



Personalized Ontology Model for Web Information Gathering

Ms. Amruta Y. Kokate

*Department of Information Technology,
JSPM'S R.S.C.O.E, Pune.*

Mr. Bhushan D. Nighot

*Department of Information Technology,
JSPM'S R.S.C.O.E, Pune.*

Ms. Chetana N. Sadare

*Dept. of Information Technology,
JSPM'S R.S.C.O.E, Pune.*

Ms. Priyanka P. Mangalkar

*Dept. of Information Technology,
JSPM'S R.S.C.O.E, Pune.*

Prof.G.S.Mate

*Dept. of Information Technology,
JSPM'S R.S.C.O.E, Pune.*

ABSTRACT: Ontology is the model for knowledge description and formalization, which are widely used to represent user profiles in personalized web information gathering. When representing user profiles, many models have utilized only knowledge from either a global knowledge base or user local information. In this paper, 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. User profiles represent the concept models by user when gathering web information. A concept model is possessed by users and is generated from their background knowledge. 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.

KEYWORD: Ontology, Personalization, World knowledge, Local instance repository, User profile, Semantic relations, Web information gathering.

I. INTRODUCTION:

The amount of web based information available has increased dramatically. How to gather useful information from the web it became a challenging issue for users. Current web information gathering systems attempt to satisfy user requirements. It will capture 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. Concept models are implicitly possessed by users. They are generated from their background knowledge. While this concept model cannot be proven in laboratories, 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 users' concept models can be simulated, then a superior representation of user profiles can be built.

To simulate these models, ontologies as a web knowledge description and formalization model are utilized in personalized web information gathering. Such ontologies are called ontological user profiles or personalized ontologies. Commonly

used knowledge bases include generic ontologies (e.g. WorldNet), thesauruses (e.g. digital libraries), and online knowledge bases (e.g., online categorization and Wikipedia). To represent user profiles, many researchers have attempted to discover user background knowledge through global or local repositories. Global analysis uses existing global knowledge bases for user background knowledge. The global analysis techniques produce effective performance for user background knowledge extraction.

However, global analysis is limited by the quality of the used knowledge. For example, WorldNet was reported as helpful in capturing user interest in some areas but useless for others. Local analysis observes user behavior in user profiles and also investigates user local information. For example, Li and Zhong discovered taxonomical patterns from the user's local text documents for learning ontologies user profiles. Other researchers learned personalized ontologies from user's browsing history. Alternatively, Sekin and Suzuki analyzed query logs to discover user background. Another way to get the user background is to ask some question to the user. In some other works, such as, users were provided with a set of documents and asked for relevance feedback for user profiles. However, classification techniques for knowledge discovery and local analysis techniques rely on data mining. Mostly, the discovered information contains noisy and uncertain information. As a result, local analysis suffers from ineffectiveness at capturing knowledge. From this, we can discover that user background knowledge can be discovered better and represented if we can integrate global and local analysis within a model.

II. Literature Review & Related work:

Literature survey performs an important role in the software development process. Before developing, it is necessary to determine the Time factor, economical company strength. Next, things to determine which operating system is going to be

used and also the languages used for software development. Lots of external support is needed. This support is get from the senior developers, websites and books. Before that analysis and survey can be done.

We have to analysis the Data Mining survey,

Data Mining

The manual extraction of patterns from data has occurred for centuries. Early methods of identifying patterns in data include Bayes' theorem (1700s) and regression analysis (1800s). The proliferation, ubiquity and increasing power of computer technology has dramatically increased data collection, storage, and manipulation ability. As data sets have grown in size and complexity, direct "hands-on" data analysis has increasingly been augmented with indirect, automated data processing, aided by other discoveries in computer science, such as neural networks, cluster analysis, genetic algorithms (1950s), decision trees (1960s), and support vector machines (1990s). Data mining is the process of applying these methods with the intention of uncovering hidden patterns in large data sets. It bridges the gap from applied statistics and artificial intelligence (which usually provide the mathematical background) to database management by exploiting the way data is stored and indexed in databases to execute the actual learning and discovery algorithms more efficiently, allowing such methods to be applied to ever larger data sets.

III. Existing System

A. Golden model: TREC model

Golden model was used to demonstrate the interviewing user profiles, which reflect user concept models. Users contain set of user documents to read and judge the each as relevant or non relevant to the topic. The golden model perfectly reflects the user interests from user profiles.

B. Baseline Model: Category Model

Baseline model perform the important role in finding user interests in topics. A user interests and preferences are described by a set of weighted subject learned from the user browsing history. These subjects are specified with the semantic relations with super class and subclass in ontology. In this category model, set of positive subjects are manually feedback by that user via the OLE and from the WKB.

C. Baseline Model: Web mode

The web model was the implementation of semi-interviewing user profiles. It collect the user profiles from the web by employing a web search engine by using the feature of the noisy terms, the Google API was employed to perform two diff searches for the given topics:

1. Query for positive documents.
2. Query for negative documents.
- 3.

IV. Limitations of Existing systems:

The topic coverage TREC model was limited, with poor recall performances. Web information has much noise and uncertainties. As a result user profiles were weak in terms of precision.

V. Conclusion:

Every user has a distinct background and a specific goal when searching for information on the Web. The goal of Web search personalization is to tailor 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 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 user context as ontological profiles by assigning implicitly derived interest scores to existing concepts in a domain ontology. A spreading activation algorithm is used to maintain the interests scores based on the user's ongoing behavior. Our experiments show that re-ranking the search results based on the interest scores and the semantic evidence in an ontological user profile is effective in presenting the most relevant results to the users.

References:

- [1] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. Addison Wesley, 1999.
- [2] G.E.P. Box, J.S. Hunter, and W.G. Hunter, *Statistics For Experimenters*. John Wiley & Sons, 2005.
- [3] C. Buckley and E.M. Voorhees, "Evaluating Evaluation Measure Stability," *Proc. ACM SIGIR '00*, pp. 33-40, 2000.
- [4] Z. Cai, D.S. McNamara, M. Louwerse, X. Hu, M. Rowe, and A.C. Graesser, "NLS: A Non-Latent Similarity Algorithm," *Proc. 26th Ann. Meeting of the Cognitive Science Soc. (CogSci '04)*, pp. 180-185, 2004.
- [5] L.M. Chan, *Library of Congress Subject Headings: Principle and Application*. Libraries Unlimited, 2005.
- [6] P.A. Chirita, C.S. Firan, and W. Nejdl, "Personalized Query Expansion for the Web," *Proc. ACM SIGIR ('07)*, pp. 7-14, 2007.
- [7] R.M. Colomb, *Information Spaces: The Architecture of Cyberspace*. Springer, 2002.

- [8] 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.
- [9] 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.
- [10] 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.