

Study on Ontology Model for Web Information Gathering

Mrs. N.Kanagavalli, Dr. R.Meenakshi, Mr. N.Velmurugan

Department of IT, Saveetha Engineering College, Chennai, Tamilnadu, India

Abstract— *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 capturing 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. A concept model is implicitly possessed by users and is generated from their background knowledge. While this concept model cannot be proven in laboratories, many web oncologists have observed it in user behavior. 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 introduced: Specificity and Exhaustively. Specificity (denoted spe) describes a subject's focus on a given topic. Exhaustively 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. Our work assumes that all user local instance repositories have content-based descriptors referring to the subjects; however, a large volume of documents existing on the web may not have such content-based descriptors. For this problem, strategies like ontology mapping and text classification/clustering were suggested. 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.*

Keywords—*Ontology, Web information, Web model, ontology mining, knowledge base.*

I. INTRODUCTION

The introduction of the paper should explain the nature of 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 capturing 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. A concept model is implicitly possessed by users and is generated from their background knowledge. While this concept model cannot be proven in laboratories, many web 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 a user's concept model can be simulated, then a superior representation of user profiles can be built. To simulate user concept models, ontology—a knowledge description and formalization model—are utilized in personalized web information gathering. Such ontology are called ontological user profiles or personalized ontology. 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 (e.g., Word Net), 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, WorldNet was reported as helpful in capturing user interest in some areas but useless for others. Local analysis investigates user local information or observes user behavior in user profiles. For example, Li and Zhong discovered taxonomical patterns from the users' local text documents to learn ontology's for user profiles. Some groups learned personalized ontology adaptively from user's browsing history. Alternatively, Sekine and Suzuki analyzed query logs to discover user background knowledge. In some works, such as, users were provided with a set of documents and asked for relevance feedback. User

background knowledge was then discovered from this feedback for user profiles. However, because local analysis techniques rely on data mining or classification techniques for knowledge discovery, occasionally the discovered results contain noisy and uncertain information. 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 formalized in a global knowledge base will constrain the background knowledge discovery from the user local information. Such a personalized ontology model should produce a superior representation of user profiles for web information gathering. An ontology model to evaluate this hypothesis is proposed. This model simulates users' concept models by using personalized ontology 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 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. The users' LIRs are then used to discover background knowledge and to populate the personalized ontology. 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.

II. EXISTING SYSTEM

2.1. 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 non relevant 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.

2.2. 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.

2.3. 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.

Limitations of Existing System

- The topic coverage of TREC profiles was limited. The TREC user profiles had good precision but relatively poor recall performance.
- Using web documents for training sets has one severe drawback: web information has much noise and uncertainties. As a result, the web user profiles were satisfactory in terms of recall, but weak in terms of precision. There was no negative training set generated by this model

III. PROPOSED SYSTEM

The world knowledge and a user's local instance repository (LIR) are used in the proposed model.

- 1) World knowledge is commonsense knowledge acquired by people from experience and education
- 2) An LIR is a user's personal collection of information items. From a world knowledge base, we construct personalized ontology 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. The users' LIRs are then used to discover background knowledge and to populate the personalized ontology.

Advantages of Proposed System:

Compared with the TREC model, the Ontology model had better recall but relatively weaker precision performance. The Ontology model discovered user background knowledge from user local instance repositories, rather than documents read and judged by

users. Thus, the Ontology user profiles were not as precise as the TREC user profiles.

The Ontology profiles had broad topic coverage. The substantial coverage of possibly-related topics was gained from the use of the WKB and the large number of training documents.

Compared to the web data used by the web model, the LIRs used by the Ontology model were controlled and contained less uncertainties. Additionally, a large number of uncertainties were eliminated when user background knowledge was discovered. As a result, the user profiles acquired by the Ontology model performed better than the web model.

IV. SYSTEM DESIGN

MODULES DESCRIPTION

A) World knowledge base

The world knowledge base must cover an exhaustive range of topics, since users may come from different backgrounds.

1. Broader term- The BT references are for two subjects describing the same topic, but at different levels of abstraction (or specificity). In our model, they are encoded as is-a relation in the world knowledge base.
2. Is-a, are used for many semantic situations, including broadening the semantic extent of a subject and describing compound subjects and subjects subdivided by other topics.
3. Related term- The RT references are for two subjects related in some manner other than by hierarchy. They are encoded as the related-to relations in our world knowledge base.

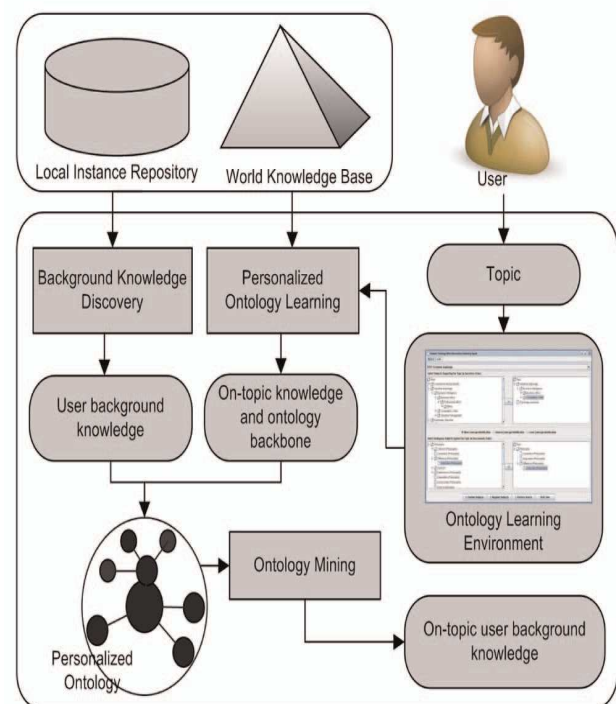
B) Ontology learning environment

The subjects of user interest are extracted from the WKB via user interaction. A tool called Ontology Learning Environment (OLE) is developed to assist users with such interaction. Regarding a topic, the interesting subjects consist of two sets: positive subjects 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 topic, the OLE provides users with a set of candidates to identify positive and negative subjects. These candidate subjects are extracted from the WKB. Who are not fed back as either positive or negative from the user, become the neutral subjects to the given topic.

C) Ontology mining

Ontology mining discovers interesting and on-topic knowledge from the concepts, semantic relations, and instances in ontology. Ontology mining method is

introduced: 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. In User Local Instance Repository, User background knowledge can be discovered from user local information collections, such as a user's stored documents, browsed web pages, and composed/received emails.



V. CONCLUSION & FUTURE ENHANCEMENTS

An ontology model is proposed for representing user background knowledge for personalized web information gathering. The model constructs user personalized ontology by extracting world knowledge and discovering user background knowledge from user local instance repositories. In evaluation, the standard topics and a large test bed were used for experiments. The model was compared against benchmark models by applying it to a common system for information gathering. The experiment results demonstrate that our proposed model is promising. A sensitivity analysis was also conducted for the ontology model.

In this investigation, shows the result that the combination of global and local knowledge works better than using any one of them. In addition, the ontology model using knowledge with both is-a and part-of

semantic relations works better than using only one of them. When using only global knowledge, these two kinds of relations have the same contributions to the performance of the ontology model. While using both global and local knowledge, the knowledge with part-of relations is more important than that with is-a.

The proposed ontology model provides a solution to emphasizing global and local knowledge in a single computational model. The findings of our proposed work can be applied to the design of web information gathering systems. The model also has extensive contributions to the fields of Information Retrieval, web Intelligence, Recommendation Systems, and Information Systems.

In our future work will investigate the methods that generate user local instance repositories to match the representation of a global knowledge base. The present work assumes that all user local instance repositories have content-based descriptors referring to the subjects; however large volume of documents existing on the web may not have such content-based descriptors. For this problem, strategies like ontology mapping and text classification/clustering were suggested. 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.

REFERENCES

- [1] 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.
- [2] 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.
- [3] L.M. Chan, Library of Congress Subject Headings: Principle and Application. Libraries Unlimited, 2005.
- [4] E. Frank and G.W. Paynter, "Predicting Library of Congress Classifications from Library of Congress Subject Headings," J. Am Soc. Information Science and Technology, vol. 55, no. 3, pp. 214-227, 2004.
- [5] 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.