Ontology for Semantic Web using Personalization

K. Anitha Kumari, Assistant Professor, anitha.psgsoft@gmail.com
K.Saranya, PG Student, saranyamurthy88@gmail.com
B.Babu, PG Student, babumvc@gmail.com
Dr.G.Sudha Sathasivam, Professor, gss@cse.psgtech.ac.in
PSG College of Technology.

***Abstract*— Semantic web using personalization support to provide relevant search results for searcher. Ontology is an effective method for modelling digital collections and user contexts. User profiling is commonly employed nowadays to enhance usability as well as to support personalization, adaptivity and other user-centric features. This work aims at creating a user profile ontology that incorporates concepts and properties used to model the user profile. So far we create ontology for contents and location based on user profile. The inbuilt ontology helps to avoid the sparecity problem in the web and retrieve the user relevant information.**

***Keywords*— Sparecity.**

I. INTRODUCTION

The current web needs to evolve into more than just a platform that provides humans with access to information and more than just an infrastructure that allows individuals and organisations to host websites. What is required is for it to become a mechanism that permits machines to meaningfully collect content from miscellaneous sources, process the information and exchange the results with other agents and human users. So, we move onto the semantic web.

An information resource on the Semantic Web will not only contain data, but will also consist of metadata which describe what the data are about. This will allow agents and their human users to identify, collect and process suitable information sources by interpreting the semantic metadata based on the task at hand. They are also free to exchange results and to communicate by sharing such resources. The semantic foundation mentioned above will be provided by ontologies.

Ontology is an explicit specification of concepts and relationships that can exist between them. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge[5].

Finding desired information quickly and exactly becomes crucial and tracking user search behaviour is also difficult in the web search. This motivated to personalization of web search where customizing the user's search environment according to their interests and thus aids the user to identify their information need without much difficulty.

Here we proposed an approach is based on concepts to profile the user interests and preferences of a user. Therefore we have the problem of addressing how to extract and represent concepts from search results of the user. We have in this paper ontology based profiling method in which concepts can be further classified into different types, such as content concepts, location concepts.

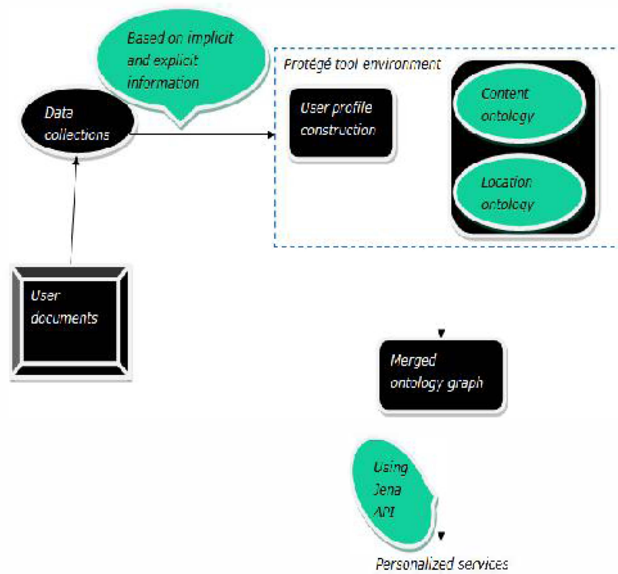

Fig 1 personalized services-flow diagram

The figure1 shows the clear idea about the personalized services. As our first step to create ontology for user profile. Next we focus on two major types of concepts, namely, content concepts and location concepts. A content concept, like a keyword or key-phrase in a Web page, defines the content of the page, whereas a location concept refers to a physical location related to the page.

Creating and manipulating large scale ontologies is a task best left to various tools, and this of course is one of the goals in creating standard metadata languages. We use protégé tool to visualize the meta data and JENA API support to retrieve the semantic information.

The rest of the paper is organized as follows. We review the related work in Section II. In Section III, we present our method for building the content and location ontologies with the user profile. In Section IV, describes how the ontology tool search personalization.

II. RELATED WORK

However, different users may have different information needs even for the same query. For example, a user who is looking for a laptop may issue a query apple to find products from Apple Computer, while a housewife may use the same query apple to find apple recipes. The objective of personalized search is to disambiguate the queries according to the users' interests and to return relevant results to the users.

Zhou[9] proposed a hybrid index structure to handle both content and location-aware queries. The system first detects geographical scopes from web documents and represents the geographical scopes as multiple minimum bounding rectangles (MBRs) based on geographical coordinates. A hybrid index structure is used to index the content and location information of the web documents. A user is required to present their content and location interest in their search queries. A ranker is then employed to rank the search results according to the content and location relevance's using the hybrid index.

Philip[2] propose a model for personalizing search results in a local search engine using a hybrid of profile- and click-based user modelling methods. User profiles are used to compare local search results to the topical interests of users and the specific businesses in which they have shown interest by way of search result "clicks".

Gan[8] suggested that search queries can be classified into two types, content (i.e., non-geo) and location (i.e.,geo). Typical examples of geographic queries are hotels hong kong building codes in seattle and virgina historical sites. A classifier was built to classify geo and non-geo queries, and the properties of geo queries were studied in detail. It was found that a significant number of queries were location queries focusing on location information.

Stein L. Tomesan[6] suggested an approach every ontology entity has an associated FV with a set of relevant terms extracted from the text corpora (the Web). An ontology entity can either be a class or an individual. FVs are built considering both semantics encoded in an ontology and a dominant lexical terminology surrounding the entities in a text corpora. Therefore, a FV constitutes a rich representation of the entities and is related to actual terminology used in the text corpora.

Ao-Jan Su[7] focus on the Google ranking algorithm and design, implement, and evaluate a ranking system to systematically validate assumptions others have made about this popular ranking algorithm. They demonstrate that linear learning models, coupled with a recursive partitioning ranking scheme, are capable of reverse engineering Google's ranking algorithm with high accuracy.

As an example, they manage to correctly predict 7 out of the top 10 pages for 78% of evaluated keywords. Moreover, for content-only ranking, their system can correctly predict 9 or more pages out of the top 10 ones for 77% of search terms. Their ranking system can be used to reveal the relative importance of ranking features in Google's ranking function, provide guidelines for SEOs and webmasters to optimize their web pages validate or disapprove new ranking features, and evaluate search engine ranking results for possible ranking bias.

Developing an automated ranking system that directly queries Google, collects search results, and feeds the results to its ranking engine for learning.

The ranking engine incorporates several learning algorithms, based on training both linear and polynomial models. Using their ranking system, they show that a linear model trained with linear programming and accompanied with recursive partitioning algorithm is able to closely approximate Google's ranking algorithm. In addition, they use their ranking system to analyze the importance of different ranking features to provide guidelines for SEOs and webmasters to improve Google's ranking of web pages.

Our work concentrates to provide relevant results by creating user profile ontology and also create content and location ontology to improve the search accuracy[2]. User interest can explicitly get from the user and implicitly mined from click through data. The content ontology and location ontology is build from the search query. i.e. the extracted web snippets is classified into content concepts and location concepts. This will help to avoid the problem of sparecity in the web search activity.

TABLE 1 SOME USER PROFILE CLASSES

Class	Descriptions of class
Person	User Information like name, e-mail .
Preference	User preferences, For example "likes blue colour" or "dislikes classical music"
Interest	User hobby or work-related interests. For example, "interested in sports", "interested in cooking"
Education	User education issues, including for example university diplomas and languages
Profession	The user's profession

III. OUR WORK

Our approach is based on concepts to profile the interests and preferences of a user. Therefore, an issue we have to address is how to extract and represent concepts from search results of the user. We propose in this paper an ontology-based, profiling method in which concepts can be further classified into different types, such as content concepts, location concepts, name entities, dates etc. As an important first step, we create ontology for user profile.

Next we focus on two major types of concepts, namely, content concepts and location concepts. A content concept, like a keyword or key-phrase in a Web page, defines the content of the page, whereas a location concept refers to a physical location related to the page.

From the figure 1 , we know that the data collection is our first step which can be done by implicit as well as explicit. The click through data history can improve the personalization. But it is our future work. Next we move into profile upgrading which is explained in the following sections.

A. *User profile:*

The "Person" class is the central one in the ontology, as it contains all the user profile characteristics. These may be of simple type, like the user "name" or "date of birth", or may be instances of other ontology classes, like "contacts", etc.

The other classes are used to describe the complex user characteristics. "Living conditions", "Contact", "Education", "Expertise", "Activity" and "Profession" include a set of slots describing the respective aspects of the user's life as well as a time period which represents the duration of that particular aspect. The slot "person" of the "Contact" class has as type an instance of the class "Person". This way, relations between different users may be modelled as well. "Interest", "Preference" also included. In the case of interests, apart from the "type" slot, which is a String, a slot named "interest type" of type "Interest" has been added to allow the creation of interest hierarchies. Table 1 shows an example of user profile classes.

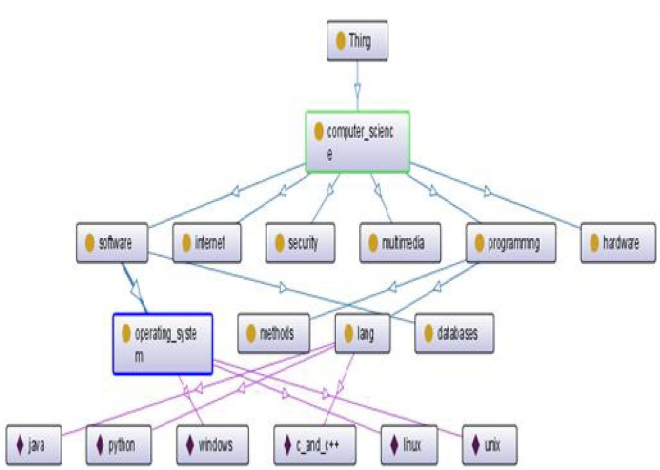

Fig 2 Simple Ontology Representation.

B. *Profile Updating:*

In our personalization framework, we adopt ontologies to model the concept space because they not only can represent concepts but also capture the relationships between concepts. Due to the different characteristics of the content concepts and location concepts, we use different ontology for content and location.

1) *Content ontology:*

The figure 2 shows a simple ontology graph which describes the objects and properties of computer science domain in the protégé tool environment. Similarly we define ontology for all web snippets.

This ontology to maintain concepts and their relationships extracted from search results. We capture the following two types of relationships for content concepts. The first one is similarity relationship which describes the same topical interest. For example the ipad and the iphone is under the apple computer store for the query apple.

The another one is parent child relationship(i.e) more specific concepts often appear with general terms, while the reverse is not true.

2) *Location ontology:*

First, a document usually embodies only a few location concepts. As a result, very few of them co-occur with the query terms in web snippets. To alleviate this problem, we extract location concepts from the full documents. So we create predefined ontology for location concepts.

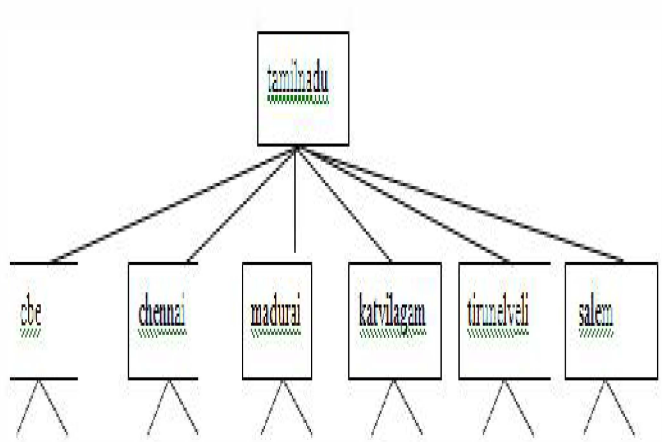

Fig 3 Example for location ontology

The location ontology helps to avoid the sparecity problem(i.e)whenever the user search the query based on the location it may have possible to empty results because that location interior to the city or town. In this situation the location ontology helps to retrieve the nearby location results.

IV. ONTOLOGY TOOLS

Tool support is really important both for the ontology development process (ontology building, annotation, merge, etc.) and for the ontology usage in applications, such as electronic commerce, knowledge management, the Semantic Web, etc.

We have grouped tools in the following clusters:

A. *Ontology development tools*

This group includes tools, environments and suites that can be used for building a new ontology from scratch or reusing existing ontologies. Apart from the common edition and browsing functionality, these tools usually include ontology documentation, ontology exportation and importation from different formats, graphical views of the ontologies built, ontology libraries, attached inference engines, etc.

B. Ontology merge and integration tools

These tools have appeared to solve the problem of merging or integrating different ontologies on the same domain. This need appears when two companies or organizations are merged together, or when it is necessary to obtain better quality ontology from other existing ontologies in the same domain.

C. Ontology evaluation tools

They appear as support tools that ensure that both ontologies and their related technologies have a given level of quality. Quality assurance is extremely important to avoid problems in the integration of ontologies and ontology-based technology in industrial applications. For the future, this effort might also lead to standardized benchmarks and certifications.

D. Ontology-based annotation tools

These tools have been designed to allow users inserting and maintaining (semi)automatically ontology-based mark-ups in Web pages. Most of these tools have appeared recently, along with the emergence of the Semantic Web. Most of them are already integrated in an ontology development environment.

E. Ontology storage and querying tools

These tools have been created to allow using and querying ontologies easily. Due to the wide acceptance and use of the Web as a platform for communicating knowledge, new languages for querying ontologies have appeared in this context.

F. Ontology learning tools

They are used to (semi)automatically derive ontologies from natural language texts. However, these tools will not be revised in this deliverable, since there will be a specific deliverable from Onto Web WP1 (D1.5) concerning them.

Our model takes the protégé and Jena API ontology tool which is used to build domain ontology and retrieval of the semantic information correspondingly.

Protégé is a free, open-source Java-based platform that provides a growing user community with a suite of tools to build domain models and knowledge-based applications with ontologies.

Protégé implements a rich set of knowledge-modelling structures and actions that support the creation, visualization, and manipulation of ontologies in various representation formats. Protégé can be customized to provide domain-friendly support for creating knowledge models and entering data. Further, Protégé can be extended via a plug-in architecture.

Various facilities could be programmatically available via the provided Jena API. Protégé gives support for building the ontologies that are frame-based, in accordance with the Open Knowledge Base Connectivity protocol (OKBC). Also, the tool provides support for OWL ontologies. A special plug-in can be used to generate graph representations of the editing ontologies.

The following figure 4 shows the merged ontology structure of user profile, content and location ontology in the protégé tool environment. Here we use only sample content ontology (computer science) and location ontology (Coimbatore) merged with user profile ontology. Protégé support to merge this simple ontology. In case of large ontology this option is not suitable. So in that case we can use merge and integration ontology tools.

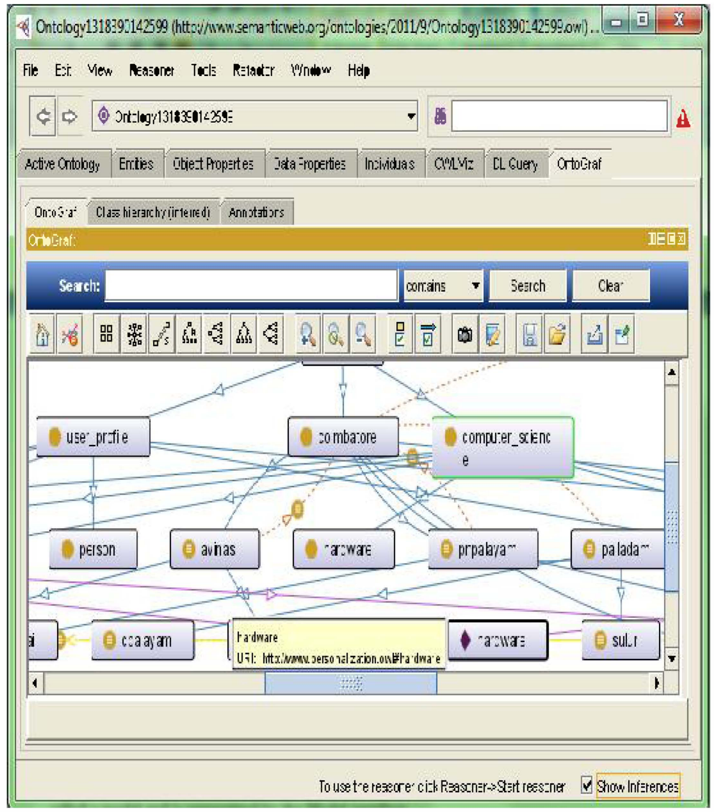

Fig 4 shows the merged ontology structure in the protégé tool environment

Jena is a Java API which can be used to create and manipulate RDF graphs. Jena has object classes to represent graphs, resources, properties and literals. The interfaces representing resources, properties and literals are called Resource, Property and Literal respectively. In Jena, a graph is called a model and is represented by the Model interface. From the graph using Jena API we can retrieve the semantic information.

The meta –knowledge within an ontology can assist an intelligent search engine with processing our query. For example , if the query returns no results ,then the ontology could be used to automatically generalize the query to find nearest partial matches. Even though the cost and space is considered the ontology cab provide the better results.

V. CONCLUSION

Ontology based personalization framework provides the search results based on user preferences. The location ontology support to avoid sparecity problem and the content ontology helps retrieve user interested information with the help of click through data. User profile ontology helps to get user intention explicitly.

The future enhancement that could be made in this project is:

- To include the click history in the user profile.
- To create ontology for dynamic data.

REFERENCES

[1] Kenneth Wai-Ting Leung. Dik Lun Lee., Wang-Chien Lee., "Personalized web search with location preferences", IEEE 26th International Conference on Data Engineering (ICDE). p.p 701 – 712.2010.

[2] Philip O'Brien,Xiao Luo, Tony Abou-Assaleh, Weizheng Gao, Shujie Li, 2009 "Personalization of content ranking in the context of local search" IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT '09. Vol.3, No September/2009.,p.p 532-539., 2009.

[3] Cheqian Chen, Kequan Lin, Heshan Li, Shoubin Dong, "Personalized search based on learning user click history"in Proc. 9th IEEE Int. Conf. on Cognitive Informatics (ICCI'10).,p.p 490-495.

[4] Maria Golemati, Akrivi Katifori, Costas Vassilakis, George Lepouras, Constantin Halatsis, "Creating an ontology for the user profile: method and applications".

[5] Stein L. Tomassen, Darijus Strasunskas"Construction of ontology based semantic-linguistic feature vectors for searching:the process and effect".WI-IAT'09 IEEE/WIC/ACM International Join Conference On web intelligence and intelligent agent technologies,vol 3.,PP 133-138.,2009.

[6] Ao-Jan Su, Y. Charlie Hu, Aleksandar Kuzmanovic, and Cheng-Kok Koh" How to improve your google ranking myths and reality" IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.,vol 1,pp 50-57.,2010.

[7] Q. Gan, J. Attenberg, A. Markowetz, and T. Suel, ."Analysis of geographic queries in a search engine" log,. in Proc. of the International Workshop on Location and the Web, 2008.

[8] Y. Zhou, X. Xie, C. Wang, Y. Gong, and W.-Y. Ma, ".Hybrid index structures for location-based web search",. in Proc. of CIKM Conference,2005.

[9] Georges Dupret,Vanessa Murdock, and Benjamin Piwowarski,"Web search engine evaluation using clickthrough data and a user model",2007.

[10] Ojitha Kumanayaka ,and Nalin Ranasinghe,"Ontology based Web Service Personalization",International Conference on Information and Automation ,pp 69 – 74.,2006.

[11] Ahu Sieg , Bamshad Mobasher and Robin Burke ,"Web search personalization with ontological user profiles".

[12] Mariam Daoud, Lynda Tamine-Lechani and Mohand Boughanem "Using a concept-based user context for search personalization", In Proceedings of WCE'08., vol 1.,pp293-298., 2008.

[13] M. Cristani, R. Cuel, "A Survey on ontology creation methodologies", International Journal on Semantic Web and Information Systems, Vol. 1, No. 2, 49 – 69, 2005.

[14] Jena: http://jena.sourceforge.net/.

[15] Protégé: http://protege.stanford.edu/.

[16] J. Teevan, S. T. Dumais, and E. Horvitz,"Personalizing search via automated analysis of interests and activities", In Proceedings of SIGIR '05,pages 449–456, 2005.

[17] M. Speretta and S. Gauch," Personalized search based on user search histories", In Proceedings of WI '05,pp 622–628, 2005.