Efficient Information Retrieval Using Domain Ontology

Miss Pratibha S. Sonakneware
Dept. of Computer Technology
Yashvantrao Chavhan college of Engineering
(An Autonomous Institution affiliated to RTMNU)
Nagpur(Maharashtra),India
psonakneware@gmail.com

Mr. Shivkumar J. Karale
Dept. of Computer Technology
Yashvantrao Chavhan college of Engineering
(An Autonomous Institution affiliated to RTMNU)
Nagpur(Maharashtra),India
sjkarale@in.com

Abstract— **Currently most of information is accessed from world wide web from many search engines. These search engine search information on the basis of keywords, and that keyword based search engine searches enormous amount of data. User takes more time to access relevant information. So, in order to overcome the limitation of keyword based search technic of conceptual search is done, i.e. search by meaning instead of search by literal string. Search engine interpret the meaning of user's query and the relations among the concepts that a document contains with respect to a particular domain. In this paper, we proposed Information Retrieval system in which, user enters a query and for that query meaningful concepts are Extracted using domain ontology, For all the terms (expanded and initial query terms), SPARQL query is formed and then it is fired on the knowledge base(ontology) that finds appropriate RDF triples in knowledge Base. In ou r proposed system we are converting the natural query to sparql query using quepy tool and creating the knowledge base system by using Protégé tool.**

Keywords- **RDF; SPARQL; Quepy; Ontology; Information Retrieval.**

I. INTRODUCTION

In today's world phenomenal growth of World Wide Web resources has made the improvement in information retrieval technic has extremely important[4]. Presently used search engines for information retrieval are keywords based. User is provided with more results, from that finding the relevant data is challenging task. The reason behind this lies in the inability to handle the words with ambiguous meanings. For example consider the meaning of word "can", taken as either "container for storing water" or " be able to". In keyword based system, retrieved data for all meanings and in ontological concept retrieval concept for our domain. By using ontology based search above problem is solved[1].User can use ontological concept to search conceptual and semantic information. Ontology has a vital role in access and interchange of information, use and reuse of knowledge, sharing of information, and common understanding of specific domain are communicated among people for developing their applications. Ontology can describe things and their properties and interrelations in a way that computers can process and automate.

Thomas Gruber, defines ontology as "explicit specification of conceptualization", means that ontology describe a relationship between different semantic concept and their properties.Ontology based semantic search means ability to access the most relevant data for user query from our knowledge base.

In our proposed system, ontological databse is in RDF data format. For RDF data format SPARQL is a query language.So that we are converting the natural input query to SPARQL query. After that SPARQL query is fire on RDF database and retrieve the efficient information.

II. GENERALIZED ARCHITECTURE OF IR SYSTEM

In our proposed system , a user enters a query and for that query meaningful concepts are Extracted domain ontology, For all the terms , SPARQL query is formed and then it is fired on the knowledge base(ontology) that finds appropriate RDF triples in knowledge Base. Web documents related to the requested query and individuals specified in these triples are then retrieved. The flow of sequence sequence of our IR system are as shown in figure 1.

In figure 1 shows that input query is tokenizing first and then the key words of domain terms are expanded and that words are convert into SPARQL query language[15] . Then, that SPARQL query is map on ontlogical database in the RDF format and extract the relevent concept from knowledge base and retrieve the relevant information.

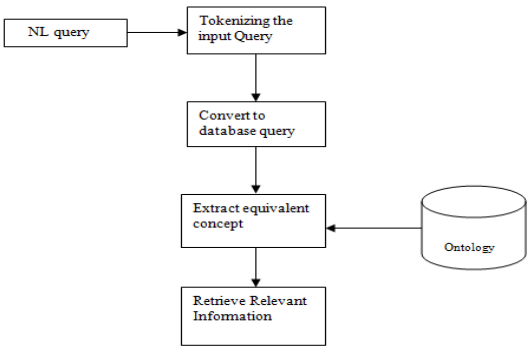

Figure 1: Generalized view of IR System

As is depicted in figure 1, input query is tokenizing first and then the key words of domain terms are expanded and that words are convert into sparql query language[15] . Then, that sparql query is map on ontlogical database and extract the relevent concept and retrieve the relevant information.

III. Proposed System

We proposed a framework for information retrieval system that follows a query expansion. .We have presented a ontology based technic for query expansion that has been integrated with the help of specified tools, to our search scheme. To sum up, the retrieval process includes following steps:

- The IR system converts user input query is converted into SPARQL query.
- The SPARQL is a language for RDF database.Thus generated SPARQL query is map on the knowledge base, which has been stored in RDF format.
- Sparql query find the semantic relation in ontology .
- Semantic information is retrieved to user.

We have presented the overall architecture for our system shown as Fig.2. The system consists of four basic modules: User Interface, Query Handling and Semantic Search Module, Domain Ontology construction.

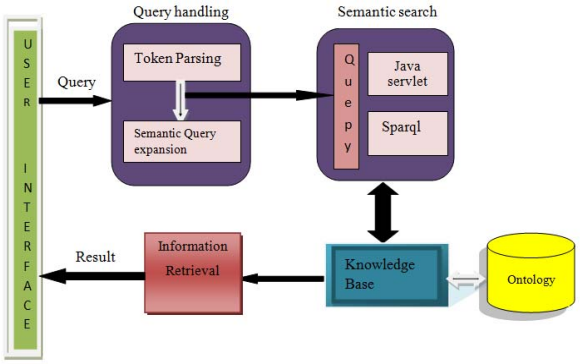

Figure 2: Proposed IR System

A. User Interface Module

The user give the input query for accessing what information he wants to get. After the processing of other modules the user will get the relaevent information accordingly our database.

B. Query Handling Module

The query given by the user through user interface is handled by query handling module. The meaningful concepts are extracted through the tokenization of the input query and expanded through semantic query expansion by using wordnet[15].WordNet is used for find the synonyms of the words. The natural language question given by the user is then semantically expanded to SPARQL query language by using a quepy tool. The quepy convert query to sparql semantically by using NLTK_DATA contains WordNet. .

C. Semantic Search Module

This module is able to interpreting the user's input query . In this module input query converted into sparql query which find the semantic RDF data in domain ontology. SPARQL is a query language for RDF data. It's work like a SQL query language for Relational data[10]. The difference is that SQL accessing the data in Tables and SPARQL accessing the data in given namespace. Namespace is also called prefix which is in the URI link, Means that accessing the data from that resources. So, SPARQL is more advanced than SQL because SPARQL access relational as well as diverse data stored in link.

SPARQL is a RDF query language for semantically map on a RDF data source and retrieve semantic data from knowledge base. SPARQL show the semantic meaning of user query for retrieve semantic data from RDF. SPARQL match the prefix i.e. namespace with RDF namespace . then search data from that URL properties. SPARQL query is map with RDF through JENA and retrieve the answer. We are taking the example: " What is the population of india?", Query generated as follow:

```
PREFIX dbpedia: <http://dbpedia.org/ontology/>
PREFIX dbpprop: <http://dbpedia.org/property/>
SELECT  population  ?x1
WHERE {
 ?x0 rdf:type dbpedia-owl:Country.
 ?x0 rdfs:label "India"@en.
 ?x0 dbpprop:populationCensus ?x1.
```

Prefix means taking the data from that URL and for that particular URL find the data properties. In select clause select population and where clause search population for India.

978-1-4799-3759-2/14/$31.00©2014 IEEE

D. Domain Ontology Construction Module

Ontology has been created for a particular domain and is used to model the knowledge for this domain in terms of Concepts (various terms of a specific domain) Relationships between concepts. Ontology show the hierarchical relationship between different classess and their subclasses in graphical pattern as shown in figure 3:

Figure 3 show the ontology for sport domain. It show the relationship between different classess,subclasses and theit attributes.

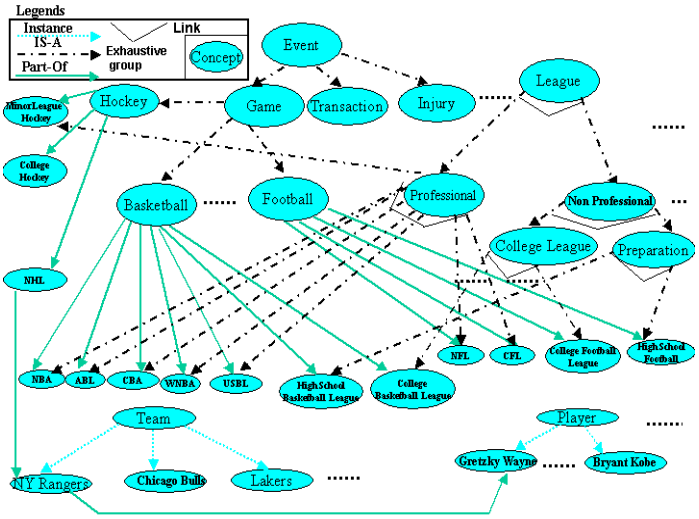

Figure 3: Ontology for sport domain

For developing an ontology for particular domain ,find all the classess in the domain. In the next step, find the properties of each of these classess. And find the relationship between them.The next step would be identifying the characteristics of the properties like Transitive, Symmetric, Functional and Inverse Functional [10].
Design the domain specific ontology following steps are consider.

Step 1: Collect all details regarding the domain.

Step 2: List important terms in ontology.

Step 3: Identify the classess and subclassess can be created .

Step 4: Third step is to identify the properties and characterstics exists in classess and subclassess.

Step 5: Bind the appropriate properties to approprite class.

Step 6: Develped ontology.

IV. IMPLEMENTATION DETAILS OF IR SYSTEM

The proposed system is implemented with appropriate tools and the proposed techniques. The implementation details of our proposed system are described in short in the next sections.

A. User Interface for IR System

Figure 4 show the user interface for IR system . In that type our question that question is fire on knowledge base and retrieve the relevant information on output window.

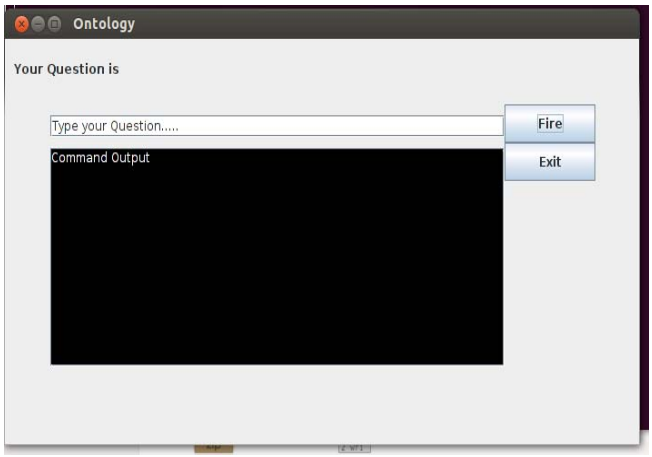

Figure 4: User Interface of IR system

B. Semantic Search Using Quepy

In our proposed system semantic search of user input query is done by using SPARQL query. Sparql ia a query language for retrieving the semantic data from RDF database.

For creating the ontology , system use the database in RDF data format. RDF format shows the relationship between different object and their attributes. For RDF database language is SPARQL. SPARQL language is difficult to understand user as compared to natural language. So it is important to convert natural language query into sparql query[4]. So for converting user input query, taken as natural language query into SPARQL query by using “Quepy” tool. Interfacing between sparql query and RDF data by using jena API.

Quepy is a tool for converting natural language input queries into SPARQL query. Quepy tool for generating database queries to be access through the Dbpedia database. For accessing our own system database instead of dbpedia give the path of our own database . For accessing different more types of question we are adding the POS tagging in the different form of question templates of quepy program then accessing the our own database. Quepy is a python

978-1-4799-3759-2/14/$31.00©2014 IEEE

framework transfer natural language question into database language as follows.

Quepy is installed on ubuntu 12.04. For checking installation successfully to show the quepy version .

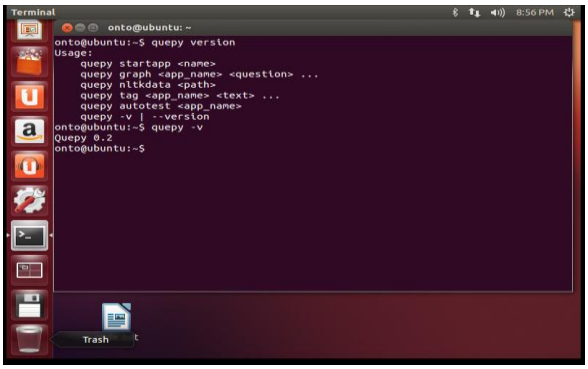

Figure 5: Installation of Quepy

As shown in figure 6 show the Input query is converted into saparql query.After giving the path of database input query converted to sparql query. In figure 7 shown that quepy is interfacing with UI.

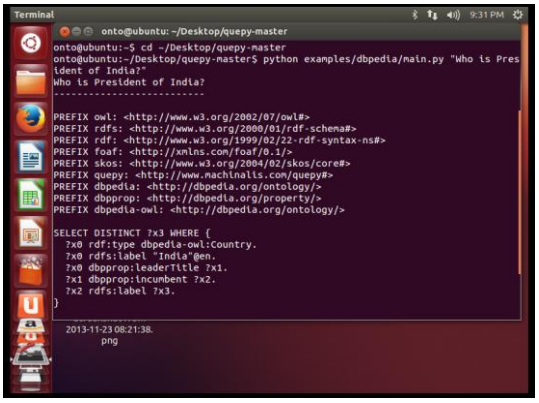

Figure 6: Input query converted to Sparql

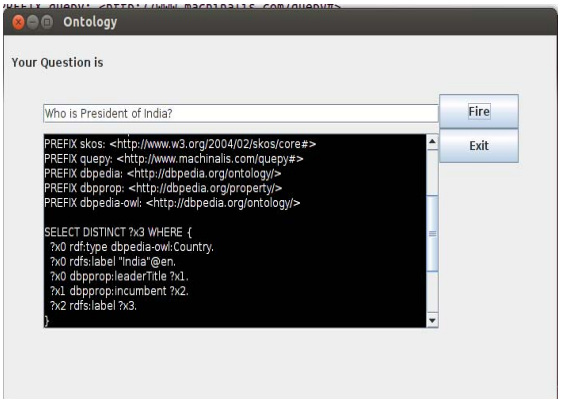

Figure 7: Quepy interfacing with Sparql

C. *Creating Domain Ontology by using Protege*

Protégé is a tool which convert data into RDF data format. We have used Protégé tool to create ontology for particular domain. The ontology for storing information about of various classes, individuals and properties with their attributes and relationships.

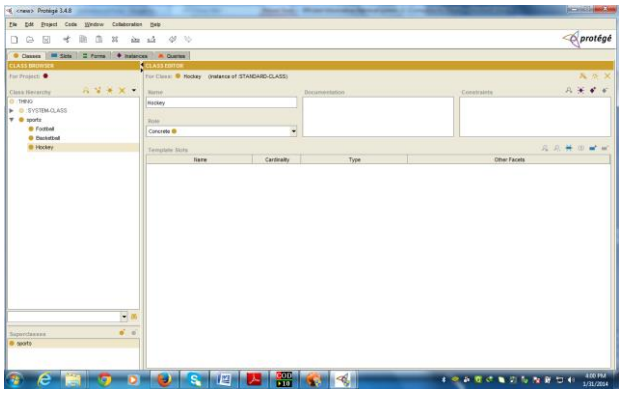

Figure 8 : Creation of sport class by using protégé tool

Figure 4 show the creation of attributes of sport class and their subclassess.Sport class have attributes national or international and show the relationship between them.

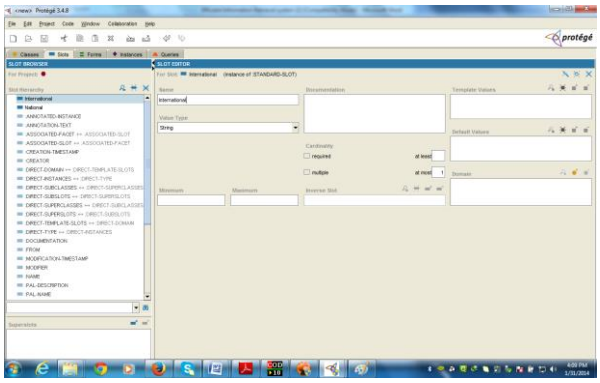

Figure 9: Creation of instances of sport class

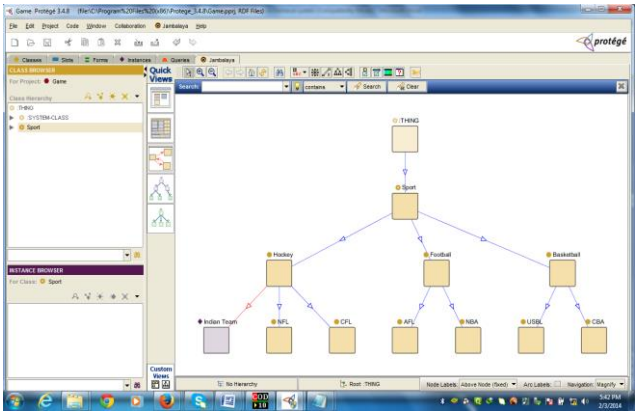

Figure 10: Hierarchical relationship of sport class with subclassess

978-1-4799-3759-2/14/$31.00©2014 IEEE

Figure 9 shows the class and their properties for creation of ontology. Figure 10 shows the hierarchical relationship of sport class with their subclasses. Ontology create in RDF data format as follows:

```
<rdfs:Class rdf:ID="sports"/>
<rdf:Property rdf:ID="Basketball">
<rdfs:domain rdf:resource="#sports"/>
 </rdf:Property>
 <rdf:Property rdf:ID="Football">
 <rdfs:domain rdf:resource="#sports"/>
 </rdf:Property>
<rdf:Property rdf:ID=”Hockey”>
 <sports rdf:about="#sport">
  <Football  >National</Football>
 <Basketball><National></basketball>
<Hockey><National></Hockey>
 </sports>
</rdf:RDF>
```

In this way by using protégé creating the ontological database in RDF data format. Sparql query is mapped with RDF database by using jena API then sparql query is fired on RDF database then retrieve the relevent infotmation.

V. CONCLUSION

We have proposed the approach of Information Retrieval system for solving the limitation of keyword-based system using the ontological concept. IR system extract relevant information instead of giving list of document containing related information. The IR system has been implemented using appropriate tools. Quepy tool is used for converting natural query to sparql query.Sparql query is used to extract the RDF data with respect to user input query. Protégé tool is used for creating ontology in RDF data format. Experimental result shows that our system retrieve some type of queries answered. Future work will extend as IR system gives the answer of all possible types of questions.

.

REFERENCES

[1] Sandhya Revuri, Sujatha R Upadhyaya , P Sreenivasa Kumar,” Using Domain Ontologies for Efficient Information Retrieval”, Department of Computer Science and Engineering Indian Institute of Technology Madras Chennai - India.

[2] Anatoly Gladun, Julia Rogushina, Victor Shtonda,” Ontological Approach To Domain Knowledge Representation For Information Retrieval In Multiagent Systems”, International Journal "Information Theories & Applications" Vol.13.

[3] Miriam Fernández , Iván Cantador , Vanesa López , David Vallet , Pablo Castells , Enrico Motta ,”Semantically enhanced Information Retrieval: An ontology-based approach “Web Semantics: Science, Services and Agents on the World Wide Web (2011).

[4] Rashmi Chauhan, Rayan Goudar, Robin Sharm, Atul Chauhan,” Domain Ontology based Semantic Search for Efficient Information Retrieval through Automatic Query Expansion”, Dept. of Computer Science and Engineering GEU Dehradun, India, 2013 International Conference.

[5] Hany M. Harb,Khaled M. Fouad, Nagdy M. Nagdy,”Semantic Retrieval Approach for Web Document”,Inrwenational Journal of Advance Computer Science and Application 2011.

[6] Lakshmi Tulasi R., Goudar R.H., Sreenivasa Rao M., Desai P.D.” Domain Ontology Based Knowledge Representation for Efficient Information Retrieval”, Journal Of Information Systems And Communication Issn: 0976-8742 & E-Issn: 0976-8750, 2012.

[7] J.Uma Maheswari, Dr. G.R.Karpagam,” A Conceptual Framework For Ontology Based Information Retrieval”, Professor, Department of Computer Science and Engineering, PSG College of Technology, Coimbatore, Tamilnadu, International Journal of Engineering Science and Technology , 2010.

[8] Latifur Khan, Dennis McLeod, Eduard Hovy,” Retrieval Effectiveness of an Ontology-Based Model for Information Selection” Department of Computer Science University of Texas at Dallas Richardson.

[9] Liu Xinhua, Zhang Xutang, Li zhongkai ,” A Domain Ontology- based Information Retrieval Approach for Technique Preparation “International Conference on Solid State Devices and Materials Science 2012.

[10] IT .Kanimozhi, Dr.A.Christy,” Incorporating Ontology and SPARQL for Semantic Image Annotation” Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT 2013).

[11] Xiangsheng Yang, Ruoman Zhao, Chuan Zhang,” An Ontology-based Framework for the Construction of Teaching Resource Library”, Ningbo University.

[12] Swathi Rajasurya , Tamizhamudhu Muralidharan , Sandhiya Devi, Nwe Ni Aung, Thinn Thu Naing,” Sports Information Retrieval with Semantic Relationships of Ontology”, University of Computer Studies, Yangon , 3rd International Conference on Information and Financial Engineering IPEDR (2011).

[13] Swathi Rajasurya , Tamizhamudhu Muralidharan , Sandhiya Devi,Dr.S.Swamynathan,” Semantic Information Retrieval Using Ontology In University Domain”, Department of Information and Technology,College of Engineering,Guindy, Anna University,Chennai.

[14] Christian Paz-Trillo , Renata Wassermann,” An Information Retrieval application using Ontologies”, Departament of Computer Science Institute of Mathematics and Statistics University of Paulo, Brazil, 3, 2005.

[15] Thinn Mya Mya Swe,” Intelligent Information Retrieval Within Digital Library Using Domain Ontology”, Computer University, Mandalay, Myanmar. International Conference on Applied Computer Science.

[16] Tang Lijun,Chen Xu,”The Study of Semantic Retrieval Based on the Ontology of Teaching Management”,Advanced in Control Engineering and Information Science CEIS 2011.

978-1-4799-3759-2/14/$31.00©2014 IEEE