Learning and Inferencing in User Ontology for Personalized Semantic Web Services

Xing Jiang and Ah-Hwee Tan

School of Computer Engineering
Nanyang Technological University
{jian0008,asahtan}@ntu.edu.sg

ABSTRACT

Domain ontology has been used in many Semantic Web applications. However, few applications explore the use of ontology for personalized services. This paper proposes an ontology based user model consisting of both concepts and semantic relations to represent users' interests. Specifically, we adopt a statistical approach to learning a semantic-based user ontology model from domain ontology and a spreading activation procedure for inferencing in the user ontology model. We apply the methods of learning and exploiting user ontology to a semantic search engine for finding academic publications. Our experimental results support the efficacy of user ontology and spreading activation theory (SAT) for providing personalized semantic services.

1. INTRODUCTION

In the Semantic Web, domain ontology is commonly used to describe web resources. Containing semantics in the form of concepts, relations and axioms, domain ontology enables software agents to perform more sophisticated tasks automatically. Specifically, many applications have been developed for information retrieval. For instance, Guha et al. [2] used ontology to improve traditional web search by augmenting search results with related concepts in the ontology.

Although there have been many applications of domain ontology, relatively few are concerned with providing personalized information services. In this paper, we propose using an ontology based user model for representing a personalized view of the target domain to capture a user's interests and a set of statistical methods for learning the user ontology. We further incorporate the proposed user ontology model and the SAT [1] based inferencing procedure into a semantic search engine for searching academic publications.

2. USER ONTOLOGY MODEL

Considering the sample domain ontology given in Figure 1, that represents a basic conceptualization of the Italian soccer teams. We see that "AC Milan" and "Inter Milan" are Italian soccer teams belonging to different leagues. But this domain ontology may be too general for individual's interests. For instance, I can be a big fan of the AC Milan team. Therefore, the concept "AC Milan" is more important to me than the concept "Inter Milan". Meanwhile, joining Champion League is more important to me than joining the Serie A League. The existing user modelling methods only consider the importance of the concepts for capturing user's interests. A user ontology, on the other hand, can capture all necessary semantics from a domain ontology for user modelling. Specifically, each concept and relation in the domain ontology will be given certain values for indicating user's interests. It is a personalized view of the conceptualization and is more comprehensive than the existing types of user models. An illustration of the user ontology is given in Figure 2, in which concepts and relations have been given specific values to indicate their relevance to a user.

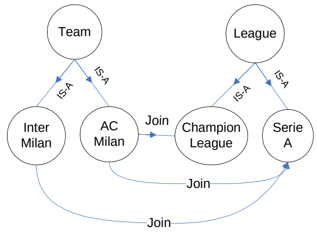

Figure 1: A partial domain ontology for the Italian soccer teams.

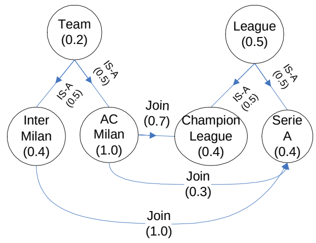

Figure 2: An illustration of the user ontology.

A user ontology can be defined formally as a structure $\Theta = (C, R, \theta, \mathcal{C}, \mathcal{R})$ consisting of

- two disjoint sets C and R, whose elements c_x and r_{xy} are the *concepts* and *relations* in the domain ontology,
- a function $\theta : \theta(C|R)$, which assigns weights to concepts and relations in the domain ontology, representing an individual's view of the particular domain,
- a vector $\mathcal{C} = [\mathcal{C}_1, \ldots, \mathcal{C}_n]$, in which $\mathcal{C}_x$ represents a user's interests to concept c_x, and
- a matrix $\mathcal{R} = [\mathcal{R}_{xy}]$, in which $\mathcal{R}_{xy}$ represents a user's interests to relation r_{xy} and $\sum_y \mathcal{R}_{xy} = 1$.

WWW2006, May 22–26, 2006, Edinburgh, UK.
.

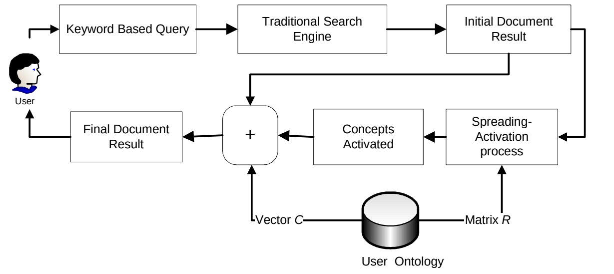

Figure 3: The procedure for exploiting user ontology in document retrieval.

3. LEARNING USER ONTOLOGY

3.1 Learning Concepts of Interests

Estimating the interest factor C_x of a user on a concept is relatively straightforward. For instance, we can record the concepts of interests to the user and their frequencies when a user searches information in the web. Meanwhile, we use a decay function [1], given by $C_x(t_{i+1}) = C_x(t_i) \times \delta^{-b}$, to prevent saturation of the interest factor C_x in the user ontology.

3.2 Learning Relations of Interests

Learning relations of interests to a user is similar to learning concepts of interests. Initially, an estimated value R^0_{xy} is assigned to each relation r_{xy}. Then, an empirical value is computed for each relation by analyzing the historical record. We used a Bayesian solution to compute a weighted average of the initial value and the empirical value as follows:

$$\mathcal{R}_{xy} = \frac{a \times \mathcal{R}^0_{xy} + F(r_{xy})}{a + \sum_y F(r_{xy})}, \qquad (1)$$

where a is a constant to normalize the empirical value and the initial estimation, and $F(r_{xy})$ is the frequency of the relation r_{xy} obtained from the user's historical record.

4. EXPLOITING USER ONTOLOGY

We present a procedure (Figure 3) wherein a user ontology is used to re-rank the search results of a search engine below.

Similar to that of a traditional search engine, a user submits a query consisting of keywords to the system. The search engine then returns an initial list of documents obtained using the classical keyword based search method. With the documents pre-annotated with concepts, we can obtain a set of associated concepts besides the documents retrieved. These concepts together with their occurrence frequencies form a vector $I = [I_1, I_2, \ldots, I_n]^T$ as the input for inferencing in the user ontology, where I_x, the input to the concept c_x, is calculated by $I_x = \frac{F(c_x)}{\sum_{c_x} F(c_x)}$, where $F(c_x)$ represents the frequency of the concept c_x in the initial document list.

Upon receiving the input vector I, the spreading activation process is performed on the user ontology to infer the concepts of relevance. Using simplified SAT in which the output of a concept c_y at time t_i is the input of the concept c_y at time t_i, $O_{c_y}(t_i) = I_{c_y}(t_i)$, the spreading activation process can be expressed using the following formula:

$$O = \left[\mathcal{E} - (1-\alpha)\mathcal{R}^T\right]^{-1} I, \qquad (2)$$

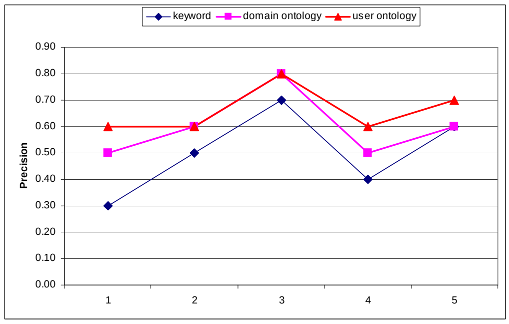

Figure 4: Average precision of the semantic search engine with and without the use of user ontology in document retrieval compared with keyword based method.

where $\mathcal{R}$ is the relation matrix of the user ontology, α is the decay factor, $\mathcal{E}$ is an $n \times n$ identity matrix, and $O = [O_1, \ldots, O_n]^T$ is the final output vector of the spreading-activation process in which O_x is the value of concept c_x obtained from the spreading-activation process.

Next, the relevance factor O_x is combined with the user's long term interest factor $\mathcal{C}_x$ to derive a final score S_x for the concept c_x. The score strikes a balance between long time interest and current relevance. In our application, the score S_x is computed by $S_x = O_x + \mathcal{C}_x \times \delta^{-b}$, where δ represents the time interval since the last query and b is a real-valued constant to simulate the decay function.

Finally, documents with high rankings in the initial list and annotated with concepts with high S score values are moved towards the top of the list for presentation to the user.

5. EXPERIMENT

A semantic search engine that incorporates user ontology and SAT has been developed for searching academic publication in a database. All documents collected are annotated using the ACM Computing Classification System, which also serves as the domain ontology.

5 users are involved in evaluating the user ontology's ability for providing personalized services. Each user provides two sets of queries, one for training the model and the other for testing. We experiment with the semantic search engine, first using the traditional keyword based method, then augmented with domain ontology, and finally enhanced with user ontology to provide recommendation for the test queries. The performance of the search engine, in terms of the average precision of the top 10 documents retrieved, is summarized in Figure 4. We see that the user ontology based system consistently outperforms or produces equivalent performance compared with the two methods, validating our approach of using user ontology as user models in the Semantic Web.

6. REFERENCES

[1] ANDERSON, R. J. A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior 22* (1983), 261–295.

[2] GUHA, R., MCCOOL, R., AND MILLER, E. Semantic search. In *WWW '03*, ACM Press, pp. 700–709.