ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 2, February 2013

Personalized Ontology Construction: A Review Study

G.S.Deokate[1], Prof. Mrs. V.M.Deshmukh[2]

Final Year Master of Engineering, Department of CSE, PRMIT & R, Bandera (M.S), India [1]

Associate Professor, Department of CSE, PRMIT & R, Badnera (M.S), India [2]

ABSTRACT: **Personalized ontologies are a conceptualization model that formally describes and specifies user background knowledge. We know that, web users might have different expectations for the same search query. For example, for the topic "New York", business travellers may have demand for different information from leisure travellers. Same user may have different expectation for the same search query if applied in the different conditions. A user may become a business traveller when planning for a business trip, or a leisure traveller when planning for a family holiday. From this observation an assumption is formed that web users have a personal concepts model for their information needs, a user's concept model may change according to different information needs. In this paper we study construction of personalized ontology from world knowledge base.**

Keywords: World knowledge base, Ontology, Web personalization, Semantic Relations

I. INTRODUCTION

Ontologies are means of knowledge sharing and reuse. They are semantic containers. The term 'Ontology' has various definitions in various texts, domains and applications. In Philosophy and Linguistics ontology is 'The study of existence', 'a theory of what there is in the world', or 'taxonomy of the world Concepts '.In this paper we study personalized ontology. The subject of user interest is extracted from world knowledge base via user interaction. The extracted subjects consist of positive subject and negative subjects. The remaining candidates which are not fed back as either positive or negative from the user are neutral subjects. Ontology is then constructed for the given topic using these users feed back subjects.

II. RELATED WORK

Many existing knowledge bases are used by many models to learn ontologies. Gauch et al. [1] and Sieg et al. [2] learned personalized ontologies from the Open Directory
Project to specify user's preferences and interests in web search. Open Directory Project (ODP) is a human-edited Index of Web sites, also known as DMOZ, an acronym for "Directory Mozilla." ODP is hosted and administered by Netscape Communications and Weblogs, Inc. and is associated with the Mozilla browser. According to the official Web site, ODP hosts the largest and most comprehensive Web site directory in the world. The purpose of the ODP is to list and categorize sites, not to rank or promote them. Although ODP offers a search feature, it differs from conventional search engines that build massive lists of Web sites by automatically searching pages. Each entry is carefully reviewed before inclusion in the directory. The user profile is created automatically and implicitly while the users browse. The user profile is essentially a reference ontology in which each concept has a weight indicating the user's perceived interest in that concept. Profiles are generated by analysing the surfing behaviour of the user, specifically the content, length, and time spent on each Web page they visit. The Web pages the user visits are automatically classified into the concepts contained in the reference ontology and the results of the classification are accumulated. This causes the concepts in the reference ontology to receive weights based on the amount of related information the user has browsed. No user feedback is necessary. Since user profile is essentially a weighted ontology, first goal was to locate or create reference ontology on which to base user profile. Rather than create own ontology, a time consuming process, they chose ontology on already existing subject hierarchies. Online portals such as Yahoo.com and About.com, provide manually-created online subject hierarchies and a set of Web pages

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 2, February 2013

manually associated with each subject designed to organize Web content for easy browsing by end-users. One of the advantages of this approach is that system can work a reference ontology created from any subject hierarchy that has associated textual information. They have based reference ontology on subject hierarchies and associated Web pages from Yahoo, Magellan, Lycos, and the Open Directory Project.

King et al. [3] developed *IntelliOnto* based on the basis of the Dewey decimal classification. The Dewey Decimal Classification (DDC) system is a general knowledge organization system that is continuously revised to keep pace with knowledge. The DDC is used around the world in 138 countries; over sixty of these countries also use Dewey to organize their national bibliographies. Over the lifetime of the system, the DDC has been translated into more than thirty languages.

The system has value because of its well-defined categories, well-developed hierarchies, and rich network of relationships among topics, worldwide use, and language-independent representation of concepts. The existence of interoperable translations, mappings to other subject schemes, and the large amount of categorized content already associated with the system.

Wikipedia was used by Doweney et al. [4] to help understand underlying user interests in queries. They describe results from Web search log studies aimed at elucidating user behaviours associated with queries and destination URLs that appear with different frequencies. They note the diversity of information goals that searchers have and the differing ways that goals are specified. They examine rare and common information goals that are specified using rare or common queries. They identify several significant differences in user behaviour depending on the rarity of the query and the destination URL. They find that searchers are more likely to be successful when the frequencies of the query and destination URL are similar. They also establish that the behavioural differences observed for queries and goals of varying rarity persist even after accounting for potential confounding variables, including query length, search engine ranking, session duration, and task difficulty. Finally, using an information-theoretic measure of search difficulty, they show that the benefits obtained by search and navigation actions depend on the frequency of the information goal.

These works effectively discovered user background knowledge but their performance was limited by the quality of the global knowledge bases.

Li and Zhong [5] used pattern recognition and association rule mining techniques to discover knowledge from user local documents for ontology construction. Zhong [6] proposed a domain ontology learning approach that employed various data mining and natural language understanding techniques.

Navigli et al. [7] developed *OntoLearn* to discover semantic concepts and relations from the web documents. *OntoLearn* system is an infrastructure for automated ontology learning from domain text. It is the only system, as far as we know, that uses natural language processing and machine learning techniques. *OntoLearn* extracts terminology from a corpus of domain text, such as specialized Web sites and warehouses or documents exchanged among members of a virtual community.

Jiang and Tan [8] use content mining techniques to find semantic knowledge from domain-specific text documents for ontology learning.

Using a fuzzy domain ontology extraction algorithm, a mechanism was developed by Lau et al. [9] in 2009 to construct concept maps based on the posts on online discussion forums.

Doan at al. [10] proposed a model called GLUE and used machine learning technique to find similar concepts in different ontologies. For given two ontologies,

Each concept in ontology, GLUE finds the most similar concept in the other ontology. GLUE can work with all of them. Another key feature of GLUE is that it uses multiple learning strategies, each of which exploits well a different type of information either in the data instances or in the taxonomic structure of the ontologies.

III. WORLD KNOWLEDGE REPRESENTATION

World knowledge is most important for gathering information. World knowledge is commonsense knowledge possessed by people and acquired by through the experience and education. In this proposed model, user background knowledge is extracted from a world knowledge base encoded from the Library of Congress Subject Heading (LCSH).The Library of congress subject Heading (LCSH) is ideal for world knowledge base construction. The LCSH system is a thesaurus developed for organizing and retrieving information from a large volume of library collections. For over a hundred years the knowledge constrained in the LCSH has undergone continuous revising and enriching. The LCSH system represents the natural growth and distribution of human intellectual work and covers comprehensive and exhaustive topics of world knowledge. In addition, the LCSH system is the most comprehensive non-specialized controlled vocabulary in English. In many respects, the system has become a standard for subject cataloging and indexing, and is used not as major subject access tool in

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 2, February 2013

library catalogs but also means for enhancing subjects access to knowledge management system. [11]
The LCSH system is better than other world knowledge taxonomies used. Table1. shows a comparison of the LCSH with Library of Congress Classification (LCC) used by Frank and Paynter [12], the Dewey Decimal Classification (DDC) used by Wang and Lee [13], and the reference categorization (RC) developed by Gauch et al.[1] using online categorizations. As shown in table1 LCSH has more topics, has more specific structure and more semantic relations.

	LCSH	LCC	DDC	RC
# of Topics	394,070	4,214	18,462	100,000
Structure	Directed Acyclic Graph	Tree	Tree	Directed Acyclic Graph
Depth	37	7	23	10
Semantic Relations	Broader, Used-for, Related-to	Super- and Sub-class	Super- and Sub-class	Super- and Sub-class

Table1. – Comparison of World Taxonomies [11]

The structure of world knowledge base used is encoded from the LCSH references. The LCSH system contains three types of references: Broader terms (BT), Used-for (UF) and Related term (RT). The subjects in the knowledge base are linked to each other by semantic relation. Fig.1 [11] shows word knowledge base dealing with topic "Economic Espionage".

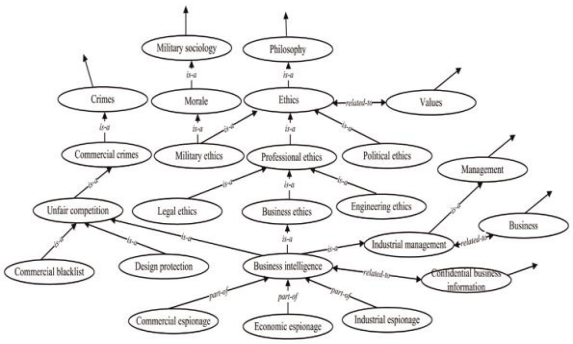

Fig.1. Part of World Knowledge Base [11]

IV. ONTOLOGY CONSTRUCTION

The Library of Congress Subject Headings is stored in MARC 21 records for use in computational systems. MARC stands for Machine-Readable Cataloging, Which are standard formats for the representation and communication of bibliographic and related information in machine-readable form. The subjects of user interest are extracted from the *WKB* via user interaction. A tool called ontology Learning Environment (OLE) Fig.2 [11] is developed to assists users with such interaction. Related to the topic, the interesting subjects consist of two sets; positive subjects and negative subjects.

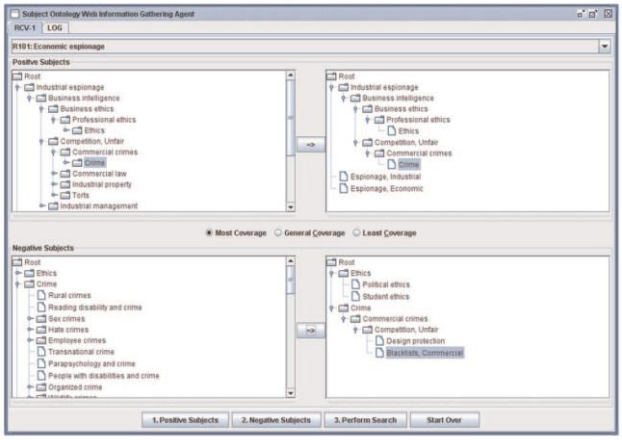

Fig.2 Ontology Learning Environment [11]

The positive subject is the concepts relevant to the information need, and negative subjects are the concepts resolving paradoxical or ambiguous interpretation of the information need. Thus for a given topic, the OLE provides users with a set of candidates to identity positive and negative subjects. These candidate subjects are extracted from the *WKB*. [11]

V. CONCLUSION

The constructed ontology is personalized because the user selects positive and negative subjects for personal preference and interest. If user search for topic "New York " and plans for a business trip, the user could have different subjects selected and a different ontology constructed when comparing with which select and constructed by a leisure user planning for a holiday. User's requirements are different and thus concept model

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 2, February 2013

changes every time according to the user requirement needs.

REFERENCES

[1] S. Gauch, J. Chaffee, and A. Pretschner, "Ontology-Based Personalized Search and Browsing" Web Intelligence and Agent System,Vol.1 no.3/4,pp.219-234,2003

[2] A. Sieg, B.Mobasher, and R.Burke," Web Search Personalization with Ontological User Profiles", Proceedings of 16th ACM Conference on Information and Knowledge Management,pp.525-534,2007

[3] J.D.King,Y.Li,X.Tao,and R.Nayak,"Mining World Knowledge for Analysis of Search Engine Contents, "Web Intelligence and Agent System",vol.5,no.3,pp.233-253,2007

[4] D.Doweney,S.Dumais, D.Liebling,and E.Horvitz, "Understanding the relationship between Searches, Queries and Information Goals." Proceedings of 17th ACM Conference on Information and Knowledge Management ,pp.449-458, 2008

[5] Y.Li and N.Zhong, "Mining Ontology for Automatically Acquiring Web Uer Information Needs," IEEE Transaction on Knowledge and Data Engineering,Vol.18,no.4,pp.554-568,2006

[6] N. Zhong , "Representation and Construction of Ontologies for web Intelligence", International Foundation of Computer Science,Vol.13,no.4,pp.555-570,2002

[7] R. Navigli, P.Velardi,and A. Gangemi, "Ontology Learning and its Application to Automated Terminology Translation", IEEE Transaction on Intelligent System,vol.18,no.1,pp.22-31,2003

[8] X.Jiang and A.H. Tan, " Mining Ontological Knowledge from Domain –Specific Text Documents", Proceedings of Fifth International Conference on Data Mining ,pp.665-668,2005

[9] R.Y.K. Lau, D.Song,Y.Li,C.H. Cheung and J.X. Hao, " Towards a Fuzzy Domain Ontology Extraction Method for Adaptive e-Learning", IEEE Transaction on Knowledge and Data Engineering,Vol.21,no.6, pp.800-813, 2009

[10] A.Doan, J.Madhavan,P.Domingos,and A.Halevy, " Learning to Map Between Ontologies on the Semantic Web", Proceedings of 11th International Conference on World Wide Web, pp.662-673,2002

[11] Xiaohui Tao,Yuefeng Li,and Ning Zhong, " A Personalized Ontology Model for web Information Gathering", IEEE Transaction on Knowledge and Data Engineering ,Vol.23,no.4,pp.496-509,2011

[12] E. Frank and G.W. Paynter, " Predicting Library of Congress Classification from Library of Congress Subject Heading", Information Science and Technology,vol.55,no.3 pp.214-227,2004

[13] J. Wang and M.C. Lee, " Reconstructing DDC for Interactive Classification", Proceedings of .16th ACM Conference on Information and Data Management ,pp.137-146,2007

Bibolography

G.S.Deokate is currently pursuing his Master of Engineering from Prof. Ram Meghe Institute of Technology, Badnera (India) from Sant Gadge Baba Amaravati University, Amaravati. His subjects of interests are data mining, Intelligent System, and Knowledge Engineering.

Prof. Mrs. V.M.Deshmukh is currently pursuing her PhD from Amaravati University. She is working as Associate Professor in Department of Computer Science & Engineering, at Prof. Ram Meghe Institute of Technology and Research, Badnera (India).