Ontology Based Information Retrieval System using Multiple Query

M.P. Navale[1], Rohini Kasar[1], Sneha Kasbe[1], Priyanka Shahane[1]

Department of Computer Engineering, NBN Sinhgad School of Engineering, Pune[1]

Abstract: In traditional information retrieval system searching operation is performed based on the keywords which result in large amount of data due to this, it becomes to difficult to find out essential data from this large data. In order to overcome this drawback of traditional information system the new web architecture known as semantic web was introduced. In this paper we described the ontology based semantic information retrieval and jena semantic web framework. In this system input query entered by the user is parsed by Standford parser after which we apply triplet extraction algorithm in order to retrieve the desired information.

Keywords: Ontology Based Information Retrieval system: Semantic Web, Jena API, SPARQL, Query Processing.

1. INTRODUCTION

Ontology formally describes a list of terms represent important concepts such as, classes of objects and the relationships between them. Information retrieval search engine is engine where the user can search answers from ontology is the concepts exist in the knowledge based the system can answer the question quickly. Natural Language processing technique is mostly implemented in question answering (QA) system for answering users question. In the previous work the search engine which performs searching operation which based on keywords which result into large amount of data to the user, this make user difficult to select the proper option.

In Computer science and information science, an ontology is a formal naming and definition of the types, properties and interrelationships of the entities that really or fundamentally exist for a particular domain of discource. It is thus a practical application of philosophical ontology with taxonomy.

The Objective of This system is:
Ontological database is developed or create for storage which is baesd on domain specific knowledge.
SPARQL query is generating from user input query. Semantic search of natural language query from user input query.
Retrieve or extract the answer from domain specific knowledge of ontological database.

2. METHODOLOGY

Semantic Web:
The Semantic web is nothing but the extension of web through standards by the World Wide Web Consortium (W3C). Semantic web technologies enable people to create data stores on web, build vocabularies and write rules for handling data. Linked data are enpowered technologies. Such as RDF, SPARQL, OWL, SKOS.

Ontology:
Ontology is a formal naming an definitions of any types, properties and interrelationships of the entities that really or fundamentally exists for particular domain of discource.

SPARQL:
SPARQL is query language for RDF W3C. RDF is directed labelled graph data format for representing information in the web. The SPARQL is used to express the queries across various data sources, whether the data is stored natively or viewed as RDF via middleware.

3. JENA API

It is used to map SPARQL query on RDF. It stores the information as RDF triplets in directed graph and allows our code to add, delete, change, store and publish information. RDF API (Application Programming Interface) provides to an and remove triplets to graph and find the triplets that match with specified pattern.

4. SYSTEM ARCHITECTURE

In our proposed system, user interface is created when user enters a input query in natural language. Input query may be in formal representation or in SQL statement. Then Standford parser processes, for all the input query parse tree that is tree bank structure is constructed by Standford parser.

Using ontology on to triplets are constructed. Then SPARQL query is formed and it is fired on the knowledge base that is find the appropriate RDF triplets in knowledge and retrieve the relevant information using jena semantic web framework.

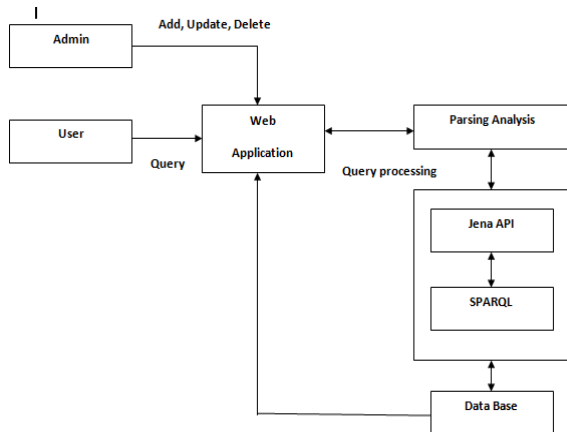Sample output for query parsing by Standford parser and then get subject, predicate, object by using triplet.

Fig: System Architecture

5. PROPOSED ALGORITHM

Aim of the proposed algorithm is to extract the meaningful concept is extracted from the user input query. Using tis concept query expansion is performed. QA system using Ontology technique have hire result precision that the system than the keyword based the information system. The proposed algorithm is consist of three main steps.

Algorithm Extract Subject (String):
1. Perform a Breath First Search(BFS) of the parse tree obtained by using Standford core NLP library.
2. The NP subtree contains the subject, and it is the first noun in the tree when traversed BFS.

Algorithm Extract Predicate (String):
1. Perform a Depth First Search of the VP subtree, the verb that is deepest in the tree is the predicate.

Algorithm Extract Object (String):
1. Perform a search of the PP, ADFP subtree, and extract the first noun the tree. This is the object.

We have implemented an essential version of the above algorithm using java and Standford NLP package, we find our implementation to work for simple sentences (Sentences that have a single subject, single predicate and single object), and need to be enhanced for more complex sentences.

6. PSEUDO CODE

Step 1: Enter the input query in natural language.
Step 2: Standford parser is processes all the input queries.
Step3:SPARQL query is formed and is fired on knowledge bases that appropriate RDF triplets in Knowledge base.
Step 4: Then Jena API map the SPARQL query on RDF.
Step 5: Apply Triplet Extraction Algorithm.
Step 6: If match is found then
      Display book list
       Else
      Display Book Not Found message.
Step 7: End.

8. CONCLUSION

We believe and argue that our suggested methodology is quite generic, and can be easily adapted to any new domain such as through we used the library domain to validate our approach. We believe that the methodology of defining rules for pattern is quite reasonable and simple. An application of this type becomes valuable when used across large number of website to extract relevant information and it to ontology. The updated ontology can then be used to handle semantic queries which otherwise would not be possible. This system is capable of handling different multiple queries. This methodology is domain specific so in future this method can be applied for different domain.

REFERENCES

[1] XiangbinXu, Gonzalez Moctezuma Luis, Andrei Lobov, Jose L. Martinez Lastra . "Multiple Ontology Workspace Management and Performance Assessment" International Conference on Industrial informayion(INDIN)2015 IEEE 13th .
[2] Yinghui Huang, Guanyu Li, Oiang piang Li" Rough Ontology Based semantic informayion Retrival" ,international Symposium on Computational intelligence and Design(ISCID),2013 16th.
[3] T.Kanimozhi,A. Christy , "Incorporating ontology and SPARQL for semantic image annotation"2013 IEEE Conference on Information and Communication Technologies (ICT)
[4] "Domain ontology based semantic search for information retrieval through automatic query expansion", Dept. of Computer science and Engineering GEU Deharadun, India,2013 International Conference
[5] Jun Zhai, Kaitao Zhou "Semantic Retrieval for sports information based on ontology and SPARQL" information Science and management Engineering (ISME),international conference of. 2010
[6] RashmiChauhan, Ryan Goudar, Robin Sharma, AtulChauhan, "Domain ontology based semantic search for information retrieval through automatic query expansion", Dept. of Computer Science and Engineering GEU Deharadun, India,2013 International Conference