Available online at www.internationaleJournals.com

International eJournals

International ejournal of Mathematics and Engineering 198 (2012) 1889 - 1895

ISSN 0976 – 1411

INTERNATIONAL eJOURNAL OF MATHEMATICS AND ENGINEERING

www.InternationaleJournals.com

A Multi Dimensional Ontology Model for Personalized Web Information Gathering

A.Koteswara Rao [#1,] D.Madhavi[#2]

[#]Computer Science and Engineering, JNTUK
Gokul Institute of Technology and Sciences, Piridi, Andhra Pradesh, India.
[1]koti093@gmail.com
[2]DMadhavi3@gmail.com

Abstract— **As a model for knowledge description and formalization, ontologies are widely used to represent user profiles in personalized web information gathering. However, when representing user profiles, many models have utilized only knowledge from either a global knowledge base or user local inforamation. In this paper, a personalized ontology model is proposed for knowledge representation and reasoning over user profiles. This model learns ontological user profiles from both a world knowledge base and user local instance repositories. The ontology model is evaluated by comparing it against benchmark models in web information gathering. The results show that this ontology model is successful.**

Keywords— **Multidimensional, Ontology, personalization, world knowledge, local instance repository, user profiles, web
Information gathering.**

I. Introduction

The amount of web-based information available has increased dramatically. How to gather useful information from the web has become a challenging issue for users. Current web information gathering systems attempt to satisfy user requirements by capturing their information needs. For this purpose, user profiles are created for user background knowledge description .User profiles represent the concept models possessed by users when gathering web information. A concept model is implicitly possessed by users and is generated from their background knowledge. While this concept model cannot be proven in laboratories, many web ontologists have observed it in user behavior. When users read through a document, they can easily determine whether or not it is of their interest or relevance to them, a judgment that arises from their implicit concept models. If a user’s concept model can be simulated, then a superior representation of user profiles can be built. To simulate user concept models, ontologies—a knowledge description and formalization model—are utilized in personalized web information gathering. Such ontologies are called ontological user profiles or personalized ontologies. To represent user profiles, many researchers have attempted to discover user background knowledge through global or local analysis. Global analysis uses existing global knowledge bases for user background knowledge representation. Commonly used knowledge bases include generic ontologies (e.g., WordNet), thesauruses (e.g., digital libraries), and online knowledge bases (e.g., online categorizations and Wikipedia). The global analysis techniques produce effective Performance for user background knowledge extraction. However, global analysis is limited by the quality of the used knowledge base. For example, WorldNet was reported as helpful in capturing user interest in some areas but useless for others. Local analysis investigates user local information or observes user behavior in user profiles. For example, Li and Zhong discovered taxonomical patterns from the users' local text documents to learn ontologies for user profiles. Some groups learned personalized ontologies adaptively from user's browsing history. Alternatively, Sekine and Suzuki analyzed query logs to discover user background knowledge. In some works, such as, users were provided with a set of documents and asked for relevance feedback. User background knowledge was then discovered from this feedback for user profiles. However, because local analysis techniques rely on data mining or classification techniques for knowledge discovery, occasionally the discovered results contain noisy and uncertain information. As a result, local analysis suffers from ineffectiveness at capturing formal user knowledge. From this, we can hypothesize that user background Knowledge can be better discovered and represented if we can integrate global and local analysis within a hybrid model. The knowledge formalized in a global knowledge base will constrain the background knowledge discovery from the user local information. Such a personalized ontology model should produce a superior representation of user profiles for web information gathering. In this paper, an ontology model to evaluate this hypothesis is proposed. This model simulates users' concept models by using personalized ontologies and attempts to improve web information gathering performance by using ontological user profiles. The world knowledge and a user's local instance repository (LIR) are used in the proposed model. World knowledge is commonsense knowledge acquired by people from experience and education an LIR is a user's personal collection of information items. From a world knowledge base, we construct personalized ontologies by adopting user feedback on interesting knowledge. A multidimensional ontology mining method, Specificity and Exhaustivity, is also introduced in the proposed model for analyzing concepts specified in ontologies. The users' LIRs are then used to discover background knowledge and to populate the personalized ontologies. The proposed ontology model is evaluated by comparison against some benchmark models through experiments using a large standard data set. The evaluation results show that the proposed ontology model is successful.

II. RELATED WORK

1. Ontology Mining

Ontology mining discovers interesting and on-topic knowledge from the concepts, semantic relations, and instances in ontology. Ontology mining method is introduced: Specificity and Exhaustively. Specificity (denoted spe) describes a subject's focus on a given topic. Exhaustively (denoted exh) restricts a subject's semantic space dealing with the topic. This method aims to investigate the subjects and the strength of their associations in ontology. In User Local Instance Repository, User background knowledge can be discovered from user local information collections, such as a user's stored documents, browsed web pages, and composed/received emails.

2 .Ontology Learning

Global knowledge bases were used by many existing Models to learn ontologies for web information gathering. For example, learned personalized ontologies from the Open Directory Project to Specify users'

preferences and interests in web search. On the basis of the Dewey decimal classification, King ET al.Developed IntelliOnto to improve performance in
Distributed web information retrieval. Wikipedia was used By Downey et al. help understand underlying user Interests in queries. These works effectively discovered user Background knowledge; however, their performance was limited by the quality of the global knowledge bases. Aiming at learning personalized ontologies, many works Mined user background knowledge from user local information. Used pattern recognition and Association rule mining techniques to discover knowledge from user local documents for ontology construction. Tran Et al. translated keyword queries to Description Logics' Conjunctive queries and used ontologies to represent user Background knowledge. Proposed a domain Ontology learning approach that employed various data mining and natural-language understanding techniques. Developed onto Learn to discover semantic Concepts and relations from web documents. Web content Mining techniques to Discover semantic knowledge from domain-specific text Documents for ontology learning. Finally, Shehata et al. captured user information needs at the sentence level rather than the document level, and represented user profiles by the Conceptual Ontological Graph. The use of data mining techniques in these models leads to more user background knowledge being discovered. However, the knowledge discovered in these works contained noise and uncertainties. Additionally, ontologies were used in many works to improve the performance of knowledge discovery. Using a fuzzy domain ontology extraction algorithm, a mechanism was developed by Lau et al. [19] in 2009 to construct concept maps based on the posts on online discussion forums. Quest and Ali [31] used ontologies to help data mining in biological databases. Jin et al. [17] integrated data mining and information retrieval techniques to further enhance knowledge discovery. Doan et al. [8] proposed a model Called GLUE and used machine learning techniques to find Similar concepts in different ontologies. Dou et al. [9]
Proposed a framework for learning domain ontologies using
Pattern decomposition, clustering/classification, and association Rules mining techniques. These works attempted to explore a route to model world knowledge more efficiently.

3. User Profiles

User profiles were used in web information gathering to
Interpret the semantic meanings of queries and capture user
Information needs User profiles. Interesting topics of a user's information need. They also categorized user profiles into two diagrams: the data diagram user profiles acquired by analyzing a database or a set of transactions; the information diagram user profiles acquired by using manual techniques, such as questionnaires and interviews automatic techniques, such as information retrieval and machine learning. Van der Sluijs and Huben proposed a method called the Generic User Model Component to improve the quality and utilization of user modeling.

Wikipedia was also used to help discover user interests. In order to acquire a user profile, Teevan et al. collection of user desktop text documents and emails, and cached web pages to explore user interests. Makris et al. [24] acquired user profiles by a ranked local set of categories, and then utilized web pages to personalize search results for a user. These works attempted to acquire user profiles in order to discover user background knowledge. The users read each document and gave a positive or negative judgment to the document against a given topic. Because, only users perfectly know their interests and preferences, these training documents accurately reflect user background knowledge. Semi-interviewing user profiles are acquired by semi automated techniques with limited user involvement. These techniques usually provide users with a list of categories and ask users for interesting nonintersecting categories No interviewing techniques do not involve users at all, but ascertain user interests instead. They acquire user profiles by observing user activity and behavior and discovering user background knowledge.

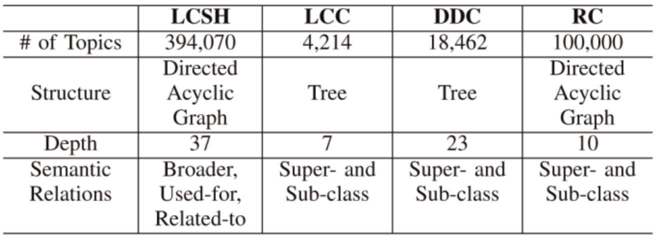
III. ONTOLOGY CONSTRUCTION

From observations in daily life, we found that web Users might have different expectations for the same search query. For example, for the topic “Den Mark,” business travelers may demand different information from leisure travelers. Sometimes even the same user may have different expectations for the same search query if applied in a different situation. A user may become a business traveler when planning for a business trip, or a leisure traveler when planning for a family holiday. Based on this observation, an assumption is formed that web users have a personal concept model for their information needs. A user’s concept model may change according to different information needs. In this section, a model constructing personalized ontologies for web users’ concept models is introduced.

1. World Wide Web Knowledge Representations

World knowledge is necessary for lexical and referential disambiguation, including establishing co reference relations and resolving ellipsis as well as for establishing and maintaining connectivity of the discourse and adherence of the text to the text producer’s goal and plans.” In this proposed model, user background knowledge is extracted from a world knowledge base encoded from the Library of Congress Subject Headings (LCSH). We first need to construct the world knowledge base. The world knowledge base must cover an exhaustive range of topics, since users may come from different backgrounds. For this reason, the LCSH system is an ideal world knowledge base. The LCSH was developed for organizing and retrieving information from a large volume of library collections.

Table I



	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

Comparison of different World Taxonomies

IV. MULTIDIMENSIONAL ONTOLOGY MINING

Ontology mining discovers interesting and on-topic knowledge from the concepts, semantic relations, and instances in an ontology. In this section, a 2D ontology mining method is introduced: Specificity and Exhaustivity. Specificity (denoted spe) describes a subject’s focus on a given topic. Exhaustivity (denoted exh) restricts a subject’s semantic space dealing with the topic. This method aims to investigate the subjects and the strength of their associations in an ontology. We argue that a subject's specificity has two focuses: 1) on the referring-to concepts (called semantic specificity), and 2) on the given topic (called topic specificity). These need to be addressed separately.

1. Semantic Specificity

The semantic specificity is investigated based on the structure of OðT Þ inherited from the world knowledge base. The strength of such a focus is influenced by the subject's locality in the taxonomic structure taxS of OðT Þ (this is also argued by [42]). As stated in Definition 4, the taxS of OðT Þ is a graph linked by semantic relations. The subjects located at upper bound levels toward the root are more abstract than those at lower bound levels toward the "leaves." The upper bound level subjects have more descendants, and thus refer to more concepts, compared with the lower bound level subjects. Thus, in terms of a concept being referred to by both an upper bound and lower bound subjects, the lower bound subject has a stronger focus because it has fewer concepts in its space. Hence, the semantic specificity of a lower bound subject is greater than that of an upper bound subject.

V.SPECIFICITY

1. Local Instance Repository

User background knowledge can be discovered from user local information collections, such as a user's stored documents, browsed web pages. The ontology OðT Þ constructed in Section 3 has only subject labels and semantic relations specified. In this section, we populate the ontology with the instances generated from user local information collections. We call such a collection the user's local instance repository (LIR).

However, many documents do not have such direct, clear references. For such documents in LIRs, data mining techniques, clustering, and classification in particular, can help to establish the reference, as in the work conducted by [20], [49]. The clustering techniques group the documents into unsupervised (non predefined) clusters based on the document features. These features, usually represented by terms, can be extracted from the clusters. They represent the user background knowledge discovered from the user LIR.

VI. WEB INFORMATION GATHERING SYSTEM

The information gathering system, IGS, was designed for common use by all experimental models This model was the implementation of the proposed ontology model. The input to this model was a topic and the output was a user profile consisting of positive documents (Dþ) and negative documents (D_). Each document d was associated with a supportðdÞ value indicating its support level to the topic. were also extracted and encoded as the semantic relations of is-a, art-of and related-to in the WKB,respectively . The authors played the user role to select positive and negative Subjects for ontology construction, following the descriptions and narratives associated with the topics. On average, each personalized ontology contained about 16 positive and 23 negative subjects. For each topic T , the ontology mining method was performed on the constructed OðT Þ and the user LIR to discover interesting concepts The user LIRs were collected through searching the subject catalog of the QUT library by using the given topics. The catalog was distributed by QUT library as a 138 MB text file containing information for 448,590 items. The information was pre-processed by removing the stop words, and stemming and grouping the terms. Librarians and authors have assigned title, table of content, summary, and a list of subjects to each information item in the catalog. These were used to represent the instances in LIRs. The semantic relations of is-a and part-of were also analyzed in the ontology mining phase for interesting knowledge discovery.

VII .CONCLUSIONS AND FUTUREWORK

The model constructs user personalized ontologies by extracting world knowledge and discovering user background knowledge from user local instance repositories. An ontology model is proposed for representing user background knowledge for personalized web information gathering. In evaluation, the standard topics and a large test bed were used for experiments. The model was compared against benchmark models by applying it to a common system for information gathering. The experiment results demonstrate that our proposed model is promising. A sensitivity analysis was also conducted for the ontology model.

Here In this observation, we found that the combination of global and local knowledge works better than using any one of them. In addition, the ontology model using knowledge with both is-a and part-of semantic relations works better than using only one of them. When using only global knowledge, these two kinds of relations have the same contributions to the performance of the ontology model. While using both global and local knowledge, the knowledge with part-of relations is more important than that with is-a. The proposed ontology model in this paper provides a solution to emphasizing global and local knowledge in a single computational model. The findings in this paper can be applied to the design of web information gathering systems. The model also has extensive contributions to the fields of Information Retrieval, web Intelligence, Recommendation Systems, and Information Systems. In our future work, we will investigate the methods that generate user local instance repositories to match the representation of a global knowledge base. The present work assumes that all user local instance repositories have content-based descriptors referring to the subjects, however large volume of documents existing on the web may not have such content-based descriptors. For this problem, strategies like ontology mapping and text classification/clustering were suggested. These strategies will be investigated in future work to solve this problem. The investigation will extend the applicability of the ontology model.

VIII. REFERENCES

[1].J. Han and K.C.-C. Chang, “Data Mining for Web Intelligence, “Computer, vol. 35, no. 11, pp. 64-70, Nov. 2002

[2].Data Communications and Networking, by Behrouz A Forouzan.

[3] L.M. Chan, Library of Congress Subject Headings: Principle and Application. Libraries Unlimited, 2005.

[4] P.A. Chirita, C.S. Firan, and W. Nejdl, “Personalized query Expansion for the Web,” Proc. ACM SIGIR (’07), pp. 7-14, 2007.

[5] R.M. Colomb, Information Spaces: The Architecture of Cyberspace. Springer, 2002.

[6] A. Doan, J. Madhavan, P. Domingos, and A. Halevy, “Learning to Map between Ontologies on the Semantic Web,” Proc. 11th Int’l Conf. World Wide Web (WWW ’02), pp. 662-673, 2002.

[7] D. Dou, G. Frishkoff, J. Rong, R. Frank, A. Malony, and D. Tucker, "Development of Neuroelectromagnetic Ontologies(NEMO): A Framework for Mining Brainwave Ontologies," Proc. ACM SIGKDD ('07), pp. 270-279,2007.
[8] D. Downey, S. Dumais, D. Liebling, and E. Horvitz, "Understanding the Relationship between Searchers' Queries and Information Goals," Proc. 17th ACM Conf. Information and Knowledge Management (CIKM '08), pp. 449-458, 2008.
[9] E. Frank and G.W. Paynter, "Predicting Library of Congress Classifications from Library of Congress Subject Headings," J. Am.Soc. Information Science and Technology, vol. 55, no. 3, pp. 214-227,2004.
[10] S. Gauch, J. Chaffee, and A. Pretschner, "Ontology-Based Personalized Search and Browsing," Web Intelligence and Agent Systems, vol. 1, nos. 3/4, pp. 219-234, 2003.
[11] R. Gligorov, W. ten Kate, Z. Aleksovski, and F. van Harmelen, "Using Google Distance to Weight Approximate Ontology Matches," Proc. 16th Int'l Conf. World Wide Web (WWW '07), pp. 767-776, 2007.
[12] J. Han and K.C.-C. Chang, "Data Mining for Web Intelligence, "Computer, vol. 35, no. 11, pp. 64-70, Nov. 2002.
[13] B.J. Jansen, A. Spink, J. Bateman, and T. Saracevic, "Real Life Information Retrieval: A Study of User Queries on the Web," ACM SIGIR Forum, vol. 32, no. 1, pp. 5-17, 1998.
[14] X. Jiang and A.-H. Tan, "Mining Ontological Knowledge from Domain-Specific Text Documents," Proc. Fifth IEEE Int'l Conf. Data Mining (ICDM '05), pp. 665-668, 2005.
[15] W. Jin, R.K. Srihari, H.H. Ho, and X. Wu, "Improving Knowledge Discovery in Document Collections through Combining Text Retrieval and Link Analysis Techniques," Proc. Seventh IEEE Int'l Conf. Data Mining (ICDM '07), pp. 193-202, 2007.
[16] J.D. King, Y. Li, X. Tao, and R. Nayak, "Mining World Knowledge for Analysis of Search Engine Content," Web Intelligence and Agent Systems, vol. 5, no. 3, pp. 233-253, 2007.
[17] R.Y.K. Lau, D. Song, Y. Li, C.H. Cheung, and J.X. Hao, "Towards aFuzzy Domain Ontology Extraction Method for Adaptive e-Learning," IEEE Trans. Knowledge and Data Eng., vol. 21, no. 6,pp. 800-813, June 2009.
[18] K.S. Lee, W.B. Croft, and J. Allan, "A Cluster-Based Resampling Method for Pseudo-Relevance Feedback," Proc. ACM SIGIR '08, pp. 235-242, 2008.
[19] D.D. Lewis, Y. Yang, T.G. Rose, and F. Li, "RCV1: A New Benchmark Collection for Text Categorization Research," J. Machine Learning Research, vol. 5, pp. 361-397, 2004.
[20] Y. Li and N. Zhong, "Web Mining Model and Its Applications for Information Gathering," Knowledge-Based Systems, vol. 17, pp. 207-217, 2004.