ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

Ontology Ranking Algorithms on Semantic Web: A Review

K. SRIDEVI[1] and Dr. R. UMARANI[2]

Assistant Professor, Department of Computer Science, Nehru Memorial College, Puthanampatti, Trichy District, Tamilnadu, India[1]

Associate Professor, Department of Computer Science, Sri Sarada College for Women, Salem, Tamilnadu, India[2]

Abstract: Semantic Web is the representation of knowledge which consists of a huge amount of ontologies. Ontologies provide an efficient way to reduce the amount of information overload by defining the structure of a specific domain and enabling easier access to the information. Since the demand for the use of ontology has been increased, similar to the Web searching, ontologies on the Semantic Web are to be searched in an efficient way. Ontology search can prove its excellence only when the retrieval involves with highly relevant information based on the user query. The ranking method increases the scope of the knowledge searching and makes the users to view the relevant need for the query on the top most. There are a number of Semantic search engines available to aid in the discovery and ranking of ontologies, but with benefits and pitfalls. This paper reviews most of the ontology ranking methods used, which will help the researchers to proceed further.

Kewords: Semantic Web, Semantic Search, Ontology, Ontology Ranking

I. INTRODUCTION

Keyword technology (integrated with a series of statistical elements such as PageRank) has the enormous advantage of being simple, easily applicable to many languages and very fast. When applied to the web, keyword technology took advantage of the free and voluntary labor hours of hundreds of millions of people. People, who by searching and clicking on one or more results, provide creators with an enormous quantity of information every day. This kind of information is priceless and helps to re-organize search results in the best possible way.

Most knowledge on the Web is encoded as natural language text, which is convenient for human users but very difficult for software agents to understand. Even with increased use of XML-encoded information, software agents still need to process the tags and literal symbols using application dependent semantics. The Semantic Web offers an approach in which knowledge can be published by and shared among agents using symbols with a well defined, machine-interpretable semantics. At the core, a semantic search engine has the ability to understand the relationships between keywords, phrases or parts of speech within a search phrase, therefore allowing it understand the underlying meaning of the entire phrase. For example, a semantic search engine would be able to easily distinguish the differences between the following phrases made up of the same 'keywords' but with obvious different implications:

- *How to burn a dress?*
- *How to dress a burn?*

In the example above, the phrases are made up of the same keywords, while the subject/action relationships are reversed. In traditional web search, which are based on ranking algorithms, since the relationships between the sentence parts are unknown, the engines would return identical or nearly identical results, even though it was being asked two completely different questions. Additional problems with web search also arise when the keywords are too specific, producing few or no results, or too general, in which case the results are overwhelming and irrelevant.

Alternatively, since semantic search technology understands the meaning of the above sentences, it would be able to produce highly relevant *answers* to the questions. The goal of semantics is to always provide the direct insights and answers needed to complete research tasks, rather than burying those ideas among scores of irrelevant documents.

The idea with this terminology is to offer more relevant results without limiting searches to just keywords (traditional Google search would be called "keyword search" as opposed to a semantic search).

Semantic search is the process of typing something into a search engine and getting more results than just those that feature the exact keyword you typed into the search box. Semantic search will take into account the context and meaning of your search terms. It's about understanding the assumptions that the searcher is making when typing in that search query.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

The Semantic Web aims to achieve better data automation, reuse and interoperability. The main advantage of Semantic Web is to enhance search mechanisms with the use of Ontology's. Ontology is a general description of all concepts as well as their relationship. The Resource Description Framework /Schema (RDF(S)) and Web Ontology Language (OWL) are W3C recommended data representation models which are used to represent the ontology's. The basic method for constructing the Semantic Web is to use the terms defined in ontology as metadata to markup the Web's content. It is generally accepted that ontology refers to a formal specification of conceptualization.

Ontologies have been shown to be beneficial for representing domain knowledge, and are quickly becoming the backbone of the SemanticWeb. Building ontologies, however, represents a considerable challenge for a number of reasons. It takes a considerable amount of time and effort to construct an ontology, and it necessitates a sophisticated understanding of the subject domain. Thus it is an even greater challenge if the ontology engineer is not familiar with the domain. However, one of the major advantages claimed of ontologies is the potential for the "reuse" of knowledge.

A number of ontology libraries currently exist, hosting various ontology files. Examples of such libraries include Ontolingua, the DAML library, the Protege OWL library, etc. However, the ontology search facilities provided by these libraries are at best limited to term search, making it difficult for the user to select the relevant ontologies from others than happened to contain a class with the desired label. As the number of publicly available ontologies increases, this problem is bound to get worse. Thus there is a contradiction in this situation. For a variety of purposes, including the Semantic Web, there is a need for more and more ontologies to be constructed and made available. However, as this occurs, so the re-use of this knowledge becomes an ever greater problem.

In order to achieve an effective level of knowledge reuse, it is required that search engines capable of helping to find the ontologies the users are looking for. Some ontology search engines have been developed that can provide lists of ontologies that contain specific search terms, such as Swoogle [1] and OntoSearch [2]. Such search engines are a good step forward, but more is required in terms of ontology search if re-use is to become a reality.

The Semantic Web's distributed nature raises significant data access problems – how can an agent discover, index, search and navigate knowledge on the Semantic Web? Swoogle [1] was developed to facilitate semantic web data access by providing these services to both human and software agents. Swoogle is a search engine for Semantic Web ontologies, documents, terms and data published on the Web. Swoogle employs a system of crawlers to discover RDF documents and HTML documents with embedded RDF content. It focuses on two levels of knowledge granularity. URI based *semantic web vocabulary* and *semantic web documents* (SWDs), i.e., RDF and OWL documents.

Swoogle Architecture:
As shown in figure 1, Swoogle's architecture can be broken into four major components: SWD discovery, metadata creation, data analysis, and interface. This architecture is data centric and extensible; components work independently and interact with one another through a database.

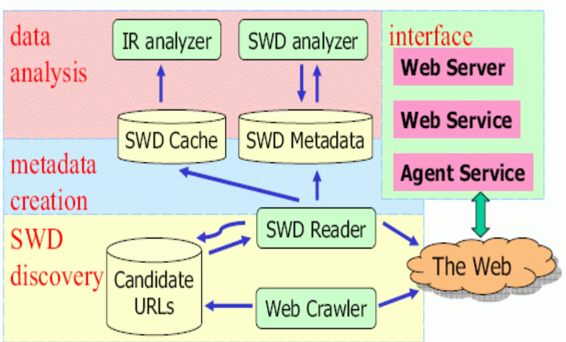

Figure 1: The architecture of Swoogle

The *SWD discovery* component discovers potential SWDs throughout the Web and keeps up-to-date information about SWDs. The *metadata creation* component caches a snapshot of a SWD and generates objective metadata about SWDs at both the syntax level and the semantic level. The *data analysis* component uses the cached SWDs and the created metadata to derive analytical reports, such as classification of SWOs and SWDBs, rank of SWDs, and the IR index of SWDs. The *interface* component focuses on providing data services to the Semantic Web community.

Google has surpassed other search engines because of the effectiveness of its page ranking approach, and most likely the same will happen in the near future for ontology search engines. As the number of ontologies that such search engines can find increases, so will the need increase for a proper ranking method to order the returned lists of ontologies in terms of their relevancy to the query. A proper ranking of ontologies could save the user a lot of time and effort. It would reduce the need to examine in detail each and every ontology returned to find out how well it suits the needs of the knowledge engineer.

II.ONTOLOGY RANKING ON SEMANTIC WEB

Ranking has always been at the heart of information retrieval. This became even more apparent given the enormous size of the web and its continuous expansion. Google uses the PageRank [3] method to rank documents based on hyperlink analysis. Swoogle [1] and OntoKhoj [4] rank ontologies also using a PageRank like method that

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

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.

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. For example, supposing an engineer was looking for an ontology about "students," there could be an ontology about the academic domain that is well connected, and thus popular. If this ontology contains a concept named "Student", then this ontology will show up high on the list of candidates. However, it could very well be the case that the "Student" class is very weakly represented. That ontology might have become popular due to its coverage of publications and research topics, rather than for it's coverage of student related concepts. Similarity measures have often been used in information retrieval systems to provide better ranking of query results.

Ontologies can be viewed as semantic graphs of concepts and relations, and hence similarity measures can be applied to explore these conceptual graphs. Resnik applied a similarity measures to WordNet to resolve ambiguities [5]. The measure he used is based on the comparison of shared features. Another common-feature based similarity is the shortest-path measure, introduced by Rada [6]. He argues that the more relationships objects have in common, the closer they will be in an ontology. Rada used this measure to help rank biomedical documents which were represented in a semantic knowledge-base.

Probability-based measures to explore concept similarities over the Gene ontology was investigated in [7]. Jones and colleagues developed a number of measures to estimate similarity between geographical entities, based on analysing non-common super-classes of concepts in a geographical ontology [8]. Most of the measures above are based on pairwise comparison of concepts (or sets of concepts). However, experiments on measuring similarity between whole ontology structures have also been reported [9][10]. To the best of our knowledge none of such measures have been applied to ranking ontologies, even though some work has been reported on ranking semantic queries using ontologies [11].

As ontology ranking has only been attempted using link-based analysis (eg Swoogle and Ontokhoj), [12] describes AKTiveRank, a system for ranking ontologies by aggregating a number measures that look into certain structural features of concepts, such as their centrality of the terms in a hierarchy, structural density, and semantic similarity to other concepts of interest. The Ranking Approach in AKTiveRank applies four types of assessments (measures) for each ontology to measure the rankings. Each ontology is examined separately. Once those measures are all calculated for an ontology, the resulting values will be merged to produce the total rank for the ontology. The four measures are described in the following.

AKTiveRank [13] is a prototype system for ranking ontologies by aggregating a number of graph- nalysis measures that use certain structural features of concepts, such as their hierarchical centrality, structural density, and semantic similarity to other concepts of interest.

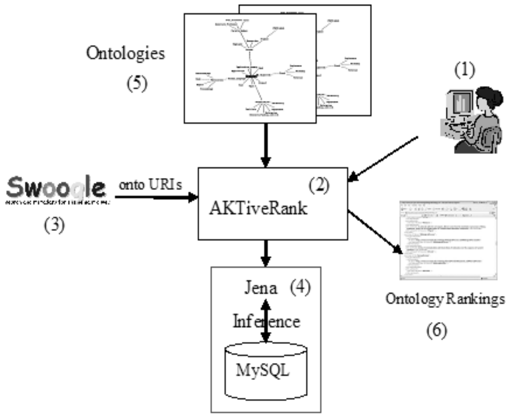

Figure 2: AKTiveRank Architecture

As represented in figure 2, When a query is received, AKTiveRank queries Swoogle for the given search terms and scrapes the ontology URIs from the results page returned by Swoogle. Once a list of ontology candidates is gathered from Swoogle, AKTiveRank starts to check whether those ontologies are already stored in the Jena MySQL database backend, and if not, load them from the web and add them to the database. The Jena API is used here to read the ontologies and handle the database storage. All the analysis of ontology structures that AKTiveRank performs for the ranking is also undertaken using Jena's API. An inference engine (Racer4 for OWL Lite and DL, and Jena's own inference engine for OWL FULL) is applied to every ontology loaded from the database before analysed in AKTiveRank. AKTiveRank then analyses each of the ontology candidates to determine which is most relevant to the given search terms. This analysis will produce a ranking of the retrieved ontologies, and the results are returned to the user as an OWL file containing the ontology URIs and their total ranks.

AKTiveRank applies four types of assessments (measures) for each ontology to measure the rankings:

1. Class Match Measure (CMM): Evaluates the coverage of an ontology of the given search terms. AKTiveRank looks for classes in each ontology that have labels matching a search term either exactly (class label identical to search term) or partially (class label "contains" search term). An ontology that contains all search terms will obviously score higher than others, and exact matches are

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

regarded as better than partial matches. For example if searching for "Student" and "University", then an ontology with two classes labelled exactly as the search terms will score more in this measure than another ontology which contains partially matching classes, eg labelled "UniversityBuilding" and "PhDStudent".

2. Centrality Measure (CEM): The Centrality Measure (CEM) is aimed to assess how representative a class is of an ontology. Several approaches have been proposed for the task of defining classes when building an ontology, such as top-down, bottom-up, outside-in, or middle-out approaches. Even though all those approaches are valid, but psycholinguistic evidence has shown that middle level concepts tend to be more detailed and prototypical of their categories than classes at higher or lower hierarchical level. Thus we here assume that the more central a class is in the hierarchy, the more likely it is for it to be well analysed and fully represented. The Centrality Measure is meant to estimate just that.

3. Density Measure (DEM): When searching for a "good" representation of a specific concept, one would expect to find a certain degree of detail in the representation for the target concept. This may include how well the concept is further specified, how many attributes and siblings the class has, etc. DEM is intended to approximate the representational-density of classes and consequently the level of detail for concepts.

4. Semantic Similarity Measure (SSM): This measure calculates the semantic similarity between the classes that were matched in the ontology with the search terms. The motivation here is that it might be preferred for the search terms to be closely related to each other in the ontology than otherwise. SSM formula based on the shortest path measure defined in [14].
The Total Score of an ontology is calculated once the four measures are applied to all the returned ontologies. Total score is the aggregation of all the measures' values, taking into account their *weights*, which are used to determine the importance of each measure in the ranking.

OntoSearch, which is a tool for capturing and searching ontologies on the Semantic web. OntoSearch[15][16] has grown from a system which used the Google API and provided additional filtering and information on the results returned to a hybrid system which searches a local repository and only reverts to Google when it does not have local information. This functionality was developed to fulfil several requirements defined during user evaluations.

- The ability to specify the type of file(s) to be returned (OWL, RDF, all)
- The ability to specify the type of entities to be matched by each keyword (concept, attribute, values, comments, all)
- The ability to specify partial or exact matches on entities. So in partial match mode CHEMICAL would match CHEMICALS, CHEMICAL_AGENTS, etc; and of course in exact matching mode, only CHEMICAL would be matched.
- The ability to specify a sub-graph to be searched for. For example, concept Animal with concept Pig within 3 links; animals with particular attributes would be a further variant.

This required the implementation of a more advanced architecture with a triple store to provide a repository of Ontological information. Two search strategies are currently possible using Onto-Search. Searching for structure using a simple query language which allows all the requirements identified to be covered or searching for classes using a keyword based search which is currently more restrictive.

In order to rank ontologies, Content based Ranking [17] system attempts to find a corpus that relates to the domain that the user requires an ontology to represent. This method is inspired by [18], but differs from it in that the corpus is selected based on the user query, rather than the ontology itself. The corpus will then be analysed to identify domain-related terms to use for evaluating the existing ontologies in terms of how well they cover the domain of interest. Using a representative corpus allows terms to be extracted using term frequency measures (tf-idf [19]). The terms which get the highest Tf-idf score from this corpus can then be considered as potential concept labels. This system uses the top 50 words of such an analysis. An ontology which has more class labels that match these words is deemed more suitable by the system and is therefore ranked higher than others. The following sections demonstrate our ranking method.

The content-based ontology ranking algorithm obtains a list of ontologies from a search engine. Based on the term given by the knowledge engineer the retrieved ontologies are ranked. The ranking is done according to the number of concept labels in those ontologies which matches a set of terms extracted from a WordNet. It is done related to the domain of knowledge identified by the knowledge engineer‟s original search terms. Each ontology is then ranked according to how many of these new terms match class labels within them. The class match score (CMS) is used.

The OntoRank algorithm [20] 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 [21]. The reference relations are directional and transitive. It evaluates the importance of ontology in a static manner and doesn‟t consider the user query as an effective factor in ranking the results.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

OS_Rank algorithm called Ontology Structure Ranking (OS_RANK) [22] ranks the ontologies based on its semantic relation and structure. The overall ranking criteria are based on the three ranking scores:

- Ranking based on class name
- Ranking based on semantic relation
- Ranking based on ontology structure.

These measures are applied to retrieved ontology from search engine based on the user query and ranking is performed. The user can decide the weights of the ranking measure according to the needs and importance of their applications.

Another work named Semantic-aware Importance Flooding (SIF RANK) [23] retrieves the OWL ontology and converts them into directed graph. The iteration fix point computation is done in each graph to calculate the importance of nodes. It is based on the nine kinds of patterns, semantically treated correct. This computation reaches the maximum number of iterations and the normalization is done to neglect the nodes which are not semantically linked.

As a result of conclusion after reviewing number of related papers it conforms that AKtive Rank does ranking based on the concept covered in the internal structure of ontology. It has pitfall of increasing time complexity. Content-based Ontology Rank 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 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 OntoRank is that most ontologies are poorly inter-referenced and this will be reflected in the quality of the ontology retrieval. OS_Rank 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.

III. CONCLUSION

As an extension of the current Web, Semantic web provides a structured data and knowledge representation framework for Web information. Ontologies play an important role in framing the Semantic Web. Searching through ontologies for a domain facilitates the retrieval of relevant information on Semantic Web. As the number of ontologies available online is increasing rapidly, searching a relevant ontology of knowledge source becomes essential. Lot of research contributions are available in this regard. This paper has reviewed the methods to rank the ontologies that are retrieved as the result of user query. Ranking approach is the method which places the highly relevant ontology for the query on the top rank list. This enables the searchers to meet their need at the earliest stage without wasting their time by going through the long list of retrieved items.

REFERENCES

[1] 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.

[2] Y. Zhang, W. Vasconcelos, and D. Sleeman. Ontosearch: An ontology search engine. In Proc. 24th SGAI Int. Conf. on Innovative Techniques and Applications of Artifial Intelligence, Cambridge, UK, 2004.

[3] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Uni., 1999.

[4] 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.

[5] P. Resnik. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. Journal of Artificial Intelligence Research, 11:95–130, 1999.

[6] R. Rada, H. Mili, E. Bicknell, and M. Blettner. Development and application of a metric on semantic nets. IEEE Transactions on Systems Management and Cybernetics, 19(1):17–30, 1989.

[7] P. Lord, R. Stevens, A. Brass, and C. Goble. Semantic similarity measures as tools for exploring the gene ontology. In Proc. Pacific Symposium on Biocomputing, 2003.

[8] C. B. Jones, H. Alani, and D. Tudhope. Geographical Terminology Servers: Closing the Semantic Divide, chapter 11, pages 201–218.

[9] A. Maedche and S. Staab. Measuring similarity between ontologies. In Proceedings of the European Conference on Knowledge Acquisition and Management (EKAW), pages 251–263, Madrid, Spain, 2002. Springer.

[10] P. C. Weinstein and W. P. Birmingham. Comparing concepts in differentiated ontologies. In Proc. 12th Workshop on Knowledge Acquisition, Modeling and Management (KAW' 99), Banff, Alberta, Canada, 1999.

[11] N. Stovanovic, R. Studer, and L. Stovanovic. An approach for the ranking of query results in the semantic web. In Proc. 2nd Int. Semantic Web Conf. (ISWC03), pages 500–516, Sanibel Island, FL, USA, 2003.

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

[13] Alani H, Brewster C. Ontology Ranking based on the Analysis of Concept Structures. Third Int. Conf. On Knowledge Capture (K-Cap), Banff, 2005.

[14] Rada R, Mili H, Bicknell E, Blettner M. Development and application of a metric on semantic nets. *IEEE Trans. Sys. Man. and Cybernetics*, 19(1):17–30, 1989.

[15] Zhang Y, Vasconcelos W, and Sleeman D. Onto- Search: An Ontology Search Engine (AI-2004). *The 24th SGAI Int. Conf. on Innovative Techniques and Applications of AI, Cambridge*, 2004.

[16] 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.

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

[18] Brewster, C., Alani, H., Dasmahapatra, S. and Wilks, Y. Data driven ontology evaluation. *In International Conference on Language Resources and Evaluation, Lisbon, Portugal*, 2004.

[19] Salton, G. and Buckley, C. 1988 Term-weighting approaches in automatic text retrieval. *Information Processing & Management* 24(5): 513–523 .

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

[21] 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*.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

[22] Wei Yu, Junpeng Chen. 2009. Ontology Ranking For the Semantic Web. Third International Symposium on Intelligent Information Technology Application.
[23] Yixi Chen, Keting Yin and Xiaohu Yang. 2010. A Semantic-Aware Algorithm to Rank Concepts in an OWL Ontology Graph. Information Technology Journal.

BIOGRAPHIES

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 4 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 has acted as Resource Person in various National and International Conferences. She has produced 2 Ph.D. scholars and is currently guiding 18 Ph.D. scholars. She has guided 34 M.Phil. scholars and currently guiding 4 M.Phil. Scholars. Her areas of interest include Information Security, Data Mining, Fuzzy Logic and Mobile Computing.