RESEARCH ARTICLE OPEN ACCESS

Area Information Ontology for web page

Mohammad Mateen[1], Asra Begum[2]
[1,2](Computer Science and Engineering, Lords Institute of Engineering and Technology, Hyderabad)

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.***

I. INTRODUCTION

Today, Global analysis uses existing global knowledge bases for user background knowledge representation. Commonly used knowledge bases include generic ontologies 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, Word 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 behaviour 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, analysed 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 common sense 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 exhaustively, is also introduced in the proposed model for analysing 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.

I. PREVOIUS 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 ozntology 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 specifically, 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.

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.

II. 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.

In this module the first step is to Collect the terms, In order to collect the terms, we will collect the Web log file from the Web server of the website for a period of time, run a pre-processing unit to analyze the Web log file and produce a list of URLs of Web-pages that were accessed by users, run a software agent to crawl all the Web-pages in the URL list to extract the titles, and apply an algorithm

to extract terms from the retrieved titles, i.e., single tokens are extracted first by removing stop words from the titles, some single tokens are then combined into composite terms if these single terms often occur at the same time and there is never any token appears between these tokens, and the remaining single tokens will become single word terms. It is possible for some extracted terms to share the same features, so it is better for them to be instances of a concept, rather than standalone concepts. In this step, the domain concepts will be defined for the given website based on the extracted terms.

$$O_{man} := < \bar{T}_{man}, D, A, \dot{A} >$$

There are three possible approaches to develop the taxonomic relationships, such as, a top-down development process starts from the most general concepts in the domain and then identifies the subsequent specialization of the general concepts, a bottom-up development process starts from the most specific concepts as the leave nodes in the concept hierarchical structure/tree structure, then groups these most specific concepts into more general concepts, a hybrid development process is the combination of the top down and bottom-up approaches. We identify the core concepts in the domain first and then generalize and specialize them appropriately.

B. Term Net WP Construction

In this module in order to construct Term Net WP, we apply the procedure consisting of the following steps: Collect the titles of visited Web-pages In order to collect the titles, we will collect the Web log file from the Web server of the website for a period of time (at least seven days), run a pre-processing unit to analyse the Web log file and produce a list of URLs of Web pages that were accessed by users, and run a software agent to crawl all the Web-pages in the list to extract the titles.

$$T_{auto} = \{t_i : 1 < i < p\}$$

$$D = \{d_j : 1 \le j \le q\}$$

$$X_j = t_1 t_2 \ldots t_n, t_k \in T_{auto}$$

Extract term sequences from the Web-page titles we apply the algorithm used in the domain ontology construction to extract the terms from the retrieved titles. The extracted terms are organized in the order as they appear in each title, namely they are collected as term sequences. Build the semantic network Term Net WP in Term Net WP, each node represents a term in the extracted term sequences and the order of the terms in sequences determines the 'from-Instance' and 'to-Instance' relations of a term between other terms. By scanning all the term sequences extracted from the previous step, we can build the Term Net WP.

C. Conceptual Prediction Model (CPM)

In order to obtain the semantic Web usage knowledge that is efficient for semantic-enhanced Web-page recommendation, a conceptual prediction model (CPM) is proposed to automatically generate a weighted semantic network of frequently viewed terms with the weight being the probability of the transition between two adjacent terms based on FVTP.

$$\rho_{S,x} = \frac{\partial_{S,x}}{\sum_{y=1}^{N} \partial_{S,y}}$$

According to the model, a kind of model efficient to represent a collection of navigation records, CPM is developed as a self-contained and compact model. It has two main kinds of elements: state nodes, and the relations between the nodes. One node presents the current state, e.g. current viewed term, and may have some previous state nodes and

some next state nodes. By scanning each term pattern F ∈ F, each term becomes a state in the model. There are also two additional states: a start state, S, representing the first state of every term pattern; and a final state, E, representing the last state of every term pattern. There is a transition corresponding to each pair of terms in a pattern, a transition from the start state S to the first term of a term pattern, and a transition from the last term of a term pattern to the final state E. The model is incrementally built by processing the complete collection of FVTP.

D. Semantic Enhanced Recommendation

In this module four recommendation strategies, that apply the semantic knowledge base of a given website, which includes the domain ontology of Web-pages (Domain Onto WP) or the semantic network of Web-pages (Term Net WP) and the weighted semantic network of frequently viewed terms of Web-pages within the given website (Term Nav Net), to make Web-page recommendations. These recommendations are referred to as semantic enhanced Web-page recommendations. For a given current Web-page or a combination of the current and previous Web-pages, the next Web-pages could be recommended differently depending on which knowledge representation model and the order of CPM are used. The following steps are used: builds Domain On to WP; generates FWAP using PLWAP-Mine; builds FVTP; builds a 1st-TermNavNet given FVTP; identifies a set of currently viewed terms using query Topic man on Domain On to WP; infers next viewed terms given each term using query on the Term Nav Net; recommends pages mapped to each term using query on Domain On to WP.

III. RESULT

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.

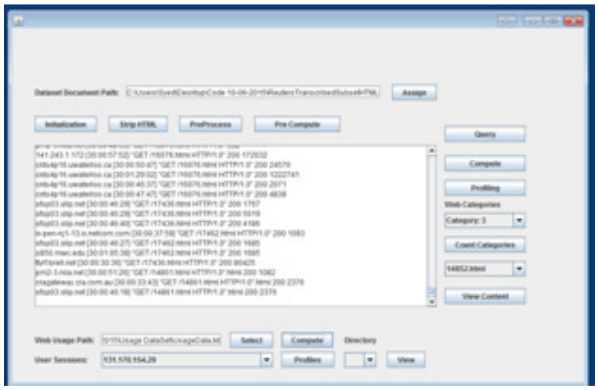

Fig. 1 Ontology Generation

Fig 2: Recommendations

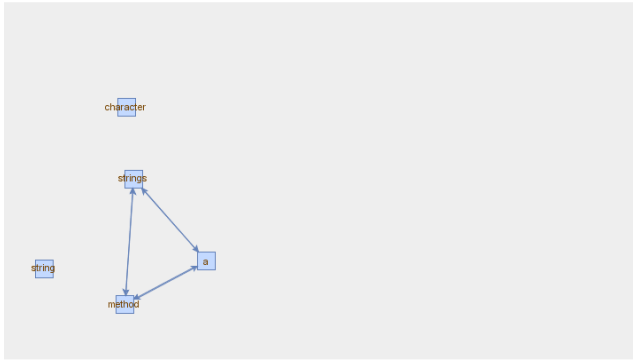

Fig. 3: Displaying Constructed Ontology

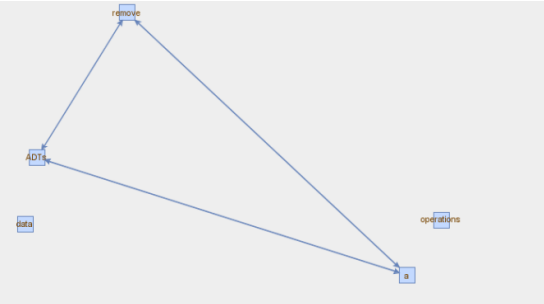

Fig. 4: Displaying Constructed Ontology

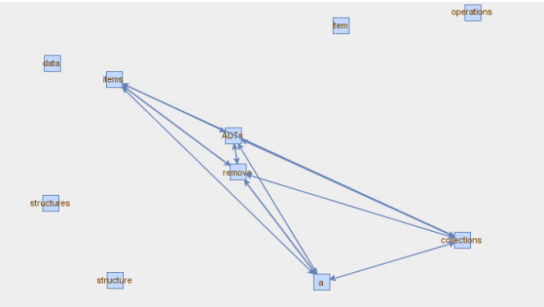

Fig. 5: Displaying Constructed Ontology

IV. CONCLUSIONS

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, exhaustively 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 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, 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.

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.