Web Personalization using Efficient Ontology Relations

Mohd. Sadik Ahamad [1], **S. Naga Raju** [2]
[1] Kakatiya University, Kakatiya Institute of Technology and Science,
Warangal, Andhra Pradesh, India.
[2]Kakatiya University, Kakatiya Institute of Technology and science,
Assoc.Prof. Department of computer science engineering, Warangal, Andhra Pradesh, India.

Abstract— 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 capturing their information needs. For this purpose, user profiles are created for user background knowledge description. 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 user local information. In this project, a personalized ontology model is proposed for knowledge representation and reasoning over user profiles.

Keywords— Ontology, Semantic Relations, Web Mining

1. Introduction

Today, 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, WordNet 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, taxonomical patterns from the users' local text documents to learn ontologies for user profiles. Some groups learned personalized ontologies adaptively from user's browsing history. Alternatively, 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.

In this paper, an ontology model to evaluate this hypothesis is proposed. This model simulates users' concept models by using personalized ontologies, 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 ontologies by adopting user feedback on interesting knowledge. A multidimensional ontology mining method, Specificity and Exhaustivity, is also introduced in the proposed model for analyzing concepts specified in ontologies. The users' LIRs are then used to discover background knowledge and to populate the personalized ontologies. The proposed ontology model is evaluated by comparison against some benchmark models through experiments using a large standard data set.

2. Previous Work

Electronic learning (e-Learning) refers to the application of information and communication technologies (e.g., Internet, multimedia, etc.) to enhance ordinary classroom teaching and learning. With the maturity of the technologies such as the Internet and the decreasing cost of the hardware platforms, more institutions are adopting e-Learning as a supplement to traditional instructional methods. In fact, one of the main advantages of e-Learning technology is that it can facilitate *adaptive learning* such that instructors can dynamically revise and deliver instructional materials in accordance with learners' current progress. In general, adaptive teaching and learning refers to the use of what is known about learners, a priori or through interactions, to alter how a learning experience unfolds, with the aim of improving learners' success and satisfaction. The current state-of the- art of e-Learning technology supports automatic collection of learners' performance data (e.g., via online quiz). [1]

However, few of the existing e-Learning technologies can support automatic analysis of learners' progress in terms of the knowledge structures they have acquired. In this paper, we illustrate a methodology of automatically constructing concept maps to characterize learners' understanding for a particular topic; thereby instructors can conduct adaptive teaching and learning based on the learners' knowledge structures as reflected on the concept maps. In particular, our concept map generation mechanism is underpinned by a context-sensitive text mining method and a fuzzy domain ontology extraction algorithm.

The notion of ontology is becoming very useful in various fields such as intelligent information extraction and retrieval, semantic Web, electronic commerce, and knowledge management. Although there is not a universal consensus on the precise definition of ontology, it is generally accepted that ontology is a formal specification of conceptualization.

Ontology can take the simple form of a taxonomy of concepts (i.e., light weight ontology), or the more comprehensive representation of comprising a taxonomy, as well as the axioms and constraints which characterize some prominent features of the real-world (i.e., heavy weight ontology). Domain ontology is one kind of ontology which is used to represent the knowledge for a particular type of application domain. On the other hand, concept maps are used to elicit and represent the knowledge structure such as concepts and propositions as perceived by individuals. Concept maps are similar to ontology in the sense that both of these tools are used to represent concepts and the semantic relationships among concepts. [1]

However, ontology is a formal knowledge representation method to facilitate human and computer interactions and it can be expressed by using formal semantic markup languages such as RDF and OWL, whereas concept map is an informal tool for humans to specify semantic knowledge structure. Figure shows an example of the owl statements describing one of the fuzzy domain ontologies automatically generated from our system. It should be noted that we use the (rel) attribute of the <rdfs:comment> tag to describe the membership of a fuzzy relation (e.g., the super-class/sub-class relationship). We only focus on the automatic extraction of lightweight domain ontology in this paper. More specificially, the lightweight fuzzy domain ontology is used to generate concept maps to represent learners' knowledge structures.

With the rapid growth of the applications of e-Learning to enhance traditional instructional methods, it is not surprising to find that there are new issues or challenges arising when educational practitioners try to bring information technologies down to their classrooms. The situation is similar to the phenomenon of the rapid growth of the Internet and the World Wide Web (Web). The explosive growth of the Web makes information seekers become increasingly more difficult to find relevant information they really need [1].

This is the so-called problem of information overload. With respect to e-learning, the increasing number of educational resources deployed online and the huge number of messages generated from online interactive learning (e.g., Blogs, emails, chat rooms) also lead to the excessive information load on both the learners and the instructors. For example, to promote reflexive and interactive learning, instructors often encourage their students to use online discussion boards, blogs, or chat rooms to reflect what they have learnt and to share their knowledge with other fellow students during or after normal class time. With the current practice, instructors need to read through all the messages in order to identify the actual progress of their students.

3. Proposed System

A. *Ontology Construction*

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. For each subject, its ancestors are retrieved if the label of contains any one of the query terms in the given topic. From these candidates, the user selects positive subjects for the topic. The user-selected positive subjects are presented in hierarchical form. The candidate negative subjects are the descendants of the user-selected positive subjects. From these negative candidates, the user selects the negative subjects. These positive subjects will not be included in the negative set. The remaining candidates, who are not fed back as either positive or negative from the user, become the neutral subjects to the given topic. Ontology is then constructed for the given topic using these users fed back subjects. The structure of the ontology is based on the semantic relations linking these subjects. The ontology contains three types of knowledge: positive subjects, negative subjects, and neutral subjects.

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.

C. *Topic Specificity*

The topic specificity of a subject is performed, based on the user background knowledge discovered from user local information. 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. The ontology constructed has only subject labels and semantic relations specified. we populate the ontology with the instances generated from user local information collections. We call such a collection the user's local instance repository. The documents may be semistructured (e.g., the browsed HTML and XML web documents). In some semistructured web documents, content-related descriptors are specified in the metadata sections. These descriptors have direct reference to the concepts specified in a global knowledge base. These documents are ideal to generate the instances for ontology population. When different global knowledge bases are used, ontology mapping techniques is used to match the concepts in different representations. The clustering techniques group the documents into unsupervised clusters based on the document features. These features, usually represented by terms, can be extracted from the clusters. The documents can then be classified into the subjects based on their similarity. Ontology mapping techniques can also be used to map the features discovered by using clustering and classification to the subjects, if they are in different representations.

D. Analysis of Subjects

The exhaustivity of a subject refers to the extent of its concept space dealing with a given topic. This space extends if a subject has more positive descendants regarding the topic. In contrast, if a subject has more negative descendants, its exhaustivity decreases. Based on this, we evaluate a subject's exhaustivity by aggregating the semantic specificity of its descendants where Subjects are considered interesting to the user only if their specificity and exhaustivity are positive. A subject may be highly specific but may deal with only a limited semantic extent.

4. RESULTS

The concept of this paper is implemented and different results are shown below, The proposed paper is implemented in Java technology on a Pentium-IV PC with 20 GB hard-disk and 256 MB RAM with apache web server. The propose paper's concepts shows efficient results and has been efficiently tested on different Datasets. The Fig 1, Fig 2, Fig 3 and Fig 4 shows the real time results compared.

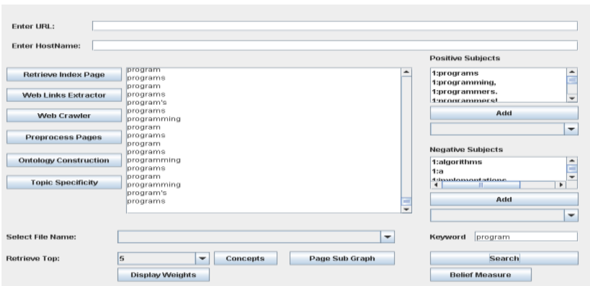

Fig. 1 Computation of Positive and Negative Subjects.

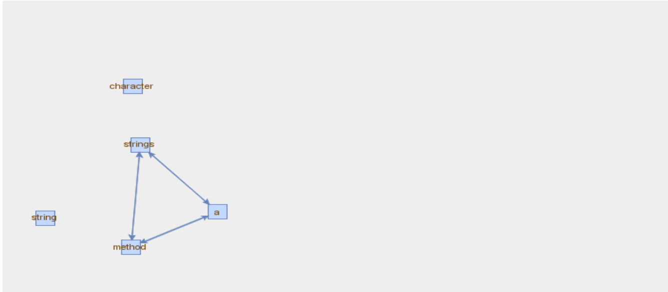

Fig. 1 Displaying Constructed Ontology

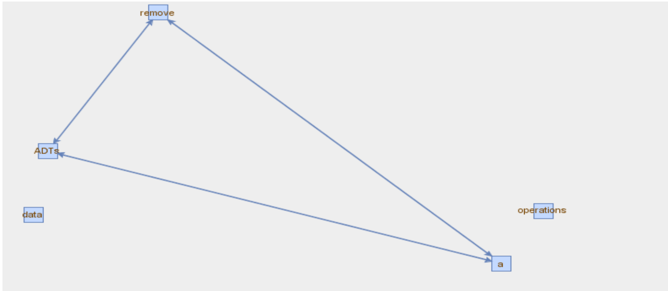

Fig. 3 Displaying Constructed Ontology

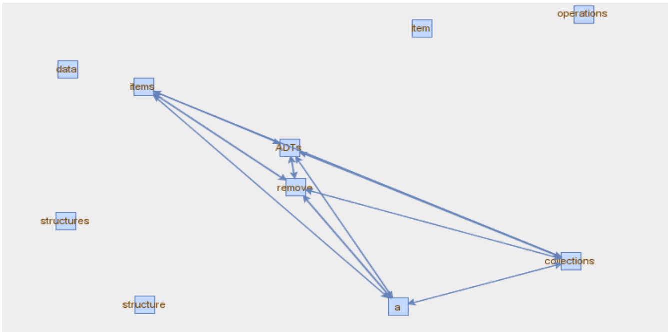

Fig. 4 Displaying Constructed Ontology

5. SYNONYM HANDLING

When we handle the retrieved ontology keywords we would drag the semantic relationships between the instance sets of subject headings. Although we get the results related to user knowledge but there may be a chance of losing data because of the synonyms. Sometimes synonyms of keywords may give us the results better that user expected.

For this reason synonyms of keywords retrieved and maintained later by using all these words we form the instance sets and retrieve more subject headings from LCSH and add to LIR.

We can derive the probability of the results before synonym handling case and after synonym handling case. For example if we got **M** words without synonym case, and the probability is **P1.** For the synonym case if we got **N** words, and calculated probability is **P2**. If we compare these two probabilities, definitely

$$P1 < P2 \qquad (M<N)$$

Finding the right direction for searching ontology related words is difficult. Ontology is vast and there could be many directions the user may not able to find the relevant results of interest according to him. The problem becomes bigger if we consider the synonyms of the words. To find the more related suitable synonyms and words we find the probability of result set for each synonym and compare them with the existing results. We consider only the synonyms that gives more results according to user and the user interest. With this approach we can refine the search.

If we could drag the relationship between the resultset and the no of synonym words added, through probability then we can predict the results for other cases. This helps to formulate the analysis. By taking some small cases we can do that and it helps to solve and predict complex cases.

6. CONCLUSION

In this project, 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.

A multidimensional ontology mining method, exhaustivity and specificity, is also introduced for user background knowledge discovery. In evaluation, the standard topics and a large testbed 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. Sensitivity analysis was also conducted for the ontology model. In this investigation, we found 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 in this project provides a solution to emphasizing global and local knowledge in a single computational model. The findings in this project 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. Synonyms will give us more directions to choose user interests. It refines the search.

REFERENCES

[1] R.Y.K. Lau, D. Song, Y. Li, C.H. Cheung, and J.X. Hao, “Towards a Fuzzy Domain Ontology Extraction Method for Adaptive e- Learning,” IEEE Trans. Knowledge and Data Eng., vol. 21, no. 6, pp. 800-813, June 2009.

[2] 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.

[3] C. Makris, Y. Panagis, E. Sakkopoulos, and A. Tsakalidis, “Category Ranking for Personalized Search,” Data and Knowledge Eng., vol. 60, no. 1, pp. 109-125, 2007.

[4] 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.

[5] 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.

[6] D.N. Milne, I.H. Witten, and D.M. Nichols, “A Knowledge-Based Search Engine Powered by Wikipedia,” Proc. 16th ACM Conf. Information and Knowledge Management (CIKM ’07), pp. 445-454, 2007.

[7] 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.

[8] M.D. Smucker, J. Allan, and B. Carterette, “A Comparison of Statistical Significance Tests for Information Retrieval Evaluation,” Proc. 16th ACM Conf. Information and Knowledge Management (CIKM ’07), pp. 623-632, 2007.

[9] 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.

[10] 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.

AUTHORS

Mohd. Sadik Ahamad received the B.E. degree in computer science Engineering from Muffakham jha College of Engineering and Technology, Osmania University in 2008 and M.Tech. Software engineering from kakatiya institute of technology and science, affiliated to Kakatiya University in 2012. During 2009-2010, he worked as an Asst. Prof in Balaji institute of technology and science.

S. Naga Raju Assoc. Prof. of Department of computer science engineering, kakatiya institute of technology and science, Warangal, researcher in web mining, member of ISTE, B.Tech. from Amaravati University, M.Tech from JNTUH.