RESEARCH ARTICLE OPEN ACCESS

A Personalized User Concept Model For Web Information Gathering

S.Swapna, Mohd.Ahmed.Ali
(Assistant Professor,IT,Aurora Technological and Research Institute,Hyd)
(M.TECH(Web Technology),Aurora Technological and Research Institute,Hyd)

ABSTRACT
Nowadays It has become very easy to access information from internet via World Wide Web. If we want any information through search engine they do not deliver the relevant information because they are programmed as one size fits all. Ironically the very vast size of this collection has become an obstacle for information retrieval. It is very logical and reasoning to personalize the retrieval system specific to the preference of a user. However,when representing user profiles,many models have utilized only knowledge from either global knowledge base or a user local information This paper work contribute in improving the accuracy of information gather to personalize the user profile by combining the knowledge base and user local instance repository with 2D ontology mining method is introduced: Specificity and Exhaustivity. This concept model cannot be proven in laboratories; many web ontologists have observed it in a user behavior the results show that this ontology model is successful.

Keywords **-** Local instance repository, Ontological user profiles, Personalization, Semantic relations, User context

I. INTRODUCTION

World Wide Web (W.W.W) is a magnificent and vast portal for acquiring , gathering and retrieval information from the infinite information present on the web. Web information is available in a great range of topics and categories. How to collect the required information from the collected information is a challenging task. 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 Rapid growth of documents in the Web makes it difficult to determine the most relevant documents for a particular user, given a general query. User performs searching to access relevant document for himself. Therefore, aspecific system should be there to personalize the information retrieved by user according to his preference. User profiles reflect the interests of users [5]. User profiles concepts are used in gathering web information to capture user information needs in order to get personalized web information for users. For this purpose only, user profiles are created for user background knowledge description.

Global analysis uses only existing global knowledge bases for user background knowledge discovery. Generally used knowledge bases include generic ontologies , thesauruses (e.g., digital libraries), and online knowledge bases (e.g., online categorizations). The global analysis techniques effectively produce good performance for user background knowledge extraction.However, global analysis is limited to the quality of the used knowledge base.

Local analysis investigates the user local information and observes user behavior in user profiles. For example, Li and Zhong [6] discovered taxonomical patterns from the users'local text documents to learn ontologies from user profiles. Alternatively, Sekine and Suzuki [14] analyzed the query logs to discover user background knowledge. In some works, such as [13],users were provided with a set of documents and asked for relevance feedback from that documents. User background knowledge was discovered from this feedback for user profiles. However, because local analysis techniques rely on data mining or old classification techniques for knowledge discovery, occasionally the discovered results contain noisy and unrelevant information. As a result, local analysis suffers from ineffectiveness at capturing the relevant user information.

From this, we can hypothesize that user background knowledge can give better result if we can collaborate global and local analysis within a hybrid model. The Information formalized in a global knowledgebase will constrain the background knowledge discovered from the user local information.

II. RELATED WORK

2.1 Learning Personalized Ontological Context

Effective personalization of relevant information access involves two important challenges: accurately identifying the user context and organizing the information in such a way that it matches the particular context[9]. Since the acquisition of user interests is an essential element in

identifying the user context, most personalized search systems employ a user concept modeling component.

Global knowledge base has been used by many previous or existing models to learn ontologies for web information gathering. For example, Gauch et al. [2] and Sieg et al. [9] learned personalized ontologies from the Open Directory Project to specify users' interests in web search. Wikipedia used by Downey et al. [12] to understand underlying user interests in queries. These works effectively discovered the user background knowledge; but , their performance was limited by the quality of the global knowledge bases.

Learning personalized ontologies, required many works mining of user background knowledge from the user local information.Li and Zhong [6] used pattern recognition and association rule mining techniques to discover the background knowledge from user local documents for ontology construction. Zhong [6] proposed a domain personalized ontology learning approach that will employed the various data mining and natural language techniques. Navigli et al. [3] developed the OntoLearn technique to discover semantic concepts and relations from web documents. Finally, captured user information at the sentence level rather than the document level, represented user profiles by the Conceptual Ontological Graph. The use of data mining techniques in these models leads to user background knowledge being discovered. However, the knowledge collected in these works contained noise and unrelated data.

Additionally, ontologies were used for many works to improve the performance of knowledge discovered.Uptil,concept based search,keyword based search is available but the URL searching is text based online searching which is different than available search.

2.2 Techniques for Generating User Concept Profile

The user's intent for information seeking. We propose to model a user's concept model for information access context by seamlessly integrating knowledge from the world knowledge and local instance repository. In our framework, context is implicitly defined through the notion of ontological user profiles, which are updated over time to reflect changes in user interests[9]. This representation our approach differentiate from previous work which depends on the context information to be explicitly defined.

When acquiring user concept profiles, the content and applications are taken into consideration ,since user interests are approximate and it is suggested that it can be represented by ontologies[15] . User profile techniques can be categorized into three groups: 1) Interviewing 2) Non-interviewing 3) Semi-interviewing.

The interviewing technique are done manually by asking the questions and by using the user trained datasets. Users will read the training sets of documents and then assign positive or negative feedback based on user's personal interests.e.g. TREC model was used to acquire training set manually[13]. The topic coverage of TREC profiles is limited. But it provides more accuracy,but weak in terms of precision.

In Non-interviewing technique there is no involvement of user, it is based on the observation at user's behavior, user's interests and preferences which are described by a set of weighted subjects, learned from the user's browsing history.When an OBIWAN agent receives the search results based on a given topic, it filters and reranks the results based on their semantic similarity with the subjects. Then the similar documents will be kept aside, then awarded and reranked higher on the result list. e.g. Category model.

Semi-interviewing technique the user involvement is very less, in which user profiles are acquired from the web by employing a web search engine. The noisy and uncertain terms will referred to the paradoxical concepts. e.g. Web mining model. Using web documents for training sets has severe drawback: web information has much more noise and uncertainties. As a result, the web user profiles are satisfactory in terms of recall, but weak in terms of precision. There is no negative training set generated by this model. In ontology model ,semi-interviewing technique is used.

Fig.1.Shows the user context model which represents the users intent for information seeking
We propose to model a user's information access context by seamlessly integrating knowl-edge from the immediate and past user activity as well as knowledge from a pre-existing ontology as an explicit representation of the domain of interest[9].This representation distinguishes our approach from previous work which depends on the context information to be explicitly defined.

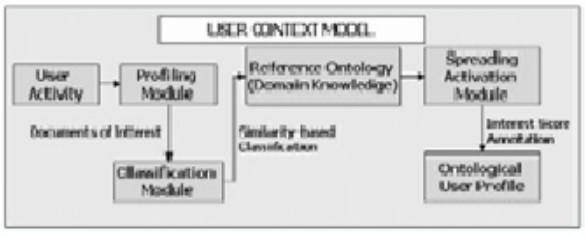

Fig .1: User Context Model[9]

III. CONSTRUCTION OF PERSONALIZED ONTOLOGICAL PROFILES

In the present framework, the user context is represented using an ontological user profile, which is an annotated instance of reference ontology[9].Each personalized ontological user profile is initially an instance of the reference ontology.Thus,the user context is maintained and updated incrementally

based on user's ongoing behavior.

User personalized ontologies are a conceptualization model, that formally describes and specifies user background knowledge. From observations in our daily life, we found that web users might have different expectations for the same search query. For example,a person may demand different information for various medical aids[15]. Sometimes even the same user may have different expectations for the same search query if applied in a different situation. Based on this observation, an assumption is made that web users have a personal concept model for their information needs. Therefore, a user concept model for different information need is suggested.

3.1 Representation Of World Knowledge Base

World knowledge base contains commonsense knowledge possessed by people and acquired by their experience and education.World knowledge base contains exhaustive range of topics,because users may come from different backgrounds. Also, "world knowledge is necessary for lexical analysis and referential disambiguation, including establishing co reference relations and resolving ellipsis as well as for establishing connectivity of the discourse and adherence of the text to the text producer's goal and plans"[15]. In this proposed model, user background knowledge is extracted from web.

3.2 Creation of Ontology Environment

Creation of Ontology learning environment(OLE) involes the subjects of user interest which are extracted from the WKB via user interaction. A tool called Ontology Learning Environment (OLE) is used to assist users with such interaction[1].Depending upon the topic, the interesting subjects consisting of two sets: positivesubjects are 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 specific topic, the OLE provides the users with a set of candidates to identify positive and negative subjects. These candidate can select the subjects which are extracted from the WKB.

Fig. 2 is a screen-shot of the OLE for the sample topic "Economic espionage." The subjects listed on the top-left panel of the OLE are the candidate subjects presented in hierarchical form.

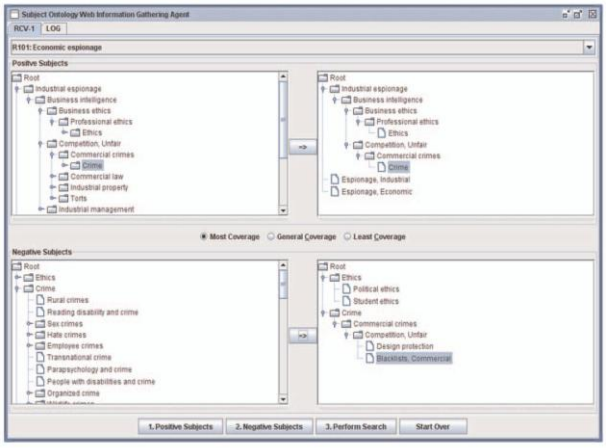

Fig. 2: Ontology learning environment[1]

From these candidates, the user selects positive subjects for the topic. The user-selected positive subjects are presented on the top-right panel in hierarchical form. The candidate negative subjects are the descendants of the user-selected positive subjects. They are shown on the bottom-left panel. From these negative candidates, the user selects tnegative subjects. These user-selected negative subjects are listed on the bottom-right panel (e.g., "Political ethics" and "Student ethics")[1]. Note that for the completion of the structure, some positive subjects (e.g., "Ethics," "Crime," "Commercial crimes," and "Competition Unfair")are also included on the bottom-right panel with the negative subjects[15]. These positive subjects will not be included in the negative set.The remaining candidates, which are not fed back as either positive or negative from the user, become the neutral subjects to the given topic.

An ontology is then constructed for the given topic using these user fed back subjects. The structure of the ontology is based on the semantic relations linking these subjects in the WKB. The ontology contains three types of knowledge:positive subjects, negative subjects, and neutral subjects.

IV. TWO DIMENSIONAL ONTOLOGY MINING

Two dimension Ontology mining method is used for mining data. Specificity describes a subject's focus on a given topic where as Exhaustivity restricts a subjects semantic space dealing with the topic. This two methods aims to investigate the subjects focus and the strength of their associations in ontology.The knowledge formalized in a global knowledge base will contain the background knowledge discoverd from the user local repository. Such a personalized user concept model should produce a superior representation of user profiles for web information gathering.

Algorithm:

Algorithm works based on personalized ontology. It traverses all the nodes and identifies nodes with strongest and weakest specificity. The semantic specificity of a subject is measured based on the investigation of subject locality that comes from ontology and relationship among concepts.

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 specificity[1]

4.1 Topic Specificity Of Subject

The topic specificity of a subject is investigated, based on the user background knowledge discovered from user local history or information.

4.2 User Background Local Instance Repository

The User background knowledge can be discovered from the user collection of local information , such as a user's stored documents, browsed web pages, and composed/received emails [6]. The ontological user profiles 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 infor-mation collections. Such a collection of the user's local instance repository (LIR). The topic specificity of a subject is evaluated based on the instance topic strength of its citing URLS. With respect to the absolute specificity, the topic specificity can also be called relative specificity

V. ONTOLOGY MODEL ARCHITECTURE

The proposed user concept ontology model aims to discover user back-ground knowledge and learns personalized ontologies to represent user concept profiles.

Fig. 3 illustrates the architecture of the user concept 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 in these model. The world knowledge base provides the taxonomic structure for the personalized ontology. The user background knowledge is collected from the user local instance repository.

From the figure, 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.

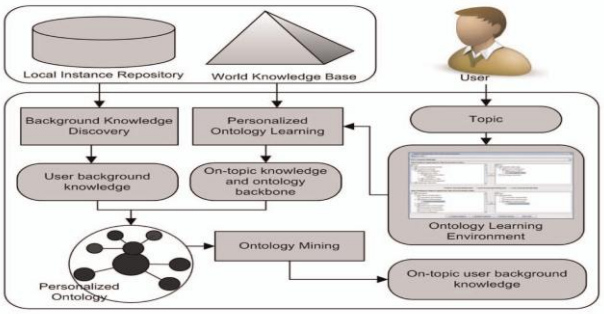

Fig 3: user concept model Architecture[15]

REFERENCES

[1] Xiaohui Tao, Yuefeng Li, and N. Zhong, ―A Personalised Ontology Model for Web Information Gathtering, *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 4, APRIL 2011.*

[2] 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.*

[3] R 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.*

[4] D. Quest and H. Ali, ―Ontology Specific Data Mining Based on Dy-namic Grammars, *Proc. IEEE Computational Systems Bioinformatics Conf. (CSB '04), pp. 495-496, 2004.*

[5] Trajkova and S. Gauch, ―Improving Ontology-Based User Profiles, *Proc. Conf. Recherche d'Information Assistee par Ordinateur (RIAO '04), pp. 380-389, and 2004.*

[6] 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.*

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

[8] G. M. Voorhees and Y. Hou, ,Vector Expansion in a LargeCollec-tion,' *Proc. First Text REtrieval Conf., pp. 343-351,*

[9] A. Sieg, B. Mobasher, and R. Burke, ‚Web SearchPersonalization with Ontological User Profiles‛ *Proc. 16th ACM Conf. Information and Knowledge Management (CIKM '07),pp. 525-534, 2007.*

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

[11] T. Tran, P. Cimiano, S. Rudolph, and R. Studer, ‚Ontology-Based Interpretation of Keywords for Semantic Search,‛ *Proc. Sixth Int'l Semantic Web and Second Asian Semantic Web Conf. (ISWC '07/ASWC '07), pp. 523-536, 2007.*

[12] 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.*

[13] S.E. Robertson and I. Soboroff, "The TREC 2002 Filtering Track Report,"*Proc.Text REtrieval Conf., 2002.*

[14] S. Sekine and H. Suzuki, "Acquiring Ontological Knowledge from Query Logs," *Proc. 16th Int'l Conf. World Wide Web (WWW '07)*

[15] Shubhangi Shindikar, Anand Deshpande,"A Personalized Ontology Model for Web Information Gathering by Domain Specific" *International Journal of Scientific & Engineering Research, volume 3, issue 7, July 2012*

Authors Profile

First Author - S.Swapna recived her M.Sc Computer Science in 2007 from Reddy women's College Narayanguda and M.Tech in Web Technologies from Aurora's Technological and Research Institute, JNTUH. Her area of expertise includes Operating system, Database and Management System (DBMS),. She is working as Assistant Professor in department of Information Technology at Aurora's Technological and Research Institute, Hyderabad. JNTUHHyderabad,

SecondAuthor – Mohd.Ahmed.Ali, M.Tech (W.T), Aurora's Technological and Research Institute, JNTUH yderabad, Andhra Pradesh, pin code-500035, India