Available at: www.researchpublications.org

# Mining Ontological User Profiles: A Review Study

Prof. Mrs.V.M.Deshmukh

Associate Professor, Department of CSE, PRMIT &R Badnera(M.S.) India
msvmdeshmukh@rediffmail.com

Mr.G.S.Deokate

Final Year Master of Engineering, PRMIT & R, Badnera (M.S) India
deokate_gajanan@rediffmail.com

***Abstract-*** **To simulate user concept models, ontologies—a knowledge description and formalization model—are utilized in personalized web information gathering. Such ontologies are called ontological user profiles. In this paper we study ontological user profiles construction and mining. World knowledge base and local instance repositories are both are used in this model. Specificity and exhaustivity methods are used in mining such ontological user profiles.**

***Keywords***— Ontology user profiles, personalization, semantic relations, world knowledge base, local instance repository, user profiles, web information gathering.

## I. Introduction

There is huge information available on the internet. The web information gathering systems before this satisfy the user requirements by capturing their information needs. For this reason user profiles are created for user background knowledge description. The user profiles represent the concepts models possessed by user while gathering the web information. A concept model is generated from user background knowledge and possessed implicitly by user. If the user concept model can be simulated then we can build a better user profile. To simulate user's concepts model, ontologies are utilized in personalized web information gathering which are called ontological user profiles or personalized ontologies [1], [2], [3].

In this paper we study mining of ontological user profiles. The specificity and exhaustivity methods are used in ontology mining.

## II. Ontological user profiles

Web users have different expectation from same search query. For example, for the topic "New York", business travellers may have demand for different information from leisure travellers. Same user may have different expectation from same query if applied in the different situation. A user may become a business traveller when planning for a business trip, or a leisure traveller when planning for a family holiday. From this observation an assumption is formed that web users have a personal concepts model for their information needs, a user's concept model may change according to different information needs.

### *A. World Knowledge Base*

The World knowledge base is very important for information gathering. World knowledge is commonsense knowledge possessed by people and acquired by through the experience and education. The Library of congress subject Heading (LCSH) is ideal for world knowledge base. The LCSH system is a thesaurus developed for organizing and retrieving information from a large volume of library collections. LCSH has undergone continuous revising and enriching. The LCSH system is better than other world knowledge taxonomies such as Library of Congress Classification (LCC) used by Frank and Paynter [4], the Dewey Decimal Classification (DDC) used by Wang and Lee [5], and the reference categorization (RC) developed by Gauch et al. [1] using online categorizations. LCSH has more topics, more specific structure and more semantic relations. For over a hundred years, the knowledge contain in LCSH has undergone continuous revision and enrichment. The LCSH is more comprehensive nonspecialized controlled vocabulary in English.

### *B. Personalized Ontology Construction*

The subjects which are interested are extracted from world knowledge base. Ontology Learning Environment (OLE) tool is used to extract interested subject for user. For a given topic, interested subjects consists of two sets; positive and negative subjects. The positive subjects consist of relevant information and negative subjects resolve ambiguous interpretation of information need. The OLE provides users with a set of candidates to identity positive and negative subjects. The ontology contains three types of subject candidates: positive, negative and neutral. The candidates which are not feedback as positive or negative are treated as neutral subjects. These candidate subjects are extracted from the world knowledge base. User selects positive and negative subject for their interests and preferences hence constructed ontology is personalized. [6], [7]

## III. Ontology Mining

Ontology mining is used to discover interesting topic from semantic relations, concepts and instances. Two methods are used which are specificity and exhaustivity. Subject's focus on a given topic is described by Specificity and subject's semantic space dealing with the topic is restricted by exhaustivity. Using this method we can investigate subject and the strength of their association in ontology. The subject's specificity has two focuses which are semantic specificity and topic specificity.

### *A. Semantic Specificity*

It is investigated based on the structure of $O(\tau)$ inherited from the world knowledge base. The lower bound subjects have a stronger focus because it has fewer concepts in its space. Hence, the semantic specificity of a lower bound subjects is greater than that of an upper bound subjects.

It is measured based on the hierarchical semantic relations (is-a and part-of) held by a subjects and its neighbours. The subjects have a fixed locality on the $tax^s$ of $O(\tau)$.It is also called as absolute specificity and denoted by $spe_a(s)$. The determination of a subject's $spe_a(s)$ described in algorithm1. [7]

**input**:a personalized ontology $O(\tau) :=$ $< tax^s, rel >$; a coefficient θ between (0,1),

**output**: $spe_a(s)$applied to specificity.

**1** Set k=1, get the set of leaves $S_0$ from $tax^s$, for $(s_0 \in S_0)$ assign $spe_a(s_0) = k$;

**2** get $S'$ which is the set of leaves in case we remove the nodes $S_0$ and the related edges from $tax^s$ ;

**3** **if** $(S' == \emptyset)$ **then** return;

**4** **foreach** $s' \in S'$ do

**5** **if** $isA(s') == \emptyset)$ then $spe_a^1(s') = k$;

**6** **else** $spe_a^1(s') = \theta \times \min\{spe_a(s) | s \in isA(s')\}$;

7 $\boldsymbol{if}(partOf(s') == \emptyset) then\ spe_a^2(s') = k$;

**8** $\boldsymbol{else}\ spe_a^2(s') = \frac{\sum_{s \in partOf(s')} spe_a(s)}{|partOf(s')|}$ ;

**9** $spe_a(s') = \min(spe_a^1(s'), spe_a^2(s'))$ ;

**10 end**

**11** $k = k \times \theta, S_0 \cup S'$,go to step2;

**Algorithm 1. Analyzing semantic relations for specificity[7]**

As the $tax^s$ of $O(\tau)$ is a graphic taxonomy, the leaf subjects have no descendants. Thus, they have the strongest focus on their referring- to concepts and highest $spe_a(s)$. The leaf subjects have the strongest $spe_a(s)$ of 1 in the range of 0 to 1.The root subjects have the weakest $spe_a(s)$ and smallest value in (0, 1). [3], [7]

### B. *Topic Specificity*

Topic specificity measures the focus of subjects on the given topic. It is investigated based on the user background knowledge discovered from user local information. User background knowledge can be discovered from user local information collections, such as user's stored documents, browsed web pages, and composed/received emails. Such collections is called Local instance Repository. The reference strength between an instance and a subject is evaluated. The subjects cited by an instance are indexed by their focus. Many subjects cited by an instance may mean loose specificity of subjects, because each subject deals with only a part of the instance. Hence, denoting an instance by *i*, the strength of *i* to a subject *s* is determined by

$$str(i, s) = \frac{1}{priority(s,i) \times n(i)}$$

Where $n(i)$ the number of subjects on the citing list of $i$ and $priority(s, i)$ is the index of *s* on the citing list. The $str(i, s)$ aims to selects the right instances to populated $O(\tau)$.

With the $str(i, s)$determined, the relationship between an *LIR* and $O(\tau)$ can be defined. [7], [8]

### C. *Evaluating Topic Specificity*

Set of positive, negative and neutral subjects is presents in an $O(\tau)$. Depending on the mapping of subject and instances, if an instances refers only to positive subjects, the instances fully supports the $\tau$ and in case of negative subjects, it is strongly against the $\tau$. [7]

### D. *Multidimensional Analysis of Subjects*

The exhaustivity of a subject is the extent of its concepts space dealing with a given topic. If subjects has more positive descendants this space extends. Otherwise its exhaustivity decreases. [3].

## IV. ONTOLOGY MODEL

The architecture of ontology model is shown in fig.1 Two knowledge resources, world knowledge base and local instance repository is are utilized by the model. Taxonomic structure is provided by world knowledge base whereas the background knowledge is discovered by user local instance repository.

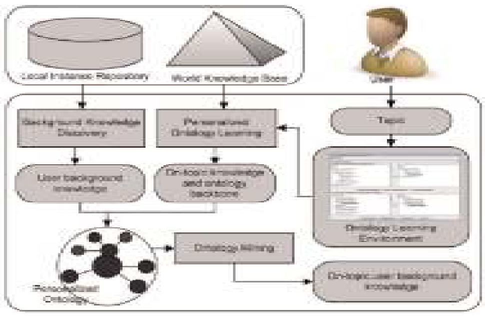

Fig.1. Architecture of Ontology Model [7]

Specificity and Exhaustivity of subjects are investigated for user background knowledge discovery. [7]

## V. CONCLUSION

In this paper, we study construction of ontological user profile and two dimensional ontology mining method.The model construct user personalized ontologies by extracting world knowledge from the LCSH system and discovering user background knowledge from user local instance repositories. Ontology mining discovers interesting and on-topic knowledge from the concepts, semantic relations, and instances in ontology

## REFERENCES

[1] S.Gauch, J. Chaffee, and A. Pretschner, "Ontology-Based Personalized Search and Browsing" Web Intelligence and Agent Systems, vol. 1, nos. 3/4, pp. 219- 234, 2003.

[2] Y. Li and N. Zhong, "Web Mining Model and Its Applications for information Gathering" Knowledge-Based Systems, vol. 17, pp. 207-217, 2004.

[3] Y. Li and N. Zhong, "Mining Ontology for Automatically Acquiring Web User Information Needs," IEEE Trans. Knowledge and Data Eng., vol. 18, no. 4, pp.554-568, Apr. 2006.

[4] E. Frank and G.W. Paynter, “Predicting Library of Congress Classification from Library of Congress Subject Headings,” J. Am. Soc. Information Science and Technology, vol. 55, no. 3, pp. 214-227, vol. 55, no. 3, pp. 214-227, 2004

[5] J. Wang and M.C. Lee, “Reconstructing DDC for Interactive Classification,” Proc. 16th ACM Conf. Information and Knowledge Management (CIKM ’07), pp. 137-146, 2007.

[6] 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.

[7] Xiaohui Tao,Yuefeng Li, and Ning Zhong.”A Personalized Ontology Model for Web Information Gathering” IEEE transaction on knowledge and data engineering,vol-23,n0-4,,pp-496-509,2011

[8] X. Tao, Y. Li, N. Zhong, and R. Nayak, “Ontology Mining for Personalized Web Information Gathering,” Proc. IEEE/WIC/ACM Int’l Conf. Web Intelligence, pp. 351-358, 2007