International Journal of Computer Science and Informatics

Volume 3 | Issue 3 Article 10

January 2014

ONTOLOGY TECHNIQUES FOR WEB DATA

A.ESWARA RAO
St.Martins Engg College,Secunderabad,India., ealapata@gmail.com

G. V. S. SOWMYA
PRRM Engg College,Hyderabad,India., sowmyagvs@gmail.com

P.N. SAILAKSHMI
Pragna Bharath Institute Of Technology,Hyderabad,India, sailakshmi1218@gmail.com

Recommended Citation

RAO, A.ESWARA; SOWMYA, G. V. S.; and SAILAKSHMI, P.N. (2014) "ONTOLOGY TECHNIQUES FOR WEB DATA," *International Journal of Computer Science and Informatics*: Vol. 3 : Iss. 3 , Article 10.
DOI: 10.47893/IJCSI.2014.1145
Available at: https://www.interscience.in/ijcsi/vol3/iss3/10

ONTOLOGY TECHNIQUES FOR WEB DATA

A.ESWARA RAO[1], G. V. S. SOWMYA[2] & P.N.SAILAKSHMI[3]

[1]St.Martins Engg College,Secunderabad,India.
[2]PRRM Engg College,Hyderabad,India.
[3]Pragna Bharath Institute Of Technology,Hyderabad,India
Email:ealapata@gmail.com[1]sowmyagvs@gmail.com[2], sailakshmi1218@gmail.com[3]

Abstract— This paper presents techniques for knowledge description and formalization, ontologies are used to represent user profiles in personalized web data. However, when representing user profiles, many models or techniques 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.

Index Terms— *Ontology, personalization, world knowledge, local instance repository, user profiles, web information gathering*

I. INTRODUCTION

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 [4] [8].

User profiles represent the concept models possessed by users when gathering web information. A concept model is implicitly possessed by users and is generated from their background knowledge. While this concept model cannot be proven in laboratories, many web ontologists have observed it in user behavior [8]. When users read through a document, they can easily determine whether or not it is of their interest or relevance to them, a judgment that arises from their implicit concept models. If a user's concept model can be simulated, then a superior representation of user profiles can be built.

To simulate user concept models, ontologies—a knowl-edge description and formalization model—are utilized in personalized web information gathering. Such ontologies are called ontological user profiles [4] or personalized ontologies [12]. To represent user profiles, many researchers have attempted to discover user background knowledge through global or local analysis.

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, Li and Zhong [12] discovered taxonomical patterns from the users' local text documents to learn ontologies for user profiles. Some groups [4] learned personalized ontologies adaptively from user's browsing history. Alternatively, Sekine and Suzuki [11] analyzed query logs to discover user background knowledge. In some works, such as [10], 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 knowl-edge. 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 knowl-edge 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.

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.

The rest of the paper is organized as follows. In Section II, we describe the model of the system. We represent multidimensional ontology mining method and proposed model in Section III. Section IV analyzes and evaluates the performance of the proposed system. Experimental results are demonstrated in Section V, and Section VI concludes the paper.

II. SYSTEM MODELING

Personalized ontologies are a conceptualization model that formally describes and specifies user background knowl-edge. 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. In this section, a model constructing personalized ontologies for web users' concept models is introduced.

A. World Knowledge 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. In this proposed model, user background knowledge is extracted from a world knowledge base encoded from the Library of Congress Subject Headings (LCSH).

We first need to construct the world knowledge base. The world knowledge base must cover an exhaustive range of topics, since users may come from different backgrounds. For this reason, the LCSH system is an ideal world knowledge base. The LCSH was developed for organizing and retrieving information from a large volume of library collections. For over a hundred years, the knowledge contained in the LCSH has undergone continuous revision and enrichment. The LCSH represents the natural growth and distribution of human intellectual work, and covers comprehensive and exhaustive topics of world knowledge [5]. In addition, the LCSH is the most comprehensive non specialized controlled vocabulary in English. In many respects, the system has become a de facto standard for subject cataloging and indexing, and is used as a means for enhancing subject access to knowledge management systems [5].

The LCSH system is superior compared with other world knowledge taxonomies used in previous works. Table 1 presents a comparison of the LCSH with the Library of Congress Classification (LCC) used by Frank and Paynter [11], the Dewey Decimal Classification (DDC) used by Wang and Lee and King et al. and the reference categorization (RC) developed by Gauch et al. [4] using online categorizations. As shown in Table 1, the LCSH covers more topics, has a more specific structure, and specifies more semantic relations. The LCSH descriptors are classified by professionals, and the classification quality is guaranteed by well-defined and continuously refined cataloging rules [5]. These features make the LCSH an ideal world knowledge base for knowledge engineering and management.

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 A is used for an action, A becomes 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 in the world knowledge base. 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.

The primitive knowledge unit in our world knowledge base is subjects. They are encoded from the subject headings in the LCSH. These subjects are formalized as follows:

Definition 1. Let $ be a set of subjects, an element s € $ is formalized as a 4-tuple s: = {label; neighbor; ancestor, descendant}, where
. label is the heading of s in the LCSH thesaurus;
. neighbor is a function returning the subjects that have direct links to s in the world knowledge base;
. ancestor is a function returning the subjects that have a higher level of abstraction than s and link to s directly or indirectly in the world knowledge base;
. descendant is a function returning the subjects that are more specific than s and link to s directly or indirectly in the world knowledge base.

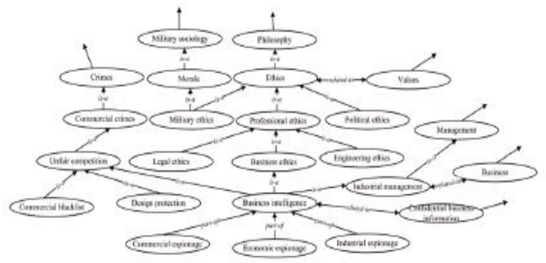

Fig. 1. A sample part of the world knowledge base

The subjects in the world knowledge base are linked to each other by the semantic relations of is-a, part-of, and related-to. The relations are formalized as follows:

Definition 2. Let IR be a set of relations, an element r € IR is a 2-tuple r :={ edge, type}, where
. an edge connects two subjects that hold a type of relation;
. a type of relations is an element of { is-a, part-of, related-to}.
With Definitions 1 and 2, the world knowledge base can then be formalized as follows:

Definition 3. Let WKB be a world knowledge base, which is a taxonomy constructed as a directed acyclic graph. The WKB consists of a set of subjects linked by their semantic relations, and can be formally defined as a 2-tuple WKB: = {$; IR}, where
. $ is a set of subjects $:={ s1, s2. . . sm} ;
. IR is a set of semantic relations IR :={ r1, r2. . . rn} linking the subjects in $.

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. These candidate subjects are extracted from the WKB.

Fig. 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 € $, the s and its ancestors are
retrieved if the

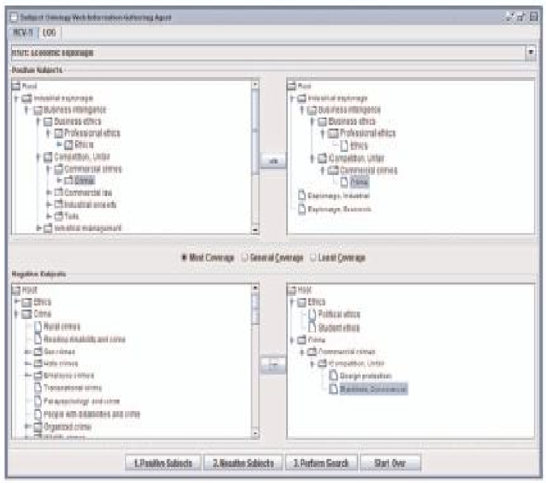

Fig. 2. Ontology learning environment

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, 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 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,” where the white nodes are positive, the dark nodes are negative, and the gray nodes are neutral subjects.

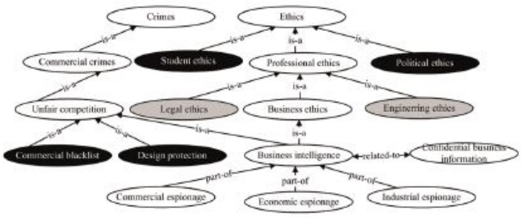

Fig. 3. An ontology (partial) constructed for topic “Economic Espionage.”

C. PROPOSED SYSTEM

In describing the proposed system we have used multi dimensional ontology mining methodologies.

MULTIDIMENSIONAL ONTOLOGY MINING

In this section, a 2D ontology mining method is introduced: Specificity and Exhaustivity. Specificity (denoted spe) describes a subject’s focus on a given topic. Exhaustivity (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.

A. Semantic Specificity

The semantic specificity is investigated based on the structure of O (T) inherited from the world knowledge base. The strength of such a focus is influenced by the subject’s locality in the taxonomic structure tax^S of O (T). As stated in above definitions, the tax^S of O (T) is a graph linked by semantic relations. The subjects located at upper bound levels toward the root are more abstract than those at lower bound levels toward the “leaves.” The upper bound 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 an upper bound and lower bound subjects, the lower bound subject has a stronger focus because it has fewer concepts in its space. Hence, the semantic specificity of a lower bound subject is greater than that of an upper bound subject.

The semantic specificity is measured based on the hierarchical semantic relations (is-a and part-of) held by a subject and its neighbors in tax^S. Because subjects have a fixed locality on the tax^S of O (T), semantic specificity is also called absolute specificity and denoted by $spe_a(s)$.

The determination of a subject’s spe_a is described in Algorithm 1. The isA(s’) and part of(s’) are two functions in the algorithm satisfying $isA(s') \cap partOf(s') = \acute{\emptyset}$. The isA(s’) returns a set of subjects s Є tax^S that satisfy tax(s -> s’) = True and type(s->s’) is a. The partOf(s’) returns a set of subjects s Є tax^S that satisfy tax(s -> s’) =True and type(s ->! s’) = part - of. Algorithm 1 is efficient with the complexity of only O(n). where n = |S|. The algorithm terminates eventually because tax^S is a directed acyclic graph, as defined in Definitions.Algorithm 1. Analyzing semantic relations for specificity

input : a personalized ontology $\mathcal{O}(T) := \langle tax^S, rel \rangle$; a coefficient θ between (0,1).
output: $spe_a(s)$ applied to specificity.

set $k = 1$, get the set of leaves S_0 from tax^S, for $(s_0 \in S_0)$ assign $spe_a(s_0) = k$;
get S' which is the set of leaves in case we remove the nodes S_0 and the related edges from tax^S;
if $(S' == \emptyset)$ **then** return;*//the terminal condition*;
foreach $s' \in S'$ **do**
  if $(isA(s') == \emptyset)$ **then** $spe_a^1(s') = k$;
  else $spe_a^1(s') = \theta \times min\{spe_a(s) | s \in isA(s')\}$;
  if $(partOf(s') == \emptyset)$ **then** $spe_a^2(s') = k$;
  else $spe_a^2(s') = \frac{\sum_{s \in partOf(s')} spe_a(s)}{|partOf(s')|}$;
  $spe_a(s') = min(spe_a^1(s'), spe_a^2(s'))$;
end
$k = k \times \theta, S_0 = S_0 \cup S'$, go to step 2.

As the tax^S of O(T) is a graphic taxonomy, the leaf subjects have no descendants. Thus, they have the strongest focus on their referring-to concepts and the highest $spe_a(s)$. By setting the spe_a range as (0, 1] (greater than 0, less than or equal to 1), the leaf subjects have the strongest $spe_a(s)$ of 1, and the root subject of tax^S has the weakest $spe_a(s)$ and the smallest value in (0, 1]. Toward the root of tax^S, the spe_a (s) decreases for each level up. A coefficient is applied to the $spe_a(s)$ analysis, defining the decreasing rate of semantic specificity from lower bound toward upper bound levels. (θ = 0:9 were used in the related experiments presented in this paper.)

From the leaf subjects toward upper bound levels in tax^S, if a subject has is-a child subjects, it has no greater semantic specificity compared with any one of its is-a child subjects. In is-a relationships, a parent subject is the abstract description of its child subjects. However, the abstraction sacrifices the focus and specificity of the referring-to concepts. Thus, we define the $spe_a(s)$ value of a parent subject as the smallest spe_a (s) of its is-a child subjects, applying the decreasing rate.

B. Topic Specificity

The topic specificity of a subject is investigated, based on the user background knowledge discovered from user local information.

User Local Instance Repository

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 O (T) constructed in Section 2 has only subject labels and semantic relations specified. In this section, we populate the ontology with the instances generated from user local information collections. We call such a collection the user's local instance repository (LIR).

Generating user local LIRs is a challenging issue. The documents in LIRs may be semi structured (e.g., the browsed HTML and XML web documents) or unstructured (e.g., the stored local DOC and TXT documents). In some semi structured 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, for example, the info set tags in some XML documents citing control vocabularies in global lexicons. These documents are ideal to generate the instances for ontology population. When different global knowledge bases are used, ontology mapping techniques can be used to match the concepts in different representations.

However, many documents do not have such direct, clear references. For such documents in LIRs, data mining techniques, clustering, and classification in particular, can help to establish the reference. The clustering techniques group the documents into unsupervised (non predefined) clusters based on the document features. These features, usually represented by terms, can be extracted from the clusters. They represent the user background knowledge discovered from the user LIR. By measuring the semantic similarity between these features and the subjects in O (T), the references of these clustered documents to the subjects in O (T) can be established and the strength of each reference can be scaled by using methods like Non latent Similarity. The documents with a strong reference to the subjects in O (T) can then be used to populate these subjects.

Because ontology mapping and text classification/ clustering are beyond the scope of the work presented in this paper, we assume the existence of an ideal user LIR. The documents in the user LIR have content-related descriptors referring to the subjects in O (T). In particular, we use the information items in the catalogs of the QUT library as user LIR to populate the O (T) constructed from the WKB in the experiments.

The WKB is encoded from the LCSH. The LCSH contains the content-related descriptors (subjects) in controlled vocabularies. Corresponding to these descriptors, the catalogs of library collections also contain descriptive information of library-stored books and documents. Fig. 4 displays a sample information item used as an instance in an LIR. The descriptive information, such as the title, table of contents, and summary, is provided by authors and librarians. This expert classified and trustworthy information can be recognized as the extensive knowledge from the LCSH. A list of content-based descriptors (subjects) is also cited on the bottom of Fig. 4, indexed by their focus on the item's content. These subjects provide a connection between the extensive knowledge and the concepts formalized in the WKB. User background knowledge is to be discovered from both the user's LIR and O (T).

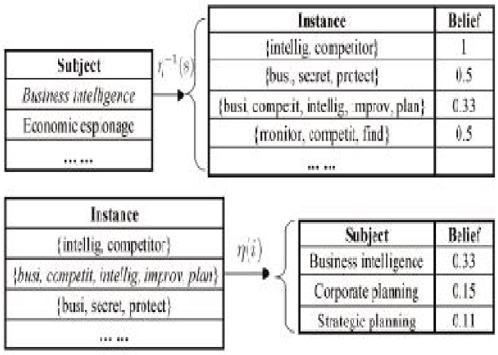

Fig. 4. Mapping of subjects and instances.

The reference strength between an instance and a subject needs to be evaluated. As mentioned previously, 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) is the number of subjects on the citing list of i and priority(s,i) is the index (starting with one) of s on the citing list. The str (i,s) aims to select the right instances to populate O(T). With the str (s,i) determined, the relationship between an LIR and O(T) can be defined.

Architecture of the ontology model

The proposed ontology model aims to discover user back-ground 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.

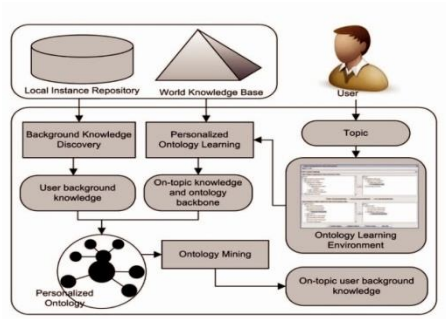

Fig. 5. Ontology Model Architecture.

III. EVALUATION

In information gathering evaluations, a common batch-style experiment is developed for the comparison of different models, using a test set and a set of topics associated with relevant judgments. Our experiments followed this style and were performed under the experimental environment set up by the TREC-11 Filtering Track. This track aimed to evaluate the methods of persistent user profiles for separating relevant and non relevant documents in an incoming stream.

User background knowledge in the experiments was represented by user profiles. A user profile consisted of two document sets: a positive document set $D^{þ}$ containing the on-topic, interesting knowledge, and a negative document set D containing the paradoxical, ambiguous concepts. Each document d held a support value support (d) to the given topic. Based on this representation, the baseline models in our experiments were carefully selected. User profiles can be categorized into three groups: interviewing, semi-interviewing, and non interviewing pro-files, as previously discussed in Section 2. In an attempt to compare the proposed ontology model to the typical models representing these three group user profiles, four models were implemented in the experiments:

1. The Ontology model that implemented the proposed ontology model. User background knowledge was computationally discovered in this model.

2. The TREC model that represented the perfect interviewing user profiles. User background knowl-edge was manually specified by users in this model.

3. The Category model that represented the non inter-viewing user profiles.

4. The Web model that represented the semi-interview-ing user profiles.

The experiment dataflow is illustrated in Fig. 7. The topics were distributed among four models, and different user profiles were acquired. The user profiles were used by a common web information gathering system, the IGS, to gather information from the testing set. Because the user profiles were the only difference made by the experimental models to the IGS, the change of IGS performance reflected the effectiveness of user profiles, and thus, the performance of experimental models. The details of the experiment design are given as follows: The TREC-11 Filtering Track testing set and topics were used in our experiments. The testing set was the Reuters Corpus Volume 1 (RCV1) corpus that contains 806,791 documents and covers a great range of topics. This corpus consists of a training set and a testing set partitioned by the TREC. The documents in the corpus have been processed by substantial verification and validation of the content, attempting to remove spurious or duplicated documents, normalization of dateline and byline formats, addition of copyright statements, and so on. We have also further processed these documents by removing the stop-words, and stemming and grouping the terms.

IV. EXPERIMENTAL RESULTS

The experiments were designed to compare the information gathering performance achieved by using the proposed (Ontology) model, to that achieved by using the golden (TREC) and baseline (web and Category) models. The performance of the experimental models was measured by three methods: the precision averages at 11 standard recall levels (11SPR), the mean average precision (MAP), and the F_1 Measure. These are modern methods based on precision and recall, the standard methods for information gathering evaluation [1], [3]. Precision is the ability of a system to retrieve only relevant documents. Recall is the ability to retrieve all relevant documents.

The MAP is a discriminating choice and recommended for general-purpose information gathering evaluation [3]. The average precision for each topic is the mean of the precision obtained after each relevant document is retrieved. The MAP for the 50 experimental topics is then the mean of the average precision scores of each of the individual topics in the experiments. Different from the 11SPR measure, the MAP reflects the performance in a non interpolated recall-precision curve. The experimental MAP results are presented in Table 2. As shown in this table, the TREC model was the best, followed by the Ontology model, and then the web and the Category models. Table 2 also presents the average macro-F_1 and micro-F_1 Measure results. The F_1 Measure is calculated by

$$F = \frac{2 * precision * recall}{precision + recall}$$

where precision and recall are evenly weighted. For each topic, the macro-F_1 Measure averages the precision and recall and then calculates F_1 Measure, whereas the micro-F_1 Measure calculates the F_1 Measure for each returned result and then averages the F_1 Measure values. The greater F_1 values indicate the better performance. According to the results, the Ontology model was the best, followed by the TREC model, and then the web and the Category models.

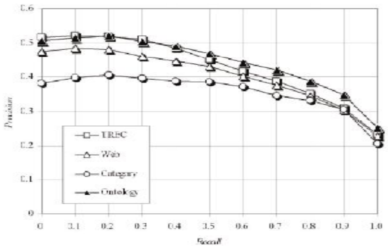

Fig. 6. The 11SPR experimental results.

V. CONCLUSION

In this paper, an ontology model is proposed for represent-ing user background knowledge for personalized web information gathering. The model constructs user persona-lized 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 test bed were used for experiments. The model was compared against bench-mark 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 paper provides a 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 contributions to the fields of Information Retrieval, web Intelligence, Recom-mendation Systems, and Information Systems.

In our future work, we will investigate the methods that generate user local instance repositories to match the representation of a global knowledge base. The present work assumes that all user local instance repositories have content-based descriptors referring to the subjects, how-ever, a large volume of documents existing on the web may not have such content-based descriptors. For this problem, we suggested strategies like ontology mapping and text classification/clustering were suggested. These strategies will be investigated in future work to solve this problem. The investigation will extend the applicability of the ontology model to the majority of the existing web documents and increase the contribution and significance of the present work.

REFERENCES

[1] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. Addison Wesley, 1999.

[2] G.E.P. Box, J.S. Hunter, and W.G. Hunter, Statistics For Experi-menters. John Wiley & Sons, 2005.

[3] C. Buckley and E.M. Voorhees, “Evaluating Evaluation Measure Stability,” Proc. ACM SIGIR ’00, pp. 33-40, 2000.

[4] 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.

[5] L.M. Chan, Library of Congress Subject Headings: Principle and application. Libraries Unlimited, 2005.

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

[7] 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.

[8] 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

[9] S. Nirenburg and V. Rasin, Ontological Semantics. The MIT Press, 2004.

[10] S.E. Robertson and I. Soboroff, “The TREC 2002 Filtering Track Report,” Proc. Text REtrieval Conf., 2002.

AUTHORS:

A.Eswara Rao did M.Sc (CS) in 2003 from MADRAS UNIVERSITY and M.Tech (CSE) in 2006 from BHARATH UNIVERSITY, CHENNAI. He has around 4 years Of Teaching Experience and now he is working as Asst.Professor in IT Dept in St.Martins Engg College,Secunderabad,India

G.V.S.Sowmya, Studying M.Tech(CSE) from PRRM Engg College, Hyderabad. She has completed B.Tech (CSE) from Rao & Naidu Engg. College, Ongole in 2008. She has worked as Asst.Prof in PBIT, Hyderabad since from Oct 2008 to Sep 2010.

P.N.SaiLakshmi did BTech from Padmasri Dr B.V.Raju Institute of Technology,Narsapur,Medak and MTech from RRS College of Engineering And Technology.Currently she is Working As Assistant Professor in Pragna Bharath Institute Of Technology,Hyderabad

◈◈◈