ISSN: 2250–3676

[IJESAT] INTERNATIONAL JOURNAL OF ENGINEERING SCIENCE & ADVANCED TECHNOLOGY Volume-2, Issue-6, 1587 – 1595

ONTOLOGICAL USER PROFILES FOR WEB INFORMATION SEARCHING

K.V. Srinivasa Rao[1], Ch. Sunitha[2], S.Srinivasulu[3], G. Pradeepini[4]

[1]*Associate Professor, Computer Science and Engineering, Prakasam Engineering College,A.P.,India,* ***srinivasa_rao_kalva@yahoo.co.in***

[2]*Student, Computer Science and Engineering, Prakasam Engineering College,A.P.,India,* ***ch.sunithareddy@gmail.com***

[3]*Head of the Department, Computer Science and Engineering, Prakasam Engineering College,A.P.,India,* ***sreenivasulusadineni@gmail.com***

[4]*Associate Professor, Computer Science and Engineerin, Rao & Naidu Engineering College, A.P.,India,* ***pradeepini.gera@gmail.com***

Abstract

Ontologies have been utilized as a model for knowledge description and formulization to show user profiles in web information . However when showing user profiles, many models have used only knowledge from either a global knowledge base or a user local information. In this paper, a ontology example is used for knowledge exhibition and reasoning over user profiles. Ontological user profiles are learnt by this model from both a world knowledge base and user local example repositories. The model of ontology is examined by comparing it against other models in web information searching. So the model become successful on seeing the result.

Index Terms: *knowledge base, Ontology, user profiles, semantic relations, Information, local data.*

---------- *** ----------

1. INTRODUCTION

Over the decades, the number of web-based information available has been increasing successfully. It becomes a challenge to users how to search useful information from the web. Current web information searching system endeavour to pacify user necessities by catching their information needs. For this purpose, a user profiles are coined for knowledge description.

User profiles show the concept models obtained by users when searching web information. Users posses a concept model and generated from their background knowledge. While this concept model cannot be proven in laboratories, many web ontologies have watched it in user behaviour. When users read a document, they are able to know whether or not it is of their interest or relevance to them, a judgement that arises from their implicit concept models. If a user's concept model is simulated, then a superior representation of user profiles can be built.

To imitate user concept models, Ontologies – a knowledge description and formalization model – are highly used in web information gathering. Such ontologies are known as ontological user profiles. To portray user profiles, many researchers have tried to fix user background knowledge through global or local analysis.

Existing global knowledge is used by Global analysis for user background knowledge exhibition. Effective performance is produced by the global analysis techniques for the benefit of user's background knowledge. However, global analysis is limited by the quality of the used knowledge base. For instance, word Net was reported as beneficial in capturing user interest in some areas.

Local analysis studies user local information or watches user behaviour in user profiles. For example, Li and Zhong found taxonomical patterns from the users' local text documents to learn ontologies for user profiles. Some groups learned ontologies adaptively from user's browsing history. Instead of that, Sekine and Suzuki gave analysis on query logs to find user background knowledge. In some works, such as users had been provided with a set of documents and asked for feedback concerned. User background knowledge was then found from this feedback for user profiles. However, because local analysis techniques believe data mining or classification techniques for knowledge finding situation ally, the found results have noisy and unclear information. As a result, local

analysis suffers from ineffectiveness at after catching formal user knowledge.

From this, we can state that user background better found and represented if we can integrate global and local analysis within a hybrid model. The knowledge formalized in a knowledge base will constrain the background knowledge discovery from the user local information. Such a ontology model must produce a great representation of user profiles for web information searching.

In this paper, an ontology model to examine this statement is proposed. This model imitates user's concept models by utilising ontologies, and tries to enhance web information searching performance by using ontological user profiles. The world knowledge and a users local example repository are utilized in the proposed model. World knowledge is commonsense knowledge got by people from experience and education. From a word knowledge base we create ontology by adapting user feedback on interesting knowledge. A multidimensional ontology mining method, speciality and Exhaustively, is also introduced in the proposed model for analyzing concepts specified in ontologies. The proposed ontology model is evaluated by comparison against some benchmark models, through experiments using a large standard data set. The evaluation consequences show that the proposed ontology model becomes successful.

The research contributes to knowledge engineering, and has the potential to improve the design of web information searching systems. The services are original and increasingly valuable, considering the rapid explosion of web information and the growing accessibility of online documents.

2. RELATED WORK

2.1. Ontology Learning

Global knowledgebase's had been utilized by many existing models to learn Ontologies for web information gathering on the basis of the Dewey Decimal classification, king et al. Developed IntlliOnto to improve performance in distributed web information retrieval. Wikipedia had been used by Downey to help understand underlying user interests in queries. These works effectively found user background knowledge. However, their performance had been limited by the quality of the global knowledge bases.

Aiming at leaning ontologies many works mined user background knowledge from user local information. Li and Zhong used pattern recognition and association rule mining techniques to find out knowledge from user local documents for ontology construction. Tran translated Keyword queries to description Logics' Conjugate queries and used ontologies to represent user background knowledge. Zhong proposed a domain ontology learning approach that employed various data mining and natural language understanding techniques. Navigli developed on to Learn to find out semantic concepts and relations from web documents. Web content mining techniques had been used by Jiang and Tan to find out semantic knowledge from domain – specific text documents for ontology learning. Finally user information needs were captured by shehata at the sentence level rather than the document level, and represented user profiles by the conceptual ontological Graph. The use of data mining techniques in these models lead to more user background being discovered. However, the knowledge found in these works have noise and the knowledge found in these works have noise and uncertainties.

Additionally, ontologies had been used in many works to enhance the exhibition in 2009 using a fuzzy domain ontology extraction algorithm to construct concept maps based on the posts on online discussion forums. Ontologies were used by Quest and Ali to assist data mining in biological data bases. Data mining and information retrieval techniques were integrated to further improve knowledge discovery. A model was proposed by Doan called GLUE and machine learning techniques were used to find similar concepts in different ontologies. A framework was proposed by DOV for learning domain ontologies using pattern decomposition clustering/classification, and association rules mining techniques. These works attempted to explore a route to model world knowledge more efficiently.

3. ONTOLOGY CONSTRUCTION

Ontologies is a conceptual model that formally describes and specifies user background knowledge. Web users might expect different views 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. A user may become a business traveler when planning for a business trip, or a leisure traveler when planning for a family holiday. Based on this observation, an assumption is formed that web users have a personal concept model for their information needs. A user's concept model may change according to different information needs. In this section, a model constructing ontologies for web users's concept models is introduced.

3.1 World Knowledge Representation

World knowledge is important for information searching. World knowledge is commonsense knowledge possessed by

people and acquired through their experience and education. Also, as pointed out by Nirenburg and Raskin, "world knowledge is necessary for lexical and referential disambiguation, including establishing co reference relations and resolving ellipsis as well as for establishing and maintaining connectivity of the discourse and adherence of the text to the text producer's goal and plans." 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 nonspecialized 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].

TABLE 1
Comparison of Different World Taxonomies

	LCSH	LCC	DDC	RC
# of Topics	394,070	4,214	18,462	100,000
Structure	Directed Acyclic Graph	Tree	Tree	Directed Acyclic Graph
Depth	37	7	23	10
Semantic Relations	Broader, Used-for, Related-to	Super- and Sub-class	Super- and Sub-class	Super- and Sub-class

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), the Dewey Decimal Classification (DDC) used by Wang and Lee and King et al., and the reference categorization (RC) developed by Gauch et al. 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 objectAis 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 SS be a set of subjects, an element s € SS 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.

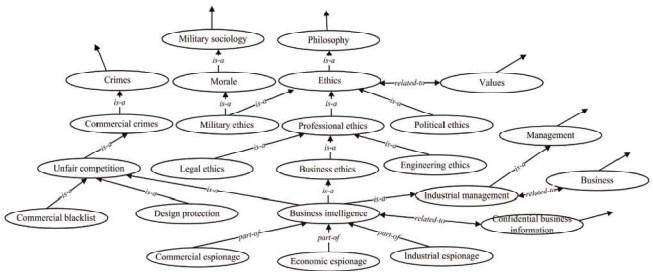


Fig. 1. A sample part of the world knowledge base.

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 =<S, R>
Where

- S is a set of subjects S = {s_1, s_2, . . . , s_m},
- IR is a set of semantic relations IR ={r_1, r_2, . . . , r_n} linking the subjects in S.

Fig. 1 illustrates a sample of the WKB dealing with the topic "Economic espionage." (This topic will also be used as an example throughout this paper to help explanation.) .

3.2 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 ∈ S, the s and its ancestors are retrieved if the 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.

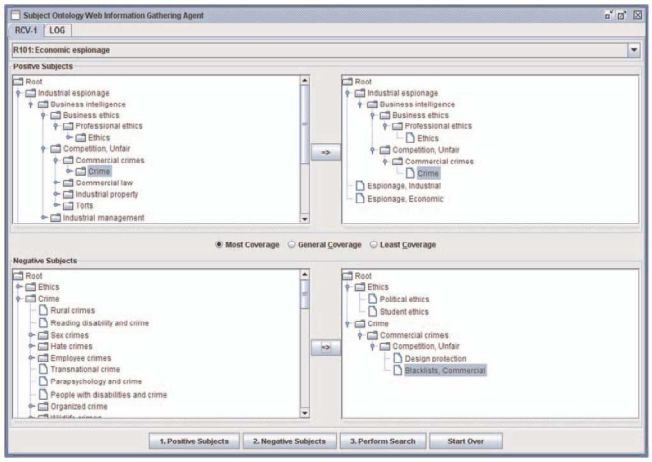
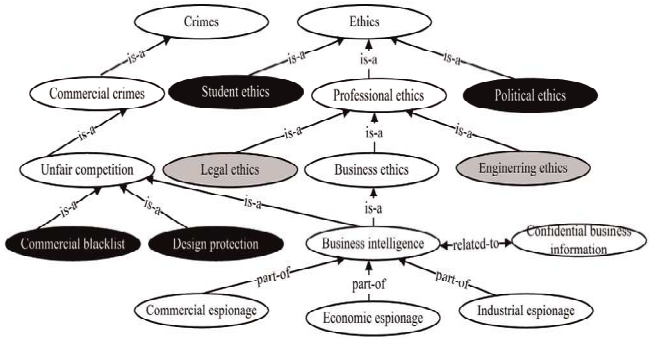


Fig. 2. Ontology learning environment.

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, which are not fed back as either positive or negative from the user, become the neutral subjects to the given topic. An ontology is then constructed for the given topic using these user 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.



ISSN: 2250–3676

[IJESAT] INTERNATIONAL JOURNAL OF ENGINEERING SCIENCE & ADVANCED TECHNOLOGY Volume-2, Issue-6, 1587 – 1595

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

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. Here, we formalize the ontology constructed for a given topic:

Definition 4. The structure of an ontology that describes and specifies topic T is a graph consisting of a set of subject nodes. The structure can be formalized as a 3-tuple O(T)= <S, taxS;, rel>, where

- S is a set of subjects consisting of three subsets S^+, S^-, and S^0, where S^+ is a set of positive subjects regarding T , $S^- \subseteq S$ is negative, and $S^0 \subseteq S$ is neutral;
- tax^S is the taxonomic structure of O(T), which is a noncyclic and directed graph (S, ∈). For each edge e ∈ E and type(e) = is-a or part-of, iff $(s_1 \rightarrow s_2) \in E$, $tax(s_1 \rightarrow s_2)$ = True means s_1 is-a or is a part-of s_2.
- rel is a boolean function defining the related-to relationship held by two subjects in S.

The constructed ontology is because the user selects positive and negative subjects for personal preferences and interests. Thus, if a user searches "New York" and plans for a business trip, the user would have different subjects selected and a different ontology constructed compared to those selected and constructed by a leisure user planning for a holiday.

4 MULTIDIMENSIONAL ONTOLOGY MINING

Ontology mining discovers interesting and on-topic knowledge from the concepts, semantic relations, and instances in an ontology. 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 an 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.

4.1 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 taxS of O(T). As stated in Definition 4, the taxS 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 both 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 taxS. Because subjects have a fixed locality on the taxS of O(T), semantic specificity is also called absolute specificity and denoted by spea(s).

The determination of a subject's spea is described in Algorithm 1. The isA(s1) and part of(s1) are two functions in the algorithm satisfying isA(s1) ∩ partOf(s1) = Ø. The isA(s1) returns a set of subjects s ∈ taxS that satisfy tax(s →s1) = True and type(s → s1) = is _ a. The part of(s1) returns a set of subjects s ∈ taxS that satisfy tax(s →s1) = True and
type(s → s1) = part _ of. Algorithm 1 is efficient with the complexity of only O(n), where n = |S| . The algorithm terminates eventually because taxS is a directed acyclic graph, as defined in Definition 4.

Algorithm 1. Analyzing semantic relations for specificity :
Input : a onotology O(T) = (taxS,rel); a coefficient θ between (0,1)
Output : spea(s) applied to specificity

1.set k =1, get the set of leaves S0from taxS, for (s0 ∈ S0) assign spea(s0) = k
2. get S1 which is the set of leaves in case we remove the nodes S0 and the related edges from taxS.
3. if (S1 = Ø) then return
4. for each s1 ∈ S1 do
{
5. if (isA(s1) = Ø) then spe1a (s1) = k
6. else spe1a (s1) = θ X min { spea(s) / s ∈ isA(s1)}
7. if (part of(s1) = Ø) then spe2a(s1) = k
8. else spe2a (s1) = $\frac{\Sigma_{s \in partOf(s^1)} spe_a(s)}{|partOf(s^1)|}$
9. spea(s1) = min(spe1a (s1), spe2a (s1));
}
10. k= k X θ, S0 = S0 ∪ S1, go to step 2.

As the taxS 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 spea(s). By setting the spea range as (0, 1] (greater than 0, less than or equal to 1), the leaf subjects have the strongest spea(s) of 1, and the root subject of taxS has the weakest spea(s) and the smallest value in (0, 1]. Toward the root of taxS, the spea(s) decreases for each level up. A coefficient θ is applied to the spea(s) analysis, defining the decreasing rate of semantic specificity from lower bound toward upper bound levels. (θ = 0.9 was used in the related experiments presented in this paper.)

From the leaf subjects toward upper bound levels in taxS, 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 spea(s) value of a parent subject as the smallest spea(s) of its is-a child subjects, applying the decreasing rate θ.

If a subject has part-of child subjects, the spea(s) of all part-of child subjects takes part of their parent subject's semantic specificity. As a part-of relation, the concepts referred to by a parent subject are the combination of its part-of child subjects. Therefore, we define the parent's spea as the average spea value of its part-of child subjects applying θ.

If a subject has direct child subjects mixed with is-a and part-of relationships, a spe1a and a spe2a are addressed separately with respect to the is-a and part-of child subjects. The approaches to calculate spe1a and spe2a are the same as described previously. Following the principle that specificity decreases for the subjects located toward upper bound levels, the smaller value of spe1a or spe2a is chosen for the parent subject.

In summary, the semantic specificity of a subject is measured, based on the investigation of subject locality in the taxonomic structure taxS of O(T). In particular, the influence of locality comes from the subject's taxonomic semantic (is-a and part-of) relationships with other subjects.

5 ARCHITECTURE OF THE ONTOLOGY MODEL

The proposed ontology model aims to discover user background knowledge and learns personalized ontologies to represent user profiles. Fig. 4 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.

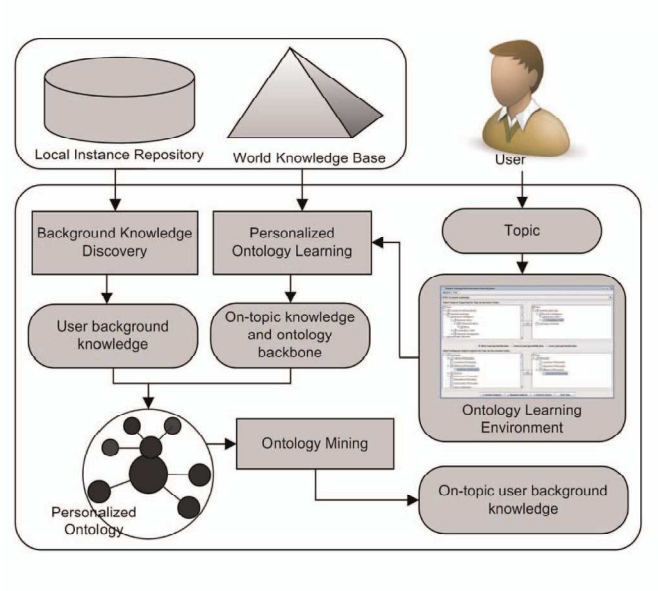


Fig. 4. Architecture of the ontology model.

5.1 Experiment Design

The proposed ontology model was evaluated by objective experiments. Because it is difficult to compare two sets of knowledge in different representations, the principal design of the evaluation was to compare the effectiveness of an information gathering system (IGS) that used different sets of user background knowledge for information gathering. The knowledge discovered by the ontology model was first used for a run of information gathering, and then the knowledge manually specified by users was used for another run. The latter run set up a benchmark for the evaluation because the knowledge was manually specified by users. Under the same experimental conditions, if the IGS could achieve the same (or similar) performance in two different runs, we could prove that the discovered knowledge has the same quality as the user specified knowledge.

The proposed ontology model could then be proven promising to the domain of web information searching. In information searching 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.3 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, such as those in the experiments and the TREC-11 Filtering Track. A user profile consisted of two document sets: a positive document set D^+ 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 noninterviewing profiles. 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 knowledge was manually specified by users in this model.
3. The Category model that represented the noninterviewing user profiles.
4. The Web model that represented the semi-interviewing user profiles.

The experiment dataflow is illustrated in Fig. 5. 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.

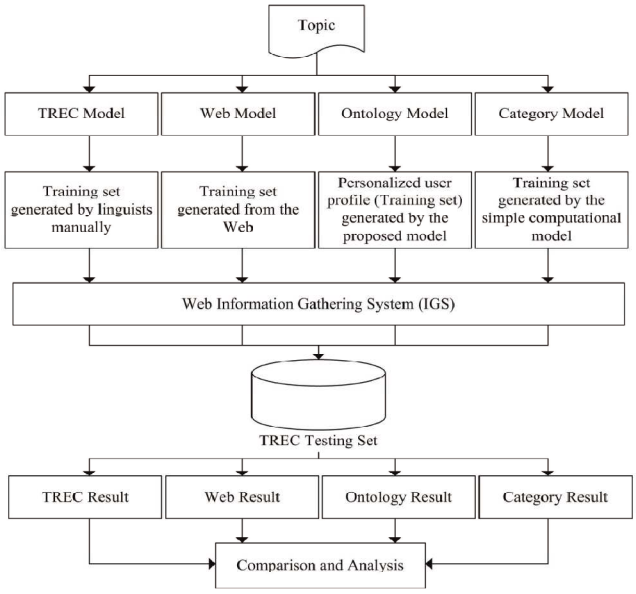


Fig. 5. Experiment design.

In the experiments, we attempted to evaluate the proposed model in an environment covering a great range of topics. However, it is difficult to obtain an adequate number of users who have a great range of topics in their background knowledge. The TREC-11 Filtering Track provided a set of 50 opics specifically designed manually by linguists, covering various domains and topics. For these topics, we assumed that each one came from an individual user. With this, we simulated 50 different users in our experiments. Buckley and Voorhees [3] stated that 50 topics are substantial to make a benchmark for stable evaluations in information searching experiments. Therefore, the 50 topics used in our experiments also ensured high stability in the evaluation.

Each topic has a title, a description, and a narrative, provided by the topic author. In the experiments, only the titles of topics were used, based on the assumption that in the real world users often have only a small number of terms in their queries.

5.2 Evaluation of Ontology

This model was the implementation of the proposed ontology model. As shown in Fig. 5, the input to this model was a topic

and the output was a user profile consisting of positive documents (D+) and negative documents (D-). Each document d was associated with a support(d) value indicating its support level to the topic.

The WKB was constructed based on the LCSH system, as introduced in Section 3.1. The LCSH authority records distributed by the Library of Congress were a single file of 130 MB compiled in MAchine-Readable Cataloging (MARC) 21 format. After data preprocessing using expression techniques, these records were translated to human-readable form and organized in an SQL database, approximately 750 MB in size. Theoretically, the LCSH authority records consisted of subjects for personal names, corporate names, meeting names, uniform titles, bibliographic titles, topical terms, and geographic names. In order to make the Ontology model run more efficiently, only the topical, corporate, and geographic subjects were kept in the WKB, as they covered most topics in daily life. The BT, UF, and RT references (referred to by "450 jwj a", "450," and "550" in the records, respectively) linking the subjects in the LCSH thesaurus, were also extracted and encoded as the semantic relations of is-a, part-of, and related-to in the WKB, respectively. Eventually, the constructed WKB contained 394,070 subjects covering a wide range of topics linked by semantic relations.

The user ontologies were constructed as described in Section 3.2 via user interaction. The authors played the user role to select positive and negative subjects for ontology construction, following the descriptions and narratives associated with the topics. On average, each personalized ontology contained about 16 positive and 23 negative subjects.

For each topic T , the ontology mining method was performed on the constructed O(T) and the user LIR to discover interesting concepts, as discussed in Section 4. The user LIRs were collected through searching the subject catalog of the QUT library by using the given topics. The catalog was distributed by QUT library as a 138 MB text file containing information for 448,590 items. The information was preprocessed by removing the stop-words, and stemming and grouping the terms. Librarians and authors have assigned title, table of content, summary, and a list of subjects to each information item in the catalog. These were used to represent the instances in LIRs. For each one of the 50 experimental topics, and thus, each one of the 50 corresponding users, the user's LIR was extracted from this catalog data set. As a result, there were about 1,111 instances existing in one LIR on average.

The semantic relations of is-a and part-of were also analyzed in the ontology mining phase for interesting knowledge discovery. For the coefficient θ in Algorithm 1, some preliminary tests had been conducted for various values (0.5, 0.7, 0.8, and 0.9). As a result, θ = 0.9 gave the testing model the best performance and was chosen in the experiments.

Finally, a document d in the user profile was generated from an instance i in the LIR. The d held a support value support(d) to the T , which was measured by

$$\text{support}(di) = \text{str}(i,T) \; X \sum_{s \in \eta(i)} \text{spe}(s,T)$$

where s □ S of O(T), str(i,T) was defined by (4), and spe(s,T) by (6). When conducting the experiments, we tested various thresholds of support(d) to classify positive and negative documents. However, because the constructed ontologies were personalized and focused on various topics, we could not find a universal threshold that worked for all topics. Therefore, we set the threshold as support(d) = 0, following the nature of positive and negative defined in this paper. The documents with support(d) > 0 formed D+, and those with negative support(d) <= 0 formed D- eventually.

ACKNOWLEDGMENTS

We would like to express our sincere thanks to Sri. Dr. Kancharla Ramaiah Secretary and Correspondent, Prakasam Engineering College, Kandukur, A.P. India for his support with providing research environment . We are extremely thankful to our collogues , friends and family members who are cooperated in this 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 Experimenters. John Wiley & Sons, 2005.
[3] C. Buckley and E.M. Voorhees, "Evaluating Evaluation Measure Stability," Proc. ACM SIGIR '00, pp. 33-40, 2000.
[4] Z. Cai, D.S. McNamara, M. Louwerse, X. Hu, M. Rowe, and A.C. Graesser, "NLS: A Non-Latent Similarity Algorithm," Proc. 26th Ann. Meeting of the Cognitive Science Soc. (CogSci '04), pp. 180-185, 2004.
[5] L.M. Chan, Library of Congress Subject Headings: Principle and Application. Libraries Unlimited, 2005.
[6] P.A. Chirita, C.S. Firan, and W. Nejdl, "Personalized Query Expansion for the Web," Proc. ACM SIGIR ('07), pp. 7-14, 2007.
[7] R.M. Colomb, Information Spaces: The Architecture of Cyberspace. Springer, 2002.
[8] A. Doan, J. Madhavan, P. Domingos, and A. Halevy, "Learning to Map between Ontologies on the Semantic Web,"

Proc. 11th Int'l Conf. World Wide Web (WWW '02), pp. 662-673, 2002.
[9] D. Dou, G. Frishkoff, J. Rong, R. Frank, A. Malony, and D. Tucker, "Development of Neuroelectromagnetic Ontologies(NEMO): A Framework for Mining Brainwave Ontologies," Proc. ACM SIGKDD ('07), pp. 270-279, 2007.
[10] 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.