RESEARCH ARTICLE OPEN ACCESS

Personalized Ontology Model for Web Information Gathering

Nilesh P. Patil*, Shrinivas R. Zanwar**, Dr. S. P. Abhang***
* M. E. Student, Department of Computer Sci.& Engg, Dr. Seema Quadri Institute of Tech, Aurangabad, India.
** Asst. Prof, Department of Electronics & Communication Engineering, Dr. Seema Quadri Institute of Tech, Aurangabad, India.
*** Asst. Prof, Department of computer Sci. & Engg., CSMSS College of Engeneering, Aurangabad.

ABSTRACT
The World Wide Web is an interlinked collection of billions of documents formatted using HTML. 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. Therefore, the new technology is to be introduced that that will be helpful for the web information gathering Ontology as model for knowledge description and formalization is used to represent user profile in personalized web information gathering. Ontology is the model for knowledge description and formalization. However the information of user profiles represents patterns either global or local knowledge base information, according to our analysis many models represents global knowledge. In this paper ontology system is used to recognize and reasoning over user profiles, world knowledge base and user instance repositories. This work also compares the analysis of existing system and ontology with other research areas are more efficient to represent.

Keywords – Local Instance Repository, Ontology, Personalization, Semantic Relations, User Profiles, Web Information gathering

I. INTRODUCTION

Today is the world of internet. The amount of the web-base information available on the internet has increased significantly. Personalization of information access indeed to face considerable growth of data heterogeneity of the roles and needs to the rapid development of mobile system becomes important to propose a personalized system able to provide user with relevant information need. 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. Ontology is best the candidate for representing knowledge about users to have a shared understanding between people or software agents of terms and their relations a controlled vocabulary. Ontology's have been proven and effective information means for modeling a user context can be very useful tool because they may present an overview of the domain related to a specific area of interest and used for browsing query refinement, provides rich semantics for humans to work with required formalism for computers to perform mechanical processing. On the last decades, 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 c a p t u r i n g their information needs. For this purpose, user profiles are created for user background knowledge description.

Data Mining

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

II. RELATED WORK

A. Architecture of proposed Ontology Model

The proposed ontology model aims to discover user background knowledge and learns personalized ontology to represent user profiles. **Figure 1** Illustrates the architecture of the ontology model. A personalized ontology is constructed, according to a given topic. Two knowledge resources, the global World knowledge base and the user's local instance repository, are utilized by the model. The world knowledge base provides the taxonomic structure for the personalized ontology. The user background knowledge 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.

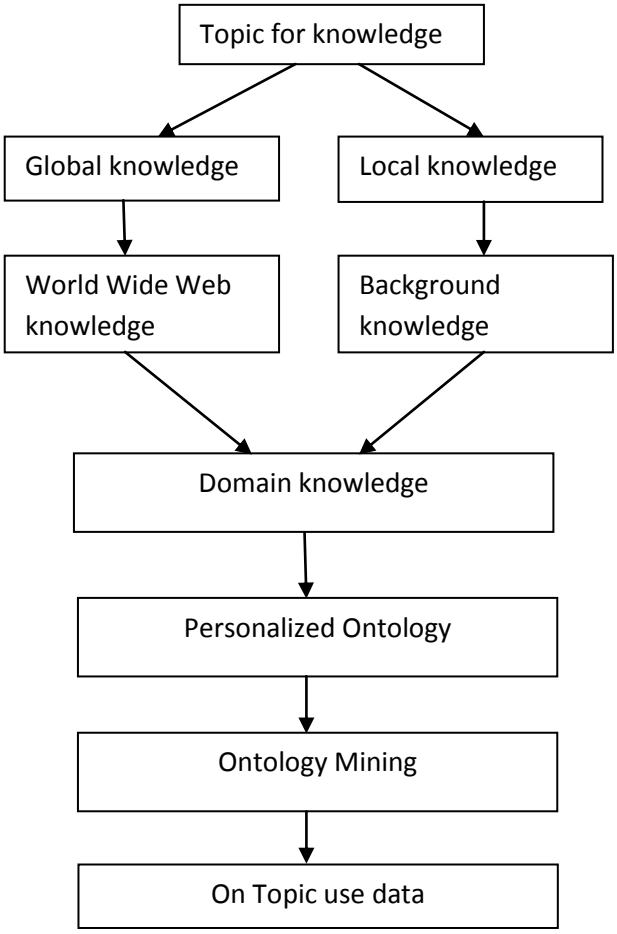

Figure3. Architecture of Ontology Model.

B. Local Profiles

For capturing the user information needs User Profiles were used in web Information gathering. A user profile is a collection of personal data associated to a specific user. A profile refers therefore to the explicit digital representation of a person's identity [11]. A user profile can also be considered as the computer representation of a user model. A profile can be used to store the description of the characteristics of person.

User profiles are categorized into three groups: Interviewing, semi-interviewing, and non-interviewing. Interviewing user profiles are considered to be perfect user profiles. They are acquired by using manual techniques, such as questionnaires, interviewing users, and analyzing user classified training sets. One typical 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 usually 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. They acquire user profiles by observing user activity and behavior and discovering user background knowledge [6].

III. Existing System

A. **Golden Model: TREC Model**

The TREC model was used to demonstrate the interviewing user profiles, which reflected user concept models perfectly. For each topic, TREC users were given a set of documents to read and judged each as relevant or nonrelevant to the topic. The TREC user profiles perfectly reflected the users' personal interests, as the relevant judgments were provided by the same people who created the topics as well, following the fact that only users know their interests and preferences perfectly.

B. Baseline Model: Category Model

This model demonstrated the non-interviewing user profiles, a user's interests and preferences are described by a set of weighted subjects learned from the user's browsing history. These subjects are specified with the semantic relations of super class and subclass in ontology. When an OBIWAN agent receives the search results for a given topic, it filters and re-ranks the results based on their semantic similarity with the subjects. The similar documents are awarded and re-ranked higher on the result list.

C. Baseline Model: Web Model

The web model was the implementation of typical semi interviewing user profiles. It acquired user profiles from the web by employing a web search engine. The feature terms referred to the interesting concepts of the topic. The noisy terms referred to the paradoxical or ambiguous concepts.

IV. ALGORITHM: ANALYZING THE SEMANTIC RELATIONS:

Here we have combined both semantic and KMP searching algorithm for retrieving webpage content. Semantic search technique is used to retrieve a WebPages by finding relations between the texts given. KMP algorithm is used to find a partial match for given input. Knuth-Morris-Pratt algorithm is for pattern recognition. Semantic search is used to identify specificity.

A. Multidimensional Ontology Mining

Ontology mining discovers interesting and on-topic knowledge from the concepts, semantic relations, and instances in ontology. In this section, a 2D ontology mining method is in- traduced: 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. Subject's specificity has two focuses: 1) on the referring-to concepts (called semantic specificity), and 2) on the given topic (called topic specificity).

input : a personalized ontology $\mathcal{O}(\mathcal{T}) := \langle tax^{\mathcal{S}}, rel\rangle$; a coefficient θ between (0,1).
output: $spe_a(s)$ applied to specificity.

1 set $k = 1$, get the set of leaves S_0 from $tax^{\mathcal{S}}$, for $(s_0 \in S_0)$ assign $spe_a(s_0) = k$;
2 get S' which is the set of leaves in case we remove the nodes S_0 and the related edges from $tax^{\mathcal{S}}$;
3 **if** $(S' == \emptyset)$ **then** return;//*the terminal condition*;
4 **foreach** $s' \in S'$ **do**
5 **if** $(isA(s') == \emptyset)$ **then** $spe_a^1(s') = k$;
6 **else** $spe_a^1(s') = \theta \times min\{spe_a(s)|s \in isA(s')\}$;
7 **if** $(partOf(s') == \emptyset)$ **then** $spe_a^2(s') = k$;
8 **else** $spe_a^2(s') = \frac{\sum_{s \in partOf_{(s')}} spe_a(s)}{|partOf(s')|}$;
9 $spe_a(s') = min(spe_a^1(s'), spe_a^2(s'))$;
10 **end**
11 $k = k \times \theta, S_0 = S_0 \cup S'$, go to step 2.

Algorithm 1. Analyzing Semantic Relations For Specificity.

B. **Semantic Specificity**

The semantic specificity is computed based on the structure inherited from the world knowledge base. The strength of such a focus is influenced by the subject's locality in the taxonomic structure. The subjects are graph linked by semantic relations. The upper 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 and lower subjects, the lower subject has a stronger focus because it has fewer concepts in its space. Hence, the semantic specificity of a lower subject is greater than that of an upper subject. The semantic specificity is measured based on the hierarchical semantic relations (is-a and part-of) held by a subject and its neighbors. The semantic specificity of a subject is measured, based on the investigation of subject locality in the taxonomic structure. In particular, the influence of locality comes from the subject's taxonomic semantic (is-a and part- of) relationships with other subjects.

V. KNOWLEDGE REPRESENTATION

A. Global Knowledge Base

Global knowledge is the knowledge possessed by people acquired from experience and education. A global knowledge base is a global ontology that formally describes and specifies world knowledge. With a global knowledge base, user-interesting concepts are extracted, including both the relevant and non-relevant concepts according to user information needs.

The Library of Congress Subject Headings (LCSH) classification is a system developed for organizing large volume of information stored in a library. The LCSH system specifies the semantic relation in the subject heading and the user's perspective in accessing the information in a library catalogue. Based on the LCSH system, a global knowledge base is constructed by defining each subject heading as a class node and using the specified semantic relations as the links between the nodes

B. Local Instance Repository

User background knowledge can be discovered from user local information collections, such as user's stored documents, browsed web pages and composed/received emails. Generating user local instance repository (LIR) is a challenging issue. The documents in LIRs may be semi structured (e.g. the browsed HTML and XML web documents) or unstructured (e.g., the stored local DOC and TXT documents)

VI. METHODOLOGY

The LGSM (Local Global search methodology) it is used to calculate the hit/miss rate. For calculating hit ratio,

$$\text{Hit Ratio} = \frac{\text{Number of Hits}}{(\text{Number of Hits} + \text{Number of Miss})}$$

The performance of memory is frequency measured in terms of quantity is called hit ratio. When cpu needs to find the word in cache, if word is found in cache then it produces a hit. If the word is not found in the cache, it is in main memory it is counted as miss. If it retrieves information from the local repository it is considered as hit. If it retrieves data directly from global it is considered as miss [16].

VII. CONCLUSION

In this paper, an Ontology model is proposed for representing user background knowledge for personalized web information gathering. The ontology model provides a solution to emphasizing global and local knowledge in a

single computational model. . This model constructs the global search from the world knowledge base and local search from local instance repository. In addition, the ontology model using knowledge with both is-a and part-of semantic relations works better than using only one of them.

REFERENCES

[1]. Claudia Marinica and Fabrice Guillet,"Knowledge based interactive post-mining of association rules using Ontology" IEEE Transactionson Knowledge and data engineering, Vol. 22, NO. 6, June 2010.

[2]. T. Tran , P. Cimiano, S. Rudolph, and R. Studer,"Ontology-Based Interpretation of KeywordsforSemantic Search, " Proc. Sixth Int'l Semantic Web and Second Asian Semanti Web Conf.(ISWC '07/ASWC '07), pp.523-536, 2007.

[3]. D.D. Lewis, Y. Yang, T.G. Rose, and F. Li, "RCV1: A New Benchmark Collection for 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.

[4]. 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.

[5]. S. Shehata, F. Karray, and M. Kamel, "Enhancing Search Engine Quality Using Concept-Based Text Retrieval," Proc. IEEE/WIC/ ACM Int'l Conf. Web Intelligence (WI '07), pp. 26-32, 2007.

[6]. A. 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.

[7]. J.D. King, Y. Li, X. Tao, and R. Nayak, ,Mining World Knowledge for Analysis of Search Engine Content,˚ Web Intelligence and Agent Sy s- tems,vol.5,no.3,pp.233-253,2007