ISSN : 2231-0711

Nandini P.Wasnik et al | International Journal of Computer Science Engineering and Technology(IJCSET) | April 2014 | Vol 4, Issue 4,103-107

Gathering Web Information for Personalized Ontology Model

Nandini P.Wasnik , Prof.Mahip M.Bartere

Dept. of Computer science and Engineering
G.H. Raisoni College of Engineering
Amravati, India

***Abstract*— In a personalized web information gathering, for the knowledge description ontology term is use. Mainly ontology used for acquires knowledge, share, reuse and increase relations description of knowledge. Paper shows different problems and searching techniques also related work shows working of different authors on ontology. Main work of ontology is to gather web information based on keywords that may be local repository or global repository. Initialization of information gathering is beginning according to user profile. Also section covers basic architecture of ontology which focuses on overall information gathering. Learning concept extract the information in structured format for unstructured input. Ontology as model for knowledge description and validation is used to represent user profile in personalized web gathering information. While presenting user profiles most of the models used a global knowledge bases or user local information for representing user profiles. Ontology is the model for knowledge description and validation, which are largely used to represent user profile in personalized web information gathering. When representing user profiles, most of the models have access only knowledge from either a global knowledge base or user local information. This paper include, a personalized ontology model is proposed for knowledge representation and reasoning over user profiles. It will contain user profiles from both world knowledge base and user local instance repository. The ontology model is evaluated against benchmark models in web information gathering. The concept models of the user profile represent by user when gathering web information. A concept model is possessed by users and is generated from there background knowledge. This concept model cannot be proven in laboratories; many web ontologists have observed it in a user behaviour the results show that this ontology model is successful.**

***Keywords*— Ontology, personalization, semantic relation, world knowledge, local instance repository, user profile,web information gathering.**

1. INTRODUCTION

Today is the internet world. The total of the available web-base information on the internet has increased dramatically. But collecting the information uesful from the internet has become the most challenging job today's scenario. People are interested in the relevant and interested information from the world web.The information gathering from web systems before this satisfy the user requirements by capturing their need of information. Because of this reason user profiles are developed for user background knowledge description. The user profiles represent the concepts models possessed by user while gathering the web information. A concept model is generated from user background knowledge and possessed by user implicitly. But it is observed by many ontologists that when user read a document they can easily determined whether or not it is of their interest or relevance to them .If the concept model of user can be simulated, and then representation is better of the user profile can be build. To Simulate use the concepts model,are utilized ontologists in personalized web information gathering which are called ontological user profiles or personalized ontologies [1],[2],[3].

In Global analysis, global knowledge bases are used for user background knowledge representation. Local user use user analysis information. Global analysis is limited by quality of knowledge base whereas local analysis is not sufficient for capturing user knowledge. If we integrate global and local analysis within a hybrid model the global knowledge will be constrain the background knowledge discovery form the user general information. Such an model of ontology will give the better representation of user profiles [4].For representing the user profiles, the knowledge must be gathered by user background by using global or local analysis. Global analysis uses worldwide knowledge base for background knowledge representation. The commonly used knowledge bases include generic ontologies e.g. Word net, Thesauruses, digital libraries. Compared with other benchmark models ontology model is successful. The research contributes to knowledge engineering, and has the potential for design to improve of personalized web gathering information systems.The involmeant are pure and growingly significant, considering the rapid explosion of web information and the growing accessibility of online documents.

Local analysis investigates user local information or observes user behavior in user profiles. For example, Li and Zhong [23] discovered taxonomical patterns from the users' local text documents to learn ontologies for user profiles. Some groups [12],learned personalized ontologies adaptively from browsing history of the user.. Alternatively, Sekine and Suzuki analyzed query logs to discover user background knowledge. In some works, such as [29], users were provided with a set of documents and asked for relevance feedback. Then the user background knowledge was discovered from this feedback for profile of the user. However, because of the techniques of local analysis rely on data mining or classification techniques for discovery of knowledge, occasionally the founded results contain noisy and information which is uncertain. 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 discovered in a global knowledge base will constrain the background knowledge discovery from local information of user. Such ontology model which is personalized should produce a superior representation of user profiles for web information collecting. In this paper, to evaluate an ontology model this hypothesis is proposed. The users concept models simulated model 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 ontology's by adopting user feedback on interesting knowledge. A multidimensional ontology mining method, Specificity and Exhaustively, is also introduced in the proposed model for analyzing concepts specified in ontology's. The users' LIRs are then used to discover background knowledge and to populate the personalized ontology's.

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 [12],or personalized ontologies [29].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 ontology's (e.g., WordNet [26]), 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, Word Net was reported as helpful in capturing user interest in some areas but useless for others.

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. Additionally, ontology's 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 [28] 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 toexplore a route to model world knowledge more efficiently.

The paper is organized as follows: Section 2 discusses the related work; in Section 3, we introduce how personalized ontologies are constructed for users. After that, Section 4 gives the architecture of the proposed model. Finally, Section 5 makes conclusions and addresses our future work.

2. LITERATURE REVIEW

2.1 Ontology Learning:

Ontologists are means of knowledge sharing and reuse. They are semantic containers. The term Ontology" has various definitions in many texts, field and applications. Many current knowledge bases are used by many models to learn ontologies.Gauch et al. [1] and Sieg et al. [5] learned personalized ontologies from the Open Directory Project to specify users" preferences and interests in web search. King developed *IntelliOnto* based on the basis of the Dewey decimal classification. Doweney et al. [7] used Wikipedia which helps in understanding user attention in queries. The user background knowledge is discovered but the performance is limited by quality of the global knowledge base. Much work has been done for discovering user background knowledge from user local information. Aiming at learning personalized ontologies, many works sourced user background knowledge from user local information. Li and Zhong [23] used pattern recognition and association rule mining techniques to discover knowledge from user local documents for ontology construction.Tran et al. [25] transformed keyword of queries to 'Description Logics' conjunctive queries and used ontologies to represent user background knowledge. Zhong[23] proposed a domain ontology learning approach that employed various data mining and natural-language understanding techniques. Additionally, ontologies were used in many works to improve the knowledge discovery performance.Using a fuzzy domain ontology removal algorithm, a mechanism was developed by Lau et al. [19] in 2009 to construct maps based on the concept the posts on online discussion forums.

2.2 User Profiles:

In the web information gathering, user profiles were used to know the semantic meanings of queries and capture user needs of information. User used user profile modelling and personalization, it is used to reflect the notice of user. Li and Zhong declared user profiles as the interesting topics of a user's information required. The profiles of user are divided into two diagrams: the data diagram and which are acquired by analysing a database or a set of transaction whereas the information diagram user profiles developed by using manually such as questionnaires and interviews or automatic techniques such as information retrieval and machine learning. Profiles of user are categorized into three groups: interviewing, semi-interviewing, and non-interviewing. [1], [3], [12], [13]. Interviewing user profiles are considered to be perfect user profiles. They are developed by using manual techniques, such as questionnaires, interviewing users, and classified training sets by analysed user. One usual example is the TREC

Filtering Track training sets, which were generated manually [4]. The users read each document and gave a positive or negative judgment to the document against a given topic. Semi-interviewing user profiles are acquired by semi automated techniques with limited user involvement. These techniques normally provide users with a list of categories and ask users for interesting or non interesting categories. One typical example is the web training set acquisition model introduced by Tao et al. [5], which extracts training sets from the web based on user fed back categories. Non interviewing techniques do not involve users at all, but ascertain user interests instead. User profiles is acquired by analysing user activity and behaviour and discovering user background knowledge [6]. The interviewing, semi-interviewing, and non interviewing user profiles can also be viewed a manual semiautomatic, and automatic profiles, respectively.

3. PERSONALIZED ONTOLOGY CONSTRUCTION

Personalized ontology's are a conceptualization model that formally describes and specifies user background knowledge. From observations in daily lives, we found that web users might have different expectations for the same search query. For example, for the topic "New York," business travellers may demand different information from vacation travellers.

Most of the times the same user even may have different expectations for the same search query if applied in a different situation. A user may become a business traveller when planning for a business trip, or a vacation traveller when planning for a family holiday. Observation based on this, an assumption is formed that web users have a personal concept model for their information needs. A concept model of user's may be change according to needs of different information. A model constructing personalized ontology's for web users' concept models is introduced.

3.1 World Knowledge Representation:

For the information gathering the world knowledge is very important. World knowledge is commonsense knowledge possessed by people and acquired by through the experience and education. 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. The LCSH system is a thesaurus developed for organizing and retrieving information from a large volume of library collections. LCSH has undergone continuous revising and enriching. The LCSH system is better than other world knowledge taxonomies used. Following table shows a comparison of the LCSH with Library of Congress Classification (LCC) used by Frank and Paynter [16], the Dewey Decimal Classification (DDC) used by Wang and Lee [17], and the reference categorization (RC) developed by Gauch et al. [1] using online categorizations that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

	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

Table – Comparison of World Taxonomies

They are encoded from the subject headings in the LCSH. Let SS be a set of subjects, an element s 2 SS is formalized as a 4-tuples s:=< hlabel; neighbor; ancestor; descendanti, >

- label is the heading of s in the thesaurus of LCSH;
- neighbour function return the subjects that have direct links to s in the world knowledge base;
- the subjects return by ancestor function that have a higher level of abstraction than s and link to s directly or indirectly in the world knowledge base;
- descendant is a function returning the subjects that are more specific than s and link to s directly or indirectly in the world knowledge base. The subjects in the world knowledge base are linked to each other by the semantic relations are is-a relation, part-of relation, and related-to relation.

3.2 Ontology Construction:

The subjects based on user interest are extracted from the WKB via user interaction. A tool known as Ontology Learning Environment (OLE) is developed to assist users with such interaction. Regarding to topic, the interesting subjects consist of two sets: positive subjects are the concepts relevant to the information need, and negative subjects are resolving the concept paradoxical or ambiguous interpretation of the need of information. Thus, for a given topic, the OLE gives users with a set of candidates to identify positive and negative subjects. These candidate subjects are retrieving from the WKB. User interested subjects are extracted from the *WKB* via user interaction. Ontology Learning Environment (OLE) tool is developed to assists users with such interaction. Related to the topic, the interesting subjects consist of two sets; positive subjects and negative subjects. The subjects which are relevant to the information need are positive subjects and the subjects who resolve ambiguous interpretation of information need are negative subjects The OLE provides with a set of candidates for user to identity positive and negative subjects. The constructed ontology is personalized because the user selects positive and negative subjects for personal preferences and interests. Thus, if a user searches "New York" and plans for a business trip, the different subjects have to be selected from user and a different ontology constructed, compared to those selected subject sand constructed by a leisure user planning for a holiday.

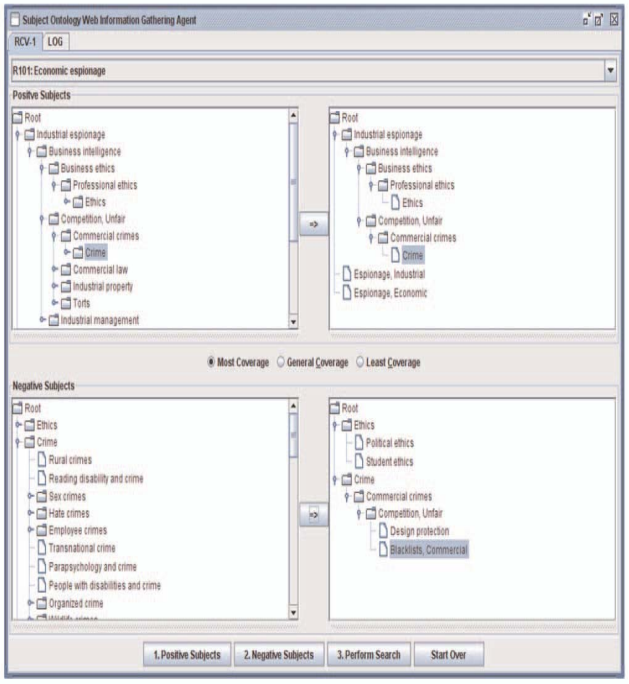

Fig. OLE (Ontology Learning Environment)

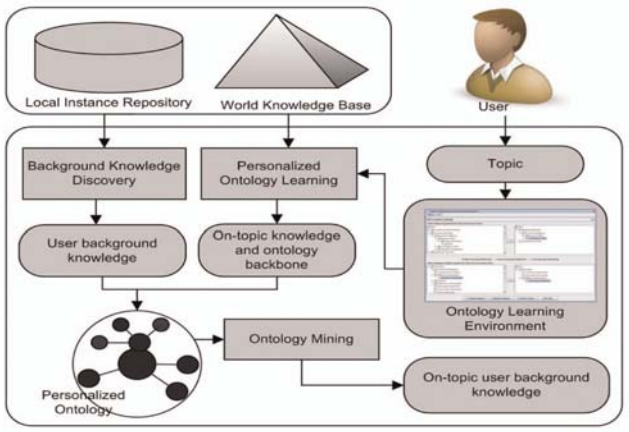

Fig. Architecture of Ontology model

4. ARCHITECTURE OF THE ONTOLOGY MODEL

The proposed ontology model aims to discover user background knowledge and learns personalized ontologies for representing user profiles. Following fig. illustrates the architecture of the ontology model. A personalized ontology model is constructed, according to the topic. The two resources of knowledge, the global world knowledge base and the user's local instance repository, are utilized by the ontology model. The world knowledge base provides the taxonomic structure for the personalized ontology. The background knowledge of user is discovered from the user local instance repository. Against the given topic, the specificity and exhaustively of subjects are investigated for user background knowledge discovery. Application of ontology to search information that info may be local or may be global. Mainly global repository searching finds overall information for keyword. That information may be finding directly using URL in web browser or entering keyword in search engine. Also for global repository user's feedback is consider to improvement in system performance. Local repository ontology concept is mainly for standalone window application where user searches only for local keywords which are saved on database of system. For detail flow, following diagram shows overall architecture [2]. Where,

LIR - Local Instance Repository.
WKB- World Knowledge Base
OLE- Ontology Learning Environment.

5. CONCLUSION AND FUTURE WORK

In this paper we covers overall concept of ontology model which is mainly used for web information gathering. An ontology model is proposed for representing user background knowledge for personalized web information gathering. The model constructs ontologies for user in personalized manner by extracting world knowledge from the LCSH system and discovering user background knowledge from user local instance repositories. The model ontology in this paper include 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 large contributions to the fields of retrieving information, Intelligence web, 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 guess that all user local instance repositories have content-based descriptors referring to the subjects however, documents having large volume existing on the web may not have such content-based descriptors. These strategies will be investigated in future work to solve this problem The investigation will extend the applicability of the ontology model to the majority of the existing web documents and increase the contribution and significance of the present work. For easily data gathering for single computational model we can use global and local repository at a time. At the same time our aim will be increase system performance.

REFERENCES

[1]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.

[2] Y. Li and N. Zhong, "Web Mining Model and ItsApplications for information Gathering" Knowledge-Based Systems, vol. 17, pp. 207-217, 2004.

[3] Y. Li and N. Zhong, "Mining Ontology for Automatically Acquiring Web User Information Needs," IEEE Trans. Knowledge and Data Eng., vol. 18, no. 4, pp.554-568, Apr. 2006.

[4] X. Tao, Y. Li, N. Zhong, and R. Nayak, "Ontology Mining for Personalized Web Information Gathering," Proc. IEEE/WIC/ACM Int"l Conf. Web Intelligence, pp. 351-358, 2007

[5]. X. Tao, Y. Li, N. Zhong, and R. Nayak, "Automatic Acquiring Training Sets for Web Information Gathering," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence, pp. 532-535, 2006.

[6]. J. Trajkova and S. Gauch, "Improving Ontology-Based User Profiles," Proc. Conf. Research 'Information Assistee par Ordinateur (RIAO '04), pp. 380-389, 2004.

[7] 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
[8] L.M. Chan, Library of Congress Subject Headings: Principle and Application, Libraries Unlimited,2005
[9]Miss. Deshmukh Rupali R., Prof. Keole R.R.,"Ontology Mining For Personalized Web Information Gathering,," International Journal of Engineering And Computer Science, Issn:2319-7242 Volume 1 Issue 3 Dec 2012.
[10] Qiuyu Zhang, Fengman Miao, Zhanting Yuan, QikunZhang, Zhi Fan, "Construction of A Dynamic Trust Ontology Model," 2008 International Conference On Computational Intelligence And Security.
[11] Xiaohui Tao, Yuefeng Li, And Ning Zhong, " A Personalized Ontology Model For Web Information Gathering," IEEE Transactions on Knowledge And Data Engineering, Vol. 23, No.4, April 2011
[12]S. Gauch, J. Chaffee, and A. Pretschner "Ontology-Based PersonalizedSearch and Browsing," Web Intelligence and Agent Systems, vol. 1, nos. 3/4, pp. 219-234, 2003.
[13] 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.
[14] J.Han and K.C.-C. Chang,"Data Mining for Web Intelligence," Computer, vol. 35, no. 11, pp. 64-70, Nov. 2002.
[15] 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.
[16] 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.
[17] D.D. Lewis, Y. Yang, T.G. Rose, and F. Li, "RCV1: A New Benchmark Collection for Text Categorization Research," Machine Learning Research, vol. 5, pp. 361-397, 2004.
[18] Y. Li and N. Zhong, "Web Mining Model and Its Applications for Information Gathering," Knowledge-Based Systems, vol. 17, pp. 207- 217, 2004.
[19] Y. Li and N. Zhong, "Mining Ontology for Automatically Acquiring Web User Information Needs," IEEE Trans. Knowledge and Data Eng., vol. 18, no. 4, pp. 554-568, Apr. 2006.
[20] C. Makris, Y. Panagis, E. Sakkopoulos, and A. Tsakalidis, "Category Ranking for Personalized Search," Data and Knowledge Eng., vol. 60, no. 1, pp. 109-125, 2007.
[21]S.E.Middleton, N.R. Shadbolt, and D.C. De Roure,"Ontological User Profiling in Recommender Systems," ACM Trans. Information Systems (TOIS), vol. 22, no. 1, pp. 54-88, 2004.
[22]G.A. Miller and F. Hristea, "WordNet Nouns: Classes and Instances," Computational Linguistics, vol. 32, no. 1, pp. 1-3, 2006.
[23]D.N.Milne, I.H. Witten, and D.M. Nichols, "A Knowledge-Based Search Engine Powered by Wikipedia," Proc. 16th ACM Conf.Information and KnowledgeManagement(CIKM '07), pp. 445-454, 2007.
[24] R. Navigli, P. Velardi, and A. Gangemi, "Ontology Learning and Its Application to Automated Terminology Translation," IEEE Intelligent Systems, vol. 18, no. 1, pp. 22-31, Jan./Feb. 2003.
[25]S. Nirenburg and V. Rasin Ontological Semantics.The MIT Press, 2004.
[26]A.-M. Popescu and O. Etzioni,"Extracting Product Features and Opinions from Reviews," Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing (HLT '05), pp. 339-346, 2005.
[27] D. Quest and H. Ali, "Ontology Specific Data Mining Based on Dynamic Grammars," Proc. IEEE Computational Systems Bioinformatics Conf. (CSB '04), pp. 495-496, 2004.
[28] S.E. Robertson and I. Soboroff, "The TREC 2002 Filtering Track Report," Proc. Text REtrieval Conf., 2002.
[29]S. Sekine and H. Suzuki, "Acquiring Ontological Knowledge from Query Logs," Proc., pp. 1223 1224, 2007.