A Review on Ontology Ranking Algorithms

R.Subhashini
Senior Lecturer
Department of Computer Science and Engineering
Sona College of Technology, Salem – 600 35

Dr. J.Akilandeswari
Head of the Department
Department of Information Technology
Sona College of Technology, Salem – 600 35

V.Sinthuja
PG Scholar
Department of Computer Science and Engineering
Sona College of Technology, Salem – 600 35

ABSTRACT
Semantic web is based on knowledge representation which contains a large number of ontologies. Nowadays the increasing demand for ontology has triggered a growing number of usable ontology in web. Similar to the web page searching, ontology searching results are to be ranked. The ranking method increases the scope of the knowledge searching in ontology-driven searches. This paper gives an overview of different types of ranking algorithm their methodology and the constraint. It is used to help the researchers to select the most suitable algorithm for their application based on the efficiency.

Keywords
Semantic web, Knowledge representations, ontology, ontology ranking

1. INTRODUCTION
Semantic Web is the extension of Web, aimed to ensure better understanding of the information and provides a good quality knowledge representation based on the user query [8]. It contributes several mechanisms that are used to classify information and its context retrieving information on web. It has a number of Resource Descriptions Framework (RDF) which overcomes the difficulties in understanding and integrating the information [1].

Ontology is an explicit specification of shared conceptualization **[6]**. A number of ontology libraries and search engines are in existence to facilitate retrieval of potentially relevant ontologies and provides a domain-related ontology to depict the real world applications. There is a set of standard web ontology language (OWL) which is based on RDF model to describe the concepts explicitly with their relationship [1].

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.

2. ONTOLOGY RANKING
The search engines play a vital role in retrieving the information required by the user. However, the retrieved web pages also contain ineffective or irrelevant information. The latest web architecture, represented by semantic web, overcomes this limitation by applying the ranking algorithms. The ranking algorithm extract the information on the user queries from the semantic search engine and provides the desired result. However, in order to rank results, most of the existing solutions need to work on the whole annotated knowledge base.

2.1 Algorithms Overview
Ontology ranking algorithms can be of different types as follows:

- AKTive Rank Algorithm
- Content-based Ontology Rank Algorithm
- OntoRank Algorithm
- Ontology Structure Rank Algorithm.
- SIF Rank Algorithm

2.1.1 AKTive Rank Algorithm
AKTiveRank [2] is a technique for ranking ontology based on different analytical measures that assess the ontology in terms of depth of coverage. Users can use ontology search engine (e.g. Swoogle) for searching. The query submitted to the search engine is used by AKTive Rank to identify the concepts that match the user's request. The ranking measures applied by AKTive Rank will be based on the representation of those concepts and their neighbourhoods.

2.1.2 Content-based Ontology Rank Algorithm
The content-based ontology ranking algorithm [3] 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.

2.1.3 OntoRank Algorithm
The OntoRank algorithm [5] 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 [1]. 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.

2.1.4 OS_Rank Algorithm
Ontology Structure Ranking (OS_RANK) [15] ranks the ontologies based on its semantic relation and structure. The overall ranking criteria are based on the three ranking scores:

1. Ranking based on class name
2. Ranking based on semantic relation
3. 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 [15].

2.1.5 SIF Rank Algorithm

The Semantic-aware Importance Flooding (SIF RANK) [16] 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 [16].

3. FUNCTIONAL PROCESS

This section describes the methodologies and the functional process of all the above mentioned algorithms.

3.1 AKtive Rank Algorithms

The steps involved in the AKTive rank algorithm [2] are as follows:

1. The system receives search query from the user and forwards to Swoogle.
2. Retrieve the ontology URI's and convert each URI's into graph format.
3. Applies four analytical measures [2] to the URIs. The analytical measures are

- Class match measure (CMM).
- Centrality measure (CEM).
- Density measure (DEM).
- Semantic similarity measure (SSM).

4. Finally the rank score is calculated by combining all the measured scores.

3.1.1 Class Match Measure

In CMM [2], the ranking algorithm searches the class labels to match the user query either exactly or partially. If the ontology classes match all the query terms then it will be ranked higher.

$$CMM\ (o, T) = \alpha * E\ (o, T) + \beta * P\ (o, T) \qquad (1)$$

Where,

E (o, T) and P (o, T) are the number of classes of ontology "o" that have labels which matches any of the search terms "t" either exactly or partially.

C[o] is a set of classes in ontology "o", and T is the set of search terms.

3.1.2 Centrality Measure

CEM [2] measures the closeness of a concept in an ontology class which is placed in the middle level of its hierarchy. The search terms in the middle of the hierarchy is ranked higher than, when it is placed at the top. But little ontology was placed based on the concept of interest near the top of the hierarchy. These few ontologies were entirely focused around the concept the user searches for, as it is not in centre they scored very low.

$$\mathrm{CEM}[0] = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Cem}[c] \qquad (2)$$

Where,

n = number of matches of class.

Cem[c] – Centrality measure of particular class.

3.1.3 Density Measure

DEM [2] is intended to approximate the structural-density of classes and consequently the depth of detail covered for concepts. This measure is limited to calculate the score only for the direct relations such as subclasses, super classes, relations and siblings.

$$DEM[O] = \frac{1}{n}\sum_{i=1}^{n} dem(c) \qquad (3)$$

$$n = E\ (0, T) + P\ (0, T) \qquad (4)$$

Where,

Dem(c) = density measure of a class.

W_i = weight factor set to a default value of 1.

3.1.4 Semantic Similarity Measure

SSM [2] calculates the semantic similarity between the classes that matches the ontology with the search terms. It finds the shortest path between them to calculate the distance between relation sets to predict the closeness.

$$SSM[0] = \frac{1}{k\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} ssm\,(c_i,c_j)} \qquad (5)$$

Where,

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

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

3.1.5 Combined Score of Analytical Measures

The overall score for the AKTive rank algorithm is a combination of all the four analytical measures discussed. The weight for each measure is given to resolve the importance of measures relatively [2]. The formula is

$$Total\ Score\ (o \in O) = \sum_{i=1}^{4} w_i \frac{M[i]}{max_{\,1\le j\le |o|} M[j]} \qquad (6)$$

Let M = {M [1], M [2], M [3], M [4] } = {CMM, CEM, SSM , DEM}

W_i – weight factor.

O – The set of ontology to rank.

3.2 Content-Based Ontology Rank Algorithm

The content-based ontology rank [3] algorithm ranks the ontology by extracting the query term's related words through the WordNet along with two scores:

1. Class match score (CMS)
2. Literal match score (LMS)

In the class match score, each ontology matches with class labels and in the literal text match score, the literal text like comments are matched with the class labels.

$$CMS = [o \in O] = \sum_{i=1}^{n} I(P_i, o) * 5 \log(n + 2 - i) \quad (7)$$

$$\text{Total} = \alpha * \text{CMS} + \beta * \text{LMS} \quad (8)$$

Where,

O = set of ontologies to be ranked.

P_i = set of potential class labels obtained from the corpus.

n = number of terms collected from corpus.

α and β are the weight factors.

3.3 OntoRank Algorithm

The OntoRank algorithm [5] 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. In this accessed probability of ontology 'a' is calculated as follows:

$$OntoRank(a) = wPR(\propto) + \sum_{x \in OTC(a)} wPR(x) \quad (9)$$

Where,

wPR (α) = accessed probability of Semantic Web document (SWD) of 'a' of itself.

$\sum_{x \in OTC(a)} wPR(x)$ = accessed probability of all imported SWD.

3.4 OS_Rank Algorithm

The OS_Rank algorithm [15] calculates the structural and semantic relation of an ontology .The final ranking is calculated after all the three measures are considered. The measures are:

3.4.1 Ranking based on class name (C_{Ai}) [15]

After the user submits the query, the ontology is converted into graph structure. The class matches with the query either exactly or partially that has been calculated using the formula:

$$C_{Ai} = M_a * \mu + P_a * \nu, \quad 0<\mu, \nu<1 \quad (10)$$

Where,

M_a – total number of exact match found in ontology

P_a – total number of partial match found.

μ and ν are weight factors.

3.4.2 Ranking based on ontology structure (S_{Ai}),

The hierarchy and the concept coverage of the class based on the super classes and the sub classes of ontology O_j is calculated using the formula [15]:

$$S_B = \sum_{1}^{m=\beta} \frac{1}{\Omega_{Bm}} \quad (11)$$

$$S_{Ai}(O_j) = \sum_{1}^{k=M_a} S_{Bk} * \mu + \sum_{1}^{j=P_a} S_{Bj} * \vartheta \quad (12)$$

Where,

B_m – super classes of search term A_i.

3.4.3 Ranking based on semantic relation (E_{Ai}),

Number of semantic arcs and their semantic association is calculated using the formula [15].

$$S_D = \sum_{1}^{n=\lambda} \left(\frac{d_n}{\Omega_{Dn}} * d_n + \frac{d'_n}{\Omega'_{Dn}} * d'_n \quad (13)$$

d_n – number of outgoing semantic arcs.

Ω_{Dn} – total number of outgoing semantic arcs to other classes in O_j.

The values of three measures normalized and the final normalized values are C'_{Ai}, S'_{Ai}, E'_{Ai}.

$$OSRank_{ij} = w_1 * C'_{Ai} + w_2 * S'_{Ai} + w_3 * E'_{Ai} \quad (14)$$

$$w_1 + w_2 + w_3 = 1 \quad (15)$$

The final ranking value of ontology O_i can be computed by formula:

$$F_{oj} = z_1 * OSRank_{1j} + z_2 * OSRank_{2j} + z_3 * OSRank_{3j} + \dots\dots + z_t * OSRank_{tj} \quad (16)$$

$$Z = \{ z_1, z_2, z_3 \dots\dots z_t \} . \quad (17)$$

Z - Set of weight factors

3.5 SIF RANK ALGORITHM [16]

In this algorithm, normalization is done after the formation of transfer matrix and relocation matrix. The initial importance of any node is based on the semantically correct paths. The normalization of nodes is calculated as follows:

$$r^{k+1}(t) = \sum_{for\ each\ e_{s,t}} (r^k(s) * TM(type(e_{(s,t)}) * RMtypees,t) \quad (18)$$

$r^{k+1}(t)$ = normalize ($r^{k+1}(t)$).

type ($e_{(s,\ t)}$) – returns the semantic patterns of $e_{(s,t)}$ and the semantic pattern belongs to the set { U , D , H , N}.

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.

5. EVALUATION ANALYSIS

The experimental results for the above discussed ranking algorithms were done by obtaining the relevant ontologies from the Swoogle for the domain "student and university". Table 1 shows the list of sample ontologies. The algorithms of each ranking techniques are applied on those ontologies to get a score value and their results were obtained as numerical data's in table 2. Figure 1 shows the chart comparison for all the described ranking techniques where the X-axis denotes the owl files and the Y-axis denotes the corresponding score values based on the results obtained. To evaluate the semantic web two indicators named recall *rate* and *precision call* are necessary.

Table 1. Ontology URLs

s.no	ONTOLOGY URL
1.	http://www.mondeca.com/owl/moses/ita.owl
2.	http://protege.standford.edu/plugins/owl/owl-lib/ka.owl

Table 2. Results

RANKING ALGORITHM	ita.owl	ka.owl
AKTive rank	0.684	0.573
Content-based rank	5.2	5.9
OntoRank	0	0.525
OS_Rank	0.35	0.725
SIF Rank	0.75	0.65

Fig 1: Analysis Result

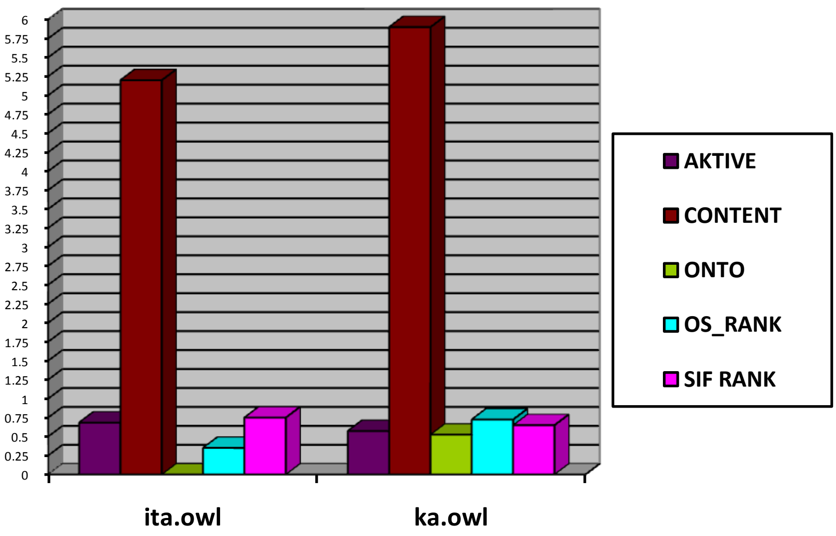

Table 3. Criteria, Advantages and Disadvantages of Ranking Algorithms

s.no	Algorithms	Criteria used for ranking ontologies	Advantages	Disadvantages
1.	AKtive Rank	1.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	1. 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	1. 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.

6. CONCLUSION

The paper describes about the different ranking algorithm used to rank the ontologies in semantic web to give the best ontology URI's to the users. The methodology, advantages and disadvantages of the various ranking algorithms is discussed. The calculation is done taking two ontology URLs. The experimental results are evaluated by using the *recall rate* and *precision call*. The precision is evaluated at a given cutoff rank, considering only the topmost result returned by the system. Recall in information retrieval is the fraction of documents that are relevant to the query that are successfully retrieved. Thus the overview of the ranking algorithms proposes the researchers to choose the efficient ranking algorithm suitable for their domains which reduces the development time.

7. REFERENCES

[1] 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.

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

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

[4] Harith Alani. 2006. Ontology Construction from Online Ontologies. In proc. 15th International World Wide Web Conference.

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

[6] Thomas R. Gruber .1993. Towards a principles for the Design of Ontologies used for Knowledge sharing.

[7] 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.

[8] Huang Xiao-xi, Zhou chang-le. 2007. An OWL-based WordNet lexical ontology. Journal of Zhejiang university science.

[9] Chintan Patel, Kaustubh Supekar, Yugyung Lee, E.K.Park. A Semantic Portal for Ontology Searching, Ranking and Classification.

[10] Sanjay Kumar Malik, Nupur Prakash, S.A.M. Rizvi. 2010. Ontology Merging using prompt plug-in of Protégé in Semantic Web, International Conference on Computational Intelligence and Communication Networks.

[11] Iván Cantador, Miriam Fernández, Pablo Castells. 2007. Improving Ontology Recommendation and Reuse in WebCORE by Collaborative Assessments.

[12] Andrés Montoyo1, Manuel Palomar1 and German Rigau. Method and Interface for WordNet Enrichment with Classification Systems.

[13] Wei Yu, Jiaheng Cao, Junpeng Chen. 2006. A Novel Approach for Ranking Ontologies on the Semantic Web. 1st International Symposium on Pervasive Computing and Applications.

[14] J.Anitha Josephine, S.Sathiyadevi. 2011. Ontology Based Relevance Criteria for Semantic Web Search Engine" IEEE conference.

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

[16] Yixi Chen, Keting Yin and Xiaohu Yang. 2010. A Semantic-Aware Algorithm to Rank Concepts in an OWL Ontology Graph. Information Technology Journal.