European Journal of Computer Science and Information Technology
Vol.1, No. 3, pp. 21-29, December 2013
Published by European Centre for Research Training and Development UK (www.ea-journals.org)

A HYBRID AND PERSONALIZED ONTOLOGY RANKING MODEL USING U-MEANS CLUSTERING AND HIT COUNT

K. SRIDEVI[1] and Dr. R. UMARANI[2]
1. Assistant Professor, Department of Computer Science, Nehru Memorial College, Puthanampatti, Trichy District, Tamilnadu, India[1]
2. Associate Professor, Department of Computer Science, Sri Sarada College for Women, Salem, Tamilnadu, India[2]

ABSTRACT:
Semantic Web is an extension of current Web which offers to add structure to the present Web. Ontologies play an important role in Semantic Web development and retrieval of relevant ontology. Ontology is being represented as a set of concepts and their inter-relationships relevant to some knowledge domain. As the number of Ontology repositories are more on Semantic Web, the problem of retrieving relevant ontologies of the scope arises. Even though there are Semantic Web search engines available, a major problem is that the huge number of results returned and which gives overhead to the searcher to find their need by themselves after going through the long list. This makes time consumption in search and creates dissatisfaction. One solution for this problem is that of maintaining the history of already analyzed, highly relevant and quality results in a log, which can used quickly to respond to the users of the similar type. This places highly relevant results analyzed and stored on the top list when results are presented to the searcher. Personalization and ranking takes care of these approaches. Another solution is the integration of clustering approach which helps in retrieving results from the history or log faster. This paper proposes a hybrid approach that creates the log and retrieves from log when the query is known and there are sufficient entries in the log. This approach imparts convenience to users and reduces the time complexity in finding their relevant needs.

Kewords: Semantic Web, Semantic Search, Ontology, Ontology Ranking, Personalization, Clustering.

1. INTRODUCTION

Conventional direct keyword based information retrieval mechanism cannot meet the growing user retrieval need. The keyword based information retrieval technology fails to integrate information spread over different resources. This technology does not use the semantics, to overcome this problem in Web, the next-generation Web, which Tim Berners-Lee and others call the "Semantic Web," [1] aims at allowing machines to process information automatically and gives focus on semantics of the content. Ontologies [2] offer and efficient way to reduce the amount of information overload by encoding the structure of a specific domain and offering easier and meaningful access to the information for the users. There are number of ontology search engines with which, it is possible to search for the need. The search engines also employ ranking mechanism which makes the user to get their more relevant ontology. But still there are researches to improve the time spent on searching and

relevancy of getting results. Two solutions that help to attain the above said needs are of injecting personalization and clustering approaches.

Web Personalization is a multi discipline area for putting together data and producing personalized output for individual users or groups of users. This approach helps the researchers to improve the efficiency of Information Retrieval (IR) systems.

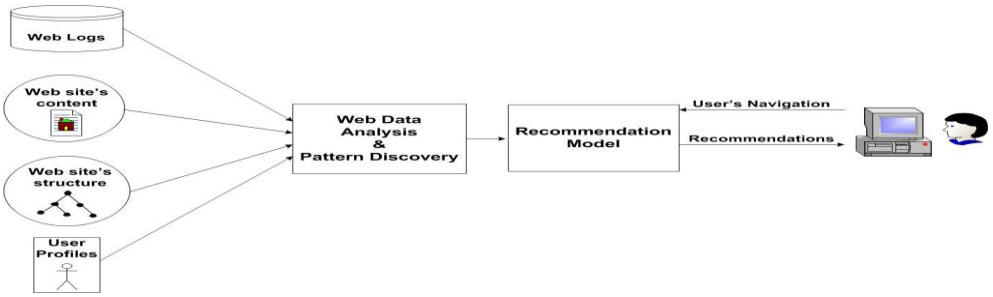

Figure 1. Web Personalization System Process

Figure 1 shows the architecture on how personalization can be done. Personalization aims to find a subset of Web data that matches the interest profile of a user or a group of users. This can be achieved by recommending Web pages or Websites to the users, or by filtering Web pages that are of interest to the users [3]. Clustering is one of the main data analysis techniques and deals with the organization of a set of objects in a multidimensional space into unified groups, called clusters. Each cluster contains objects that are very similar to each other and very dissimilar to objects in other clusters. Cluster analysis aims at discovering objects that have some representative behaviour in the collection. The Information Retrieval community has explored document clustering as an alternative method of organizing retrieval results. Grouping similar documents together into clusters will help the users find relevant information quicker and will allow them to focus their search in the appropriate direction. Various web document clustering techniques are now being used to give meaningful search result on web. There are many document clustering approaches available [4].

The remainder of this paper is organized as follows. This next section reviews with related works carried out on ontology based personalization and clustering approaches in ranking ontologies. Section III describes the proposed system used to rank ontologies with the help of U-Means clustering with personalization. Section IV presents an implementation and experiments done to attain the recommendations. Section V presents the results obtained and analyzes the importance of the proposed system. Section VI explores the conclusion made on implementing the proposed system.

2. RELATED WORKS

There are number of ranking approaches to help the users to reach their need quickly. Swoogle [5] and OntoKhoj [6] rank ontologies also using a PageRank like method that analyses links and referrals between ontologies in the hope of identifying the most popular ontologies. However, the majority of ontologies available on the Web are poorly connected, and more than half of them are not referred to by any other ontologies at all. Poor connectivity would certainly produce poor results. Furthermore, a popular ontology does not necessarily indicate a good representation of all the concepts it covers. Popularity does not necessarily correlate with ‘good’ or appropriate representations of knowledge. Similarity

measures have often been used in information retrieval systems to provide better ranking of query results.

AKTiveRank [7] is a prototype system for ranking ontologies by aggregating a number of graph- analysis measures that use certain structural features of concepts, such as their hierarchical centrality, structural density, and semantic similarity to other concepts of interest. OntoSearch[8] is a hybrid system which searches a local repository and only reverts to Google when it does not have local information. In order to rank ontologies, Content based Ranking [9] system attempts to find a corpus that relates to the domain that the user requires an ontology to represent. The ranking is done according to the number of concept labels in those ontologies which matches a set of terms extracted from a WordNet. The OntoRank algorithm [10] applies the link analyze method. Here two concepts are considered as a reference relationship "if and only if" a relationship exists between the two classes in a relation set [11]. OS_Rank algorithm called Ontology Structure Ranking (OS_RANK) [12] ranks the ontologies based on its semantic relation and structure.

Most of the earlier research efforts in Web Personalization deal with Web Usage Mining [13]. There have been a number of research studies that integrate the web site's content in order to enhance the Web Personalization process [14]. When Web Personalization approaches were embedded with Semantic Web, it yields more effective search response and user satisfaction. User profiles can be used to enrich queries and to sort results at the user interface level [15]. Another approach uses extraction of information on users' navigations from system log files can be used [16]. Middleton et.al. [17] explore the use of ontologies in the user profiling process within collaborative filtering systems. The paper numbered [18] in the reference proposes a method for ranking of search results using fuzzy networks that have been developed using enriched extended user profile. A personalize ontology model [19] is proposed for knowledge representation and reasoning over user profiles. The reference paper [20] proposes a rough k-means clustering algorithm based on properties of rough variable to group gained fuzzy web access patterns and enables users to effectively mine web logs records to discover interesting user access patterns. The paper [21] focuses on web usage mining, the key process of extracting knowledge of user access pattern from web servers. As the result of conclusion, Table 1 gives the benefits and drawbacks of various existing algorithms along with exploring the gains of proposed system.

Table 1. Comparison of Ranking Algorithms with Proposed System

Algorithm	**Benefits**	**Drawbacks**	**Proposed System**
Swoogle's OntoRank	Analyzes semantic web link structure and gives priorities to files with link relationship.	Most onotlogies are poorly inter referenced, link does not reflect the quality of the ontology.	Not only considers link analysis but also uses semantics and personalization with clustering.
AKtive Rank	Analyzes the internal structure of the ontology.	Increases time complexity.	Uses an effective mechanism to reduce time complexity with the help of personalization and clustering.
Content based Ontology Rank	Analyzes internal structure with content similarity	Compares with only class labels and keyword and finds no link analysis.	Computes SCM with Wordnet and analyzes link structure with log file content

	of the ontology.		and grouping.
OS_Rank	Performs semantic analysis on the internal structure of the file	Time consuming	Proposed mechanism provides efficiency in time taken to process.

3. PROPOSED SYSTEM

Ranking is a convenient solution when the searcher met with huge list of results returned. It becomes mandatory to present the accurate results on the top list quickly. The proposed system depends on Personalization and Clustering techniques to return the processed results of Swoogle search engine from the log. Figure 2 shows the architecture of the proposed system.

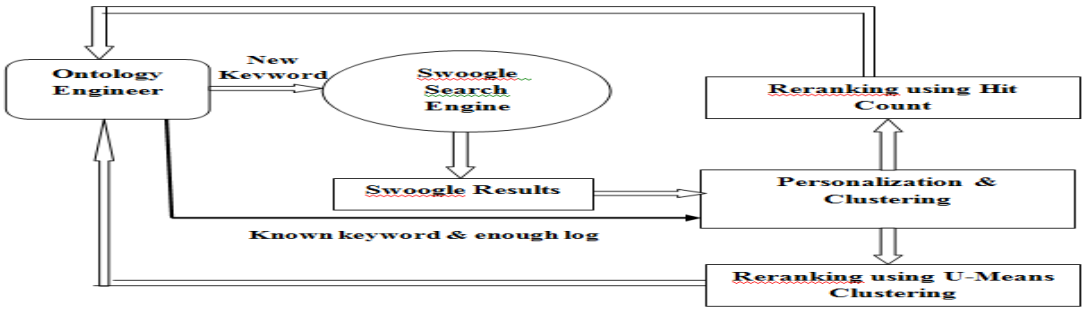

Figure 2. An Architecture of the Proposed System

This system receives the keyword for searching the ontology from the ontology engineer, if the keyword is new, which will be submitted to the Swoogle search engine and when hit is made, personalization and clustering are carried out. Reranking is done using U-Means Clustering and hit Hit Count. Finally, the revised result is returned back to the ontology engineer. The algorithms used for U-Means clustering and ranking based on hit count are presented below.

Algorithm for U-Means Clustering:

1. Get Keyword from Ontology Engineer.
2. If keyword is new, submit to Swoogle.
3. Retrieve results from Swoogle.
4. When hit is made, Create log to personalize searcher's need.
5. If keyword is known & there is enough log, access to log.
6. Retrieve URLs of similar users from log.
7. Get links and download URLs.
8. Compute Class and Semantic Closeness Measure.
9. Rerank and rearrange the URLs.
10. Return results(URLs) to Ontology Engineer.

Algorithm for Ranking based on Hit Count:

1. Get Keyword from Ontology Engineer.
2. If keyword is new, submit to Swoogle.
3. Retrieve results from Swoogle.
4. When hit is made, Create log to personalize searcher's need.
5. If keyword is known & there is enough log, access to log.
6. Compute hit count for the URLs that match with the keyword.
7. Rerank and rearrange the URLs based on hit count.
8. Return results(URLs) to Ontology Engineer.

This system performs a hybrid analysis to rank the ontologies by using U-Means Clustering and Hit Count. This work differs from other existing ranking model by giving more importance to user convenience and to provide them an ease way of attaining their needs quickly.

4. RESULTS AND FINDINGS

This system is implemented in Net Beans environment using Java. Wordnet database is used to supply the synonym sets for the keyword and class labels. The ontologies are processed using Jena API. Experiment has been done by using number of keywords. Figure 3 is the screen shot of the first page result returned by Swoogle for the keyword "networking".

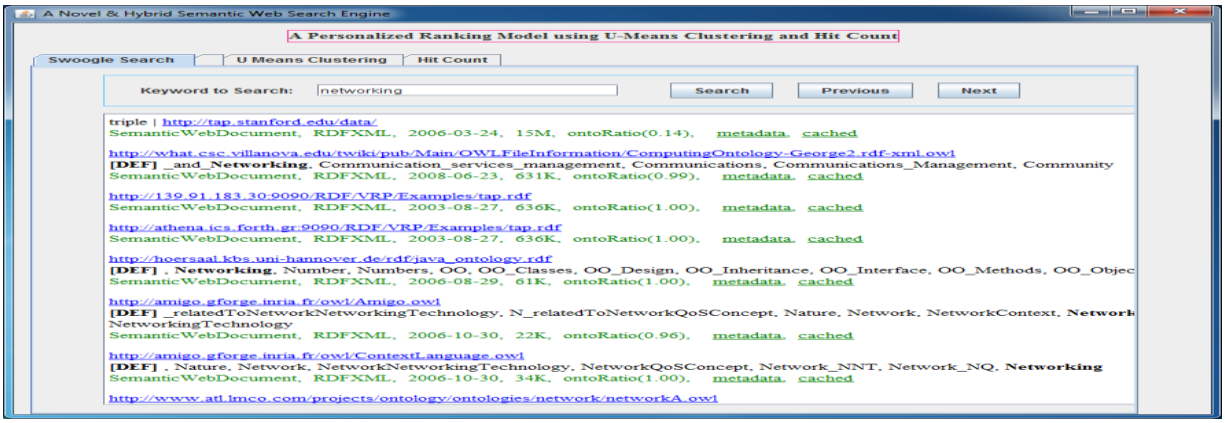

Figure 3. Result of Swoogle for the keyword 'networking'

When the hit is made, the corresponding URL and its metadata are stored in the log. Log data are being retrieved when the request comes from the similar type user and the screen shot of a sample request to a keyword "networking" is shown in figure 4 and the corresponding graphical representation is displayed in the figure 5.

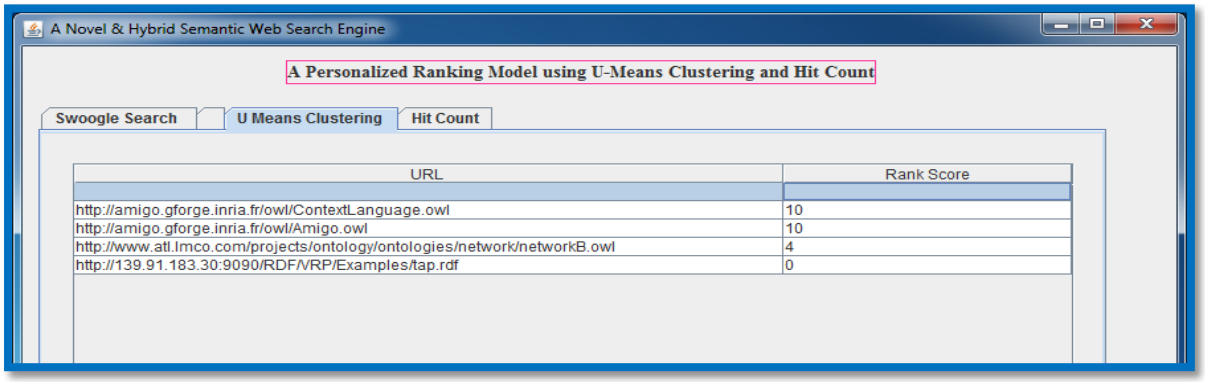

Figure 4. Computed Ranking Measure using U-Means Clustering

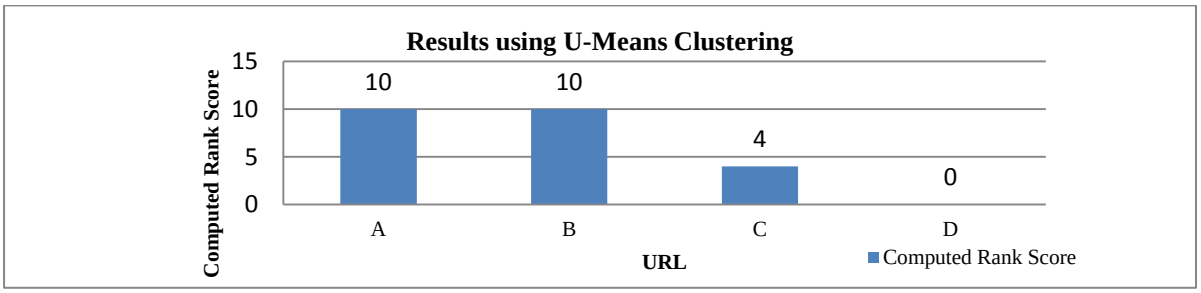

Figure 5. Graphical Representation of Computed Ranking Measure using U-Means Clustering

The owl files on the horizontal line are presented in table 2.

Table 2. Owl files from Swoogle used for Ranking

URL ID	URL	Rank Score
A	http://amigo.gforge.inria.fr/owl/ContextLanguage.owl	10
B	http://amigo.gforge.inria.fr/owl/Amigo.owl	4
C	http://www.atl.lmco.com/projects/ontology/ontologies/network/networkB.owl	4
D	http://139.91.183.30:9090/RDF/VRP/Examples/tap.rdf	0

The screen shot of the URLs based on the hit count along with date and time when last hit was made are displayed in figure 6. It shows the result after sorting based hit count. The graphical view of the same is shown in figure 7. The URLs at the horizontal line of the graph is explored in table 3.

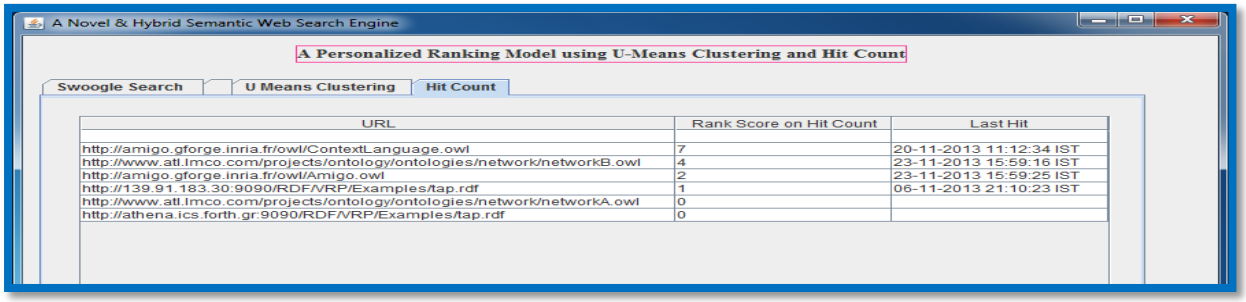

Figure 6. Results on Hit Count

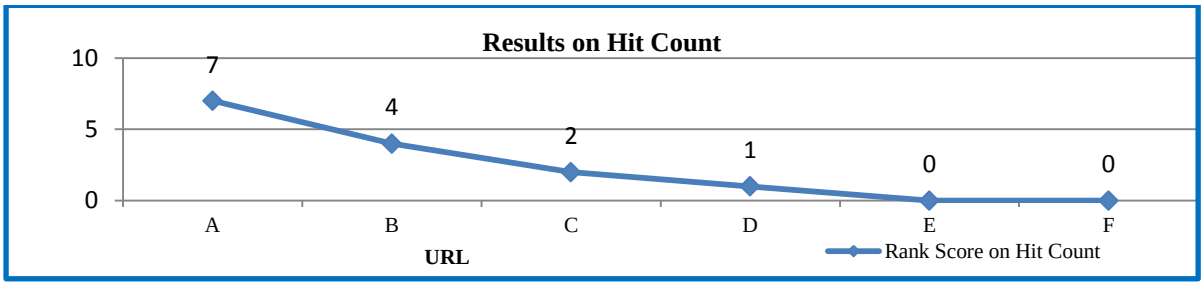

Figure 7. URLs Hit Count

Table 3. URLs used in Hit Count graph

URL ID	URL
A	http://amigo.gforge.inria.fr/owl/ContextLanguage.owl
B	http://www.atl.lmco.com/projects/ontology/ontologies/network/networkB.owl
C	http://amigo.gforge.inria.fr/owl/Amigo.owl
D	http://139.91.183.30:9090/RDF/VRP/Examples/tap.rdf
E	http://www.atl.lmco.com/projects/ontology/ontologies/network/networkA.owl

F	http://athena.ics.forth.gr:9090/RDF/VRP/Examples/tap.rdf

5. DISCUSSION

Time taken to retrieve the ranked result URLs for the number keywords such as networking, publication, research and university are observed. Figure 8 exposes the graphical representation.

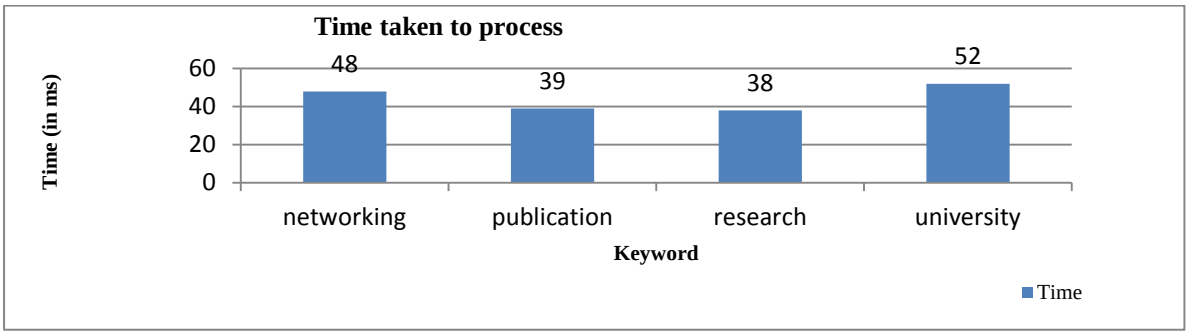

Figure 8. Time taken to process proposed system

The proposed system is tested using with four groups, having different number of users in each group. The system responds well even the number of user increases in count. Figure 9 depicts the average time taken to show the results for the keyword "networking" for different count of users.

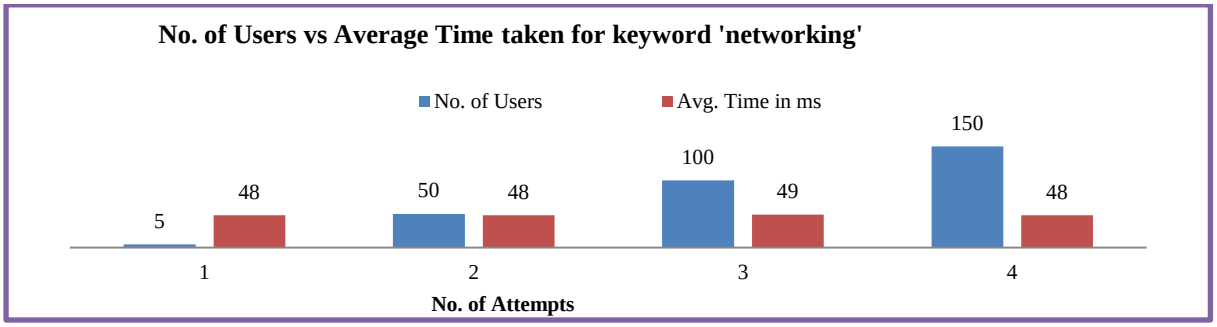

Figure 9. Average time taken to respond for the keyword "networking" for N number of users

The result shows that the only relevant and reliable results are presented to the ontology engineer and time taken to display the results is more or less same and quicker even though the request is made from any number of users. Thus the proposed system increases the search engine system throughput.

6. CONCLUSION & FUTURE RESEARCH

Even though there are number of Semantic Web search engines to search for the ontology, the long list of results makes the user inconvenient and forces them to go through the results that match with the needs. This paper reveals out a solution to this problem by presenting them the highly relevant documents on the top list which makes the searcher to catch their needs quickly. This saves the time consumed by the user. To gain this, this paper uses Personalization and U-Means Clustering approaches to store the processed results of previous users and form clusters of similar users to speed up the ease of access. This work also explores results based on hit count. This gives another option for searchers to view the results

based on the popularity of the page. This proposed system is implemented by taking user convenience and response as primary aspect. This work can be extended by improving the relevancy of the results and reducing the time taken to process.

REFERENCES

[1] Stumme.G, Hotho.A, Berendt.B, "*Semantic WebMining: State of the art and future directions*", Web Semantics: Science, Services and Agents on the World Wide Web 4(2) 2006 124-143 Semantic Grid – The Convergence of Technologies.

[2] http://en.wikipedia.org/wiki/Ontology

[3] Xiaohui Tao, Yuefeng Li, and Ning Zhong, "Senior Member, IEEE, "*A Personalized Ontology Model for Web Information Gathering*", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 4, APRIL 2011.

[4] Oikonomakou, Nora, and Michalis Vazirgiannis. "*A Review of Web Document Clustering Approaches.*" Data Mining and Knowledge Discovery Hankbook.

[5] L. Ding, T. Finin, A. Joshi, R. Pan, R. S. Cost, Y. Peng, P. Reddivari, V.C. Doshi, and J. Sachs, 2004. "*Swoogle: A search and metadata engine for the semantic web.*", In Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management.

[6] C. Patel, K. Supekar, Y. Lee, and E. Park. "*Ontokhoj: A semantic web portal for ontology searching, ranking, and classification*", In Proc. 5th ACM Int. Workshop on Web Information and Data Management, pages 58–61, New Orleans, Louisiana, USA, 2003.

[7] Harith Alani, Christoper Brewster and Nigel Shadbolt, 2007. "*Ranking Ontologies with AkTive Rank"*.

[8] Thomas E, Zhang Y, Sleeman D, Preece A, McKenzie C and Wright J. OntoSearch: a Service to Support the Reuse of Ontologies. *Demos and Posters of the 2nd European Semantic Web Conference (ESWC)*, 2005.

[9] Matthew jones, Harith Alani. July 2006. "*Content-based Ontology Ranking*". 9th International Protégé Conference.

[10] Edward Thomas, Harith Alani, Derek Sleeman, Christopher Brewster. 2007. "*Searching and Ranking Ontologies on the Semantic Web*".

[11] Zhiguo Ding, Bing-ru Yang, De-quan ZhengjieDuan. 2010. "*Improved Ontology Ranking Algorithm Based on Semantic Web"*,3rd IEEE International conference on Digital Object Identifier.

[12] Wei Yu, Junpeng Chen. 2009. "*Ontology Ranking For the Semantic Web"*. Third International Symposium on Intelligent Information Technology Application.

[13] M. Albanese, A. Picariello, C. Sansone, L. Sansone, "*A Web Personalization System based on Web Usage Mining Techniques"***,** in Proc. of WWW2004, May 2004, New York, USA.

[14] B. Mobasher, H. Dai, T. Luo, Y. Sung, J. Zhu, "*Integrating web usage and content mining for more effective Personalization"*, in Proc. of the International Conference on Ecommerce and Web Technologies (ECWeb2000), Greenwich, UK, September 2000.

[15] Koutrika, G., Ioannidis, Y.: "*A Unified User Profile Framework for Query Disambiguation and Personalization*", in "Workshop on New Technologies for Personalized Information Access", held in conjunction with the 10th International Conference on User Modeling, pp. 44–53 (2005).

[16] Paulakis, S., Lampos, C., Eirinaki, M., Vazirgiannis, M.: Sewep, "*A Web Mining System Supporting Semantic Personalization*", in "15th European Conference on Machine Learning

and 8th European Conference on Principles and Practice of Knowledge Discovery in Databases", LNCS, vol. 3202, pp. 552-554, Springer (2004).
[17] S.E. Middleton, N.R. Shadbolt, D.C. De Roure, "*Ontological User Profiling in Recommender Systems"*, ACM Transactions on Information Systems (TOIS), Jan. 2004/ Vol.22, No. 1, 54-88.
[18] Bhaskara B Rao and Valli Kumari Vatsavayi, "*Concept based Ranking of Results using an Ontology and Fuzzy Network for a Personalized Web Search Engine*", International Journal of Computer Applications 81(13):17-24, November 2013, published by Foundation of Computer Science, New York, USA.
[19] Ms. Amruta Y. Kokate, Mr. Bhushan D. Nighot, Ms. Chetana N. Sadare, Ms. Priyanka P. Mangalkar, Prof.G.S.Mate, "*Personalized Ontology Model for Web Information Gathering*", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 2, February 2013.
[20] Peilin Shi, "*An Efficient Approach for Clustering Web Access Patterns from Web Logs"*, International Journal of Advanced Science and Technology, Volume 5, April, 2009.
[21] Akshay Kansara, Swati Patel, "*Improved Approach to Predict user Future Sessions using Classification and Clustering"*, International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064, Volume 2 Issue 5, May 2013.

BIOGRAPHY

K. Sridevi received her M.Phil. degree from Bharathidasan University, Trichy in Dec 2004. She is presently working as Assistant Professor in Computer Science Department, Nehru Memorial College, Puthanampatti, Trichy District, Tamil Nadu, India. She is currently guiding M.C.A., M.Phil. research scholars of Bharathidasan University. She has published 5 papers in various International Journals as the part of Web mining research activity. Her area of research is Data and Web Mining.

Dr. R. Uma Rani received her Ph.D. degree from Periyar University, Salem in the year 2006. She is a rank holder in M.C.A. from NIT, Trichy. She has published around 70 papers in reputed journals and National and International conferences. She has received the best paper award from VIT, Vellore, Tamil Nadu in International Conference. She has done one MRP funded by UGC. She is rendering her service as Resource Person in various National and International Conferences. She has produced 2 Ph.D. scholars. She has guided 34 M.Phil. scholars. Her areas of interest include Information Security, Data Mining, Fuzzy Logic and Mobile Computing.