IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

Comparison of Ontology Ranking Algorithms on Semantic Web

S.Sudha[1], Lakshmi.S[2]

Head, Dept. of Computer Science, J.J College of Arts & Science (Autonomous), Pudukkottai, Tamilnadu, India[1]

Asst. Professor, Dept. of Computer Science, Sri Adi Chunchanagiri Women's College, Cumbum, 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. Ontology search can prove its excellence only when the retrieval involves highly relevant information based on the user's query. The ranking method increases the scope of knowledge searching in ontology-driven searches. This paper reviews most of the ontology ranking methods used, which will help researchers to proceed further.

Keywords: Semantic web, Semantic search, ontology, ontology ranking.

1.INTRODUCTION

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

The Semantic Web[1] aims to achieve better data automation, reuse and interoperability. The main advantage of Semantic Web is to enhance search mechanisms with the use of ontologies. Ontologies have been shown to be beneficial for representing domain knowledge, and are quickly becoming the backbone of the Semantic Web. One of the major advantages of ontologies is the potential for reuse of knowledge.

In order to achieve an effective level of knowledge reuse, it is required that search engines are capable of helping to find the ontologies the users are looking for. Some ontology search engines such as Swoogle and OntoSearch have been developed that can provide lists of ontologies that contain specific search terms.

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, URL based semantic web vocabulary and Semantic Web documents (SWDs) i.e. RDF and OWL documents.

As shown in the Figure1, 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.

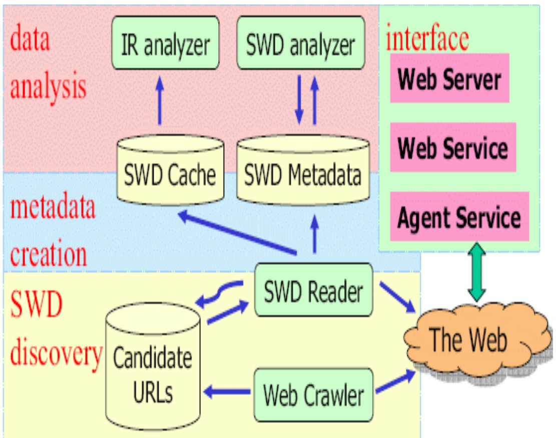

Figure 1. Swoogle Architecture Diagram

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 SWDs, rank of SWDs and the IR index of SWDs. The Interface component focuses on providing data services to the Semantic Web community.

This paper gives a detailed survey on the various ontology ranking algorithms with the description about their functional process and the constraint that impact the efficiency of the algorithm. The rest of the paper is organized as follows; section 2 describes the importance of ontology ranking and its overview. In section 3 the various ontology ranking algorithms and their functional process are described .In section 4 comparisons of various algorithms are carried out. Finally section 5 provides the conclusion.

IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

2.RANKING ALGORITHMS

Swoogle[5] is currently dominating the knowledge representation[1] area of development indexing which leads to an increasing number of ontologies covering a wide range of domains. The suitable ontologies for the particular domain application are retrieved by ranking ontologies.

Some of the techniques used to rank ontologies are:

2.1 AKTive Rank Algorithm:

AKTiveRank[2] is a technique for ranking ontologies based on different analytical measures that assess the ontology in terms of depth of coverage. Users can use an ontology search engine such as Swoogle for searching. The query submitted to the search engine is used by AKTiveRank to identify the concepts that match the user's request. The ranking measures applied by AKTiveRank will be based on the representation of those concepts and their neighborhoods. It increases the time complexity.

2.2 Content-based Ontology Rank Algorithm:

The content-based ontology rank algorithm[4] 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 match a set of terms extracted from WordNet.

2.3 Onto Rank Algorithm:

The Onto Rank algorithm[6] applies the link analyze method. Here two concepts are considered as reference relationship "if and only if" a relationship exists between the two classes in a relation set[1]. The reference relations are directional and transitive. It evaluates the importance of ontology in a static manner and does not consider the user query as an effective factor in ranking the results.

2.4 OS_Rank Algorithm:

Ontology Structure Ranking[7] (OS_Rank) ranks the ontologies based on their 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 the ontology retrieved from the search engine based on the user query and ranking is performed.

2.5 SIF Rank Algorithm:

Semantic-aware Importance Flooding (SIF) Rank algorithm[8] 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. The computation reaches the maximum number of iterations and the normalization is done to neglect the nodes which are not semantically linked.

3. AKTIVERANK – RANKING APPROACH

AKTiveRank[2] applies four types of assessments (measures) for each ontology to measure the rankings. Each ontology is examined separately. Once the measures are all calculated for an ontology, the resulting values will be merged to produce the total rank for the ontology. The measures are as follows:

3.1 Class Match Measure:

The Class Match Measure (CMM) simply 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 regarded as better than partial matches. For example, if searching for "Student" and "University", then an ontology with two classes labeled exactly as the search terms will score more in this measure than another ontology which contains partially matching classes, e.g. labeled "UniversityBuilding"and "PhDStudent".

Definition: Let c[o] be a set of classes in ontology o, and T is the set of search terms.

$$CMM[o, T] = \alpha|E[o, T]| + \beta|P[o, T]|$$

where E[o, T] and P[o, T] are the sets of classes of ontology o that have labels that match any of the search terms t exactly or partially, respectively. CMM[o, $_T$] is the Class Match Measure for ontology o with respect to search term $_T$. α and β are the exact matching and partial matching weight factors respectively. Exact matching is favored over partial matching if α > β.

3.2 Centrality Measure:

The Centrality Measure (CEM) is aimed to assess how representative a class is of an ontology. The more central a class is in the hierarchy, the more likely it is for it to be well analyzed and fully represented. The Centrality Measure is meant to estimate just that.

Definition:

$$CEM[o] = \frac{1}{n} \sum_{i=1}^{n} cem[c]$$

3.3 Density Measure:

The Density Measure (DEM) is intended to approximate the representational-density of classes and consequently the level of knowledge detail.

IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

Definition:

$$DEM[o] = \frac{1}{n} \sum_{i=1}^{n} dem[c]$$

n = E (o, T) + P (o, T)

Where, Dem(c) = density measure of a class.

3.4 Semantic Similarity Measure:
The Semantic Similarity Measure (SSM) calculates how close the classes that match the search terms are in an ontology.

Definition:

$$SSM[o] = \frac{1}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} ssm[c_i, c_j]$$

n = number of matched classes in ontology "O".

ssm (c_i , c_j) = shortest path between the classes c_i and c_j

3.5 Total Score:
The total score of an ontology can be calculated once the four measures are applied to all ontologies. Total score is calculated by aggregating all the measures values, taking into account their weights, which are used to determine the importance of each measure in the ranking.

Definition: Let M = {M[1], …M[i], M[4]} = {CMM, CEM, DEM, SSM}, w_i is a weight factor, and O is the set of ontologies to rank.

$$Score[o \in O] = \sum_{i=1}^{4} w_i \frac{M[i]}{\max_{1 \le j \le |O|} M[j]}$$

Measure values are normalized to be in the range (0 - 1) by dividing by the maximum measure value for all ontologies (M[j]).

4. COMPARISON

This section provides the comparison among all the ranking algorithms, its advantages and disadvantages in table1. It also provides the criteria used by all the ranking algorithms.

Table 1. Criteria, Advantages and Disadvantages of Ranking Algorithms

s.no	Algorithms	Criteria used for ranking ontologies	Advantages	Disadvantages
1.	AKtive Rank	Based on the internal structure of the ontology with analytical measures	1. Ranking is done based on the concept covered in the internal structure of ontology.	1. Increases the time complexity. 2. Low CEM value, when the concept of interest is placed in the top of hierarchy.
2.	Content-based Ontology Rank	Based on the internal structure with content similarity of the ontology related with corpus.	1. The ontology which has more class labels matches the words in the corpus and is ranked higher than the others.	1. If the search term is very specific, retrieval of suitable corpus is difficult.
3.	Onto rank	Based on semantic web link structure which gives priorities for different link relationship.	1. The concept dictionary enlarges the scope of the synonym and related words in terms of connotation & extension. This overcomes the limited connotation of user keywords.	1.Most ontologies are poorly inter-referenced 2. It will be reflect in the quality of the ontologies.
4.	OS_Rank	Based on both the internal structure and the semantic analysis with the three normalized measures.	1. Executes either in local ontology repository or connected to ontology search engine. 2. Method is based on both ontology structure and semantic analysis.	1. The user can give the weights of the measures applied for ranking the total score. 2. Process is time consuming and very tedious.
5.	SIF Rank	Based on the semantic meaning of either concept or relation and also the ontology structure.	The importance of concepts reinforces one another in an iterative manner. The semantically correct paths can flood relevant components of importance vector from a concept to its neighbors in ontology graphs.	1. If the iteration point is maximum, it is difficult to retrieve the concept importance if the domain has large ontology. 2. No two user can give the can same importance for an ontology when it is large.

5. CONCLUSION

As a result of conclusion after reviewing number of related papers it confirms that AKTiveRank does ranking based on the concept covered in the internal structure of ontology. It has disadvantage 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 matching the words in the retrieved documents. But if the search term is very specific, the retrieval of the 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 disadvantage of OntoRank is that most ontologies are poorly inter-referenced and this will be reflected in the quality of the ontology retrieved. OS_Rank method is based on searching both ontology structure and semantic analysis. The disadvantage is that this process is time consuming and very tedious.

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 thorough the long list of retrieved items.

REFERENCES

[1] Alani, H., and Brewster, C.,"Metrics for Ranking Ontologies", Proceedings of the 4th Int. Workshop on Evaluation of Ontologies for the Web (EON'06), at the 15th Int. World Wide Web Conference (WWW'06).Edinburgh,UK, 2006
[2] Alani.H and Brewster.C (2005), "Ontology Ranking based on the Analysis of Concept Structures". In Proceedings of the 3rd International Conference on Knowledge Capture, Banff, Alberta, Canada., 2005
[3] Matthew jones, Harith Alani. July 2006. Content-based Ontology Ranking. 9th International Protégé Conference
[4] Chintan Patel, Kaustubh Supekar, Yugyung Lee, E.K.Park, " A Semantic Portal for Ontology Searching, Ranking and Classification".
[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] Harith Alani, Christoper Brewster and Nigel Shadbol ,"Ranking Ontologies with AkTive Rank",2007.
[7] Subhashini.R, J.Akilandeswari, Sinthuja.V, "A review on ontology ranking algorithms", International journal of Computer Application, nov 2011.
[8] Jinsoo Park, Sunjoo Oh, Joongho Ahn, "Ontology Selection Ranking Model for Knowledge Reuse," Expert Systems with Applications 38 (2011) 5133–5144