Middle-East Journal of Scientific Research 15 (12): 1675-1679, 2013
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2013.15.12.1052

Improving Web Information Gathering for Personalised Ontology in User Profiles

[1]K.P. Thooyamani, [2]V. Khanaa and [3]R. Udayakumar

[1]Professor of School of Computing, Bharath University, Chennai-600073, India
[2]Dean, Information Technology, Bharath University, Chennai-600073 India
[3]Information Technology, Bharath University, Chennai-600073, India

Abstract: As a model for knowledge description and formalization, ontologies are widely used to represent user profiles in personalized web information gathering. However, when representing user profiles, many models have utilized only knowledge from either a global knowledge base or a user local information. In this paper, a personalized ontology model is proposed for knowledge representation and reasoning over user profiles. This model learns ontological user profiles from both a world knowledge base and user local instance repositories. The ontology model is evaluated by comparing it against benchmark models in web information gathering. The results show that this ontology model is successful.

Key words: Ontologies · A global knowledge · Learns ontological · Benchmark

INTRODUCTION

On the last decades, the amount of web-based information available has increased dramatically.

How 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.Global analysis uses existing global knowledge bases for user background knowledge representation. Commonly used knowledge bases include generic ontologies (e.g., WordNet), 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, World Net 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 ontologies for user profiles. Some groups learned personalized ontologiesadaptively 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.

A Multidimensional ontology mining method, Specificity and exhaustivityn, is also introduced in the proposed model for analyzing concepts specified in ontologies.The user's LIR are then used to discover background knowledge and to populate the personalized ontologies.

The research contributes to knowledge engineering,and has the potential to improve the design of personalized web information gathering systems. The contributions are original and increasingly significant, considering the rapid explosion of web information and the growing accessibility of online documents.

Corresponding Author: R. Udayakumar, Information Technology, Bharath University, Chennai-600073, India.

Related Works:

Ontology Learning: Global knowledge bases were used by many existing models to learn ontologies for web information gathering. For example, Gauch *et al*. [1] and Sieg *et al*. [2] learned personalized ontologies from the Open Directory Project to specify users' preferences and interests in web search. On the basis of the Dewey decimal classification, Kingal. [3] Developed IntelliOnto to improve performance in distributed web information retrieval. Wikipedia was used by Downey *et al*. [4] to help understand underlying user interests in queries. These works effectively discovered user background knowledge; however, their performance was limited by the quality of the global knowledge bases.

User Profiles: User profiles were used in web information gathering to interpret the semantic meanings of queries and capture user information needs [5-9]. User profiles were defined by Li and Zhong as the interesting topics of a user's information need. They also categorized user profiles into two diagrams: the data diagram user profiles acquired by analyzing a database or a set of transactions the information diagram user profiles acquired by using manual techniques, such as questionnaires and interviews or automatic techniques, such as information retrieval and machine learning. Van der Sluijs and Huben proposed a method called the Generic User Model Component to improve the quality and utilization of user modeling. Wikipedia was also used by to help discover user interests. In order to acquire a user profile, Chirita *et al*. and Teevan *et al*. used a collection of user desktop text documents and emails and cached web pages to explore user interests. Makris *et al*. acquired user profiles by a ranked local set of categories and then utilized web pages to personalize search results for a user. These works attempted to acquire user profiles in order to discover user background knowledge.

User profiles can be categorized into three groups: interviewing, semi-interviewing no interviewing. Interviewing user profiles can be deemed 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. The users read each document and gave a positive or negative judgment to the document.

Personalized Ontology Construction: Personalized on tologies area conceptualization model that formally describes and specifies user background knowledge. From observations in daily life, we found that web users might have different expectations for the same search query. For example, for the topic "New York," business travelers may demand different information from leisure travelers. Sometimes even the same user may have different expectations for the same search query if applied in a different situation.

A user may become a business traveler when planning for a business trip, or a leisure traveler when planning for a family holiday. Based on this observation, an assumption is formed that web users have a personal concept model for their information needs. A user's concept model may change according to different information needs. In this section, a model constructing personalized ontologies for web users' concept models is introduced.

World Know Ledge Representation: World knowledge is important for information gathering. According to the definition provided by, world knowledge is commonsense knowledge possessed by people and acquired through their experience and education. Also, as pointed out by Nirenberg and Raskin, "world knowledge is necessary for lexical and referential disambiguation, including establishing co reference relations and resolving ellipsis as well as for establishing and maintaining connectivity of the discourse and adherence of the text to the text producer's goal and plans." In this proposed model, user background knowledge is extracted from a world knowledge base encoded from the Library of Congress Subject Headings (LCSH).

The structure of the world knowledge base used in this research is encoded from the LCSH references. The LCSH system contains three types of references: Broader term (BT), Used-for (UF) and Related term (RT) [5]. 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 the is-a relations in the world knowledge base.

The UF references in the LCSH are used for many semantic situations, including broadening the semantic extent of a subject and describing compound subjects and subjects subdivided by other topics. The complex usage of UF references makes them difficult to encode. During the investigation, we found that these references are often used to describe an action or an object. When object is used for an action, Abecomes a part of that action (e.g., "a fork is used for dining"); when A is used for another object, B, A becomes a part of B (e.g., "a wheel is used for a car"). These cases can be encoded as

the part-of relations. Thus, we simplify the complex usage of UF references in the LCSH and encode them only as the part-of relations kno. 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.

Ontology Construction: The subjects of user interest 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 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.

Figure 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. For each s 2 SS, the s and its ancestors are retrieved if the label of s contains any one of the query terms in the given topic (e.g., "economic" and "espionage"). 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 the negative subjects. These user-selected negative subjects are listed on the bottom-right panel (e.g., "Political ethics" and "Student ethics"). 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.

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 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. Fig. 3 illustrates the ontology (partially) constructed for the sample topic "Economic espionage,"

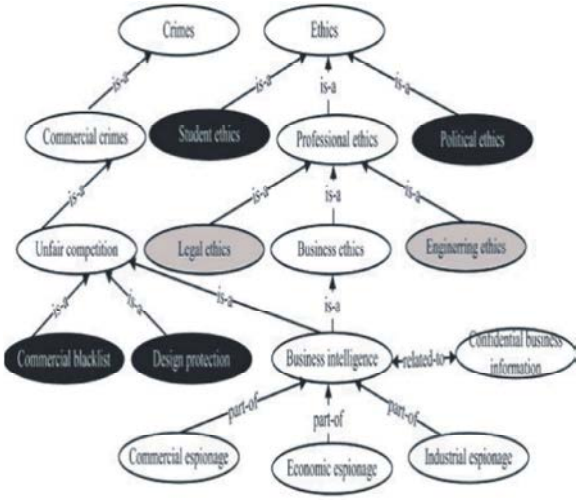

Fig. 1: An ontology construction for personalized model

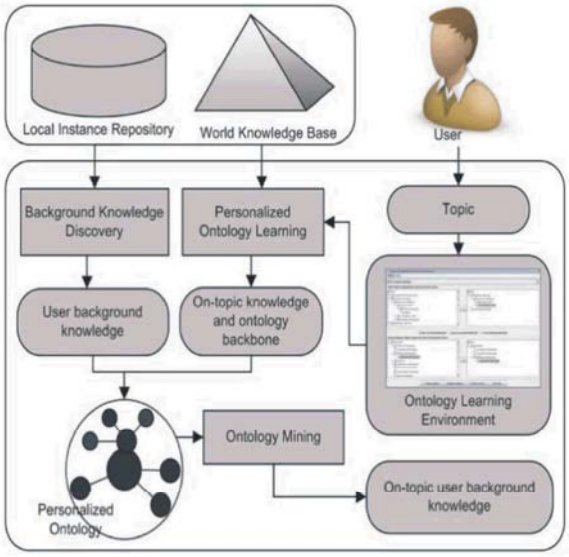

Fig. 2: Architecture of the ontology model

where the white nodes are positive, the dark nodes are negative and the gray nodes are neutral subjects. here, we formalize the ontology constructed for a given topic.

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 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. We argue that

a subject's specificity has two focuses: 1) on the referring-to concepts (called semantic specificity) and 2) on the given topic (called topic specificity). These need to be addressed separately.

Architecture of the Ontology Model: The proposed ontology model aims to discover user background knowledge and learns personalized ontologies to represent user profiles. Fig. 6 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 exhaustivity of subjects are investigated for user background knowledge discovery.

Web Information Gathering System: The information gathering system, IGS, was designed for common use by all experimental models. The IGS was an implementation of a model developed by Li and Zhong that uses user profiles for web information gathering. The input support values associated with the documents in user profiles affected the IGS's performance acutely. Li and Zhong's model was chosen since not only is it better verified than the Rocchio and Dempster- Shafer models, but it is also extensible in using support values of training documents for web information gathering

Ontology Model: This model was the implementation of the proposed ontology model. The input to this model was a topic and the output was the user profiles consisting of positive documents and negative documents. Each document was associated with a support value indicating its support level to the topic [10-14].

The user personalized were constructed as described in user interaction.The authors played the user role to select positive and negative subjects for ontology construction, following the description and narratives associated with the topics. On average, each personalized ontology contained about 16 positive and 23 negative subjects.

Trec Model: The TREC model was used to demonstrate the interviewing user profiles, Which reflected user concept models perfectly. The TREC user profiles perfectly reflected the users personal interests,As the relevant judgment were provided by the same people who created the topics as well, following the fact that only users know their interests and preference perfectly. Hence,the TREC model was the golden model for our proposed model to be measured against. The modeling of a user's concepts model could be proven if our proposed model achieved the same or similar performance to the TREC model.

Catagory Model: This model demonstrated the noninterviewing user profiles, in particular Gauch *et al*. In the model, 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 superclass and subclass in an ontology. When an OBIWAN agent receives the search results for a given topic, it filters and reranks the results based on their semantic similarity with the subjects. The similar documents are awarded and reranked higher on the result list. In this Category model, the sets of positive subjects were manually fed back by the user via the OLE and from the WKB, using the same process as that in the Ontology model.

The Category model differed from the Ontology model in that there were no is-a, part-of and related-to knowledge considered and no ontology mining performed in the model. The positive subjects were equally weighted as one, because there was no evidence to show that a user might prefer some positive subjects more than others. The training sets in this model were extracted through.

Web Model: The web model was the implementation of typical semiinterviewing user profiles. It acquired user profiles from the web by employing a web search engine. For a given topic, a set of feature terms ftjt 2 Tþg and a set of noisy terms ftjt 2 T_g were first manually identified. The feature terms referred to the interesting concepts of the topic. The noisy terms referred to the paradoxical or ambiguous concepts. Also identified were the certainty factors CFðtÞ of the terms that determined their supporting rates ([-1, 1]) to the topic. By using the feature and noisy terms, the Google4 API was employed to perform two searches for the given topic. The first search used a query generated by adding "þ" symbols in front of the feature terms and "-" symbols in front of the noisy terms. By using this query, about 100 URLs were retrieved for the positive training set. The second search used a

query generated by adding "_" symbols in front of feature terms and "þ" symbols in front of noisy terms. Also, about 100 URLs were retrieved for the negative set. These positive and negative documents were filtered by.

CONCLUSIONS

In this paper, an ontology model is proposed for representing user background knowledge for personalized web information gathering. The model constructs user personalized ontologies by extracting world knowledge from the LCSH system and discovering user background knowledge from user local instance repositories ontology mining method, exhaustivity and specificity, is also introduced for user background knowledge discovery. In evaluation, the standard topics and a large tested were used for experiments.

The model was compared system for against benchmark models by applying it to a common information gathering. The experiment results demonstrate that our proposed model is sensitivity analysis was also conducted for the ontology model.

This proposed ontology model in this paper provides the solution to emphasizing global and local knowledge in a single computational model.The findings in this paper can be applied to the design of web information gathering systems.

The model also has extensive contribution stothe field of information Retrivel, Web Intelligence, Recommendation Systems and Information Systems.

REFERENCES

1. Baeza-Yates, R. and B. Ribeiro-Neto, 1999. Modern Information Retrieval. Addison Wesley.
2. Box, G.E.P., J.S. Hunter and W.G. Hunter, 2005. Statistics for Experimenters. John Wiley & Sons.
3. Buckley, C. and E.M. Voorhees, 2000. Evaluating Evaluation Measure Stability, Proc. ACM SIGIR '00, pp: 33-40.
4. Cai, Z., D.S. McNamara, M. Louwerse, X. Hu, M. Rowe and A.C. Grassers, 2004. NLS: A Non-LatentSimilarity Algorithm, Proc. 26th Ann. Meeting of the Cognitive Science Soc. (CogSci '04), pp: 180-185. TAO *et al*.: a Personalized Ontology Model for Web Information Gathering 509 Table 6 T-Test Statistic Results for Sensitivity Test.
5. Chan, L.M., 2005. Library of Congress Subject Headings: Principle and Application. Libraries Unlimited.
6. Chirita, P.A., C.S. Firan and W. Nejdl, 2007. Personalized Query Expansion for the Web, Proc. ACM SIGIR ('07), pp: 7-14.
7. Colomb, R.M., 2002. Information Spaces: The Architecture of Cyberspace. Springer.
8. Doan, A., J. Madhavan, P. Domingos and A. Halevy, 2002. Learning to Map between Ontologies on the Semantic Web, Proc. 11th Int'l Conf. World Wide Web (WWW '02), pp: 662-673.
9. Dou, D., G. Frishkoff, J. Rong, R. Frank, A. Malony and D. Tucker, 2007. Development of Neuroelectro magnetic Ontologies (NEMO): A Framework for Mining Brainwave Ontologies," Proc. ACMSIGKDD ('07), pp: 270-279.
10. Abou-Deif, M.H., M.A. Rashed, M.A.A. Sallam, E.A.H. Mostafa and W.A. Ramadan, 2013. Characterization of Twenty Wheat Varieties by ISSR Markers, Middle-East Journal of Scientific Research, 15(2): 168-175.
11. Kabiru Jinjiri Ringim, 2013. Understanding of Account Holder in Conventional Bank Toward Islamic Banking Products, Middle-East Journal of Scientific Research, 15(2): 176-183.
12. Muhammad Azam, Sallahuddin Hassan and Khairuzzaman, 2013. Corruption, Workers Remittances, Fdi and Economic Growth in Five South and South East Asian Countries. A Panel Data Approach Middle-East Journal of Scientific Research, 15(2): 184-190.
13. Sibghatullah Nasir, 2013. Microfinance in India Contemporary Issues and Challenges. Middle-East Journal of Scientific Research, 15(2): 191-199.
14. Mueen Uddin, Asadullah Shah, Raed Alsaqour and Jamshed Memon, 2013. Measuring Efficiency of Tier Level Data Centers to Implement Green Energy Efficient Data Centers, Middle-East Journal of Scientific Research, 15(2): 200-207.