontologies are based on RDF/OWL/SKOS, with an emphasis on representing subject thesauri, such ontologies can be sensitized to facetization. This warrants extension of OWL or SKOS which can be really a faceted KOS or ontology schema that could accommodate fundamental categories as enunciated by the analytico-synthetic approach in classic library systems.

References

All URL valid as of 30 July 2009

Austin, D. (1974) *PRECIS: A manual of concept analysis and subject indexing*. London: BNB.

Baker, T.; Lynch, C. (1998) EU-NSF Working Group on Metadata. *ERCIM News*, 33 (April). Available at: http://www.ercim.org/publication/Ercim_News/enw33/baker.html.

Baldacci, M. B., et al. (1998) Implementing the Common User Interface for a Digital Library: the ETRDL experience. In *Proceedings of the EIGHTH DELOS WORKSHOP: User Interface in Digital Libraries*. Stockholm, Sweden, 21-23 October 1998. Available at: http://www.ercim.org/publication/ws-proceedings/DELOS8/baldacci.html.

Berners-Lee, T.; Hendler, J.; Lassila, O. (2001) The Semantic Web. *Scientific American*, May 2001, pp. 29-37.

Bhattacharyya, G. (1981) Subject indexing language: its theory and practice. In *Proceedings of the DRTC Refresher Seminar – 13, New Developments in LIS in India*. Bangalore: DRTC.

DESIRE (1997) *The role of classification schemes in Internet resource description and discovery*. Available at: http://www.ukoln.ac.uk/metadata/desire/classification/classification.pdf.

Haynes, D. (2004) *Metadata for Information Management and Retrieval*. London: Facet Publishing.

Hill, L. L.; Buchel, O.; Janée, G.; Zeng, M. L. (2002) Integration of Knowledge Organization Systems into Digital Library Architectures. In J.-E. Mai: C. Beghtol; J. Furner; B. Kwasnik, eds. *Advances in Classification Research, Proceedings of the 13th ASIST SIG/CR Workshop on Reconceptualizing Classification Research, Philadelphia, USA*, pp. 62-68. Available at: http://www.alexandria.ucsb.edu/paper_drafts/KOSpaper7-2-final.doc.

Hodge, G. (2000) *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Council on Library and Information Resources. Available at: http://www.clir.org/pubs/reports/pub91/contents.html.

IFLA (2004) *UNIMARC: a brief overview*. Available at: http://www.unimarc.net/brief-overview.html.

Isaac, A.; Phipps, J.; Rubin, D. (2007) *SKOS Use Cases and Requirements: W3C Working Draft 16 May 2007*. Available at: http://www.w3.org/TR/skos-ucr/.

Library of Congress (2006) *MARC 21 concise format for bibliographic data*. Available at: http://www.loc.gov/marc/bibliographic/.

Madalli, D. P. (2001) *A Frame Based Knowledge Representation Model for Management of Analytico-Synthetic Scheme: A case study in Colon Classification*. Mysore, India: University of Mysore.

Miles, A.; Bechhofer, S. (2009) *SKOS Simple Knowledge Organization System Reference: W3C Proposed Recommendation 15 June 2009*. Available at: http://www.w3.org/TR/2009/PR-skos-reference-20090615/.

Miles, A. (2005) *Extensions for faceted classification schemes*. Discussion thread. Available at: http://markmail.org/message/j76cfs4xb4qvzvw3.

Panigrahi, P.; Prasad, A.R.D. (2007) Facet sequence in analytico synthetic scheme: a study for developing an AI based automatic classification system. *Annals of Library and Information Studies*, 54(1), pp. 37-43.

Prasad, A.R.D. (1995) *Application of Natural Language Processing Tools and Techniques in Developing Subject Indexing Languages*. Dharwar, India: Karnataka University.

Prasad, A.R.D.; Madalli, D. P. (2008) Faceted infrastructure for semantic digital libraries. *Library Review*, 57(3), pp. 225-234.

Ranganathan, S. R. (1954a) Chain procedure and dictionary catalogue. *Annals of Library Science and Documentation*, 1(4), pp. 216-221.

Ranganathan, S. R. (1954b) Subject heading and facet analysis. *Journal of Documentation*, 20, pp. 109-119.

Ranganathan, S. R. (1960) *Colon Classification*. 6th. ed. Bangalore, India: SRELS.

RUBRIC (2007) *ASRC Subject Codes for DSpace*. Available at: http://rubric.edu.au/techreports/tech_report-asrc_for_dspace.pdf.

Smith, J. R.; Schirling, P. (2006) Metadata standards roundup. *IEEE Multimedia*, 13(2), pp. 84-88.

Soergel, D. (2008) *Digital Libraries and Knowledge organization*. Available at: http://www.dsoergel.com/NewPublications/SoergelDigitalLibrariesandKnowledgeOrganization.pdf.

Witten, I. H. (2003) Customizing digital library interfaces with Greenstone. *TCDL Bulletin*, Summer 2003. Available at: http://www.ieee-tcdl.org/Bulletin/v1n1/witten/witten.html.

Zeng, M. L.; Fan, W.; Lin, X. (2008) SKOS for an Integrated Vocabulary Structure. In *Metadata for Semantic and Social Applications, Proceedings of the International Conference on Dublin Core and Metadata Applications*, Berlin, Germany, 22-26 September 2008. Berlin: Universitätsverlag Göttingen, pp. 200-201.