IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

An Ontological Approach for the Semantic Web Search and the Keyword Similarity Metrics

R. Deepa[1], Dr. R Manicka Chezian[2]

Research Scholar, NGM College Pollachi, Coimbatore, India[1]

Associate Professor, NGM College Pollachi, Coimbatore, India[2]

Abstract: The Semantic web search is the structured and meaningful search where the keywords with context meaning and relations are taken as important features. The service of web which provides an easier path to find, reuse, share and combines information a user is in need. Semantic Web is the new generation Web which makes possible to express information in precise, machine-interpretable form. It enables intelligent services such as information brokers, search agents and information filters, and also offers greater functionality and interoperability. Semantic Web promotes Web based applications with both semantic and syntactic interoperability. The users search the web by querying method. The words used as query gain importance to find a relevant information. This paper is focused on an ontological approach where the meaning of the words and semantic relations are given importance. The web ontological tool protégé 4.1 is used to find the semantic relation of the web query. Finally the similarity of the ontological keyword with the document is analyzed with keyword weight and relevance based approach.

Keywords: Semantic Web, Ontology, Protégé, Similarity Measure.

1. INTRODUCTION

Semantic web is the extension of the normal web with difference that focuses importance to meaningful conceptual data. The data is represented using the XML technologies and RDF that provides the bendable approach. The Semantic Web provides common formats for the interchange of data. It also provides a common language for recording how data relates to real world objects, allowing a person or a machine to start off in one database and then move through an unending set of databases which are connected not by wires but by being about the same thing (1). The structure of semantic web is a Layered Architecture. Semantic Web is the new generation Web that tries to represent information such that it can be used by machines, not just for display purposes, but for automation, integration, and reuse across applications (2). Furthermore, semantic Web is about explicitly declaring the knowledge embedded in many Web based applications, integrating information in an intelligent way, providing semantic based access to the Internet, and extracting information from texts.

Habitually, HTML provides the standard of structured document published on the Internet. The simplicity of HTML promotes the growth of the Web that it seriously hampered advanced applications such as processing, understanding and semantic interoperability of information contained in several documents.

The explicit representation of meta-information, accompanied by domain theories (i.e. ontologies), will enable a Web to provide a qualitatively new level of service. This process may ultimately create extremely knowledgeable systems with various specialized reasoning services. .The semantic Web technologies offer a new approach to managing information and processes, the fundamental principle of which is the creation and use of semantic metadata.

2. RELATED WORK

This section is classified into two categories, one is about the web language used to generate ontology and the other is focusing on ontology-based techniques. An ontology is defined as “a logical theory that accounts for the intended meaning of a formal vocabulary.” A common feature in ontology languages is the ability to extend preexisting ontologies. Thus users can customize ontologies to include domain specific information while retaining the interoperability benefits of sharing terminology where possible(3). In particular, the OWL Guide (4) is a very good, comprehensive tutorial. The book A Semantic Web Primer (5) also provides a readable introduction to XML, RDF and OWL in one volume. Ontology describes a domain in terms of classes, properties and individuals and may include rich descriptions of the characteristics of those objects (6). Knowledge extraction has become key semantic technology and also to the semantic web [7].

2.1 ONTOLOGY-BASED TECHNIQUES

Ontology is a collection of concepts and their interrelationships, which provide an abstract view of an application domain. It is an explicit specification of a conceptualization. Over the recent years, people who are mentioned below have often held the hypothesis that ontology-based approaches should perform better than traditional ones on IR, since ontologies are more discriminative and arguably carry more “semantics”. As a result, many research concentrate on how to use ontology techniques. Zhong [19] proposes a learning approach for task (or domain-specific) ontology, which employs various mining techniques and natural language understanding methods. Li and Zhong [9] present an automatic ontology learning method, in which a class is called a compound concept, assembled by primitive classes that are the smallest concepts and cannot be divided any further. Liu and Singh [11] develop Concept Net ontology and attempt

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

to specify common sense knowledge. However, Concept Net does not count expert knowledge. Navigli et al. [12] build an ontology called Onto Learn to mine the semantic relations among the concepts from Web documents. Gauch et al. [4] use ontology references based on the categorization of online portals and propose to learn personalized ontology for users. Developed by King et al. [6], Intelli Onto is built based on the DDC (Dewey Decimal Classification) system and attempt to describe the background knowledge. Unfortunately, the previous work on ontology learning covers only a small size of concepts, where mainly uses "Is-A" (super-class, or sub-class) relation in the knowledge backbone. They don't consider to mine and characterize knowledge in a concept level rather than domains. To extend these methods, the backbone of personalized ontologies is been determined to build a real hierarchical structure by applying information in a world knowledge repository.

Ontologies are considered one of the supports of the Semantic Web. Ontologies provide Semantic Web agents with background knowledge about domain concepts and their relationships. Ontologies can also be instantiated to create individuals that describe Semantic Web resources or real-world entities. For example, individuals of an ontology for Realestate agents could represent specific site destinations or activities. In such a scenario, a Semantic Web repository would provide instance data about these individuals, and agents can use their ontological knowledge useful for applications in which knowledge plays a key role, but they can also trigger a major change in current Web contents. This change is leading to the third generation of the Web-known as the Semantic Web-which has been defined as the conceptual structuring of the Web in an explicit machine-readable way. New ontology-based applications and knowledge architectures are developing for this new Web(5).

2.2 DOMAIN ONTOLOGY

The Operation of Ontologies includes Merging, mapping and alignment, refinement, unification, integration, inheritance. It is possible that one application uses multiple ontologies, especially when using linked design of ontologies or when we need to integrate with systems that use other ontologies. In this case, some operations on ontologies may be needed in order to work with all of them. We will summarize some of these operations. The terminology in these areas is still not stable and different authors may use these terms in a bit shifted meaning, and so the terms may overlap, however, all of these operations are important for maintenance and integration of ontologies. Not all of these operations can be made for all ontologies. In general, these are very difficult tasks that are in general not solvable automatically -- - for example because of undecidability when using very expressive logical languages or because of insufficient specification of an ontology that is not enough to find similarities with another ontology. Because of these reasons these tasks are usually made manually or semi-automatically, where a machine helps to find possible relations between elements from different ontologies, but the final confirmation of the relation is left on human. Human then decides based on natural language description of the ontology elements or decides only based on the natural language names of the ontology elements and common sense(12).

The mostly used ontology construction methods including Skeleton method, TOVE ontology, G&FOX method, KACTUS And Bernaras methods, SENSUS, IDEF5 and seven - step method. Ontolingua, Ontodaurus, WebOnto, Protégé, OntoEdit are the common tools to construct ontology models. We have practiced constructing School ontology model.

2.3 THE SEMANTIC WEB: THE ROLES OF XML AND RDF

Resource Description Framework (RDF)is a framework for representing information about resources in a graph form. Since it was primarily intended for representing metadata about WWW resources, it is built around resources with URI.

Information is represented by triples subject-predicate-object in RDF. For instance. It says that "Joe Smith has homepage http://www.example.org/~joe". All elements of this triple are resources defined by URI. The first resource http://www.example.org/~joe/contact.rdf#joesmith (subject) is intended to identify Joe Smith. Note that it precisely defines how to get to a RDF document as well as how to get the joesmith RDF node in it. The second resource http://xmlns.com/foaf/0.1/homepage (predicate) is the predicate homepage from a FOAF (Friend-of-a-friend) vocabulary. The last resource (object) is Joe's homepage http://www.example.org/~joe/.

All of the elements of the triple are resources with the exception of the last element, object that can be also a literal. Literal in the RDF sense is a constant string value such as string or number. Literals can be either plain literals (without type) or typed literals typed using XML Data types [24].

XML and RDF are the current standards for establishing semantic interoperability on the Web, but XML addresses only document structure. RDF better facilitates interoperation because it provides a data model that can be extended to address sophisticated ontology representation techniques. We explain the role of ontologies in the architecture of the Semantic Web. We then briefly summarize key elements of XML and RDF, showing why using XML as a tool for semantic interoperability will be ineffective in the long run. We argue that a further representation and inference layer is needed on top of the Web's current layers, and to establish such a layer, we propose a general method for encoding ontology representation languages into RDF/RDF schema(4).

Semantically transparent services will make it possible for clients to successfully use ser-vices that are dynamically discovered without prior negotiations between client and service developers. Such goals are important for commercial Web service environments, including business-to-business and business-to-consumer applications, grid computing, ubiquitous computing, and information management (2).

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

3. PROPOSED METHODOLOGY

Ontologies play an important role in semantic Web Query Preprocessing. Query preprocessing is a necessary step for extracting terms and aspects. The important function of this section is to eliminate the insignificant words and filter the major keywords.(6)They provide formal models of domain knowledge that provide services to access, visualize, edit, and use ontologies for intelligence support(23). The paper reveals how to use OWL – Protégé supports the owl features. The following figure1 shows the instance of Owl – Ontological graph for "Institutional" Ontology. Thus the classes, object property, Annotations are given through the protégé OWL tool and the resultant graph Fig. is taken where the diagrammatic representation reveals that the Institutional Ontology has the following classes like the Root, Person, Faculty, Teaching Faculty, Non-Teaching Faculty, Students, Bachelors Degree Students, Masters Degree Students, Course, Bachelors Degree, Masters Degree. Thus the tree hierarchy is drawn where the relation of the nodes along the path of the root node is calculated using the depth wise search and the similarity is calculated using the formula Similarity Based on Keywords or Phrases summation of the (keyword weight +Relavance Weight)/2*100.This measure directly comes from IR studies. Keywords are the words, except for function words included in a stop-list. All the keywords are stemmed using the Porter algorithm [22].

Keyword Strength = ((Words Presence * Phrase Word Count) / total words count) * 100

$$S(KW_i) = (\Sigma WP * \Sigma PWC)/\Sigma\,(TWC)) * 100$$

$$S(KW_i) = (\frac{1}{\Sigma(KWi)})$$

Where $S(KW_i)$ is Strength of keywords, $0 \le i \le 1$, WP is Number of keyword Word present in the current page TWC is the Total Number of Words in the current page.

Keyword relevance based weight for each link = The Root Value + (Level of Relevancy of Keywords /2* Height of the node)

$$S(KRBW_i) = ((Zi + Li)/2Hi * 100$$

$$TRi = \sum (S(KWi) + S(KRBW_i))/2$$

Where $S(KRBW_i)$ is the keyword relevance based weight, Zi is the root Value , Li is the Level of Relevancy of Keywords, Hi is the Height of the node, TRi is the Total Relavancy.

Thus combining the two measures we take the keyword-based similarity function is defined as follows: similarity keyword (p, q) = KN(p, q) / Max(kn(p), kn(q)) (1)

where kn(.) is the number of keywords in a query, KN(p, q) is the number of common keywords in two queries. If query terms are weighted, the following modified formula can be used instead:

Similarity $_{wk}$ (p, q) = $\sum_{i=1}^{N}(w(k\,(p))\,w(k\,(q)))/Max(\,kn(p),kn(\,q)) * 2$ where w(ki(p)) is the weight of the i-th common keyword in query p and kn(.) becomes

Algorithm 1: Relevancy or Weight based approach

Input: Extracted Web Contents
Output: weight calculation of Web Content
Step 1: Extract Search Engine results SR_i for the user query where $1 < i < N$
Step 2: Pre-process user query and extract root words RW_j where $1 < j < N$
Step 3.Construct Dictionary D for the user query RW_j
Step 4. Extract and Pre-process the keywords KWi for the search results SR_i
Step 5. Compute Keyword Strength

$$S(KW_i) = (\Sigma WP * \Sigma PWC)/\Sigma\,(TWC)) * 100$$

Step 6. Compare each keyword KW_i against Dictionary D.
Step 7. If match is found then award strength $S(KW_i)$ to particular keyword the count is incremented.
Step 8. Else award 0 as strength for particular keyword.
Step 9 Calculate Total Strength for Keyword

$$S(KRBW_i) = ((Zi + Li)/2Hi * 100$$

Step 10. Compute Total Relevancy for the particular link

$$TRi = (\sum (S(KWi) + S(KRBW_i)))/2$$

Step 11. Repeat step 4 to 10 for all Search Results (SR)

the sum of weights of the keywords in a query. There are two main methods for doing this. One is by using a noun phrase recognizer based on some syntactic rules [6]. Another way is to use a phrase dictionary. In Encarta, there is such a dictionary. It contains a great number of phrases and proper nouns that appear in Encarta documents. This dictionary provides us with a simple way to recognize phrases in queries. However, it may not be complete.

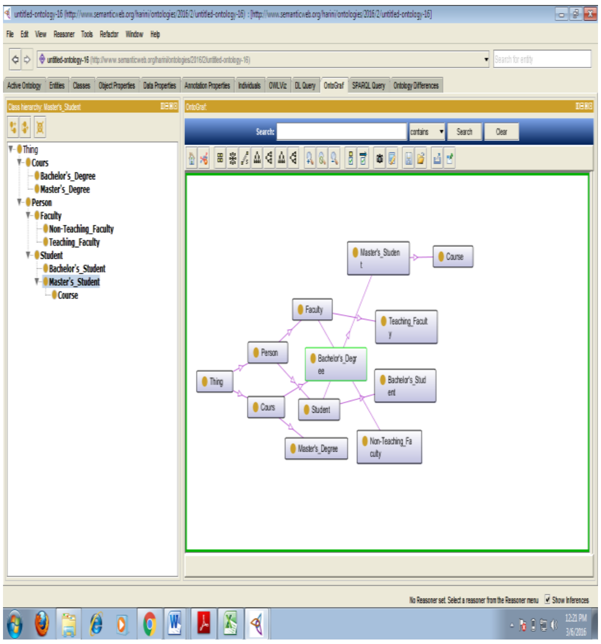

Figure 1 Onto-Graph for Institutional Domain

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

In the future, it will be supplemented by an automatic phrase recognizer based on a syntactic and statistical analysis (7).

The following figure1 shows the instance of Owl – Ontological graph for "Institutional" Ontology. Thus the classes, object property, Annotations are given through the protégé OWL tool and the resultant graph Fig. is taken where the diagrammatic representation reveals that the Institutional Ontology has the following classes like the Root, Person, Faculty, Teaching Faculty, Non-Teaching Faculty, Students, Bachelors Degree Students, Masters Degree Students, Course, Bachelors Degree, Masters Degree. I have also defined some of the data type properties for each class. Some of the data type properties for each class.

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [
   <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
   <!ENTITY xsd
"http://www.w3.org/2001/XMLSchema#" >
   <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-
schema#" >
   <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-
syntax-ns#" >
]> <rdf:RDF
xmlns=http://www.semanticweb.org/harini/ontologies/201
6/2/untitled-ontology-16#
xml:base=http://www.semanticweb.org/harini/ontologies/2
016/2/untitled-ontology-16
xmlns:rdfs=http://www.w3.org/2000/01/rdf-schema#
xmlns:owl=http://www.w3.org/2002/07/owl#
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
ns#"><owl:Ontology
rdf:about="http://www.semanticweb.org/harini/ontologies/
2016/2/untitled-ontology-16"/>
//                    Classes                    <!--
http://www.semanticweb.org/harini/ontologies/2016/2/unti
tled-ontology-16#Bachelor&apos;s_Degree              -->
<owl:Classrdf:about="http://www.semanticweb.org/harini/
ontologies/2016/2/untitled-ontology-
16#Bachelor&apos;s_Degree">
<rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/harini/ontologi
es/2016/2/untitled-ontology-16#Cours"/>
</owl:Class>
<!--
http://www.semanticweb.org/harini/ontologies/2016/2/unti
tled-ontology-16#Bachelor&apos;s_Student -->
<owl:Classrdf:about="http://www.semanticweb.org/harini/
ontologies/2016/2/untitled-ontology-
16#Bachelor&apos;s_Student">
<rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/harini/ontologi
es/2016/2/untitled-ontology-16#Student"/>
</owl:Class>
```

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

Table 1: Concept Weight and Relevancy weight table

Ont. Concept	Hierarchy	Concept Weight	Link 1 Twc (Total Word Count of a page)	Tdf (Term Frequency in a page)	Keyword Weight	TRi
Root	0	0.5	3021	3	0.000993049	0.250497
Person	1	0.625	3021	0	0	0.3125
Faculty	2	0.75	3021	7	0.002317114	0.376159
Teaching Faculty	3	0.875	3021	9	0.002979146	0.43899
Non-Teaching Faculty	3	0.875	3021	10	0.003310162	0.439155
Students	2	0.75	3021	24	0.007944389	0.378972
Bachelors Degree Students	3	0.875	3021	58	0.019198941	0.447099
Masters Degree Students	3	0.875	3021	48	0.015888779	0.445444
Course	1	0.625	3021	23	0.007613373	0.316307
Bachelors Degree	2	0.75	3021	34	0.011254551	0.380627
Masters Degree	2	0.75	3021	24	0.007944389	0.378972

The table above explains that the weight found by concept hierarchy and the weight found by the relevancy based are hardly deviated and when it is combined the metric reveals that the link of the concept keywords are matched relatively and word exact match i.e. as well as the total keyword relevancy is found key. If the metrics are combined the total relevancy is deviated much from the keyword based approach and is proportionate to the Concept based method. The following figure2 explains the graph of the total relevance table.

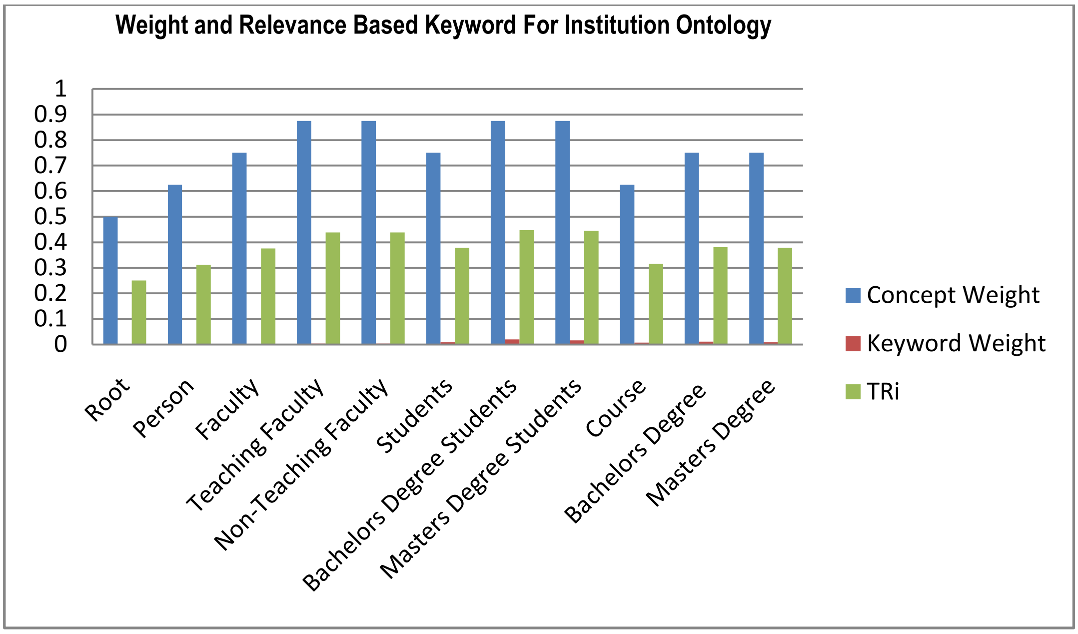

Figure 2 Graph for total Relevancy obtained combining Concept Weight and Keyword Word Weight

5. CONCLUSION

The web ontology language OWL is used to create an institutional ontology and the concept hierarchy for the ontological keywords are framed. The weight of the relational keywords is calculated using the concept hierarchy and relevance based method. The analysis states and infers that the relational keywords show a difference in the closeness to the keyword. The concept hierarchy weight shows a high correlation to the keyword whereas the relevance based weight shows the high deviation in closeness. Thus the studied approach provides more fine tuning process is needed for bringing out the best result. Then the query recommendation is displayed making use of ontological approach. By this means, the time user spends for searching the required information from the search result list can be reduced and the more conceptual keywords can be presented. The results obtained from practical evaluation are quite promising in improving the efficiency of the interactive web search engines. The experimental results clearly reveal that the proposed approach outperforms two baselines in both coverage and quality. It enhances the quality by using different query processing techniques on the results.

REFERENCES

[1] The Semantic Web: An Introduction and Issues. International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622, Vol. 2, Issue 1,Jan-Feb 2012, pp.780-786

[2] T. Berners-Lee, J. Hendler, O. Lassila et al. The semantic web. Scientific american, Volume 284, Number 5, pages 28–37, 2001.

[3] N. Guarino, Formal Ontology and Information Systems. In Proceedings of Formal Ontology and Information Systems, Trento, Italy. IOS Press, 1998.

[4] M. Smith, C. Welty, and D. McGuinness, eds. OWL Web Ontology Language Guide,W3C Recommendation, 10 February 2004, http://www.w3.org/TR/2004/REC-owl-guide- 20040210/ .

[5] G. Antoniou and F. van Harmelen. A Semantic Web Primer. MIT Press, Cambridge, MA, 2004.

[6] Bechhofer, Sean. "OWL: Web ontology language." Encyclopedia of Database Systems. Springer US, 2009. 2008-2009.

[7] Gangemi, Aldo. "A comparison of knowledge extraction tools for the semantic web." The Semantic Web: Semantics and Big Data. Springer Berlin Heidelberg, 2013. 351-366.

[8] IEEE Internet Computing, ed. Michael N. Huhns and Munindar P. Singh, Volume 9, Issue 5, 2005, pages 72-81.http://ieeexplore.ieee.org/servlet/opac?punumber=4236© 2005 by the Institute of Electrical and Electronics Engineers (IEEE)

[9] Y. Li and N. Zhong. Mining ontology for automatically acquiring web user information needs. IEEE Transactions on Knowledge and Data Engineering, pages 554– 568, 2006.

[10] R. Navigli, P. Velardi and A. Gangemi. Ontology learning and its application to automated terminology translation. IEEE Intelligent Systems, Volume 18, Number 1, pages 22–31, 2003.

[11] H. Liu and P. Singh. Commonsense reasoning in and over natural language. In Knowledge-Based Intelligent Information and Engineering Systems, pages 293–306. Springer, 2004.

[12] Pratik Agrawal, Prof. A. J. Agrawal ―Implementation of Semantic Analysis Using Domain Ontology‖ (Department of Computer Science, Shri Ramdeobaba College of Engineering & Management Nagpur, India IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727Volume 11, Issue 2 (May.-Jun. 2013), PP 72 -76.

[13] Decker, Stefan, et al. "The semantic web: The roles of XML and RDF." Internet Computing, IEEE 4.5 (2000): 63-73.

[14]Gómez-Pérez, Asunción, and Oscar Corcho. "Ontology languages for the semantic web." Intelligent Systems, IEEE 17.1 (2002): 54-60.

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

[15] Uthayan, K. R., and G. S. Anandha Mala. "Hybrid Ontology for Semantic Information Retrieval Model Using Keyword Matching Indexing System." The Scientific World Journal 2015 (2015).
[16] J.-R. Wen, J.-Y. Nie, and H. Zhang. Query clustering using user logs. ACM Trans. Inf. Syst , 20(1):59–81, 2002.
[17] Wilson, W., W. Liu and M. Bennamoun, 2012. Ontology learning from text: A look back and into the future. ACM Comput. Surv., 44: 36-36. DOI: 10.1145/2333112.2333115
[18] S. Gauch, J. Chaffee and A. Pretschner. Ontology-based personalized search and browsing. Web Intelligence andAgent Systems, Volume 1, Number 3, pages 219–234, 2003.
[19] N. Zhong. Representation and construction of ontologies for Web intelligence. International Journal of Foundations of Computer Science, Volume 13, Number 4, pages 555–570, 2002.
[20] X. Jiang and A.H. Tan. Learning and inferencing in user ontology for personalized Semantic Web search. Information Sciences, Volume 179, Number 16, pages 2794–2808, 2009.
[21] Porter, M. An algorithm for suffix stripping. Program, Vol.14(3), pp. 130137, 1980
[22] Knublauch, Holger, et al. "The Protégé OWL plugin: An open development environment for semantic web applications." The Semantic Web–ISWC 2004. Springer Berlin Heidelberg, 2004. 229-243.
[23] R. Deepa, R Manicka Chezian An Involuntary Data Extraction And Information Summarization Expending Ontology International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 10 No.44 (2015).
[24] R.Deepa, R. Manicka Chezian Keyword based web page re-ranking using relevancy and bee algorithm by incorporating user behaviour information in web mining, 103rd Indian Science Congress, 05 January 2016.
[25] R. Deepa , R. Manicka Chezian An Involuntary Data Extraction and Information Summarization Expending Ontology Artificial Intelligence and Evolutionary Computations in Engineering Systems Volume 394 of the series Advances in Intelligent Systems and Computing pp 59-68, 06 February 2016