# Ontology-Based Business Process Customization for Composite Web Services

Tulasi Krishna Kumar. K

*Computer Science and Engineering,*
*PBR Viswodaya Institute of Technology & Science,*
*Kavali, Andhra Pradesh, India*

Konda Reddy. R

*Computer Science and Engineering,*
*PBR Viswodaya Institute of Technology & Science,*
*Kavali, Andhra Pradesh, India*

Venkata Ramana. K
*Associative Professor & HOD of CSE,*
*Vaishanvi Institute of Technology, Tirupati.*
*Research Scholar in Vel Tech Dr RR & Dr SR Technical University, Chennai.*

**ABSTRACT:** . An Ontology model is proposed for representing user background knowledge and for personalized web information gathering. An Ontology model is used for extracting global knowledge from personalized LCSH system repositories and discovering user background knowledge from the user local instance repositories are searching in a multidimensional manner. But, information presents in a single hierarchy. It exists a semantic relationship for gathering global information. In existing algorithms & approaches gives more precision value and less recall value but can't successfully gather meaningful knowledge. It occurs as replication copies of knowledge. it does not eradicate data redundancy Finally, we can conclude that the Ontology model are significantly evaluates a substantial improvements achieved by a F1 measure experimental results are promising an efficiency of a knowledge discovery and it is reliable. It shows more recall valueless precision value. This model have been proved as a Benchmark model by applying it to a common system for Hierarchical Web information gathering.



**KEYWORDS:** Replication, recall, precision, world knowledge, multi dimensional mining, replication, recall, precision, local instance repository, web information gathering, single hierarchy,

## 1. Introduction

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 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 researchers have observed it in user behavior. 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, ontology a knowledge description and formalization model—are utilized in personalized web information gathering. Such ontologism is called ontological user profiles or personalized ontologism.

## 2..RELATED ENDEAVOR

### 2.1 Ontology Learning Environment

Global knowledge bases were used by many existing models to learn ontologism from web. Wikipedia to help understand underlying user interests in queries. These works effectively discovered user background knowledge; however, their performance was limited by the quality of the global knowledge bases. Aiming at learning personalized ontologism, many works mined user background knowledge from user local information. We used pattern recognition and association rule mining techniques to discover knowledge from user local documents for ontology construction. Translated keyword queries to Description Logics' conjunctive queries and used ontologism to represent user background knowledge. We proposed a domain ontology learning approach that employed various data. A List of categories and ask users for interesting or nonintersecting categories, which extracts training sets from the web based on user feedback categories.

### 2.2 USER PROFILES

User profiles are used in web information gathering to interpret the semantic meaning of queries and capture user information needs. User profiles can be categorized into three groups: interviewing, semi-interviewing, and non interviewing. Inter- viewing user profiles can be deemed perfect user profiles. They are acquired by using manual techniques, such as questionnaires, interviewing users, and analyzing user classified training sets. One typical example is the TREC Filtering Track training sets, which were generated manually . The users read each document and gave a positive or negative judgment to the document against a given topic. Because, only users perfectly know their interests and preferences, these training documents accurately reflect user background knowledge. Semi-interviewing user pro- files are acquired by semi automated techniques with limited user involvement. These techniques usually provide users with list of categories for interesting or non-interesting categories. No interviewing user profiles do not involve users at all, but ascertain user interests instead. They acquire user profiles by observing user activity and behavior and discovering user background knowledge which acquires user profiles based on users' online browsing history.



## 3. CONSTRUCTION OF PERSONALIZED ONTOLOGY MODEL

From observations in daily life, we found that web users might have different expectations for the same search query. For example, for the topic "TIRUPATHI" 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's concept model may change according to different information needs. In this section, a model constructing personalized ontology for web users' concept models is introduced.

### 3.1 A WORLD KNOWLEDGE REPRESENTATION

world knowledge is a 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. 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.

| | 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 |

**TABLE 1: Comparison of Different World Repositories**

The LCSH represents the natural growth and distribution of human intellectual work, and covers comprehensive and exhaustive topics of world knowledge . In addition, the LCSH is the most comprehensive no specialized controlled vocabulary in English. 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) and the reference categorization (RC). 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 . 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) . BT references are for two subjects describing the same topic, but at different levels of abstraction or specificity.

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 .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. 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.

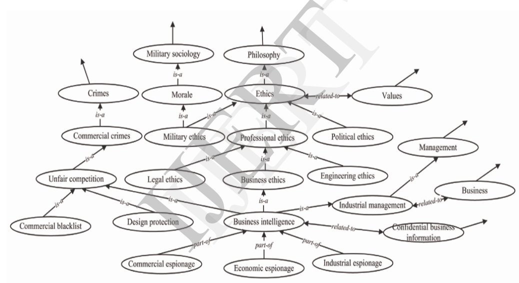


**Fig. 1. A sample part of the Global knowledge base.**

**Definition** 1. Let $ be a set of subjects, 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.

### 3.2 Ontology Construction

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.

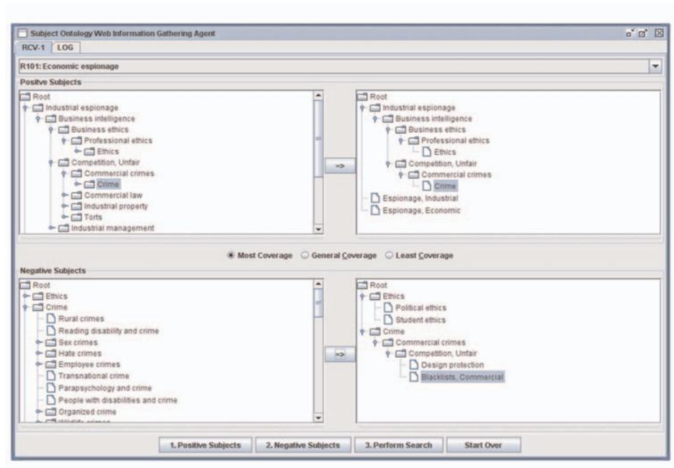

**Fig 2. An Ontology learning environment**

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. These user-selected negative subjects are listed on the bottom-right panel (e.g., "Political ethics" and "Student ethics").

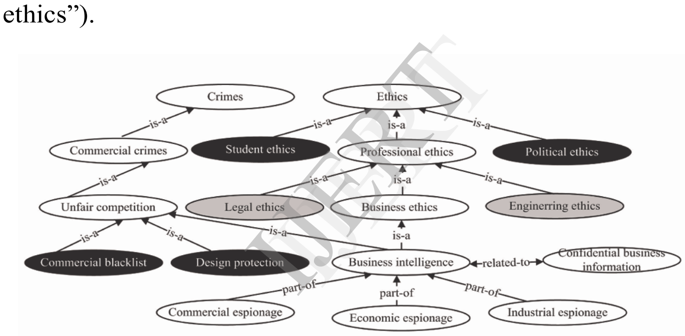


**Fig. 3. An ontology (partial) constructed for topic "Ethics"**

**Definition** 3. 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,$tax^s$,rel>,where

- S is a set of subjects consisting of three subjects $S^+$,$S^-$,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 noncyclical and directed graph (S,ϵ),for each edge e Є ϵ and type(e)=is-a or part-of, if f<$s_1 \rightarrow s_2$> Є ϵ;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.

## 4. MULTI DIMENSIONAL 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 Exhaustively. Specificity (denoted as spe) describes a subject's focus on a given topic. Exhaustively (denoted as 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) ontology 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 for world knowledge

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 taxes of O(T) (this is also argued by .as stated in definition 4,the taxes 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 root 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 Algorithm: Analyzing semantic relations for specificity is used for gathering global information 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.

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

1 set $k = 1$, get the set of leaves $S_0$ from $tax^{\mathcal{S}}$, 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 $tax^{\mathcal{S}}$;
3 **if** $(S' == \emptyset)$ **then** return;*//the terminal condition*;
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 **if** *(partOf*$(s')$ $== \emptyset)$ **then** $spe_a^2(s') = k$;
8 **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 $k = k \times \theta, S_0 = S_0 \cup S'$, go to step 2.

If a subject has part-of child subjects, the $spe_a(s)$ 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 objects. therefore, we define the parent's $spe_a$.

1. In this analysis, the related-to semantic relations are not considered because they are no taxonomic. In this paper, we assume they have no influence on each other in terms of specificity. However, this is an interesting issue and will be pursued in our future work. In summary, the semantic specificity of a subject is measured, based on the investigation of subject locality in the taxonomic structure $tax^S$ 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. The WKB is encoded from the LCSH, as discussed in Section 3.1. The LCSH contains the content-related descriptors (subjects) in controlled vocabularies.

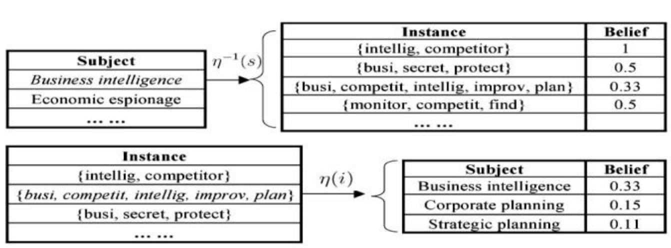


Fig. 4. QUT library catalogs

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 TOP of Fig. 4, indexed by their focus on the item's content. Because the str(i,T) from (4) could be positive or negative values, the $spe_r$ (S,T,LIR) values from (5) could be positive or negative as well.

Author Landsman, Mark, 1966-
Title Dictatorship and demand : the politics of consumerism in East Germany / Mark Landsman.
Published Cambridge, MA : Harvard University Press, 2005.

| ITEM LOCN | CALL NO | STATUS |
|---|---|---|
| Carseldine | 306.30943109045 1 | IN LIBRARY |

| Table of Contents | | |
|---|---|---|
| 1 | Production and consumption : establishing priorities | 16 |
| 2 | The contest begins : the currency reform, the Berlin blockade, and the introduction of the HO | 38 |
| 3 | The planned and the unplanned : consumer supply and provisioning crisis | 74 |
| 4 | The rise, decline, and afterlife of the new course | 115 |
| 5 | Demand research and the relations between trade and industry | 149 |
| 6 | Crisis revisited : the main economic task and the building of the Berlin Wall | 173 |

Description xii, 296 p. ; 24 cm.
Series Harvard historical studies ; 147
ISBN 067401698X (alk. paper)
Bibliography Includes bibliographical references (p. 223-287) and index.
Summary "Based on research in recently opened East German state and party archives, this book depicts a regime caught between competing pressures. While East Germany's leaders followed a Soviet model, which fetishized productivity in heavy industry and prioritized the production of capital goods over consumer goods, they nevertheless had to contend with the growing allure of consumer abundance in West Germany. The usual difficulties associated with satisfying consumer demand in a socialist economy acquired a uniquely heightened political urgency, as millions of East Germans fled across the open border." "A new vision of the East-West conflict emerges, one fought as much with washing machines, televisions, and high fashion as with political propaganda, espionage, and nuclear weapons. Dictatorship and Demand deepens our understanding of the Cold War."--BOOK JACKET.
Subject Consumption (Economics) -- Germany (East)
Socialism -- Germany (East)
Germany (East) -- Economic conditions.
Germany (East) -- Politics and government.

Fig. 5. Mappings of subjects

As discussed previously, a subject's specificity have two focuses: semantic specificity and topic specificity .Therefore ,the final specificity of a subject is composition of them and calculated by

$$spe_r(s, \mathcal{T}, \mathcal{LIR}) = \sum_{i \in \eta^{-1}(s)} str(i, \mathcal{T}).$$

4.3 Multidimensional Analysis of Subjects and its instances

The exhaustively of a subject refers to the extent of its concept space dealing with a given topic .This space extends if a subject has more positive descendants regarding the topic. In contrast ,if a subject has more negative descendants ,its exhaustively decreases .based on this ,let desc(s) be a function that returns the descendants of a s(inclusive) in O(T);we evaluate a subject 's exhaustively by aggregating the semantic specificity of its descendants:

$$exh(s, \mathcal{T}) = \sum_{s' \in desc(s)} \sum_{i \in \eta^{-1}(s')} str(i, \mathcal{T}) \times spe_a(s', \mathcal{T}). \quad (2)$$



Subjects are considering interesting to the user only ,if their specificity and exhaustively are positive. the subject sets of $S^+$,$S^-$,and $S^0$,originally defined in section 3.2,can be refined after ontology mining for the specificity and exhaustively of subjects: few theorems can be introduced based on the subject analysis of specificity and exhaustively.

$$\mathcal{S}^+ = \{s | spe(s, \mathcal{T}) > 0, exh(s, \mathcal{T}) > 0, s \in \mathcal{S}\}; \quad (3)$$

$$\mathcal{S}^- = \{s | spe(s, \mathcal{T}) < 0, exh(s, \mathcal{T}) < 0, s \in \mathcal{S}\}; \quad (4)$$

$$\mathcal{S}^\circ = \{s | s \in (\mathcal{S} - (\mathcal{S}^+ \cup \mathcal{S}^-))\}. \quad (5)$$

## 5 . ONTOLOGY MODEL ARCHITECTURE

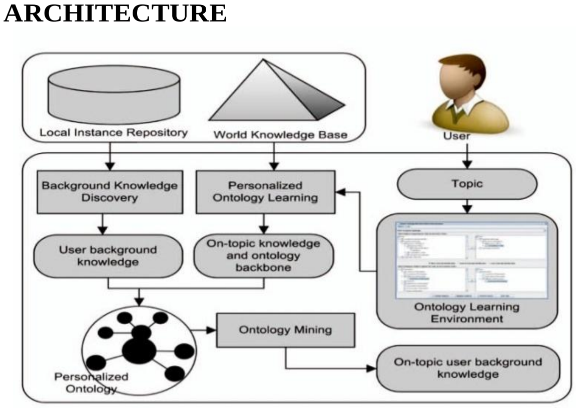


Fig. 6. Architecture of the ontology model.

The proposed ontology model aims to discover user background knowledge and learns personalized ontologism 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 repositories are utilized by model. the world knowledge base provides the taxonomic structure for the personalized ontology. The user local background knowledge is discovered from the user local instance repository. Against the given topic ,the specificity and exhaustively of subjects are the investigated for user background knowledge discovery.

## 6. PROPOSED SYSTEM

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 gathering. The semantic relations of is-a and part-of were also analyzed in the ontology mining phase for interesting knowledge discovery. 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 Twitch was measured by



$$support(d_i) = str(i, \mathcal{T}) \times \sum_{s \in \eta(i)} spe(s, \mathcal{T}), \quad (6)$$

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 ontology were personalized and focused on various topics. Therefore ,we set the threshold as support(d)=0,following the nature of positive and negative defined in this paper. the documents will support(d) $\geq 0$ formed $D^+$,and those with negative support(d) $\leq 0$ formed $D^-$ eventually.

### 6.1 EXPERIMENTAL ANALYSIS

the mean average precision(MAP) and the F1 Measure, These are modern methods based on precision and recall, the standard methods for information gathering evaluation ,.precision is the ability of system to retrieve only relevant documents. Recall is the ability to retrieve all relevant documents. An 11SPR value is computed by summing the interpolated precisions at the specified recall cutoff, and then dividing by the number of topics: Where N denotes the number of topics , and

=indicates the cutoff points where the precisions are interpolated. An average precisions then link to a curve describing the recall-precision performance. the experimental 11SPR results are plotted in fig.8,where the 11SPR curves show that the Ontology model was the best ,followed by the TREC model ,the web model ,and finally, the category model.

| | TREC | Web | Category | Ontology |
|---|---|---|---|---|
| MAP | 0.2901 | 0.2775 | 0.2612 | 0.2886 |
| Micro-FM | 0.3559 | 0.3458 | 0.3288 | 0.3622 |
| Macro-FM | 0.3875 | 0.3759 | 0.3554 | 0.3941 |

TABLE 2 F1 Measure Experimental Results Analysis

Table2 also presents the average macro-F1 and micro-F1 Measure results. The F1 Measure is calculated by

$$F_1 = \frac{2 \times precision \times recall}{precision + recall},$$

where precision and recall values are evenly weighted. for each topic, the macro-F1 Measure averages the precision and recall and then calculates F1 Measure, whereas the micro-F1 Measure calculates the F1 Measure calculates the F1 Measure for each returned result and then averages the F1 Measure values. The greater F1 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.

## 7.CONCLUSION

In this paper, 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.

## FUTURE WORK

The model also has extensive contributions to the fields of Information Retrieval, web Intelligence, Recommendation Systems, and Information Systems. 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; however, a large volume of documents existing on the web may not have such content-based descriptors. For this problem, strategies like ontology mapping and text classification/clustering were suggested. These strategies will be investigated in future work to solve this problem. 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.



## REFERENCES :

[1]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.
[2]G.A. Miller and F. Hristea, "WordNet Nouns: Classes and Instances," Computational Linguistics, vol. 32, no. 1, pp. 1-3, 2006.
[3]R. Navigli, P. Velardi, and A. Gangemi, "Ontology Learning and Its Application to Automated Terminology Translation," IEEE Intelligent Systems, vol. 18, no. 1, pp. 22-31, Jan./Feb. 2003
[4] S.E. Robertson and I. Soboroff, "The TREC 2002 Filtering Track Report," Proc. Text REtrieval Conf., 2002.
[5] A. Sieg, B. Mobasher, and R. Burke, "Web Search Personalization with Ontological User Profiles," Proc. 16th ACM Conf. Information and KnowledgeManagement (CIKM '07), pp. 525-534,2007
[6] 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.
[7] J. Trajkova and S. Gauch, "Improving Ontology-BasedUser Profiles," Proc. Conf. Recherche d'Information Assistee par Ordinateur (RIAO '04), pp. 380-389, 2004.
[8] T. Tran, P. Cimiano, S. Rudolph, and R. Studer, "Ontology-Based Interpretation of Keywords for Semantic Search," Proc. Sixth Int'l Semantic Web and Second Asian Semantic Web Conf. (ISWC '07/ ASWC '07), pp. 523-536, 2007.
[9]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.
[10]L.M. Chan, Library of Congress Subject Headings: Principle and Application. Libraries Unlimited, 2005.