Ontology and Ontological Systems for Semantic Webs in Digital Libraries

Anil Kumar Dhiman

Abstract

The amount of available information online has increased exponentially with the rapid development of the World Wide Web. A lack of standardization and common vocabulary has continued to generate heterogeneity, which strongly hinders information exchange and communication. Ontologies, which capture the semantics of information from various sources and giving them a concise, uniform and declarative description, have gained significance due to the demands in academia and industry. This paper discusses ontology and its uses in semantic webs for digital libraries.

Keywords: Ontology, Semantic Webs, Digital Libraries

1. Introduction

In the recent past, the development in information and communication technology has led the emergence of various types of library resources and services – digital libraries, blogs, weblogs, Library 2.0 services and institutional repositories etc., (Dhiman, 2003, 2007 & 2008; Dhiman and Sharma, 2008a,b). Because of these developments, we are able to give accurate and fast retrieval of information to the users. Webs and web – based services are gaining much importance in the libraries. Moreover today's, semantic webs are getting much favour in the digital environment.

It was Tim Berners-Lee, the inventor of the World Wide Web who has described his vision on Semantic Web along with his other colleagues in the year 2001. He with co-workers envisioned that the semantic web would bring structure to the content of Web pages and enable computers to perform sophisticated tasks for people (Berbers-Lee, et al., 2001). Since then, many studies on ontology and the semantic web have been carried in many parts of the world. Many efforts have been made for surveying ontology-related research studies from various aspects, including that of ontology representation languages (Corcho and Gomez-Perez, 2000), ontology development (Jones et al., 1998), and ontology learning approaches (Daconta et al., 2003). It is expected that semantically enabled technology will bring a number of benefits to the users of corporate digital libraries. In particular, the technology will help people to find relevant information more efficiently and more effectively, give better access to that information, and aid the sharing of knowledge within the user community of a digital library.

2. Semantic Relations

Semantic relations are meaningful associations between two or more concepts, entities, or sets of entities (Khoo and Na, 2006). As a new information representation system, ontology aims to substantiate the rich variety of semantic relations among the concepts it represents – a characteristic that distinguishes it from other representation and organization systems. An ontology is a rich expression of semantic relations while a term list,

7th International CALIBER-2009,
Pondicherry University, Puducherry, February 25-27, 2009

free or controlled, is a natural arrangement of word forms. Ontology and the Semantic Web strive to express and enable semantic relations among represented entities.

Hodge (2000) grouped typical information representation systems into three general categories: term lists, classifications and categories and relationship lists. Term lists emphasize lists of terms usually presented with definitions; Classifications and categories emphasize the creation of subject sets; and the Relationship lists emphasize on the connection between terms and concepts. Hjørland (2007) summarized Hodge's list of systems into the following taxonomy:

Term Lists
- Authority Files
- Glossaries
- Dictionaries
- Gazetteers

Classifications and Categories
- Subject Headings
- Classification Schemes
- Taxonomies
- Categorization Schemes

Relationship Lists
- Thesauri
- Semantic Networks
- Ontologies

In term lists, terms that having specific meanings are listed, typically in alphabetical order, so that they can be easily accessed when needed. The associations among these terms normally do not go beyond their alphabetical order. In other words, the meaning of a term does not have any relation with the meaning of a term that comes before or after it. They are related by the order of alphabetic letters, not by the meaning they contain. The relations they indicate are generally not semantic relations. In the classifications and categories, terms or concepts are arranged hierarchically. A specific type of relation among terms or concepts determines the order. Those arranged in the higher level are in a higher class or a broader category and usually more inclusive in meaning than those arranged in the lower order. Hierarchical lists indicate, if not more, at least class-subclass semantic relations among terms and concepts that are associated in meaning. In the relationship lists, relations indicated among terms or concepts normally go beyond their hierarchical order. More semantic relations are constructed and expressed in relationship lists. Terms and concepts can be meaningfully associated, for instance, in hierarchical order (class-subclass), horizontal order (synonyms), reverse order (antonyms), or causation order (cause-effect).

The understanding of different semantic relations indicated in term lists, hierarchical lists, and relationship lists provides a useful framework to explain how ontology is different from or similar to other forms of representation models. Researchers in library and information science note that ontology is associated in one way or anther with traditional library representations such as a thesaurus, taxonomy, classification scheme, controlled vocabulary, or even a dictionary (Daconta et al., 2003; Jacob, 2003). To what extent traditional library representation models and ontology are associated, can be illustrated by arranging them in the following taxonomy:

Term Lists
- Controlled vocabulary
- Dictionary

Hierarchical lists

Classification Scheme

Taxonomy

Relationship Lists

Thesaurus

Ontology

Semantically speaking, the association between an ontology and representation models in the term list category remains fairly weak. The semantic tie between an ontology and representation models in the hierarchical list category increases as hierarchical semantic relations are present in an ontology as well as in a classification scheme and a taxonomy. However, Wang et al. (2006) have pointed out, classification schemes are largely tied to a paper-based environment and more constrained within the academic community while taxonomies are largely created in a Web environment to organize digital resources that are not limited within subjects. As a result, the taxonomy bears a closer tie to an ontology than a classification scheme. Daconta et al. (2003) noted that in the model of ontological representation lies an underlying taxonomical relationship and the basic taxonomic sub-class of hierarchies acts as the framework of ontologies. Welty and Guarino (2001) identified that some notions in a taxonomy are also used to represent the most important properties in an ontology, thus indicating strong mutual relationships between these two content representation forms.

According to Daconta et al. (2003), the basic taxonomic sub-class of hierarchies acts as the skeleton of ontologies, but ontologies add additional muscle and organs – in the form of elaborate relations, properties/attributes, or property values. Ontologies thus enable people to specify the semantics of their domain in great detail. Because of their rich semantic representation power, to equate ontologies with any other type of representational structure is to diminish both the function and potential of ontologies (Jacob, 2003). Jacob thus urged the library community to make a conscious effort to rethink the traditional representational approaches in light of the changing requirements generated by Web environments.

3. Ontological Tools

The task of maintaining and re-organizing ontology in order to facilitate the re-use of knowledge is becoming

challenging as the number of different ontologies is on the increasing. A breakthrough in ontology technology would require methodological aids and tools that enable effective and efficient development. A key aspect in achieving this is successful re-use of ontologies. Being developed for supporting knowledge sharing and reuse, it is the lack of proper support of ontology re-use that hampers a broader dissemination of the ontology. To facilitate the re-use of ontology, a library system must, at the very least, support the following:

- ontology re-use by open storage, identification and versioning.
- ontology re-use by providing smooth access to existing ontologies and by providing advanced support in adapting ontologies to certain domain and task-specific circumstances,

instead of requiring such ontologies to be developed from scratch.

- ontology re-use by fully employing the power of standardization. Providing access to upper-layer ontologies and standard representation languages is one of the keys to developing knowledge sharing and re-use to its full potential.

Ontology library systems are an important tool in grouping and re-organizing ontologies for further

re-use, integration, maintenance, mapping and versioning. An Ontology library system is a library system that offers various functions for managing, adapting and

standardizing groups of ontologies. An ontology library system should be easily accessible and offer efficient support for re-using existing relevant ontologies and standardizing them based on upper-level ontologies and ontology representation languages. For this reason, an ontology library system will, at the very least, feature a functional infrastructure to store and maintain ontologies, an uncomplicated adapting environment for editing, searching and reasoning ontologies and strong standardization support by providing upper-level ontologies and standard ontology representation languages. There are available different ontology library systems for semantic webs, but in general, they should possess following characteristics:

- **Management :** This function is the most important function of an ontology library system which facilitate the re-use of knowledge (ontologies).

- **Adaptation:** Ontology library systems should facilitate the task of extending and updating ontologies. They should provide user-friendly environments for searching, editing and reasoning ontologies. Important aspects in an ontology library system include support in finding and modifying existing ontologies.

- **Standardization**: Ontology library systems should follow existing or available standards, such as standardized ontology representation languages and standardized taxonomies or structures of ontologies.

Various ontology library systems have been developed by now for developing ontology of semantic webs. Some of the important systems are listed below :

DAML Ontology library system is the part of DARPA Agent Markup Language (DAML) Program, which officially started in August 2000. The goal of the DAML effort is to develop a language and tools to facilitate the concept of the Semantic Web. The ontology library system contains a catalogue of ontologies developed using DAML. This catalogue of DAML ontologies is available in XML, HTML, and DAML formats. People can submit new ontologies via the public DAML ontology library system.

IEEE Standard Upper Ontology (IEEE) is developed by IEEE Standard Upper Ontology (SUO) Working Group, who took a tremendous effort to create a standard top-level ontology to enable various applications, such as data interoperability, information search and retrieval, automated inferencing and natural language processing. Their ontology library system is very simple and is accessible in its preliminary form on their website. It contains a group of classified ontologies, such as, ontologies in SUO-KIF, formal ontologies and linguistic ontologies/lexicons. Only the very basic hyperlinks of the ontologies are provided to help users to jump to the home pages hosted by the ontologies. However, there are no clear management, adaptation and standardization functions.

ONIONS stands for ONtological Integration Of Naive Sources, is a methodology for ontology mediation, alignment and unification, which was developed in the early 1990s to account for the problem of conceptual heterogeneity. ONIONS creates a stratified design of an ontology library system. It contains richly documented and formalized generic ontologies and a cognitively transparent top level. Moreover, intermediate modules contain the most general concepts of a domain, based on the generic ontologies and the top level.

Ontolingua was developed in the early nineties at the Knowledge Systems Laboratory of Stanford University. It consists of a server and a representation language. The server provides a repository of ontologies to assist users in generating new ontologies and amending the existing ontologies collaboratively. The ontology stored at the server can be converted into different formats.

Ontology Server is developed by the Vrije Universiteit in Brussels. It links ontology engineering to database semantics. It deploys database techniques to manage and understand ontologies. The database management system (DBMS), equipped with various syntactical constructs, enables database diagrams to present objects, sub-type taxonomies, integrity constraints, derivation rules, etc.

OntoServer (AIFB) is an ontology server to support building, maintaining and using ontologies. It has a client/server-based architecture, which integrates various types of software or tools to form tool-based support for an ontology environment.

SHOE or Simple HTML Ontology Extensions is developed by the University of Maryland (USA). It is also the first web-semantics language developed as a markup, and has been used for various applications, including for food safety for the US Food and Drug Administration and a military logistics planning system.

WebOnto is an ontology library system developed by the Knowledge Media Institute of the Open University (UK). It is designed to support the collaborative creating, browsing and editing of ontologies. It provides a direct manipulation interface displaying ontological expressions and also an ontology discussion tool called Tadzebao, which support both asynchronous and synchronous discussions on ontologies.

Thus, we see that there are available many ontology library systems, but to ascertain their priority-wise categories, a survey of various available systems was made by Ding and Fensel (2002). This survey reveals the actual picture of various systems available for ontological semantic web; these are comparatively presented in the following table.

Table : Important Ontology Library Systems

Aspect of comparison	Characteristics	WebOnto	Ontolingua	DAML library	SHOE	Ontology Server	Others (SUO, IEEE OntoServer, ONIONS)
Management	**Storage**	- client/server-based - no classification - modularity storage	- client/server-based - no classification - modular structured library	- client/server-based - classification of ontology - no modularity storage	- web accessible - classification of ontology - tree structure of ontology dependency	- database access - no classification - modularity storage	- web access (IEEE SUO), client/server-based (OntoServer), - classification of ontology (IEEE SUO, ONIONS) - stratified design (ONIONS)
	Identification	- unique name - unique unit name	- unique name	- unique URI and Identifier	- unique name	- unique name	_
	Versioning	- indirect: inherited from ancestor ontology	No versioning	No versioning	- versioning support for ontology revision	- no versioning	- no versioning
Adaptation	**Searching**	- graphical display - simple browsing	- simple browsing - idiom-based retrieval facility for simple query answering - wild-card searching - context sensitive searching - reference ontology as the index	- simple browsing	- simple browsing	- database API - DBMS - add, modify, retrieve - ontology manager - ontology browser	- simple browsing (IEEE SUO)
	Editing	TaDzeBao: - asynchronous and synchronous discussions and editing on ontologies	- simple interfaces - collaborative ontology construction - vocabulary translation - undo/redo - hyperlinked environment	No specific editing functions	- no editing	- add, modify, retrieve	- no editing
	Reasoning	- rule-based reasoning	- use situation to determine the expected properties. - ontology testing	- no reasoning	- limited reasoning support for ontology revision	- no reasoning	- no reasoning
Standardization	**Language**	OCML	KIF - ontology language translation	RDF, RDFs, DAML+OIL	SHOE	XML	-
	Upper-level Ontology	- no standard upperlevel ontology - a more fine-grained structure: based ontology, simpletime, common concepts	- public version of CYC upperlevel ontology (HPKB-UPPERLEVEL)	No standard upper-level ontology	- Base Ontology	- no standard upperlevel ontology	- IEEE SUO (upper-level ontology integration)

4. Challenges

Since the vision of the Semantic Web was laid out in Berners-Lee's 2001, a wide coverage of good quality Semantic Web has not yet appeared (McCool, 2005). The number of Web pages written in semantic markup languages is very small (Lee and Goodwin, 2005). McCool traces the root of Semantic Web challenges to the technique of knowledge representation. According to him, knowledge representation (e.g., ontology) uses Codd's mathematical theory to translate information, that humans represent with natural language, into sets of tables that use well-defined schema to define what can be entered in the rows and columns. It is a technique similar to database, but with a large number of columns and a relatively sparse set of non-empty cells. Such a complex format requires enormous cost in creation and maintenance, which makes it difficult for the Semantic Web to achieve widespread public adoption.

McCool (2006) suggested a new approach for this limitation. He said that Berners-Lee developed the Web

by taking the salient ideas of hypertext and SGML syntax and removing complexities such as backward hyperlinks, which has made authoring, sharing, and copying simple enough for people to adopt quickly. Similarly, the Semantic Web formats must be simplified in order to produce user communities. McCool claimed that instead of a Semantic Web containing classes, relations, and triples, parameters should be added to existing markup tags to generate a named-entity Web (NEW). A radical simplification would be the solution to the barriers of the Semantic Web such as limited participation. NEW would make use of existing Web technologies and provide direct benefits at a far lower participation cost.

McCool's lightweight approach to annotating existing Web data (i.e., adding some extra tags to existing Web content) might work for a small part of the Web, but would not make the original Semantic Web vision a reality. Hepp (2006) thought building the Semantic Web by means of meddling with existing Web data a flawed idea because it is based on several myths about the Web.

- First, the common assumption that everything is on the Web and one just needs to find the means to locate them is not true.
- Second, the business Web is not static and constant updates would fail any data-centric annotation.
- To further complicate, the symmetry and strategic aspects of revealing information in the business world for example, disclose information only to seriously interested parties, runs counter to the Semantic Web notion that requires data to be persistently published for an unknown audience.

Instead, he proposed a different approach and suggested that entities are more willing to expose functionality than data in business settings and urged that more research attention be paid to developing Semantic Web services (i.e., annotating computational functionality) than to annotating Web content data. He advocated a substantial shift from the data-centric approach of annotating information on Web pages to annotating exposed functionality in Semantic Web services technologies.

But more important than proposed solutions, are the inquiries focusing on the root of Semantic Web's challenges. In an attempt to tackle the uncontrollable nature of data on the Web, the Semantic Web presents a unique challenge to current knowledge and information representation techniques. Edgar Codd's seminal contributions to the theory of relational databases led to the success of modern database technology, but it is no easy task to turn information represented through natural human language into machine interpretable data. The key to the success of the Semantic Web, according to McCool (2005), lies in finding this generation's Edgar Codd to solve the representation problem. Representations to be developed under a new theoretical framework must be easy to translate to and from natural language to make semantic representation of human knowledge more a reality than a theory.

5. Conclusion

In spite of all the challenges, the dream of the Semantic Web is not only about building a Web of actionable information derived from data through a semantic theory, but also about contributing to a new Web science, which is defined as a science that seeks to develop, deploy, and understand distributed information systems, processible by both humans and computers and operating on a global scale. With technical innovations like RDF, which identifies and exchange data, and OWL, which expresses how data sources connect together, the Semantic Web will enable better data integration by allowing everyone who puts individual items of data on the Web to link them with other pieces of data using standard formats. However as stated by Berners-Lee (2007), the future of the Web lies largely in its ability to manage, integrate and analyze data, i.e., individual information elements within documents.

References

1. **Berners-lee, T., Hendler, J. And Lassila, O**. The Semantic Web. Scientific American. 2001, 284 (5), pp. 34-43.

2. **Berners-leE, T.** Testimony before the United States House of Representatives, Committee on Energy and Commerce, Subcommittee on Telecommunications and the Internet. Available on: http://dig.csail.mit.edu/2007/03/01-ushousefuture-of-the-web.html.

3. **Corcho, O. And Gomez-perez, A**. A Road Map on Ontology Specification languages. In Workshop on Applications of Ontologies and Problem solving methods, 14th European Conference on Artificial Intelligence (ECAI'00), Berlin: Germany, August 20-25, 2000.

4. **Daconta, M. C., Obrst, L. J. And Smith, K. T.** The Semantic Web: A Guide to the Future of XML, Web services and Knowledge Management. Indianapolis: Available on: http://0-library.books24x7.com.library.simmons.edu/toc.asp?bookid=6073.

5. **Dhiman, A.K**. Basics of Information Technology for Librarians and Information Scientists. 2 Vols. New Delhi.: Ess Ess Publications, 2003.

6. **Dhiman, A.K.** Library 2.0: Emerging Trend in Library Services. In Developing Library and Information Resources and Services in the Internet Era,edited by Jagtar Singh. New Delhi: Indian Library Association, 2007, pp. 274-85.

7. **Dhiman, A.K.** Digital Libraries: Futuristic Interpretations and Some Issues. In Knowledge for All:

Role of Library & Information Centres, Mumbai: Sita Publications, 2008., pp. 223-32.

8. **Dhiman, A.k. And Sharma, H.** Access of Scholarly Publications through Institutional Repositories. Pakistan Journal of Library Science. 2008, 9, pp. 97-111.

9. **Hjørland, B..** Semantics and Knowledge Organization. Annual Review of Information Science and Technology, 2007. 41, pp. 367-405.

10. **Hepp, M..** Semantic Web and Semantic Web services: Father and Son or Indivisible Twins. IEEE Internet Computing, 2006, 10 (2), pp. 85-88.

11. **Jacob, E. K.** Ontologies and the Semantic Web. Bulletin of the American Society for Information Science and Technology, 2003, 29 (4). .

12. **Jones, D., Bench-capon, T. and Visser, P.** Methodology for Ontology Development..NY: Chapman-Hall, 1998. pp, 20-35.

13. **Khoo, C. S. G. and Aa, J. C.** . Semantic Relations in Information Science. Annual Review of Information Science and Technology. 2006, 40: 157-229.

14. **Mcool, R.** Rethinking the Semantic Web: Part 1. IEEE Internet Computing, 2005, 9 (6), pp. 86- 88

15. **Mccool, R.** 2006. Rethinking the Semantic Web: Part 2. IEEE Internet Computing, 2006, 10 (1), pp. 93-95.

16. **Wang, Z. H., Chaudhary, A. S. and Khoo, C.**. Potential and Prospects of Taxonomies for Content Organization. Knowledge Organization. 2006, 33 (3), pp. 160-169.

17. **Welty, C. and Guarino, N.** Supporting Ontological Analysis of Taxonomic Relationships. Data & Knowledge Engineering. 2001, 39 (1), pp. 51-74.

About Author

Dr. Anil Kumar Dhiman, Information Scientist, Gurukul Kangri University, Hardwar.
E-mail: akvishvakarma@yahoo.com