Integration of Online Ontologies Using Combined Ranking Algorithm

C.K.Sarumathy, J.Gokulraj, P.Selvaperumal

Abstract—In the Semantic Web, knowledge representation is largely based on ontologies. Ontology should be constructed in a way such that it should meet the requirements of the users. The main difficulty involved in the construction of ontologies is the high cost incurred in building them. Gathering complete knowledge about a specific domain requires more time and it doesn’t guarantee that the resulting ontology will be better than the existing ontologies. Hence, an approach for reusing the existing ontologies to build new ontologies has been proposed. This process makes use of the following steps: identification of existing ontologies, use of combined ranking algorithm (OntoRank+AKTive), segmentation and integration. As a result, best quality ontology can be obtained.

Keywords
AKTive Rank, Extraction, Fragmentation, OntoRank, Reusability.

1. Introduction

Ontology is an explicit specification of conceptualization – Tom Gruber. As far as ontologies are concerned, they are intricate to construct[2]. This construction process is also a cumbersome task to understand for the users. Research works are going on for minimizing the cost incurred in the construction of ontologies. Many approaches have been invented for extracting ontologies from existing knowledge bases. After the extraction, those ontologies which have the rich representation of domain knowledge should be chosen by ranking them. As a result of this phase, the top ranked ontologies will be obtained. These top ranked ontologies may contain some ontologies are larger in size. In addition to this, analysing ontologies is a time consuming task. So, to raise over this difficulty, an efficient ranking system needs to be employed. This ranking system should be capable of analyzing all the available ontologies for a domain based on a certain criteria. Generally, the ontologies are ranked based on various criteria such as page citation count, frequency of occurrence of the keyword given in the query and so on.

Swoogle contains more than 10,000 ontologies and it employs the page ranking algorithm. But it doesn’t have any factors to do with the user’s query. This will result in the hiding of few links which will contain valuable concept representations. Hence, the need for ranking system arises to provide the user with the best set of links.

After ranking the ontologies, segmentation is done. The integration of different fragments segmented from several top ranked ontologies is carried out. As a result of ranking phase, the user is provided with the best set of ontologies. Each ontology contains several fragments that contain different concept representations. These dissimilar concepts are identified and finally these segments are integrated to give the better domain representation.

3. Existing System

Content based ranking algorithm

Content based ranking algorithm[3], developed by the University of Southampton ranks the ontology by analyzing how well it covers a domain of interest. According to this algorithm, an ontology which contains more class labels that match with the keyword will be given the highest rank. This algorithm makes use of the WordNet thesaurus. It helps to disambiguate the terms which contain more than one meaning. Here, the Class Match Score is calculated by the formula,

\[ CMS [o \in O] = \sum_{i=1}^{n} \log (n + 2 - 1) \]

The parameter I(Pi,O) is defined as,

1 if Pi contains a class with label matching Pi
0 if Pi contains a class with label containing Pi
0 if Pi does not contain any of class labels

where O is the set of ontologies, P represents the set of potential class labels, n is the number of terms collected from the corpus and i represents the number of search terms in a given
Integration of Online Ontologies Using Combined Ranking Algorithm

query. The values 1 and 0.4 are allotted for exact and partial matches respectively. The drawback of this algorithm is it doesn’t take any attempt to look at how well the concepts are connected.

OntoRank algorithm

OntoRank algorithm [15][16] employs the link analyze method. Here the ontology is ordered by the evaluation of their importance. This reflects the influence of one ontology exerted to another. Two concepts C1 and C2 can be used as a reference relationship if and only if a relation exists between two classes in the relationship set \{rdfs:type, rdfs:subclass, rdfs:domain, rdfs:range\}. This algorithm evaluates the ontology importance statically, but it doesn’t take user’s query as an affect factor.

AKTive rank algorithm

The AKTive rank algorithm,[4] calculates the matching degrees between ontology and multiple keywords given by user. Hence, it returns more accurate results to the user. The algorithm uses four kinds of evaluation measures as follows,

1) Class Match Measure (CMM),
2) Centrality Measure (CEM),
3) Density Measure (DEM),
4) Semantic Similarity Measure (SSM). [20]

Based on the weighted summation of these four kinds of measures, the ontology rank score is calculated. The disadvantage of this algorithm is it takes more time to respond to user’s query. Even though many ranking algorithms are in existence, they are not effective. Hence, in the proposed system a combined ranking algorithm (OntoRank + AKTive rank) is employed to overcome the drawbacks of existing algorithms.

4. Proposed System

Ontology Integration Process Using Combined Ranking Algorithm

The ontology integration from existing ontology resources requires the following semantic web technologies such as

i) Extraction of ontologies from Swoogle,
ii) Combined ranking algorithm,
iii) Extraction of fragments and
iv) Integration of fragments

4.1. Extraction of ontologies

The first step deals with the extraction of existing ontologies from web sources by using keyword-based search method in Swoogle. This process results in the retrieval of large set of ontology URI’s. Here, the process of selection of the URI’s is done by just analyzing the links.

4.2. Ranking algorithm

4.2.1 Combined Ranking Algorithm

The output of the first phase provides the user with a set of links. An apt algorithm should be applied to find and rank out the ontologies with the rich set of concept representations. To rank, a combination of OntoRank and AKTive algorithm is to be used and the algorithm comprises of the following steps,

1) Extend the concept of the Query

Assume that D is the ontology set in the semantic web, then the ontology set D = \{d1, d2,d3,…dn\}, dj ∈ D (1 ≤ n) ; xj denotes a set composed of all concept aj which has the similar semantic relation with the semantic concept xj. If x,y,z represent different concepts within dj then this will be defined as the logical view LV1,LV2,LV3 respectively. These three can be described by logical view LV in the semantic concept xj. In this way the concept x, y and z all will belong to the concept set [Xj].

The logical view LV (Logical View) of concept X is calculated as follows:

\[ LV(X) = \alpha \cdot extension(X) + \beta \cdot include(X) + X \]

Here, \( \alpha \) and \( \beta \) the percentage of concept X connotation and extension respectively. In order to extend the scope of synonyms concept dictionary is used.

According to TF-IDF (Term Frequency-Inverse Document Frequency), the weight \( w([Xi],d) \) of semantic concept \([Xi]\) can be calculated by,

\[ w([Xi],d)=\log(f([Xi],d)+1)\cdot\log(n/n_i+1) \]

where \( w([Xi],d) \) is the weight of semantic concept set to ontology d; \( f([Xi],d) \) denotes the occurrence frequency of the element \( x_i \); \( n \) denotes the total numbers of ontology in the set D; \( n_i \) denotes the number of ontologies that contain the semantic concept \( x_i \).

2) Calculation of importance of ontology

Here, the OntoRank algorithm is used to calculate the accessed probability of ontology ‘ a’. \( OntoRank(n) = wPR(n) + \sum_{X \in OTC} wPR(x) \)

wPR(a) is the accessed probability of SWD ‘a’ (Semantic Web Document) of itself and denotes the accessed probability of all imported SWD.

3) Application of AKTive Rank algorithm

After applying the OntoRank ranking algorithm, the AKTive ranking algorithm[4] is applied to calculate the score of the ontologies. This algorithm gives the score by manipulating the relevant terms of the query. The score is calculated by,

\[ Score(o \in O) = \frac{w_i \cdot M[i]}{\max_{i} w_i \cdot M[i]} \]

where \( M \) indicates four kinds of measures , \( M = \{M[1], M[2], M[3], M[4]\} = \{CMM, CEM, DEM, SSM \} \), \( O \) denotes the ontology and \( w_i \) is the weight factor and it’s normalized to have the values between zero and one.
5) Conclusion and Future Work

Ontology reusability process helps to save the development time of the researchers. Moreover, the combined ranking algorithm enables the user to find the desired ontology quickly. This algorithm grants the high quality ontology. Many of the ontologies constructed by developers are not put on the web. The idea of reusing the existing ontologies will help to tackle over this problem.

This idea is made better by allowing the user’s of the system to modify, delete from and add to the automatically built ontology. It is also possible to do improvements in the ranking phase, by modifying the parameters of the algorithm.

References

[3] Content-based Ontology Ranking, Matthew Jones, Harith Alani
[21] A Language and Algorithm for Automatic merging of ontologies,