

# **A Novel and Hybrid Ontology Ranking Framework using Semantic Closeness Measure**

**K. Sridevi**

Assistant Professor  
Department of Computer Science  
Nehru Memorial College  
Puthanampatti, Trichy District  
Tamilnadu, India

**R. Umarani, Ph.D**

Associate Professor  
Department of Computer Science  
Sri Sarada College for Women  
Salem  
Tamilnadu, India

## **ABSTRACT**

Semantic Web is a Web that adds more meaning to the Web documents in order to access knowledge instead of unstructured material and also allowing knowledge to be processed automatically. One of the methods to achieve this is of using Ontology. The Ontology defines the terms and the relations among the terms on a domain. There are number of Ontology repositories present. When this increases day by day, the need for getting relevant ontology for the search keyword also increases. Even though there are number of semantic web search engines, Swoogle is placed first, which ranks the ontologies using an adaptation of Google's Page Rank scoring method. A major drawback with this system is that many ontologies are poorly inter-referenced, which does not reflect the quality of the ontologies. This paper reviews the methodologies used in Swoogle for computing rank score and proposes Semantic Closeness Measure (SCM) which has not been employed in any other ontology ranking algorithms. This work develops a hybrid ranking system to rank the ontologies better than Swoogle and other ontology search engines. The results confirm that the proposed system places the highly relevant and quality ontologies on the top list by reranking the Swoogle's results. This ranking framework enables the searcher to get relevant results quickly and reduces time in searching the long list of results.

## **Keywords**

Semantic Web, Semantic Search, Ontology, Ontology Ranking, Semantic Closeness Measure

## **1. INTRODUCTION**

By encouraging the inclusion of semantic content in web pages, the Semantic Web [1] aims at converting the current web of unstructured documents into a web that consists of meaningful data. Most significant way on representing Web information on Semantic Web is through Ontology [2]. An ontology is a machine processable representation that contains the semantic information of a domain. This representation helps to extract accurate knowledge quickly. As the number of publicly available ontologies increases, it is required to make use an effective search engine. Some ontology search engines have been developed that can provide lists of ontology that contain specific search terms. Examples of such are Swoogle [3] and OntoSearch [4]. Swoogle is a search engine for Semantic Web ontologies, documents, terms and data present on the Web. Google is better than the other search engines because of the effectiveness of its page ranking approach. As the number of ontologies found by such search engines increases, there is a requirement for a proper ranking method to

order the returned lists of ontologies in terms of their relevancy to the keyword. A proper ranking of ontologies could save the user a lot of time and effort in searching. This work proposes a novel ontology ranking framework to gain the above said benefit.

The remainder of this paper is organized as follows. The next section reviews with related works concerning ontology ranking. Section III describes the proposed system used to rank ontologies returned by Swoogle search engine. Section IV presents an implementation and experiments carried out. Section V reveals out the results obtained and discusses with the benefits out of the experiment. Section VI explores the conclusion made on using the proposed system.

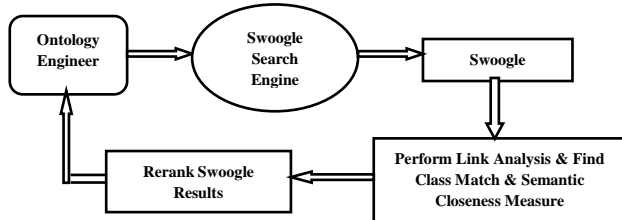
## **2. RELATED WORKS**

Ranking has always been at the heart of information retrieval. This became even more obvious with the massive size of the web and its continuous expansion. Google implements ranking with the help of PageRank [5] method based on hyperlink analysis. Swoogle and OntoKhoj [6] rank ontologies 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 PageRank results. There are various researches done on ranking ontologies. AKtive Rank [7] does ranking based on the concept covered in the internal structure of ontology. It has pitfall of increasing time complexity. Content-based Ontology Ranking [8] places highly relevant document in higher rank based on selecting the document that has more class labels matches the words in the retrieved documents. But if the search term is very specific, the retrieval of relevant document is difficult. OntoRank [9] enlarges the scope of the synonym and related words in terms of extension. This overcomes the limited search based on only the user keywords. The problem in this ranking is that most ontologies are poorly inter-referenced and this will be reflected in the quality of the ontology retrieval. OS\_Rank [10] method is based on searching both ontology structure and semantic analysis. The pitfall of this is that this process is time consuming and very tedious. An analysis of various ontology ranking algorithms is done in the paper [11]. The proposed ranking system not only considers link analysis and use of semantics but also enriches the use of semantics with the help of semantic closeness measure to improve the quality and precision of ranking results.

### 3. PROPOSED SYSTEM

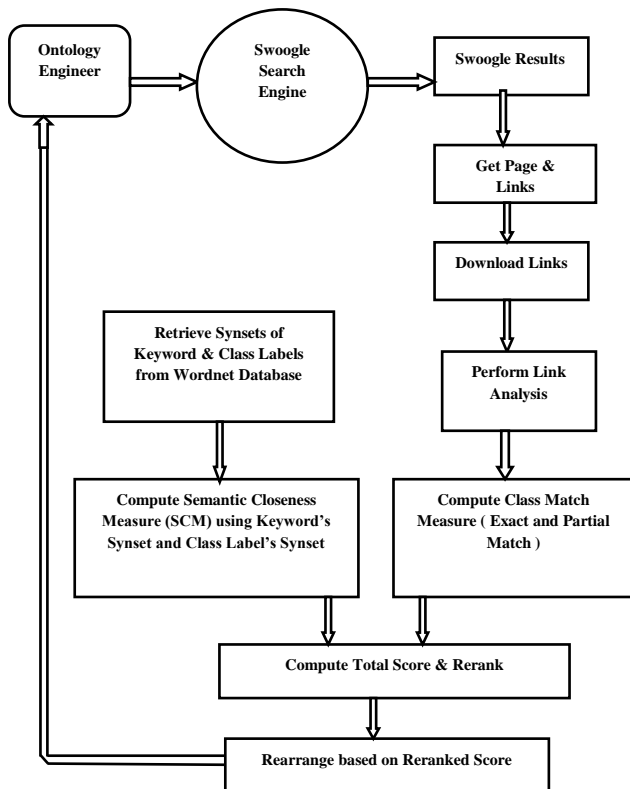
Ranking is very much needed due to the case of having enormous count of results returned by search engine. Because of getting huge set of result, it becomes important to place the highly relevant result on the top to save the time of the users and this to improve the quality of the search engine. To gain this, the proposed system revises ranking model used in Swoogle and other existing ranking models.

Figure 1 shows the overall architecture of the proposed system.



**Figure 1: Overall Proposed System Architecture**

This work is contributed to the engineers who search for the ontology on Semantic Web. This system receives the keyword for searching the ontology from the ontology engineer, which will be submitted to the Swoogle search engine. The results returned by the engine are processed to rerank the Swoogle results. Finally, the revised result is returned back to the ontology engineer. Figure 2 depicts the detailed system architecture which explores the reranking computational method used in the proposed system.



**Figure 2: Detailed Proposed System Architecture**

To rerank the ontologies, this system primarily gets the links of all files returned by Swoogle search engine and downloads the links. This system performs the method Link Analysis that analyzes links and referrals between ontologies in the hope of

finding the popular ontologies and computes Class Match Measure (CMM) for both exact and partial match. Then computes Semantic Closeness Measure (SCM) to identify and analyze the quality of the ontologies. Synonym sets are retrieved from WordNet database. Finally computes the total score by adding all the previous scores and reranks the ontology files. The ontologies are rearranged according the reranking score computed and returned to the ontology engineer.

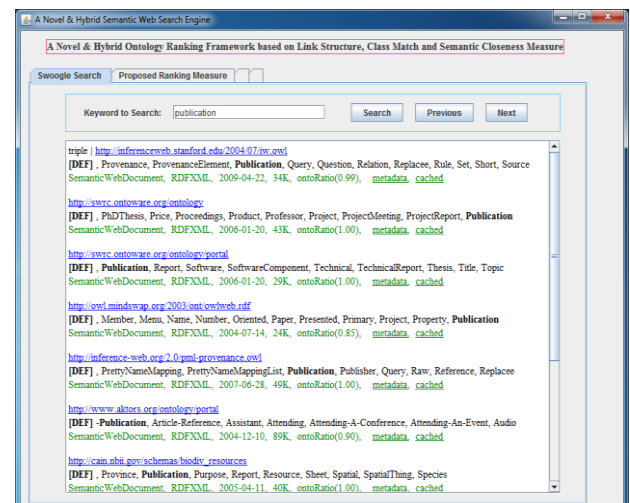
1. Get keyword from ontology engineer.
2. Submit keyword to Swoogle.
3. Receive the results.
4. Get the links and download the files.
5. Perform Link Analysis to find the links & referrals between ontologies.
6. Compute Class Match Measure (CMM).
7. Compute Semantic Closeness Measure (SCM).
8. Compute total score.
9. Rerank the resultant ontologies.
10. Rearrange and return ontologies to ontology engineer.

**Figure 3: Proposed Reranking Algorithm**

As in figure 3, this system performs a hybrid analysis to rank the ontologies by implementing both link analysis and semantics using class match measure used by Swoogle and a proposed computing measure called, Semantic Closeness Measure (SCM). This method focuses to perform in depth analysis by comparing the synonym set of class labels of ontology with synonym set of search keyword.

### 4. IMPLEMENTATION & EXPERIMENTS

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. The following is the first page result returned by Swoogle for the keyword “publication”. The screenshot is as in Figure 4.



**Figure 4: Result of Swoogle for the keyword ‘publication’**

There are some results returned by Swoogle on the first page can't be opened due to non availability of the resource, which consumes searchers time in vain. This system eliminates these files in the result and improves the reliability. Link analysis has been carried out to consider the ontology files returned from the Swoogle and also the linked referrals of each ontology files as represented below.

$$\sum_{i=1}^m O_i + \sum_{j=1}^r O_j$$

where  $m$  is the number of ontology files returned from Swoogle and  $r$  is the linked ontologies. Figure 5 given below is the pseudo code for retrieving URLs and links returned from Swoogle search engine.

1. Define String to retrieve results from Swoogle search engine.
2. Define url.
3. Open url connection using the method `openConnection`.
4. Read url files using the method `getInputStream()`.
5. Then read the content using `readLine()` method.

**Figure 5: Retrieving URLs and links returned from Swoogle**

Class Match Measure computes the occurrences of set of terms contained in the class labels of all the ontologies as represented below.

$$\Phi(t) \in \sum_{i=1}^m \sum_{j=1}^n cl_j(O_i)$$

where  $m$  is the number of ontology files,  $n$  is the class labels present in each ontology and  $t$  is the search keyword term. The pseudo code is given in figure 6.

1. Use `JenaOWLModel` to read the owl url using the method `createJenaOWLModelFromURI`.
2. Retrieve the classes using the method `getUserDefinedOWLNamedClasses`.
3. The following is the pseudo code to retrieve the class labels.  
Set iterator to classes{  
Use `next()` method of `OWLNamedClass`  
Use `getBrowserText()` and retrieve class label}
4. Compute partial and total match between class label and keyword string.

**Figure 6: Computing Class Match Measure**

WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into a synonym set, called synsets. This WordNet database is used to retrieve the synonym set of search keyword as represented in the pseudo code given in the figure 7.

1. Use `WordNetDatabase` and its method `getFileInstance()`.
2. Call `getSynsets` method by passing search keyword to get synonym sets.

**Figure 7: Retrieving synsets using WordNet**

Semantic measure is computed using this synonym set of the search keyword with all the class labels available on the

ontologies returned from the Swoogle search engine as shown below.

$$\sum_{k=1}^{qs} \Phi(t_k) \in \sum_{i=1}^m \sum_{j=1}^n cl_j(O_i)$$

where  $m$  is the number of ontology files returned from Swoogle,  $n$  is the class labels present in each ontology,  $t$  is the search keyword term and  $qs$  is the synonym set of search term. Following figure 8 depicts the pseudo code.

```

For each synsets of search keyword
    For each class label in the owl
        If class label that equals or contains synset
        of search keyword.
            Count semantic measure.
    End for
End for

```

**Figure 8: Computing Semantic Measure**

The proposed system focuses on comparing the synonym set of search term with the synonym set of every class label in all the ontology. This identifies more relevant URL. The representation is given below.

$$\sum_{k=1}^{qs} \Phi(t_k) \in \sum_{i=1}^m \sum_{j=1}^{cls} cl_j(O_i)$$

where  $m$  is the number of ontology files returned from Swoogle,  $cls$  is the synonym sets of each class label present in each ontology,  $t$  is the search keyword term and  $qs$  is the synonym set of search term.

The pseudo code of the above is represented below in figure 9.

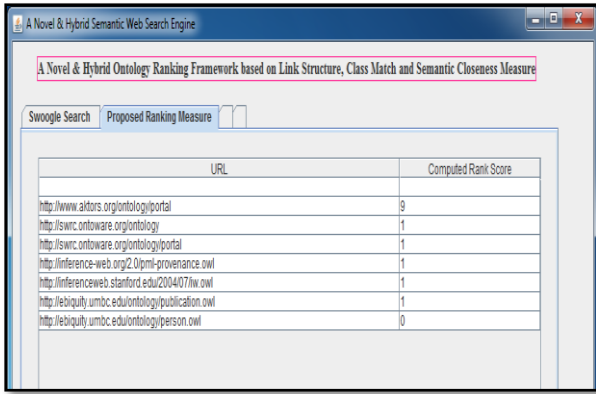
```

For each synsets of search keyword
    For each class label in the owl
        Get synset of class labels of each
        ontology using getSynsets().
        If synset of class label that equals or
        contains synset of search keyword.
            Count semantic closeness
            measure.
    End for
End for

```

**Figure 9: Computing Semantic Closeness Measure**

Computed ranking measures after rearrangement are sent to the ontology engineer. The screenshot is shown in figure 10 and the graphical representation of the rank score is presented in figure 11. The result of first page of Swoogle and the corresponding URLs ranking score of the proposed system is given.



URL	Computed Rank Score
http://www.aktors.org/ontology/portal	9
http://swrc.ontoware.org/ontology	1
http://swrc.ontoware.org/ontology/portal	1
http://inference-web.org/2.0/pml-provenance.owl	1
http://inference-web.stanford.edu/2004/07/iw.owl	1
http://ebiquity.umbc.edu/ontology/publication.owl	1
http://ebiquity.umbc.edu/ontology/person.owl	0

Figure 10: Computed Ranking Measure using SCM

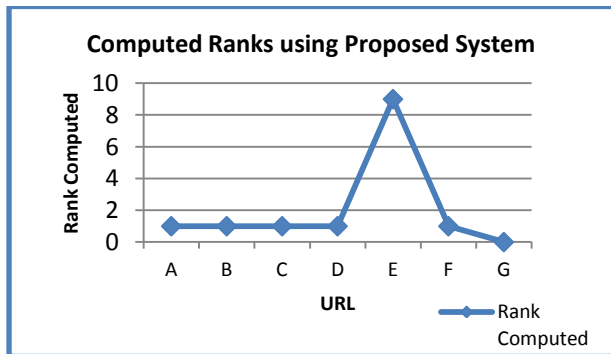


Figure 11: Graphical Representation of Computed Ranking Measure using Proposed System

The owl files on the horizontal line are given in table 1.

Table 1. Owl files from Swoogle used for ranking

URL ID	URL
A	http://inferenceweb.stanford.edu/2004/07/iw.owl
B	http://swrc.ontoware.org/ontology
C	http://swrc.ontoware.org/ontology/portal
D	http://inference-web.org/2.0/pml-provenance.owl
E	http://www.aktors.org/ontology/portal
F	http://ebiquity.umbc.edu/ontology/publication.owl
G	http://ebiquity.umbc.edu/ontology/person.owl

## 5. RESULTS AND DISCUSSION

The proposed system rearranges the ranking order of Swoogle, which places the URL E in first place because of having higher computed value received after implementing the proposed system. The content of the file shows the importance and relevancy for the search keyword. This enables to present more relevant URLs at top list which makes the search to capture the need quickly and accurately. The reranked order for the search keyword 'publication' is compared with Swoogle's rank order, the result is shown in the table 2 and graphical representation is presented in the figure 12.

Table 2. Swoogle and Proposed System Rank Order

URL ID	Swoogle Rank Order	Proposed System Rank Order
A	1	5
B	2	2
C	3	3
D	4	4
E	5	1
F	6	6
G	7	7

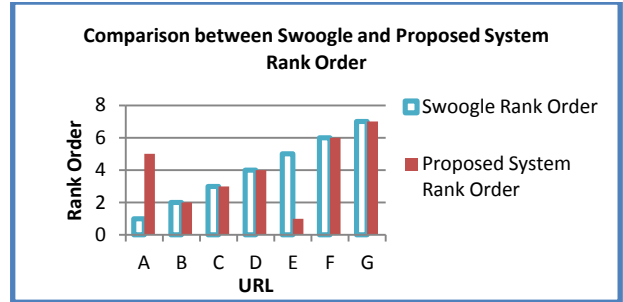


Figure 12: Comparison between Swoogle and Proposed System Rank Order

More number of analyses has been done to confirm the performance of the proposed system. Some of the other keywords are 'networking', 'research', 'university', and so on. Result page and ranking score of the proposed system for the keyword 'networking' is shown the figures 13 and 14.

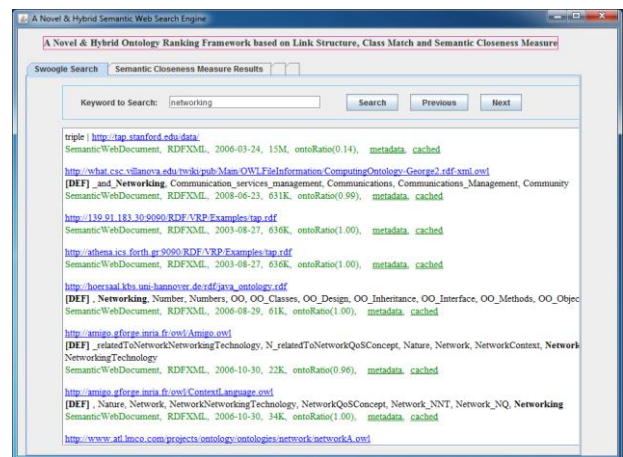
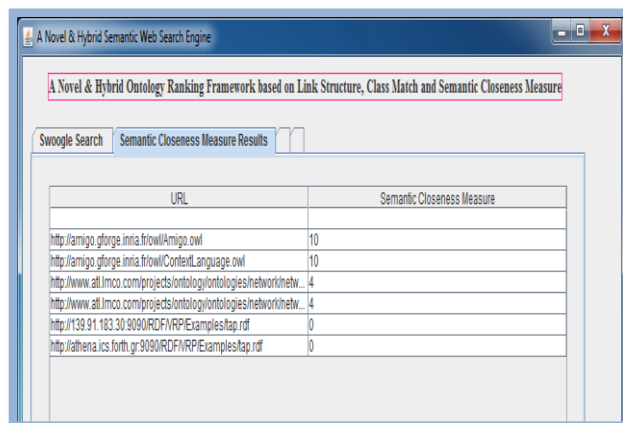


Figure 13: Result page for the keyword 'networking'.



URL	Semantic Closeness Measure
http://amigo.gforge.inria.fr/owl/Amigo.owl	10
http://amigo.gforge.inria.fr/owl/ContextLanguage.owl	10
http://www.allimco.com/projects/ontology/ontologies/networking	4
http://www.allimco.com/projects/ontology/ontologies/networking	4
http://139.91.183.30/9090/RDF/VRP/ExamplesTap.rdf	0
http://athena.ics.forth.gr/9090/RDF/VRP/ExamplesTap.rdf	0

Figure 14: Computed Ranking Measure using SCM

In the above, the synonym set of 'network' is {net, web, mesh, meshing, meshwork, electronic network}. Semantic Closeness Measure identifies the class label "world wide web" because one of the synonym set of "world wide web" is {web}. With the help of the previous semantic measure, this identification is not possible.

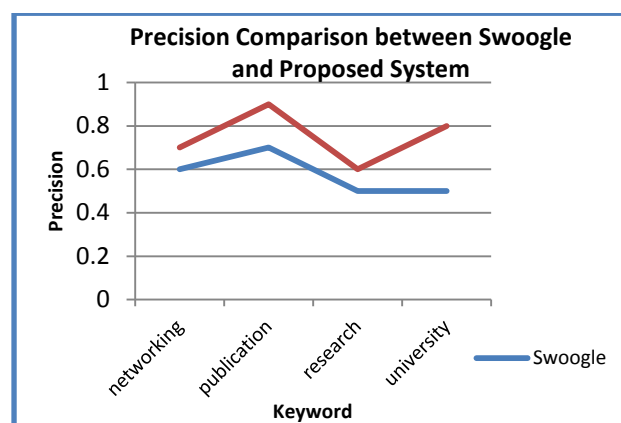
The measure ‘Precision’ is used to evaluate the performance of document retrieval. Precision ‘P’ is defined as the proportion of retrieved documents that are relevant.

$$\text{Precision} = \frac{Ra}{A}$$

where  $Ra$  is the relevant documents retrieved and  $A$  is the retrieved documents. This system also uses the same measure to evaluate the precision of results. The figure 15 shows the precision comparison between Swoogle and proposed system for keywords such as *networking*, *publication*, *research* and *university* and the result is explored in table 3. Out of 10 URLs returned from Swoogle, the URLs that respond are 6, 7, 5 and 5 respectively for the search keywords specified above. Since the proposed system focuses only on owl files, the precision of the system is as given in the table 3.

**Table 3. Precision Comparison between Swoogle and Proposed System**

Keyword	Ra/A	Swoogle	Ra/A	Proposed System
Networking	6/10	0.6	4/6	0.7
Publication	7/10	0.7	6/7	0.9
Research	5/10	0.5	3/5	0.6
University	5/10	0.5	4/5	0.8



**Figure 15: Precision Comparison between Swoogle and Proposed System**

## 6. CONCLUSION

This proposed system analyzes the pitfalls in Swoogle ranking model and other ranking systems and proposed a model that emphasis more on the quality of result, called Semantic Closeness Measure along with Class Match Measure and link analysis. The proposed system not only uses Swoogle’s link analysis and semantic measure but also improves with Semantic Closeness Measure. As AKTive rank and OS\_Rank, this work analyzes the internal structure of the ontology but with an effective mechanism as represented in the implementation section to reduce time complexity. When compared to content based ontology ranking, this system performs link analysis also along with improved semantic measure, SCM. This work enables the searchers to meet their need at the earliest stage without wasting their time by going through the long list of retrieved URLs.

## 7. 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] 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.
- [4] Y. Zhang, W. Vasconcelos, and D. Sleeman, “Ontosearch: An ontology search engine”, In Proc. 24th SGAI Int. Conf. on Innovative Techniques and Applications of Artificial Intelligence, Cambridge, UK, 2004.
- [5] L. Page, S. Brin, R. Motwani, and T. Winograd, “The pagerank citation ranking: Bringing order to the web”, Technical report, Stanford Uni., 1999.
- [6] C. Patel, K. Supekar, Y. Lee, and E. Park, “Ontokhoj: A semantic web portal for ontology searching, ranking, and classification”, In Proc. 5<sup>th</sup> ACM Int. Workshop on Web Information and Data Management, pages 58–61, New Orleans, Louisiana, USA, 2003.
- [7] Harith Alani, Christopher Brewster and Nigel Shadbolt, 2007, “Ranking Ontologies with AKTive Rank”.
- [8] Matthew jones, Harith Alani, July 2006, “Content-based Ontology Ranking”, 9<sup>th</sup> International Protégé Conference.
- [9] Edward Thomas, Harith Alani, Derek Sleeman, Christopher Brewster, 2007, “Searching and Ranking Ontologies on the Semantic Web”.
- [10] Wei Yu, Junpeng Chen, 2009, “Ontology Ranking For the Semantic Web”, Third International Symposium on Intelligent Information Technology Application.
- [11] Vandana Dhingra, Apeejay Stya University, India & Komal Bhatia, YMCA University of Science and Technology, India, “Comparative Analysis of Ontology Ranking Algorithms”, International Journal of Information Technology and Web Engineering, 7(3), 55-66, July-September 2012.

## 8. AUTHOR’S PROFILE

**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 so far. Her areas of interest include Information Security, Data Mining, Fuzzy Logic and Mobile Computing.