Annals of Library and Information Studies
Vol. 62, December 2015, pp. 274-280

Use of Ranganathan's analytico-synthetic approach in developing a domain ontology in library and information science

Shrabana Ghosh[a] and Pijushkanti Panigrahi[b]

[a]Librarian, Jadavpur University, Jadavpur, Kolkata-700032, Email: gshrabana@yahoo.co.in

[b]Professor, Department of Library and Information Science & Dean, Faculty Council of P G Studies in Education, Journalism and Library Science, University of Calcutta, Kolkata-700073, Email: panigrhipk@yahoo.com

Classification is the basis of knowledge organization. Ontology, a comparatively new concept used as a tool for knowledge organization, establishes connections between terms and concepts enhancing the scope and usefulness of library classification. Ranganathan had invented the strong theory of the analytico-synthetic method in classification and devised Colon Classification. In this study a domain ontology on library and information science has been developed by implementing Raganathan's faceted approach of classification. The hierarchical relationships among terms have been established primarily keeping conformity with that of Ranganathan's Colon Classification (7th edition). But to accommodate new vocabularies, DDC 23rd edition and UDC Standard edition are consulted. The Protégé ontology editor has been used. The study carefully examines the steps in which the analytico-synthetic method have been followed. Ranganathan's Canon of Characteristics and its relevant Canons have been followed for defining the class-subclass hierarchy. It concludes by identifying the drawbacks as well as the merits faced while developing the ontology. This paper proves the relevance and importance of Ranganathan's philosophy in developing ontology based knowledge organization.

Keywords: Aanalytio-Synthetic approach; Ranganathan; Ontology; Domian ontology; Library and information science

Introduction

Classification has long been used in library and information systems to provide guidance to the user in clarifying the information need and to structure search results for browsing. Classification (together with indexing, document description, and metadata assignment) forms the basis of knowledge organization (KO), both a practical activity and a main sub-discipline of library and information science (LIS), which focuses on improving this activity. In the context of librarianship, "classification" is often used for systems such as DDC, UDC, LCC, Colon, or Bliss, and its process involves assigning individual books or other documents to one or more classes within such a system, as opposed to verbal indexing systems. The fundamental distinction between classification systems and verbal indexing systems has been criticized by Lancaster[1]. Bhttacharyya proved that classification = indexing. Aitchison[2] supported this view by the fact that a classification system may be transformed into a thesaurus and vice versa[2]. Gunnarsson[3] further emphasizes the view: "Indexing in particular, but classification as well, stresses the assignment of labels to documents rather than the assignment of documents to classes of documents"[3].

A recent overview of knowledge organization systems have been presented by Hodge[4]. He grouped them into three general categories:

(i) term lists, which emphasize lists of terms, often with definitions; e.g. glossaries, dictionaries.
(ii) classifications and categories, which emphasize the creation of subject sets; e.g. subject headings, classification schemes, taxonomies and
(iii) relationship lists, which emphasize the connections between terms and concepts e.g. thesauri, semantic networks, ontologies.

Ontology and semantic web

The World Wide Web has changed the way people communicate with each other and the way business is conducted. The activities on the World Wide Web are not particularly well supported by software tools. In order to overcome these problems, sophisticated techniques based on artificial intelligence and computational linguistics have been developed. This alternative approach to represent Web content in a form that is more easily machine-processed and to use intelligent techniques to take advantage of these representations has revolutionized the existing web. This approach of revolutionizing the Web is referred

to as the semantic web initiative. The semantic web is propagated by the World Wide Web Consortium (W3C), an international standardization body for the Web. The driving force of the semantic web initiative is Tim Berners-Lee.

At present, the semantic web is in need of new technologies and development of tools. Ontology is one such technological tool developed for the semantic web. The term ontology originates from the discipline philosophy. In that context, it is used as the name of a subfield of philosophy, namely, the study of the nature of existence, a branch of metaphysics. It is concerned with identifying the kinds of things that actually exist, and how to describe them. However, in recent years, ontology has become one of the words of computer science. Gruber[5] defines ontology (and later refined by Studer) as, "An ontology is an explicit and formal specification of a conceptualization". "Conceptualization" refers to an abstract model of phenomena in the world by having identified the relevant concepts of those phenomena. "Explicit" means that the type of concepts used, and the constraints on their use are explicitly defined. "Formal" refers to the fact that the ontology should be machine readable. "Shared" reflects that ontology should capture consensual knowledge accepted by the communities.

Ontologies are designed to provide a set of vocabulary representing concepts and their relationships that allow searching the unstructured data and metadata in the digital environment more effectively. They add semantic contexts for searches.

Analytico-synthetic approach in knowledge organization system

Analytico-synthetic scheme, according to Ranganathan, is used 'to denote any scheme in which a compound subject is first analyzed into its facets in the idea plane and later synthesized in the verbal plane and in the notational plane respectively'. Colon Classification (7th edition) is an analytico-synthetic scheme for library classification. S.R. Ranganathan[6], the founder of the scheme had an analytical mind which found some deficiencies in Dewey Decimal Classification (DDC). DDC could not represent expressively all the aspects of a specific subject. He was also aware of the enumerative nature of DDC. It was not possible to enumerate all subjects. The combination of ideas would be a possible alternative. This was achieved by developing a Theory of Classification. This development consisted of the Postulate of Five Fundamental Categories. According to this Postulate, each isolate facet in a compound subject should be deemed to be a manifestation of one and only one of the Five Fundamental Categories (FC)—Personality, Matter, Energy, Space and Time (PMEST). Colon Classification provides provision to interpolate new main subjects. The scheme also has provision for extrapolation at the end of each species of digits. The last digit of each species was postulated to be semantically empty but to retain its ordinal value.

Development of domain ontology in library and information science

Domain ontology is only applicable to a domain with a specific view point. That is to say that this viewpoint defines how a group of users conceptualize and visualize some specific phenomenon. Domain ontologies differ according to viewpoint of a developer and specificity.

Hence the scope of the domain under study has to be determined. Since the purpose of the creation of domain ontology is to ensure better retrieval, the domain has to be classified properly so that all the major concepts of the domain are covered.

There are two approaches followed to engineer ontologies i.e. top-down and bottom-up approach. The top-down approach is always encircled by competency questions to model a domain. Domain mapping is generated at a generic level and subsequently they are refined to shape the possible restrictions with feasibility. The concepts in the ontology are derived from an analysis and study of relevant information sources about the domain (collection of attributes).

Bottom-up approach may be applied to populate the classes with lexical entities collected from standard documents.[7] But, it can be easily inferred that, both exist and work at the idea and verbal planes. Here, both the approaches have been used to design the domain ontology as merging the two approaches facilitate pragmatics at different levels of implementation and semantic navigation. So, it is a hybrid approach. Specifically, top-down approach is implemented to classify the basic subdivision of the library and information science as a subject itself. As the subject is ever-growing, new terms and concepts are added to the domain through bottom-up approach.

The methodological steps of ontology development are-

a. Determining the domain and scope of the ontology.
b. Judgment of domain classes-subclasses and properties through analytico-synthetic method.
c. Collection of important terms from schedules and literature.
d. Defining classes-subclasses hierarchies.
e. Reusing of existing classificatory systems.
f. Determining facets through properties.
g. Creating instances
h. Evaluation

The concepts used as classes in the domain of library and information science have been collected from Colon Classification (7th edition), Universal Decimal Classification (standard edition)[8], Dewey Decimal Classification (23rd edition)[9]. However, according to these three schedules most of the concepts of library science were covered but several concepts of information science were left out.

Analytico-synthetic approach to ontology design

The faceted approach allows users to search or browse with greater flexibility. Each of the facets is a mutually independent category and may contain any number of isolates (content oriented metadata/facet element), or subcategories arranged in a hierarchy. The category hierarchy can then be used to describe, organize, and access the resource by browsing or querying. Mutually independent means each facet describes one single aspect of information object and should not contain the element of other facets. In a faceted approach one term should denote one concept.[10] The faceted approach has been followed while developing this ontology. The subclasses under the class library and information science have been defined and then object properties are used to build relationship among the classes. Compound subjects can be defined to a great extent. As in the faceted scheme of Colon Classification, there are separate schedules for Space, Languages and Basic Subjects, the ontology has treated them as Countries, Languages and Subjects as separate classes. They have relationship with the class library and information science through object properties.

The analytico-synthetic method, advocated by S.R. Ranganathan, has been followed for designing the ontology. In the analytico synthetic method, two processes are used: analysis of the collected term and then synthesizing the concept into a new class. The process of analysis involves controlling the form of the term, making a choice among synonyms, taking a decision whether a concept will be treated as a class or an instance, how are the classes related to each other, what type of property is to be assigned in a particular case etc. The answers to these questions are unified or synthesized to form a new concept. The following steps are followed after the collection of terms (which denote concepts) in a subject domain.

Control vocabularies for designating concepts

Control of terminology is achieved in various ways.[11, 12] Firstly, the form of the term is controlled, whether this involves grammatical form, spelling, singular and plural forms, abbreviations or compound form of terms. Secondly, a choice has to be made between two or more synonyms available to express the same concept. Thirdly, a decision has to be taken whether to treat a proper noun is to be defined as a class or an instance. Proper nouns are generally defined as instances of a class. But in the present ontology all individual associations have been treated as classes so that relationships can be established by object properties.

Forms of terms: The terms in the subject field may take different forms to denote concepts.

Nouns and noun phrases: A term preferably consists of a noun or noun phrase. Noun phrase belong to the category of compound terms and occur in two forms:

Adjectival phrases: e.g. *legal aspects* of information

Prepositional phrases: e.g. types of libraries

Adjectives: Adjectives alone cannot be used to denote a concept. They are generally joined with nouns to denote a concept. As for example, archival materials.

Adverbs: Adverbs such as ‘very’ or ‘highly’ are not to be used alone. A phrase beginning with an adverb may be accepted if it has acquired a special meaning within a jargon.

Verbs: Verbs may be used to denote a concept if it has acquired a special meaning within a jargon. As for example, indexing, abstracting.

Abbreviations: In case of an ontology the abbreviated form or an acronym may be treated as an equivalent class of the original class. For example, BLA is an abbreviation of Bengal Library Association. BLA is treated as an equivalent class of Bengal Library Association.

Obsolete terms: While collecting terms from the literature of Library and Information Science, it is found that some terms are to be considered as obsolete from the point of connotation. As for example, the term 'reading materials' has been used in Colon Classification (7th edition) to denote the different materials that may be included in a library. But this term might not include the concepts of all types of audio visual materials which may be a part of collection in some libraries. In the present day, a film library may have CDs and DVDs on films. A more appropriate term to designate the concept would be 'library materials'. The term was collected from AACR2R 2005.

Punctuation marks: Punctuation marks should be minimized. In the Protégé ontology editor, each class has to be treated as a single term. In order to treat a compound term as a single term the '_' (underscore) has been used. As for example, library_and_information_science.

Character set: The lower case characters have been used to designate a class.

Extrapolation and Interpolation: As found in the Colon Classification, the methods of extrapolation and interpolation can be applied to the ontology by incorporating new concepts as a sibling class of an already defined class.

Ranganathan's canon of characteristics

While classifying a concept into its composite concepts a characteristic is to be applied i.e. to create two groups or subclasses. Ranganathan defines 'characteristics' as "an attribute or any attribute-complex with reference to which the likeness or unlikeness of entities can be determined and at least two of them are unlike."

Ranganathan in his *Prolegomena to Library Classification*[13] defined the canons which regulate the work in the idea plane while designing the scheme of Colon Classification. One of these canons is the Canon of Characteristics. Ranganathan further says that each characteristic of an associated scheme of characteristics should satisfy the following four canons:-

1 Canon of Differentiation;
2 Canon of Relevance;
3 Canon of Ascertainability; and
4 Canon of Permanence.

These canons are applicable to the classification of any entity.

According to Ranganathan, the Canon of Differentiation states: A characteristic used as the basis for the classification of a universe should differentiate some of its entities-that is, it should give rise at least to two classes or ranked isolates. The characteristic "height" differentiates the universe of men.

According to Ranganathan, the Canon of Relevance states: A characteristic used as the basis for the classification of a universe should be relevant to the purpose of the classification. For example, if the universe of boys are divided into graded groups for tutorial work, then mother tongue, intelligence, etc. would be relevant characteristics but not mode of dressing hair, and clothes.

According to Ranganathan, the Canon of Ascertainability states: A characteristic used as the basis for the classification of a universe should be definite and ascertainable. As for example, date of birth in the universe of poets.

According to Ranganathan, the Canon of Permanence states: A characteristic used as the basis for the classification of a universe should continue to be unchanged so long as there is no change in the purpose of classification.

In designing the present ontology, Ranganathan's Canon of Characteristics along with its associated Canons have been followed.

Defining classes-subclasses hierarchy

After the identification of relevant terms, these terms must be organized in a taxonomic hierarchy. It is important to ensure that the hierarchy superclass-subclass hierarchy. In other words, if A is a subclass of B, then every instance of A must also be an instance of B.[14] Here, in order to organize the relevant terms or concepts into a taxonomic hierarchy Ranganathan's Canon of Characteristics and its relevant canons have been followed.

There are two important issues which are encountered while defining class hierarchy. The first one relates to distinguish between classes and instances. It must be decided if some concepts represent a subclass or an instance of a certain class. This also depends on the viewpoint with which the ontology is created.

The second problem is to decide which characteristics of concept should be represented in a number of subclasses and which should be put into properties.[15]

The determination of super class-subclass hierarchy provides the organized taxonomic structure of the subject domain. This definition of super class-subclass hierarchy of the subject domain is the product of analytico-synthetic approach.

Creation of classes, subclasses and sibling classes using Protégé

Protégé originates from Stanford University. It is a free, open source ontology editor and a knowledge acquisition system written in Java language. It allows users to create ontologies in RDFs and OWL languages. Protégé allows to manipulate classes using tree-like structure (creating classes, subclasses, attaching properties etc.). Protégé 4.2 has been used for creating the OWL (Web Ontology Language) ontology.[16]

The main building blocks of an OWL ontology are classes. In Protégé, the empty ontology contains one class called 'Thing'. OWL classes are interpreted as sets of individuals (or sets of objects). The class Thing is the class that represents the set containing all individuals. Hence all classes are subclasses of Thing. Here, library and information science has been treated as a subclass of Thing. The schedule of subjects, place or countries and languages are all treated as separate classes under the class Thing**.**

The 'Classes' Tab has to be selected. This will open the class hierarchy page showing the Thing class. The 'Add Subclass' button has to be pressed. This creates a new subclass under the selected class Thing. A dialog box will appear to name the class. Here the subclass library_ and_information_science has been entered under the class Thing. Similarly, application_areas_of_information, information_evaluation_and_research, legal_aspects_ of_information, library_activities etc. are all subclasses of library_and_information_science and are created in the same way.

The 'Add Sibling class' button creates a new class in the same array of the selected class. When the selected class is library_and_information_science and 'Add Sibling class' is pressed, a dialog box appears to name the class. Here the sibling class 'countries' has to be entered. Similarly, 'languages', 'subjects' are entered as sibling classes of library_and_ information_science class. The screenshot of the class library_and_information_science, its subclasses and the sibling classes using Protégé are being shown. Fig. 1 shows the screenshot of the class library_and_information_science, its subclasses and the sibling classes.

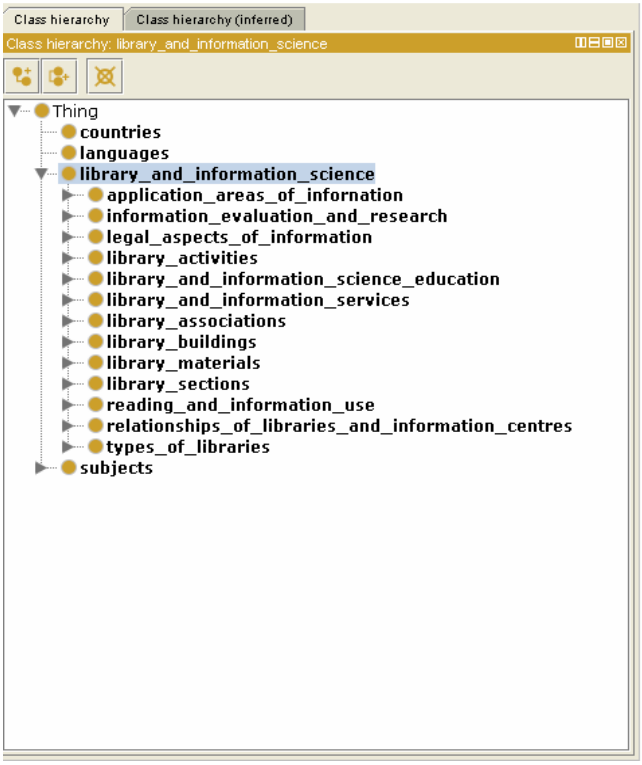

Fig. 1—The screen shot of the subclasses and the sibling classes of the class 'library_and_information_science' in Protégé.

Determine facets through properties

This step is intertwined with the previous step. Properties link the classes while organizing these classes in a hierarchy. Properties (or slots) describe internal structure of concepts. A property may point to simple value such as a string or a date (data type properties) or a class instance (object properties). The data type properties are used to describe object's physical characteristics (intrinsic properties) as well as abstract concepts (extrinsic properties). Object properties are used to represent relationships between individuals. They are also used to represent object's parts, if the object is structured.

Creation of properties using Protégé

OWL Properties represent relationships. There are two main types of properties, Object properties and Datatype properties. Object properties are relationships between two individuals.

Object properties may be created using the 'Object Properties' tab. The 'Add Object Property' button is used to create a new Object property. The property is to be added in the 'Property Name Dialog' box. In this study some of the Object properties are has Activites, has Areas, has Collection Of etc.

The screenshot of the object properties of the ontology developed in library and information science using Protégé is being shown in Fig. 2.

Property domain and range

Since the object property expresses relationship between two individuals, the two individuals are expressed as domain and range. Properties link individuals from the domain to individuals from the range. For example, in this ontology, the property 'hasActivities' would link individuals belonging to the class 'types_of_libraries' to individuals belonging to the class of 'library_activities'. In this case the domain of the 'hasActivities' property is 'types_of_libraries' and the range is 'library_activities'.

Creating instances

A class is chosen and an individual is created of the chosen class. The individual may have property values. Instances form the actual semantic description. In most of the cases, the number of instances is larger than the number of classes in an ontology.

Evaluation

Any development is incomplete without evaluation as it acts as a check-gate to verify the appropriateness and feasibility of the conceptual model and addresses the robustness of the development in order to cater different necessities of the target groups/stakeholders. Ontology evaluation is still an open area of research and cannot instruct the full solution. Generally, ontology is evaluated by judging its formal quality

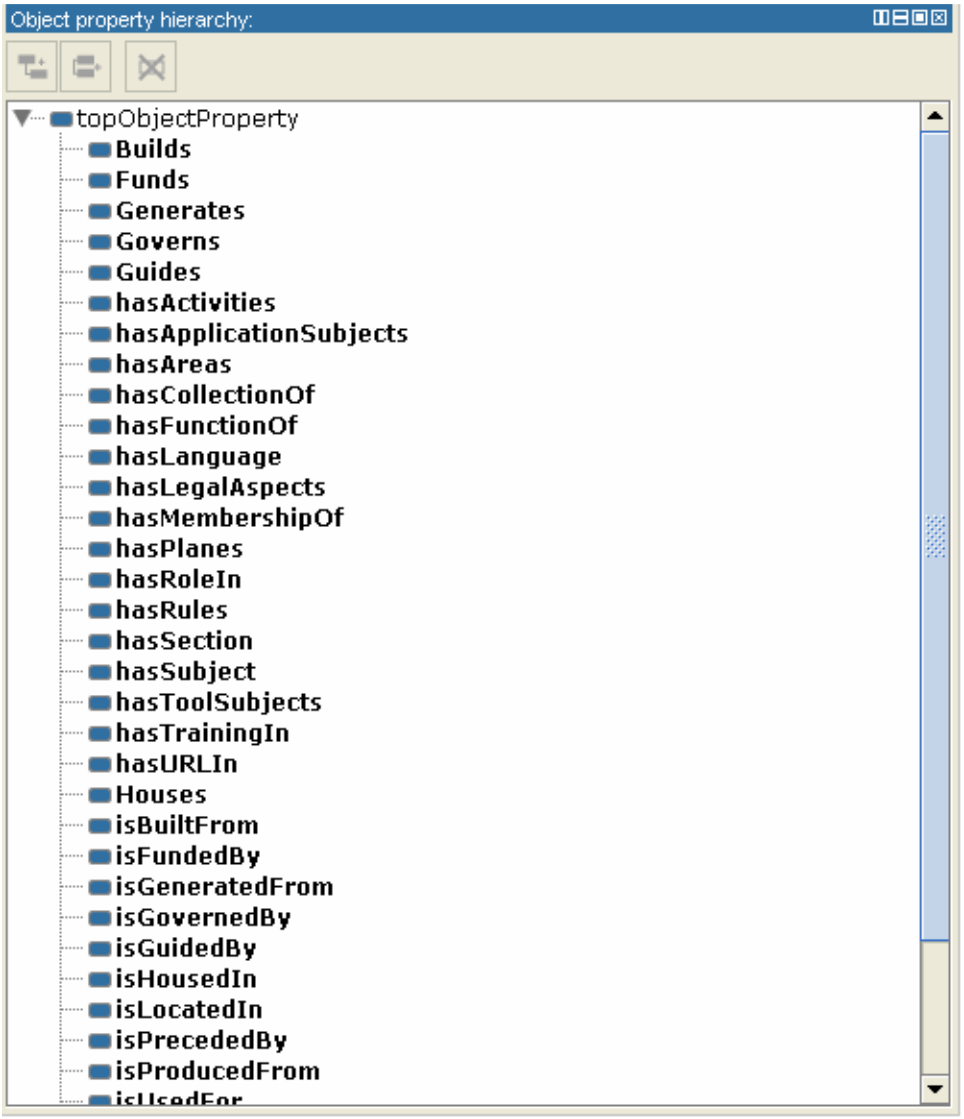

Fig. 2—The screen shot of the object properties of the ontology on library and information science in Protégé.

and technical efficiency[17]. As the work is on building domain ontology, the evaluation is based on validating two aspects. They are design feasibility and expected utility. The content evaluation process is centered on people[18], in contrast with the majority of evaluation methods available. Some competency questions arising here are:

a. What is the extent to which the ontology can give the appropriate term and associated concepts with relations?
b. Level of granularity in defining the concepts.
c. Facilitation of knowledge flow in which direction?
d. Pluripotency of the ontology.

The competency questions are taken as a baseline for the evaluation and different questions were framed by some experts to get effective answers.

Conclusion

The development of ontology is a job, which requires collaboration between IT professionals and domain experts. The present work enabled to understand the semantic mapping of the subject domain library and information science. In spite of using standard vocabularies, many of the concepts of information science were left out. These concepts were accommodated from journal articles. It is expected that constant evaluation of the ontology will be able to meet the changing needs of the users.

The fundamental difference between an ontology and a conventional representational vocabulary is the level of abstraction, relationships among concepts, the machine understandability and the expressiveness that can be provided by ontologies[19]. However, building domain ontologies based on classificatory schemes[20] has helped to a great extent in determining the taxonomic hierarchy of the subject domain.

References

1 Lancaster FW, *Indexing and abstracting in theory and practice*, (Library Association; London), 2003

2 Aitchison J, A classification as a source for thesaurus: the bibliographic classification of H.E. Bliss as a source of thesaurus terms and structure. *Journal of Documentation*, 42(3) (1986) 160-181.

3 Gunnarsson M, *Classification along genre dimensions: exploring a multidisciplinary problem*, PhD thesis, Swedish School of Library and Information Science, University of Bora°s, 2011. Available at http://bada.hb.se/handle/2320/7920 (Accessed on May 11, 2011)

4 Hodge G, *Systems of knowledge organization for digital libraries*: *beyond traditional authority files*, Council on Library and Information Resources, Washington, DC (2000). Available at www.clir.org/pubs/reports/pub91/contents.html (Accessed on 16 May 2011).

5 Gruber T R, A translation approach to portable ontology specifications, *Knowledge Acquisition,* 1993, 5(2), 199-220. Available at http://tomgruber.org/writing/ontolingua-kaj-1993.pdf (Accessed on 22 July 2015)

6 Ranganathan S R, *Colon Classification*, Gopinath M A, (Rev and Ed), 7th edn, (Sarada Ranganathan Endowment for Library Science; Bangalore), 1987

7 Davis J, Fensel D and Van Harmelen F, *Towards the semantic web: Ontology-driven knowledge management*, (Wiley; USA), 2003

8 Universal Decimal Classification, Standard edn, (British Standard Institution; London), 2005

9 Dewey Melvil, *Dewey Decimal Classification and Relative Index*, Mitchel Joan S and et al (Ed) , 23rd edn (OCLC; Dublin), 2011

10 Nasir Uddin Mohammad and Janecek Paul, Faceted classification in web information architecture, *The Electronic Library,* 25 (2) (2007) 219 – 233. Available at http://dx.doi.org/10.1108/02640470710741340 (Accessed on 12 June 2015)

11 Aitchison Jean and Gilchrist Alan, *Thesaurus Construction: a practical manual*, 2nd edn (Aslib; London), 1987, ISBN 0-85142-197-0

12 National Information Standards Organization, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, (National Information Standards Organization; Maryland) , 2005

13 Ranganathan S R, *Prolegomena to Library Classification*, Gopinath M A (Assisted), 3rd edn, (Asia Publishing House; Bombay), 1937

14 Antoniou Grigoris and Harmelen Frank van, *A semantic web primer*, 2nd edn (Massachusetts Institute of Technology; Massachusetts), 2008

15 Kruk Sebastian Ryszard and McDaniel Bill (Eds), *Semantic Digital Libraries*, (Springer; Germany), 2009

16 Horridge Mathew, Jupp Simon, Moulton Georgina, Rector Alan, Stevens Robert and Wroe Chris, *A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools*, 1.1edn, (University of Manchester; Manchester), 2007

17 Pattuelli M C, Modeling a domain ontology for cultural heritage resources: A user-centered approach, *Journal of the American Society for Information Science and Technology*, 62(2) (2011) 314-342.

18 Almeida Mauricio Barcellos and Barbosa Ricardo Rodrigues, Ontologies in Knowledge Management Support: A Case Study, *Journal of the American Society for Information Science and Technology*, 60(10) (2009) 2032–2047, doi 10.1002/asi (Accessed on 12 May 2014)

19 Ding Ying and Foo Schubert. Ontology research and development, Part 1 - a review of ontology generation, *Journal of Information Science*, 28 (2) (2002), 123–136. Available at http://dx.doi.org/ 10.1177/016555150202800204 (Accessed on 22 January 2013)

20 Soergel D, The rise of ontologies or the reinvention of classification, *Journal of the American Society for Information Science*. 50 (12) (1999) 1119–1120, doi 10.1002/(SICI)1097-4571(1999)50:12<1119::AID-ASI12>3.0.CO (Accessed on 24 January 2015)