ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

A Modified Personalized Web Information Gathering Using Ontology

Selvapriya S [1], Dr. A. Kumaravel*[2]

Assistant Professor, Dept. of IT, Jerusalem College of Engineering, Chennai, Tamil Nadu, India[1]

Dean& HOD, Department of Information Technology, Bharath University, Chennai, Tamil Nadu, India[2]

* Corresponding Author

ABSTRACT: As a model for knowledge description and formalization, ontologies are widely used to represent user profiles in personalized web information gathering. However in the existing models user profiles are represented either knowledge from local repository or global knowledge base. In this paper a personalized web information gathering model is proposed to retrieve the most relevant information from the web based on the user's profiles. This model learns ontology based personalized user profiles and discovers the user background knowledge from a knowledge base and gives most relevant information from the web. The proposed model is evaluated and the results are compared with other existing models.The result shows that personalized web information gathering model is successful.

Categories and Subject Descriptors: [Computer-Artificial Intelligence]: Ontology

General Terms: Data mining, Information retrieval

Additional Key Words and Phrases —Ontology, personalization, knowledge base, user profiles, web information gathering.

I. Introduction

The availability of web-based information gathering has increased dramatically. To gather the useful information from the web is difficult and challenging for the users. Current web information gathering system attempt to capture the information from the web based on their needs. For this purpose user profiles are created to discover the user background knowledge.

Users possess a concept when gathering the web information and user profiles represent the user concepts. To simulate the concept models, ontologies—a knowledge description and formalization model—are utilized in personalized web information gathering. These user profiles are called as ontology based personalized user profiles. Researchers represent the user profiles by discovering user background knowledge either through global or local analysis. Global analysis uses existing global knowledge bases for representing user background knowledge.The existing global knowledge bases generic ontologies (e.g., Word Net), thesauruses (e.g., digital libraries), and online knowledge bases (e.g., online categorizations and Wikipedia). Since these knowledge bases are not domain based it is difficult to discover the user background knowledge.

Local analysis discovers the user background knowledge from the user's behavior. For example, Li and Zhong [8] discovered taxonomical patterns from the users'local text documents to represent the user background knowledge. Some researchers [7], [9] discover the user background knowledge from the users browsing history. Some researchers discover the user background knowledge by providing a set of documents against a topic and asking their feedback. These works effectively discovered user background knowledge; however, their performance was limited by the quality of the global knowledge bases.From these analyses, the user background knowledge can be better discovered from the knowledge base by getting user profiles manually and preferring subjects semi-automatically.

10.15680/ijircce.2015.0304222

In this paper An Ontology based web information gathering model is developed to discover the user background knowledge and to produce a superior representation of the user profile. The objective of the paper is to retrieve the most relevant information from the web through personalized user profiles. This model simulates users' concept model and attempts to improve performance of web information gathering.The *knowledge base*is used in the proposed model. *Knowledge base* is commonsense knowledge acquired by people from experience and education. Ontology is used for developing a knowledge base for a specific domain. The user profiles are acquired by using manual technique, such as interviewing users.Porter Stemming Algorithm is used for preprocessing. An Ontology Learning Environment (OLE)is a tool, developed to collect the users' preference semi-automatically. An ontology mining method, specificity also introduced in the proposed model for retrieving concepts specified in ontology. The web information's are extracted from the knowledgebase based on their users' profile.This gives an ontology based personalized web search.The proposed model is evaluated by comparison against some benchmark models. The evaluation result shows that the proposed personalized web information gathering model is successful.

The research contributes to knowledge engineering, and has the potential to improve the design ontology based personalized web information gathering system. The contributions are original and increasingly significant, considering the rapid explosion of web information and growing accessibility of online documents.

The paper is organized as follows: Section 2 discusses the related work; in Section 3, we introduce how profile and domain ontologies are constructed for users; in section 4 we describe how ontology manager manages domain ontologyand in Section 5, present how OLE tool is constructed. After that, Section 6 gives the architecture of the proposed model; Section 7 discusses the evaluation issues, and the results are analyzed in Section 8. Finally, Section 9 makes conclusions and addresses our future work.

II. RELATED WORK

A. *Ontology Learning*

Knowledge bases were used by many existing models to learn ontologies for web information gathering. For example, Xiaohui Tao et.al[1] discovers the user background knowledge from the world knowledge base called LCSH(Library Congress Of Subject Heading). It also learns a user profiles and specify the user's preference. Wikipedia was used by Downey et al. [10] to help understand underlying user interests in queries.LCSH was used by X. Tao et.al[2] to discover the user's background knowledge from world knowledge base and specifies the user's preferences in web search.J. Trajkova et.al[3] Learns an ontological user profiles through users behavior and activity. S.E. Robertson et.al[4] Learns user profile manually and discovers the user's background knowledge from the knowledge base. Li and Zhong [8] used pattern recognition and association rule mining techniques to discover knowledge from user local documents for ontology construction.Yuefeng Li et.al[5] discovers the user background knowledge through mathematical model which is used for representing co-relation between compound classes. The above said models are not a domain specific and it is generic. So the performances are limited by the quality of knowledge base. The performance can be increased by creating a knowledge base for a specific domain.

B. *User Profiles*

User profiles can be categorized into three groups: interviewing, semi-interviewing, and non-interviewing.*Interviewing* user profiles gives a perfect user profiles. This is a manual technique and collects the user profiles through questionnaires and interviewing users. One typical example is the TRECFiltering Track training sets [4] collect the user profiles manually. For a given search topic, the users are allowed to read and judge a document as a positive and negative judgment to a document.

Semi-interviewing is a semi-automated technique with limited user involvement.Web training set acquisition model introduced by Tao et al.[6], produces a user-profiles with limited user involvement. This technique usually provides with list of categories and asks users for interesting and non-interesting categories. *Non-interviewing* is a full automatic without user involvement. They produce a user profile by observing the user behavior and activity. A typicalmodel is OBIWAN, proposed by Gauch et al. [7], which acquires user profiles based on users' online browsing history. The interviewing, semi-interviewing, and non-interviewing user profiles can also be viewed as manual, semiautomatic, and automatic profiles, respectively.

III. ONTOLOGY CONSTRUCTION

Ontologies are conceptualization model that formally describes and specifies user background knowledge. From observation we understood web users have different expectations for the same search query. For example, for the topic “India,” business travelers may expect different information from leisure travelers. Same user may have different expectation for the same search query at different situations. From this observation, an assumption is made that web users have personal concepts for their information needs. Personalized web information gathering model is introduced for web user’s concepts.

A. Knowledge Representation

Knowledge base is important for web information gathering. Knowledge base is commonsense knowledge possessed by people and acquired through their experience and education.

We first need to construct the knowledge base. The knowledge base must be a domain specific and consist wide range of topics. Since users may have different backgrounds. Knowledge base is constructed through ontology. Ontology is constructed with the semantic relations of is-a and part-of relation. *Is-a* relation defines that the subjects describing the topic but at different levels of abstraction. *Part-of* relation defines that subject belongs to particular topic i.e. when an objectA is used for an action. For e.g.“a fork is used for dining”.

The primitive knowledge unit in our knowledge base is topics. The topics are formalized as follows:

Definition 1.*Let T be a set of topic, an element t T is formalized as 4 tuples t:=< label, neighbor, ancestor, descendant>, where*

- *label is the heading of t in the knowledge base.*
- *neighbor is a function returning the topic that have direct links to t in the knowledge base.*
- *ancestor is a function returning the topic that have a higher level of abstraction than t and link to t directly or indirectly in the knowledge base.*
- *descendant is a function returning the subject that are more specific than s and link directly in the knowledge base.*

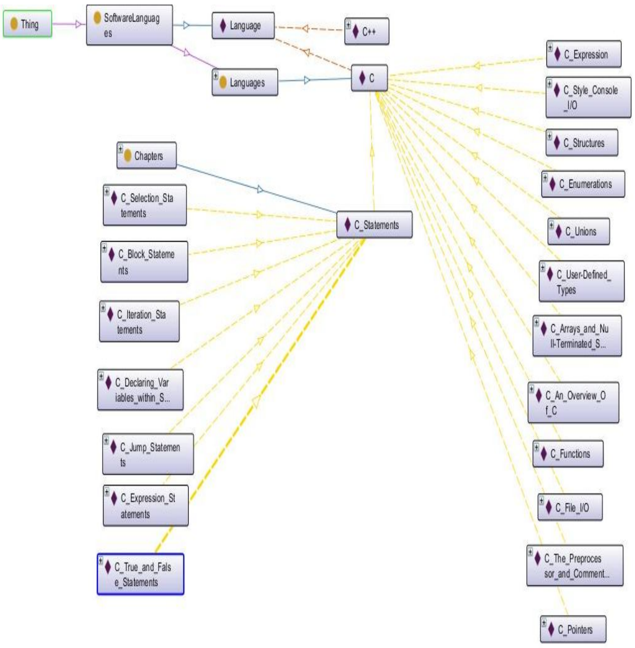

Fig.1. A sample part of the knowledge base

The topics in the knowledge base are linked to each other by semantic relations of *is-a* and part-*of*. 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 relation is an element of {is-a, part-of}.*

With Definitions 1 and 2, the knowledge can then be
formalized as follows:

Definition 3.*Let KB be the knowledge base, which is taxonomy constructed as a directed acyclic graph. The KB consists of a set of subjects linked by their semantic relations, and can be formally defined as a 2- tuple KB:=<T, IR>, where*

- *T is a set of topics T:={t1, t2,….tn};*
- *IR is a set of semantic relations IR := {r1, r2,…… } linking the topic in T.*

Fig.1 illustrates a sample of the KB dealing with the specific domain.

B. Profile representation

User profiles are represented with more user involvement. User profile is collected manually by asking questions to the users and they are represented through ontology. Thisinformation's are collected and stored in profile ontology. The collected user profiles are used for personalized web search. The user searched query items are saved into the profile ontology. The user can view their browsed history through the profile ontology.

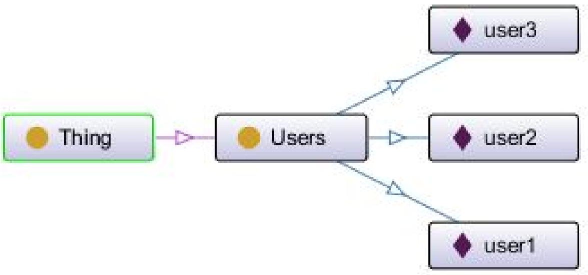

Fig. 2. A sample part of the profile ontology

The primitive knowledge unit in our profile ontology is user details. The user details are formalized as follows:

Definition 1.*Let U be the set of users, an element u Uis formalized as a 2-tuple u:=< label, childs>, where*

- *label is the user name in the profile ontology;*
- *childs are the user details.*

The user name in the profile ontology are linked by the semantic relations of has relationship. 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 the users and user details that hold a type of relation;*
- *a type of relations is an element of{has-a}.*

With the definitions of 1 and 2, the profile ontology can then be formalized as follows:

Definition 3.*Let P be a profile ontology, which is a taxonomy constructed as a directed acyclic graph. The profile ontology consists of a set of users linked by their semantic relations, and can be formally defined as a 2-tuple P:=<U, IR>, where*

- *U is a set of users U:={u1, u2…. Un};*
- *IR is a set of semantic relations IR := {r1, r2…….rn}linking the users in U*

Fig. 2 illustrates a sample of the profile ontology dealing with three users.

IV. ONTOLOGY MANAGER

Ontology Manager is designed to allow multiple users to create, enhance and browse all types of semantic model, whether they are lists, controlled vocabularies, taxonomies, thesauri or ontologies. Ontology manager manages the language that defines your space,creates the model of links and structure between language elements that can drive a new user experience and holds any term 'metadata' to drive or enhance connected applications. Ontology Manager supports a taxonomy design, initial build, validate, and enhance cycle. To populate the domain ontology, ontology manager view the users profile history which is collected in the profile ontology. If the users searched query terms are not present in the domain ontology, the ontology manager will update the query terms in the domain ontology by removing stemming and stop word.

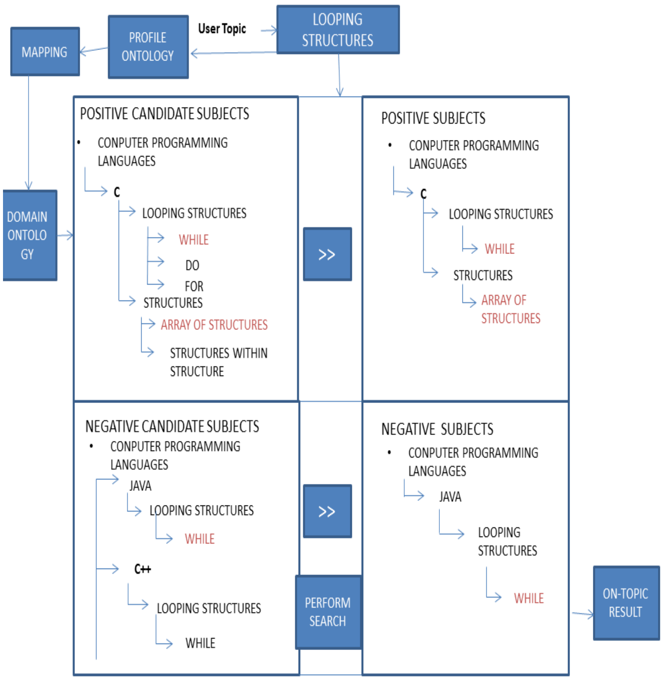

Fig. 3. Ontology Learning Environment

V. OLE TOOL CONSTRUCTIONS

The subjects of user interest are extracted from the KB via user interaction. A tool called Ontology Learning Environment (OLE) is developed to assist users with such interaction.[11]

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 which are relevant to positive subjects. Thus, for a given topic, the OLE provides users with a set of candidate’s subject to identify positive and negative subjects. These candidate subjects are extracted from the KB.The figure3shown is OLE Tool extraction for the topic Looping structures.[12] The topic search was preprocessed by removing the stop words, and stemming and grouping the terms. The subjects shown on the top left panel are the positive candidate subjects shown in the hierarchical form based on the users interesting are which is represented in the users’ profile.[13]

For each search topic the subject “s”, its ancestor and descendants are extracted from the knowledge base if any one of the query terms are present in the subject (e.g. “Looping”,“Structures”). From these positive candidate subjects, the user selects the positive subjects for the topic. The selected positive subjects are shown in top right panel in the hierarchical form.The candidate negative subjects are extracted from the knowledge base based on the user selected positive subject which are shown on the bottom left panel in the hierarchical form.The selected subjects from the negative candidate subjects become negative subjects which is shown bottom right panel in the hierarchical form.

Definition 4.For a given search topic ‘T’ can be formalized as 3-tupleole<T, T+, T-> in the OLE Tool where,

- T is a set of topics;
- T+ is a set of positive topics;
- T- is a set of negative topics.

A. Ontology mining

Algorithm1. Analyzing the topic specificity.

```
input: a search topic ‘T’.
output: links ‘l’.
    Let ‘T’ be the search topic;
    Let ‘O’ and ‘P’ be the domain profile ontology respectively;
            Map ‘O’ and ‘P’;
    Get the  user preferences;
    T+ is a set of positive topics based on mapping;
    T- is a set of negative topics based on positive topics;
    Get the links ‘l’.
```

Topic Specificity

Topic specificity of a subject is investigated, based on the user background knowledge discovered from user profile. From the Definition 4, ontology contains set of positive topics and negative topics. Positive subjects are the concepts relevant to the information need, and negative subjects are the concepts which are relevant to positive subjects. The algorithmic approach for topic specificity is described in algorithm1.

Thus the OLE is personalized because the user selects positive and negative subjects for personal preferences and interests. Thus, if a user searches “India” and plans for a business trip, the user would have different subjects selected and compared to those selected by leisure user, planning for a holiday.Ontology mining is done from knowledgebase based on the user selected positive and negative subjects through OLE Tool.

VI. ARCHITECTURE OF THE ONTOLOGY MODEL

The proposed ontology based personalized web information gathering model is developed to discover the user background knowledge from the knowledge base. The figure4 illustrates the architecture personalized web information gathering model. The ontology is constructed for the specific domain. The users interesting area is collected through questionaries’ and these are stored in the ontology.[14] The topic is preprocessed by removing the stop words, stemming and grouping the terms. The preferred subjects are collected through OLE Tool. The user selects the positive and negative subjects from the generated positive and negative candidate subjects. The web information’s are retrieved from the domain knowledgeresources according to a given topic based on the interested and preferred subjects.[15]

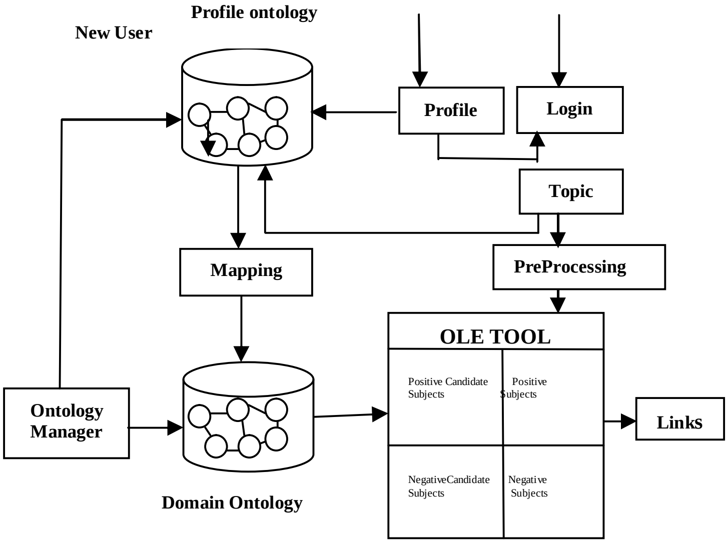

Fig. 4. Architecture of Personalized Web Information Gathering Model

VII. EVALUATION

The proposed personalized web information gathering model was evaluated through precision and recall values and compared with other models.Precision is the ability of a system to retrieve only relevant documents. Recall is the ability to retrieve all relevant documents

User profiles can be categorized into three groups: interviewing, semi-interviewing, and non-interviewing profiles, as previously discussed in Section 2. In an attempt to compare the proposed personalized web information gathering model to the typical models representing these three group user profiles.

1. The proposed web information gathering model discovers the user background knowledge computationally from the domain knowledge base both manually and semi – automatically collected user profiles.
2. The Ontology model discovers user background knowledge semi-automatically collected user profiles.
3. The TREC model that represented the perfect interviewing user profiles. User background knowledge was manually specified by users in this model.
4. The Category model that represented the semi-interviewing user.

The figure5 is a Web information gathering system which collects the web information for the users against a topic search. The proposed model results are compared with the other models.The proposed model is evaluated by collecting more than 50 users profile that have, more range of interested topics. Each user has a different interested area and user profiles are collected and stored.

VIII.RESULTS AND DISCUSSION

The proposed personalized web information gathering model was evaluated through precision, recall, F1score values and compared with other models.

Precision is the ability of a system to retrieve only relevant

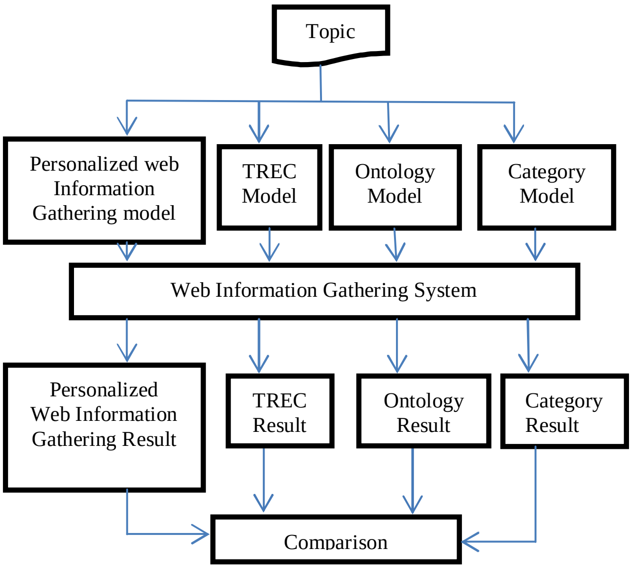
Topic
Personalized web Information Gathering model
TREC Model
Ontology Model
Category Model
Web Information Gathering System
Personalized Web Information Gathering Result
TREC Result
Ontology Result
Category Result
Comparison

Fig. 5. Web Information Gathering Systems

Precision is the ability of a system to retrieve only relevant documents.

$$Precision=\frac{|\{Relevant\}\cap\{Retrieved\}|}{|\{Retrieved\}|}$$

Recall is the ability to retrieve all relevant documents.

$$Recall=\frac{|\{Relevant\}\cap\{Retrieved\}|}{|\{Relevant\}|}$$

F1 Score is a combined measurethat assesses precision/recall trade off. The greater F1 value indicates the better performance.

The personalized ontology model is evaluated through precision, recall, and F1 measure values as follows.

True Positive(TP)=No of relevant links retrieved.
True Negative(TN)=No of irrelevant links not retrieved.
False Positive(FP)=No of irrelevant links retrieved.
False Negative(FN)=No of relevant links not retrieved.

$$\text{Precision}=\frac{No\ of\ relevant\ links\ rtrieved\ (TP)}{TP+FP}$$

$$\text{Recall}=\frac{No\ of\ relevant\ lin\text{(TP)}}{TP+FN}$$

The F1 measure is calculated by

$$F1 = \frac{2 *}{precision + recall}$$

The proposed model is compared with other models like Ontology model, TREC model and Category Model.The table 1 shows the comparison of proposed model with other benchmark models.

Table 1
The precision, recall, F1 measure experimental results

	ONTOLOGY MODEL[1]	TREC MODEL[4]	CATEGORY MODEL[7]	PROPOSED MODEL
Precision	50%	52%	45%	55%
Recall	89%	85%	80%	90%
F1 measure	39%	38%	37%	46%

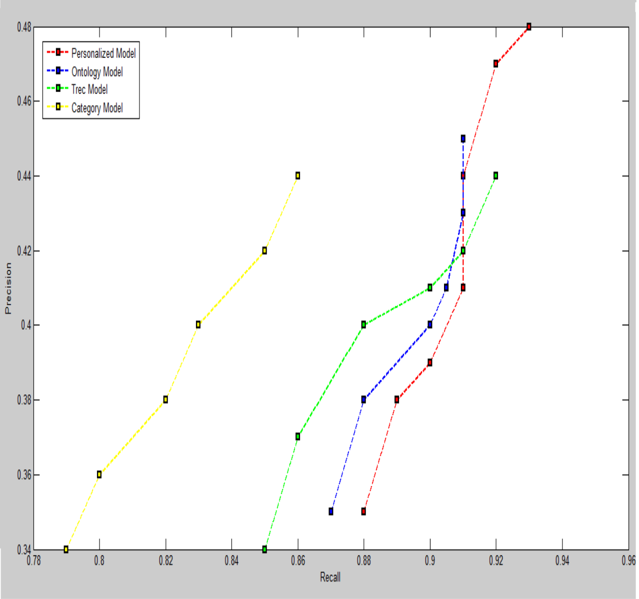

Fig. 6. The precision and recall value results.

According to the results in Fig.6 the personalized web information gathering model was the best, followed by ontology model, TREC and category model. Since the web information's are retrieved from the web based on the both manually and semi-automatically collected user profiles. The retrieved documents are more relevant against a search topic.

IX. Conclusion and Future Work

In this paper, ontology based personalized web information gathering model is proposed to discover the user background knowledge for personalized web search. The benchmark models ontology model-collected user profile semi-automatically, TREC model-collected user profiles manuallyIn the proposed model, the knowledge base is constructed through ontology. It learns the user profiles manually and semi-automatically and discovers the user background knowledge. Ontology mining method specificity is introduced for discovering user background knowledge. The present work retrieves the information from the knowledgebase only based on the user area of interest stored in the profile ontology and preferred subjects in the OLE Tool. The OLE Tool reduces the search time for the users. In future work the information's are retrieved based on the local instance repository. Many of the web documents do not contain the content related descriptors. So, to solve this problem ontology mapping, text clustering and classification is suggested.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] Xiaohui Tao, Yuefeng Li, and Ning Zhong, A "Personalized Ontology Model for Web Information Gathering" *IEEE transactions on knowledge and data engineering*, vol. 23, no. 4, april 2011

[2] X. Tao et.al "Ontology Mining for Personalized Web Information Gathering" IEEE/WIC/ACM Int'l Conf. Web Intelligence, pp. 351-358, 2007.

[3] Sree Latha R., Vijayaraj R., Azhagiya Singam E.R., Chitra K., Subramanian V., "3D-QSAR and Docking Studies on the HEPT Derivatives of HIV-1 Reverse Transcriptase", Chemical Biology and Drug Design, ISSN : 1747-0285, 78(3) (2011) pp.418-426.

[4] J. Trajkova et.al "Improving Ontology-Based User Profiles," Proc. Conf. Recherche d'Information Assistee par Ordinateur (RIAO '04), pp. 380-389, 2004.

[5] S.E. Robertson et.al "The TREC 2002 Filtering Track Report" Proc. Text Retrieval Conf., 2002.

[6] Masthan K.M.K., Aravindha Babu N., Dash K.C., Elumalai M., "Advanced diagnostic aids in oral cancer", Asian Pacific Journal of Cancer Prevention, ISSN: 1513-7368, 13(8) (2012) pp.3573-3576.

[7] Yuefeng Li et.al "Automatically Acquiring Web User Information Needs" IEEE Transactions on Knowledge and Data Engineering, Vol. 18, No. 4, April 2010

[8] X. Tao, Y. Li, N. Zhong, and R. Nayak, "Automatic Acquiring Training Sets for Web Information Gathering," Proc. IEEE/WIC/ ACM Int'l Conf. Web Intelligence, pp. 532-535

[9] Tamilselvi N., Dhamotharan R., Krishnamoorthy P., Shivakumar, "Anatomical studies of Indigofera aspalathoides Vahl (Fabaceae)", Journal of Chemical and Pharmaceutical Research, ISSN : 0975 – 7384 , 3(2) (2011) pp.738-746.

[10] 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.

[11] 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.

[12] A. Sieg, B. Mobasher, and R. Burke, "Web Search Personalization with Ontological User Profiles," Proc. 16th ACM Conf. Information and Knowledge Management (CIKM '07), pp. 525-534, 2007.

[13] 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.

[14] Reddy Seshadri V., Suchitra M.M., Reddy Y.M., Reddy Prabhakar E., "Beneficial and detrimental actions of free radicals: A review", Journal of Global Pharma Technology, ISSN : 0975-8542, 2(5) (2010) pp.3-11.

[16] B Karthik, TVUK Kumar, A Selvaraj, Test Data Compression Architecture for Lowpower VLSI Testing, World Applied Sciences Journal 29 (8), PP 1035-1038, 2014.

[17]M.Sundararajan .Lakshmi,"Biometric Security system using Face Recognition", Publication of International Journal of Pattern Recognition and Research. July 2009 pp. 125-134.

[18].M.Sundararajan," Optical Sensor Based Instrumentation for correlative analysis of Human ECG and Breathing Signal", Publication of International Journal of Electronics Engineering Research, Research India Publication, Volume 1 Number 4(2009). Pp 287-298.

[19].C.Lakshmi & Dr.M.Sundararajan**, "**The Chernoff Criterion Based Common Vector Method: A Novel Quadratic Subspace Classifier for Face Recognition**"** Indian Research Review, Vol.1, No.1, Dec, 2009.

[20].M.Sundararajan & P.Manikandan," Discrete wavelet features extractions for Iris recognition based biometric Security", Publication of International Journal of Electronics Engineering Research, Research India Publication, Volume 2 Number 2(2010).pp. 237-241.

[21].M.Sundararajan, C.Lakshmi & Dr.M.Ponnavaikko, "Improved kernel common vector method for face recognition varying in background conditions", proceeding of Springer – LNCS 6026- pp.175-186 (2010).ISSN 0302-9743.**(Ref. Jor – Anne-II)**