ANALYSING THE ACCEPTANCE CRITERIA AMONG THE CONCEPTS FOR GENERATING AUTOMATED ONTOLOGIES

S.Vigneshwari [*1]
[*1]Research Scholar, Department of Computer Science and Engineering, Sathyabama University, Chennai, Tamilnadu, India
jayam3@rediffmail.com

Dr. M. Aramudhan [#2]
[#2]Associate Professor and Head, Department of Information Technology, Perunthalaivar Kamarajar Institute of Science and Technology, Karaikal, Tamilnadu, India
aramudhan1973@yahoo.com

Abstract

The semantic search yields fruitful results while searching the concepts. Ontologies play a crucial role in searching the concepts in a semantic way. It is otherwise said that, wherever search is needed, ontology search is more preferable when compared to the text based approach. XML plays an important role in information retrieval systems. Since XML is a common format, it is easy to understand as well as easy to construct. OWL is an ontology language, which is purely based on XML, that owl files can be automatically generated using programming languages, like Java and can be viewed using OWL editors. An approach to relate the concepts, based on user search queries and to generate an automated ontology has been discussed in this paper.

Keywords: Ontology; semantic association; mutual information; ranking; OWL; automated ontology.

1. Introduction

Concept in an ontology can be identified as a 4-tuple set , comprising of label, neighbours, ancestor, descendant as given in equation 1. Fig 1. Shows sample manual ontology depicting relationships among the concepts.

$$Concept\ =(label,\ neighbours,\ ancestor,\ descendant) \qquad (1)$$

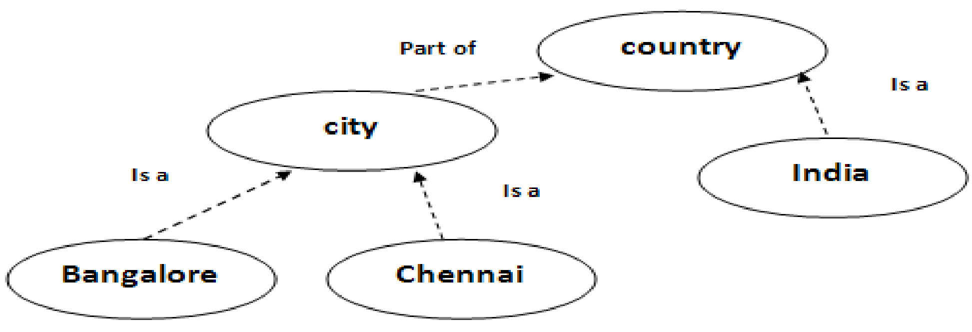

Fig 1. A sample manual ontology in depicting relationships

2. Literature Survey

X. Tao *et al* ,(2011) , used LCSH (Library of Congress subject) to generate World Knowledge Base. Tao describes about two important terminologies in finding relation between the concepts.The first one is specificity, which deals about the importance of a concept. And the second one is exhaustivity which deals about the scope of the concept. Also he classified the specificity into absolute specificity and relative specificity. Based on the specificity, with the help of any OLE tool (Ontology Learning Environment), the concepts are classified into positive and negative ones. The negative concepts are removed and the positive ones are considered for examination.

Peter D. Karp *et al*,(1999), suggested XOL ,a XML-based ontology exchange language, which is best suited for sharing ontologies in a distributed system. XOL can also be used for translating the SQL query of a relational database into XQuery of XML database.

Vigneshwari and Aramudhan(2013), devised a technique to extract the interesting measures using ontology mining. Here the balanced mutual information is used, to find out the similarity between two concepts in the same ontology.

Janez et al (2005) , prepared a survey report on ontology evaluation techniques. Here level wise evaluation is done on the ontologies. Various ontological levels like data level, taxonomy level, context level and syntactic level of evaluation were done in their paper.

Heasoo et al (2012), proposed a robust approach to organize user queries into groups dynamically and automatically. In this paper, search behavior graphs like query reformation graph, query click graph and query fusion graph are generated, with the help of which, it is experimentally proved that query automation is very much useful for a collaborative search. Dynamic query grouping has also played a significant role in organizing the user search queries, which is also important in the construction of ontologies.

Yanhui and Chong (2010) proposed a flexible mechanism to integrate ontologies in a multi ontology database system. In his paper, a framework for ontology integration which combines both ontology similarity measures and ontology integration algorithms has been suggested. The integrated ontology is evaluated and checked for consistency.

3. Techniques and definitions

3.1 *Ontology Definition*

Ontology is a representation of knowledge in a particular domain as a set of concepts and their relationships .Formal definition of ontology is given as follows. Let cls be the class, rel the relationship between the classes, attr the attributes, and ind be the individuals, then the ontology O is defined as per equation(2).

$$O=(cls ,rel, attr, ind) \qquad (2)$$

Based on the definition of ontology, the major components are identified as classes, their relationships, the attributes and individual instances.

3.2 *Ontology languages*

Early ontology languages were based on either HTML or XML. In earlier days XML based ontology exchange language (XOL) was used. XOL, which is a XML-based ontology exchange language, follows a generic approach of defining ontologies. A single set of XML tags can be used in XOL to describe the ontologies. Later it was upgraded into Ontology Integration Language (OIL), which lays the foundation for ontology integration or merging of ontologies. Then the family comprising of knowledge representations called Web Ontology Language, (OWL) family was emerged, which is purely based on XML schemas. There are two specifications of OWL like OWL 1.1(2007) and OWL 2 (2009), recommended by the World Wide Web Consortium (W3C).

3.3 *OWL specifications*

OWL1.1 has its syntax defined in XML schema language. OWL 2 is an ontology language for the semantic web with formally defined meaning. OWL2 has an XML serialization which mirrors the structural specifications defined in OWL 1.1. OWL2 found its way into the semantic editors like protégé, and semantic reasoners like Hermit.The following fig 2, is an XML schema for some sample ontology classes like authors and books, defined by OWL 1.1

```
<?xml version="1.0" encoding="UTF-8"?>
<?oxygen RNGSchema="owl1.1.xsd" type="xsd"?>
<owl11xml:Ontology
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-  instance"
    xsi:schemaLocation="http://www.w3.org/2006/12/owl11-   xml/owl1.1.xsd"
    xmlns:owl11xml="http://www.w3.org/2006/12/owl11-   xml#"
    xml:base=http://my.domain.com/myOntology
    owl11xml:ontologyURI="http://my.domain.com/myOntology">
  <owl11xml:Imports>http://my.domain.com/someOtherOntology</owl11xml:Imports>
  <!-- The ontology axioms go here -->

      <owl11xml:Declaration>
        <owl11xml:OWLClass owl11xml:URI="#author"/>
       </owl11xml:Declaration>
       <owl11xml:Declaration>
        <owl11xml:ObjectProperty owl11xml:URI="#books"/>
       </owl11xml:Declaration>
      <owl11xml:SubClassOf>
           ....
           ....
           ....
      </owl11xml:SubClassOf>
</owl11xml:Ontology>
```

Fig 2. A sample XML schema defined by OWL1.1

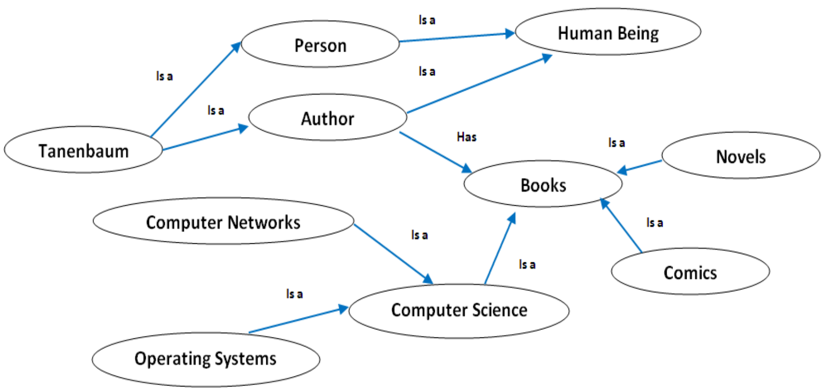

Fig.3: An example ontology describing the author of a book

In Fig 3, sample ontology is given which describes the author of the book. In the above figure, the relationships between the concepts such as is-a, has etc. are described. For example “Tanenbaum is an author (is a relationship) who has (has a relationship) written books in computer science (is a relationship)”.Multiple local ontologies can be mapped onto a global ontology.

$Person(x) \leftarrow L_1(x)$
$Book(y) \leftarrow L_2(y)$
$Human\ Being(x) \leftarrow L_3(x)$
$Author(x) \leftarrow L_4(x)$

In this example, we can consider the concepts on the left hand side are global concepts and that on the right hand side like $L_3(x)$ contain set of local concepts.

3.4 *Grouping of concepts*

Grouping up of concepts plays an important role in ontology integration. An automated, unsupervised, semantic approach is required which can be done dynamically with the help of ontologies. The algorithm is given below.

3.5 *An algorithm for grouping the concepts*

Inputs: Current concept group c_{grp} for a single user in an existing ontology

C_{cur} refers to the current concept

Set of existing concept groups $C=\{c_1,c_2,...c_n\}$

$Thr_{sim}\leq 0\leq Thr_{sim}\leq 1$, wher Thr_{sim} is the similarity threshold value

Output: The concept group c that matches the c_{grp}

(0) $c=\phi$
(1) $Thr_{max}=thr_{sim}$
(2) For i= 1 to m
  (2.1) if $sim(c_{grp},c_i)>Thr_{max}$
    (2.1.1) $c=c_i$
    (2.1.2) $Thr_{max}=sim(c_{grp},ci)$
  (2.2) End if
(3) if $c=\phi$
  (3.1) $c=c\cup c_{grp}$
  (3.2) $c=c_{grp}$
(4)End if
(5) return c_{grp}

4. Results and Discussions

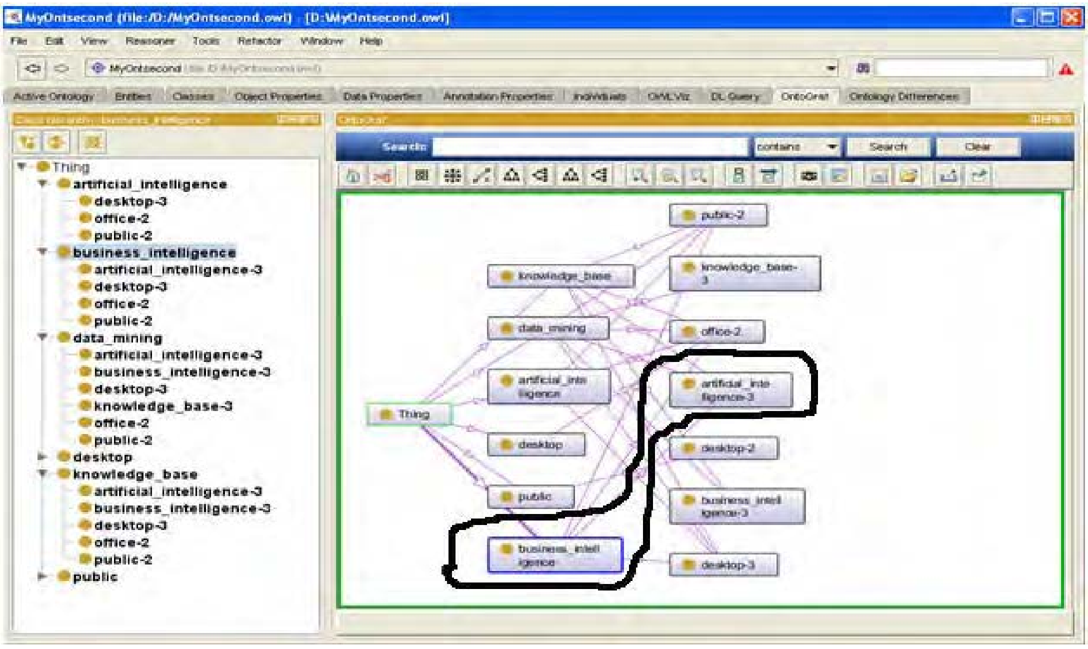

Fig.4 A sample automated ontology with highest Mutual information highlighted

Fig 4, shows a sample automated ontology. Here the mutual information between two concepts 'business intelligence' and artificial intelligence has been referred the most number of times. This will help in analysing the concepts and ranking them based on sequence mining.

4.1 Mutual Information(MI)

MI is more general and it measures the reduction of uncertainty in the concept 'y' after observing the concept 'x'. The advantage of MI over correlation is that MI can measure non-monotonic relations also. Correlation relationships and networks are purely associative.In order to calculate MI more data and computing power is needed. MI can be considered more powerful than correlations in information retrieval techniques.Two concepts are mutually related to one another, only if they have higher MI values. Mutual Information is also called transformation. It is a quantitative measurement of how much on random variable(Y) about another random variable (x), like the one given in equation(3)

$$MI(x,y)=log\,\frac{P(x,y)}{P(X)P(y)} \qquad (3)$$

Mutual information can be continuous or discrete. Mutual information results in the reduction of uncertainty about the knowledge gained. If MI is high, then it indicates a large reduction in uncertainty and if MI is low means, there will be less reduction in uncertainty. If MI between two random variables is zero, then the variables are independent.

4.2 Priority calculation based on the highest MI index

In this approach keyword pair is taken and MI is calculated. The average MI after searching 50 documents has been calculated as .62. This value is considered as a threshold value for accepting a concept. Table I shows the mutual information table, showing the mutual information index and the acceptance of a concept based on the mutual information, for generating the automated ontology. The corresponding algorithm for concept acceptance is also given in section 4.3.

Table 1 . Mutual information table

Instances(Keyword pair)	**Mutual Information index**	**Acceptance based on MI**
{city, airport,}	.77	yes
{journals, airport}	.76	yes
{journals, city}	.55	no
{journals, museum}	.51	no

4.3 Algorithm for acceptance of concepts

```
Input: keyword pair, threshold,MI
For each keyword pair
If MI>threshold
        Return 1
Else
        Return 0
End If
End
```

Keyword pair in the instances are compared with the threshold value. If the MI of the instance pair is greater than the threshold, then the documents comprising such instance pair are accepted. Else such a document is rejected. The impact of the term city and airport together is high. (.77). Using the approach of mutual information, uncertainties are greatly reduced.

We can perform the analysis between manually generated and automated ontology by finding the mutual information index between the ontologies.Let the manual ontology using mutual information approach, be MMI and the automated ontology using the mutual information approach, be AMI.The comparison between MMI and AMI is given in table 2. The concepts with high MI are selected and ranked in both the cases.

Table 2 Comparison Between Manual Ontology And Automated Ontology

Keyword pair	**MBMI (acceptance)**	**ABMI(acceptance)**	**Rank**
{city, airport,}	yes	1	High
{journals, airport}	yes	1	High
{journals, city}	no	0	Low
{journals, museum}	no	0	Low

The results show that the probability match for the acceptance criteria are approximately the same as the automated MI. In the above table, a 1 indicates that the keyword pair is accepted and a 0 indicates that the keyword pair is not accepted . This approach hopefully reduces the uncertainty in semantic matching.

5. Conclusion

Calculating the semantic relationships based on mutual information and the acceptance criteria among the concepts have been done manually and as an automated one. The results produced are invariably the same. The same technique can be applied to hybrid ontologies. This approach hopefully increases the performance based on time, memory, confidence level and support for the concepts. Also the sequence of the concepts mined can also be learned using the automated ontological approach.

Acknowledgement

I am grateful to my research supervisor, Dr. M. Aramudhan for his collaboration and support during preliminary investigations of this work. I would like to thank the reviewers for the efficient preparation of this paper.

REFERENCES

[1] Aleman-Meza B, Halaschek C, Arpinar IB, Sheth A (2005): Ranking Complex Relationships on the Semantic Web. IEEE Internet computing 9(3):37-44.
[2] Wentian Li (1990):Mutual Information Functions Versus Correlation Functions",Journal of Statistical Physics, 60(5-6):823-837.
[3] Xiaohui Tao, Yuefeng Li, and Ning Zhong, Senior Member(2011): A Personalized Ontology Model for Web Information Gathering, IEEE transactions on knowledge and data engineering, Vol. 23, No. 4, pp. 496-511.
[4] Tversky, A., (1997): Features of Similariyt, Psychological Review, 84(4):P. 327-352.
[5] M.Andrea Rodriguez and Max. J. Egenhofer(2000): Determining Semantic Similarity among entity classes from different ontologies, IEEE transactions on Knowledge and Data Engineering.
[6] Gerd Stumme, Alexander Maedche(2005): Ontology Merging for Federated Ontologies on the Semantic Web, Institute for Applied Computer Science and Formal Description Methods (AIFB)University of Karlsruhe, Vol. 3, No. 12, pp. 1-9.
[7] Chin-Ang Wu, Wen-Yang Lin, and Chuan-Chun Wu(2010):An Active Multidimensional Association Mining Framework with User Preference Ontology, International Journal of Fuzzy Systems, Vol. 12, No. 2, pp. 125-135.
[8] Peter D. Karp, Vinay K. Chaudhri, and Suzanne M. Paley(1999):A Collaborative Environment for Authoring Large Knowledge Bases, Journal of Intelligent Information Systems,Volume 13 Issue 3, Nov.-Dec. 1999 Pages 155-194.
[9] S.Vigneshwari, Dr.M.Aramudhan(2013): A Technique to Ontology Mining for Semantic Web Information Extraction ,European Journal of scientific research,Volume 94, Issue 1, PP 49-60.
[10] Heasoo Hwang, Hady W. Lauw.Lise Getoor, AlexandrosNtoulas(2012): Organizing User Search Histories,IEEE Transactions on Knowledge and Data Engineering, 24(5): 912-925.
[11] Janez Brank , Marko Grobelnik , Dunja Mladenic(2005): A survey of ontology evaluation techniques, Proceedings of the Conference on Data Mining and Data Warehouses , SiKDD 2005.
[12] Yanhui Lv ,Chong Xie(2010): A Framework for Ontology Integration and Evaluation. Intelligent Networks and Intelligent Systems (ICINIS). Proceedings of 3rd International Conference on Nov. 2010 Page(s): 521-524.
[13] S.Vigneshwari, Dr.M.Aramudhan(2013): Strengthening ontologies by mapping similarity indices with semantic relationships,Proceedings of the International Conference on Computer Networks and Information Technology, June 29,2013, Bangkok, Thailand,PP:60-63.