An Ontology Model for Knowledge Representation over User Profiles

Shaik Haseena Sultana[1], Ch. N. Santhosh Kumar[2], V. Sitha Ramulu[3]
[1]M. Tech, Swarna Bharathi Institute of Science & Technology, Khammam, A.P., India
[2]Assoc. Professor, Dept. of CSE, Swarna Bharathi Institute of Science & Technology, Khammam, A.P., India
[3]Assoc. Professor, Dept. of IT, Swarna Bharathi Institute of Science & Technology, Khammam, A.P., India

ABSTRACT: *The amount of information on the world wide web increases rapidly. But gathering the required information from the web has become the most challenging job in today's scenario. People are only interested in the relevant information from the web. The web information gathering systems before this satisfy the user's requirements by capturing their information needs. For this reason user profiles are created for user background knowledge representation and description. The user profiles represent the concepts models possessed by user while gathering the useful web information. The concept of Ontologies is utilized in personalized web information gathering which are called ontological user profiles or personalized ontologies. In this paper, an ontology model is proposed for representing the user background knowledge for personalized web information gathering. Personalized ontology attempts to improve the web information gathering performance by using ontological user profiles. The model constructs user personalized ontologies by extracting world knowledge from the Library of Congress Subject Headings system and discovering user background knowledge from user local instance repositories.*

KEYWORDS: *Ontologies, Specificity, User profiles, WWW.*

I. INTRODUCTION

Over the last decade, we have witnessed an explosive growth in the information available on the World Wide Web (WWW). Gathering useful information from the WWW has become a challenging issue for users. The Web users expect more intelligent systems or agents to gather the useful information from the huge size of Web related data sources to meet their information needs. The user profiles are created for the user background knowledge description [1][2][3]. User profiles represent the concept models possessed by users when gathering the web information. A concept model is implicitly global analysis method, which is an effective method for gathering the global knowledge. Local analysis is used for analyzing the user behavior in the user profiles. In some works, users provided with a set of documents from that background knowledge can be discovered. The user background knowledge can be better discovered if we integrate both global and the local information. It can be better improved by using the ontological user profiles. A multidimensional ontology mining method, Specificity and exhaustivity, for analyzing the concept of specified machine-readable documents.

The goal of ontology learning is to semi automatically possessed by the users and is generated from their background knowledge. This knowledge is used to gather relevant information about a user's choices and preference. World knowledge is a common sense knowledge acquired by the people from experience and education[4]. For representing the user profiles, the user background knowledge must be gathered by using local or global analysis. Global analysis uses worldwide knowledge base for representing background knowledge. The commonly used knowledge bases include generic ontologies e.g. Thesauruses, digital libraries, word net. Compared with the other benchmark models, ontology model is successful.

II. RELATED WORK

A. Ontology Learning: Ontologies are the means of knowledge sharing and reuse. They are called semantic containers. The term "Ontology" has various definitions in various domains, texts and applications. Many existing knowledge bases are used by many models for learning ontologies. In [1] and [5], the authors learned personalized ontologies from the Open Directory Project to specify users‟ preferences and interests in web search. King developed "IntelliOnto" method based on the basis of the Dewey decimal classification. The Dewey Decimal Classification (DDC) system is a general knowledge organization system that is used continuously revised to keep pace with knowledge. The DDC is used around the world in 139 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 30 languages [6]. In [7], the authors used Wikipedia which helps in understanding user interests in queries. The above work discovered user background knowledge but the performance is limited by the quality of the global knowledge base. Much work has been done for discovering user background knowledge from the user local information. Pattern reorganization and association rule mining technique to discover the knowledge from user local information is used by [3].

A domain ontology learning approach was proposed in [3] that uses various data mining and natural language understanding techniques to discover knowledge from user local documents for ontology construction. Semantic relations and concepts are discovered in [9] for which he developed a system called Ontolearn. OntoLearn system is an infrastructure for automated ontology learning from the domain text. It is the only system, as far as we know, that uses the natural language processing and machine learning techniques. In [8], the authors use content mining techniques to find semantic knowledge from domain-specific text documents for ontology learning. Much of the user background knowledge is discovered using these data mining technique. In many work, ontologies are used for getting better performance in knowledge discovery process.

B. User Profiles: In the web information gathering, user profiles were used to understand the semantic meanings of the queries and capture user Information needs. User profiles are used for user personalization and modeling. It is used to reflect the interests of users. Li and Zhong defined user profiles as the interesting topics of a users information need. The user profiles are categorized into two diagrams- the data diagram and which are acquired by analyzing a database or a set of transaction whereas the information diagram user profiles acquired by using manually such as interviews and questionnaires or automatic techniques such as information retrieval and machine learning. User profiles are categorized into three groups: they are- interviewing, semi-interviewing, and non-interviewing [1][3].

III. PROPOSED WORK

In proposed work, by using ontology mining we can discover interesting and on-topic knowledge from the concepts, semantic relations and instances in ontology. Here we discuss 2D ontology mining method callded specificity and exhaustivity. Our focus on a given topic is described by Specificity and subject's semantic space dealing with the topic is restricted by exhaustivity. Using this method we can investigate the subject and the strength of their association in ontology. The subject's specificity has two focuses which are known as semantic specificity and topic specificity.

A. Semantic Specificity: Semantic specificity is also called absolute specificity, denoted by spe_a (s). Let ∂(Γ) is a world knowledge base. The semantic specificity is measured based on the hierarchical semantic relations(is-a and part-of) held by the subject and its neighbors in tax^s. The A(s') and part of(s') are two functions in the algorithm satisfying: A(s')∩part of(s')=Ø.As the tax^s of ∂(Γ) is a graphic taxonomy, the leaf subjects have no descendants. If a subject has direct child subjects mixed with relations is-a and part-of relationships, a spe_a and spe_a are addressed separately with respect to the relations is-a and part-of child subjects. The following algorithm illustrates semantic relations for specificity:

Algorithm:

Input: a personalized ontology ∂(Γ) := < tax^s,rel> ;
A coefficient ϴ between (0,1).

Output: spe_a (s) applied to specificity.

Step 1: set k=1, get the set of leaves s_o from tax^s,
for(s_o ϵ S_o) assign spe_a (s_o) = k;

Step 2: get S' which is the set of leaves in case we remove the nodes S_o and the related edges from tax^s;

Step 3: if (S' == Ø) **then** return;// the terminal condition;

Step 4: for each s' ϵ S' do

Step 5: if (is A(s') == Ø) then spe_a^1 (s')=k;

Step 6: else spe_a^1 (s') = ϴ * min{ spe_a (s)|sϵ isA(s')};

Step 7: if (part of(s') == Ø) then spe_a^2 (s')=k;

Step 8: else spe_a^2 (s')=Σsϵ partOf(s') spe_a (s)/|partOf(s')|;

Step 9: spe_a (s')=min(spe_a^1(s'), spe_a^2 (s'));

Step 10: End

Step 11: k=k*ϴ, $S_o = S_o$ U S', go to step 2.

B. Topic Specificity: The topic specificity of a subject is investigated, based on the user background knowledge discovered from the user's local information. User background knowledge can be discovered from the user's local information collections. Populate the ontology with the instances generated from the user's local information collections. Such a collection called as user local instance repository (LIR). Generating the user's local LIRs is a challenging issue. The documents in LIRs may be semi- structured or an unstructured .In semi- structured web documents, content-related descriptors are specified in the meta-data sections. These descriptors have direct reference to the concepts specified in the global knowledge base. These documents are ideal to generate the instances for the ontology population. The documents in the user local repository have content-related descriptors referring to the subjects in O(T). The reference strength between an instance and the subject needs to be evaluated. Hence, denoting an instance by "i", the strength of i to a subject s is determined by:

$$\text{str}(i, s) = \frac{1}{priority(s,i)Xn(i)}$$

The subject str (i, s) aims to select the right instances to populate O(T). With the str(i, s) determined, the relationship between the LIR and O(T) can be defined. Let Ω= {i1, i2 ,i3,……. ik } be a finite and non- empty set of instances in an LIR, and min_str be the minimal str values for filtering out the noisy pairs with weak strengths.

C. Architecture of the Ontology Model: The proposed ontology model aims to discover the user's back-ground knowledge and learns personalized ontologies to rep-resent user profiles. Figure 1 illustrates the architecture of the ontology model. A personalized ontology is constructed using a given topic. Two knowledge resources, the global world knowledge base and the user's local instance repository, are utilized by using the model. The world knowledge base provides the taxonomic structure for the personalized ontology representation. The user background knowledge is discovered from the user's local instance repository. Against the given topic, the specificity and exhaustivity of the subjects are investigated for user background knowledge discovery.

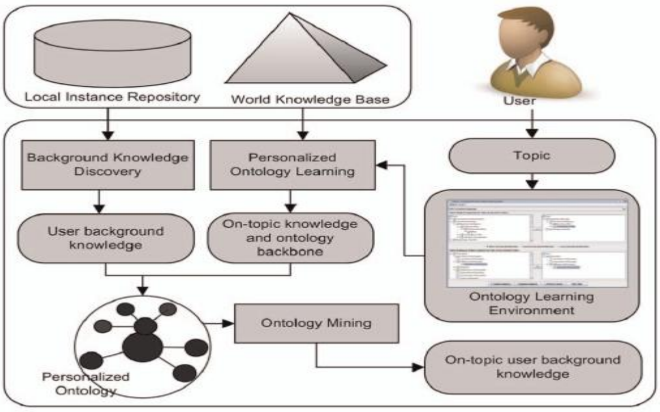

Figure 1: Architecture of the ontology model

From the figure, we can hypothesize that user background knowledge can be better discovered and represented if we can integrate global and the local analysis within a hybrid model. The knowledge formalized in the global knowledge base will con-strain the background knowledge discovery from the user local information. Such a personalized ontology model should produce a superior representation of the user profiles for web information gathering.

IV. CONCLUSION

Every user has a distinct background and a specific goal when searching for information on the World Wide Web. The goal of Web search personalization is to tailor the search results to a particular user based on that user's interests and preferences. Effective personalization of information access involves two important challenges- accurately identifying the user's context and organizing the information in such a way that matches the particular contexts. We present an approach to personalized search that involves building models of the user's context as ontological profiles by assigning implicitly derived interest scores to existing concepts in domain ontology. A spreading activation algorithm is used to maintain the interests scores based on user ongoing behaviour.

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 Its Applications 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]. S.E. Robertson and I. Soboroff, "The TREC 2002 Filtering Track Report," Proc. Text Retrieval Conf., 2002.
[5]. Sieg, B. Mobasher, and R. Burke, "Web Search Personalization with Ontological User Profiles," Proc. 16th ACM Conf. Information and knowledge Management(CIKM‟ 07),pp.525-534,2007
[6]. 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
[7]. 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
[8]. X. Jiang and A.-H. Tan, "Mining Ontological Knowledge from Domain-Specific Text Documents," Proc. Fifth IEEE Int‟ l Conf. DataMining (ICDM ‟ 05), pp. 665-668, 2005.
[9]. R. Navigli, P. Velardi, and A. Gangemi, "Ontology Learning and Intelligent Systems, vol. 18, no. 1, pp. 22-31, Jan./Feb. 2003.Its Application to Automated Terminology Translation," IEEE Intelligent Systems, vol. 18, no. 1, pp. 22-31, Jan./Feb. 2003.