International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 05 | May -2017 www.irjet.net p-ISSN: 2395-0072

Ontology Based Information Retrieval System Using Multiple Queries For Academic Library.

Sneha Kasbe[1], Priyanka Shahane[2], Rohini Kasar[3], M.P.Navale[4]

Student, Computer Department, NBN Sinhgad School of Engineering, Pune[1]
Professor, Computer Department, NBN Sinhgad School of Engineering, Pune[2]

Abstract - *Ontology is a concept which describes the list of terms that represents important concepts such as classes of objects and relationship among them. The artificial intelligence application is created by semantic web in order to make web content meaningful to computers.*

In this semantic web multiple queries entered by user are parsed by the Standford parser and then triplet extraction algorithm is applied on this parse query. Then JENA Framework is used in order to retrieve the relevant information from RDF triplet in knowledge base.

***Key Words*: Ontology, Semantic web, Jena API, SPARQL, Query Processing, Word Net, RDF.**

1. INTRODUCTION

Ontology is a concept which describes the list of terms that represents important concepts such as classes of objects and relationship among them. The artificial intelligence application is created by semantic web in order to make web content meaningful to computers.

In this semantic web multiple queries entered by user are parsed by the Standford parser and then triplet extraction algorithm is applied on this parse query. Then JENA Framework is used in order to retrieve the relevant information from RDF triplet in knowledge base.

2. METHODOLOGY

2.1 SPARQL

SPARQL is a query language for RDF W3C.RDF is directed labelled graph data format. RDF represents information in the web. The SPARQL can express the queries across various data source which is native or viewed as RDF via middleware.

SPARQL is data oriented which queries the information held in the models.

2.2 Jena API

It maps the SPARQL query on RDF. The Jena API allows adding, deleting, changing and publishing information. RDF API provides facility to find triplet that match with specified pattern.

Jena API provides SPARQL API to handle both SPARQL query and their update.

Jena is a programming toolkit, using the java programming language.

2.3 Semantic Web

It is extension of web through Standford by World Wide Web Consortium (W3C). This technology enable users to create data store on web, build vocabularies and write rules for handling data. That enables people to share content beyond the boundaries of applications ad websites. It is based on machine readable information and builds o XML technology's capability to define customising schemas and RDF's flexible approach to representing data. The semantic web provides common formats for the interchange of data. It also provide the 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.

2.4 Ontology

It represents definition of any type, properties and inters relationships of the entities that fundamentally exist for particular domain of discourse.

The most typical kind of ontology for the web possesses taxonomy and set of inference rule .The taxonomy represents classes of objects and relations between them. Classes, subclasses and relations among entities is considered as very powerful tool for web use. Ontology can express the rules on classes and their inter relationships among them so that machine can reduced some conclusion.

2.5 Word Net

Word Net is lexical database for English language. Word net can help question answering system to identify synonyms. For example, verbs "start", "begin" will be recognize as synonyms by Word Net. The synonym information can be used to help to match a question with an appropriate rule.

2.6 Resource Description Framework (RDF)

Resource Description Framework (RDF) family of World Wide Web Consortium (W3C) specification originally designed as a metadata data model. It is directed, labelled graph data format and its general purpose is to represent the information in web.

3. SYSTEM ARCHITECTURE

In this proposed system, when user enters a input query in natural language, the user interface gets created. User can enter input query as a formal representation or SQL statement. Then parse tree which is a tree bank structure is created by Standford parser by processing the input query entered by user.
Then ontology concept is used in order to construct the triplets. The SPARQL query formed after this process fired on the knowledge base to find appropriate RDF triplets from knowledge base. Then it retrieves the relevant information using Jena semantic framework.

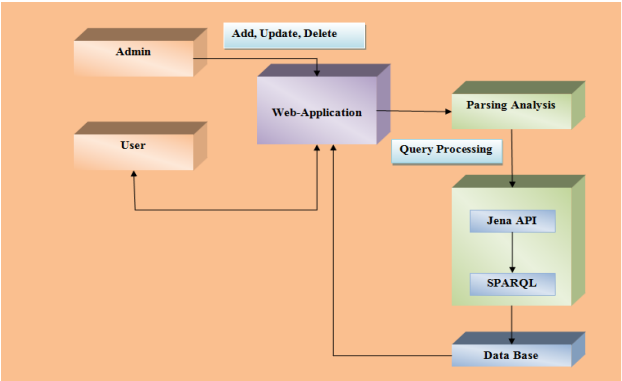


Fig -1: System Architecture

3.1 Parsing And Analysis:

At this phase we find out the analytical operation of given question. It leads to natural language processing. Standford parser processes the query .Then parse tree which is a tree bank structure is constructed for all input queries. This technique can be used to identify subject, verb, noun, phrases and adjective from the question. It takes out the token from the question and analyses the meaning of the question and reformulation of query is sent to the next page.

3.2 Classification of query:

Here the query is reformulated. Then with the help of Word Net or domain specific local dictionary is used for further expansion of the query.

3.3 Knowledge Base:

In order to retrieve the relevant answer from knowledge base the efficient storage of ontology is necessary. We used RDF database which can be easily linked in top Braid.

3.4 User Interface:

It is used to search relevant answers from ontology.

4. OVERVIEW OF IMPLEMENTATON

4.1 Database Used

Database Used Protégé is editor which creates data into RDF format. In our proposed system, Protégé is used to create ontology for Academic Library in RDF data format.

Academic Library Ontology contains several classes and subclasses. The root node, Academic Library contains classes Subject, Department, Administrator, Book_Title_Name etc. Similarly the class department has subclasses Computer Science, Mechanical, Civil etc. Those subclasses are also further divided into classes. Individual instances for those classes and subclasses are

Created for instance; the class subject has instances operating system, compiler, java etc. Properties for those classes and subclasses are created which used as Predicated in RDF for example; Book_Title_Name, book codes, rack number etc. are properties of operating system subject book. In this way hierarchical view of Academic Library Ontology is developed then exported into RDF format which are written in XML which shows RDF schema and Knowledge base KB for that schema. KB takes values of properties identified for classes and subclasses.

4.2 Triplet Extraction Algorithm

The triplet extraction algorithm is a approach for extracting subject-predicate-object triplet0 from English sentences.
A sentence (S) is represented by the parser as a tree having three children: a noun phrase (NP) , a verbal phrase (VP) and full stop (.).

Algorithm comprises of three steps:

1. In order to find subject of a sentence we have to search noun phrase (NP) sub-tree.
2. In order to find predicate of a sentence we have to search verbal phrase (VP) sub-tree.
3. In order to find object of a sentence we have to search three different sub-trees. These are prepositional phrase (PP) , noun phrase (NP) , adjective phrase (ADJP)

Standford parser generates a Treebank parse tree for the input sentence.

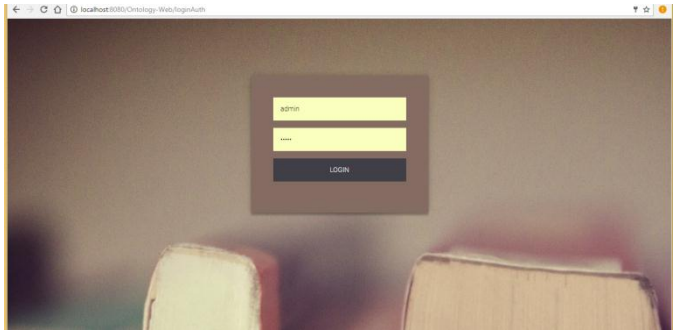
4.2 Pseudo code

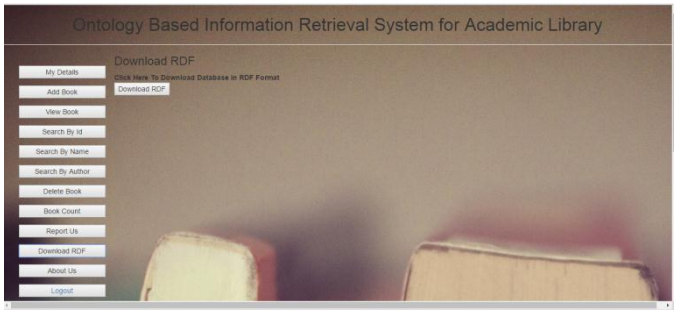
Step 1: Enter the input query in natural language.
Ste 2: Standford parser is processes all the input query.
Step 3: SPARQL query is formed and is fired in knowledge bases that find appropriate RDF triplet in knowledge base

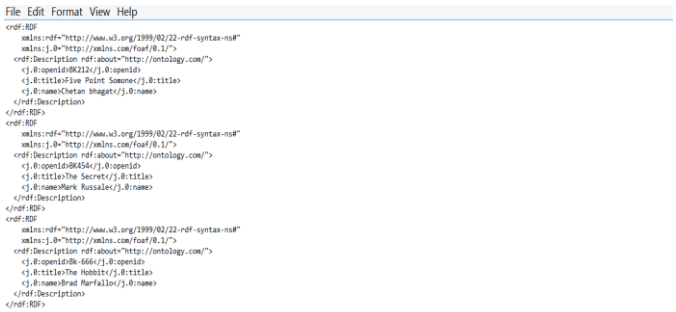
Step 4: Jena API mapped the SPARQ query on RDF.
Step 5: Apply Triple Extraction Algorithm.
Step 6: If match is found then display book
Else
Display book not found message
Step 7: End.

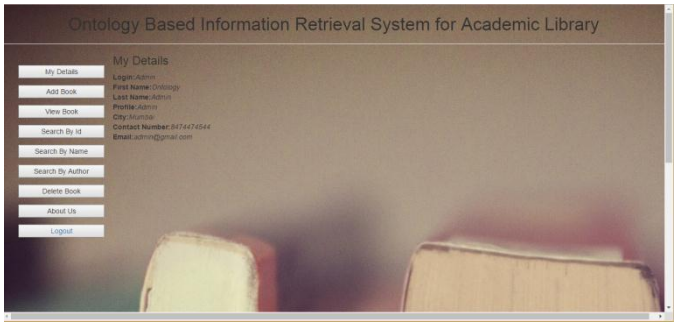
5. RESULT SET

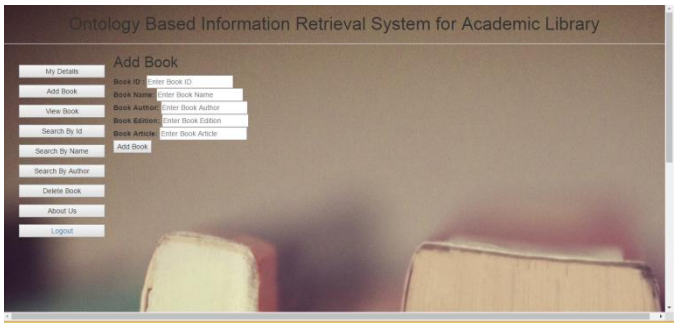
Sample output for add, search, update, and delete books from knowledge base is as follows:

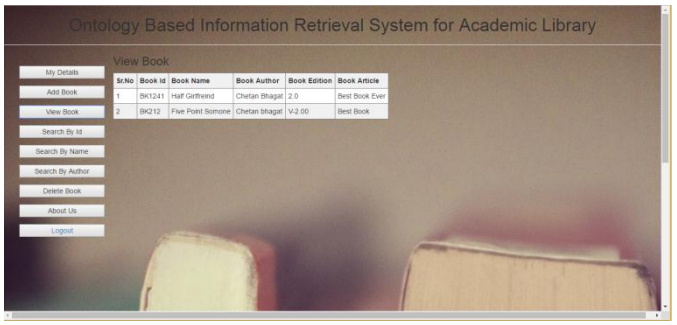


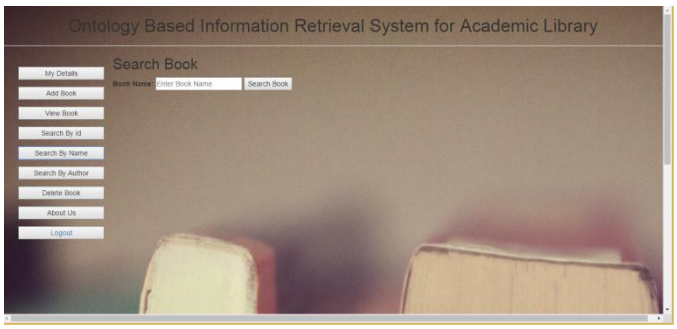












6. CONCLUSION

This system overcomes the limitation of traditional keyword based system used for query handling. The new system proposed by us is capable of extracting the relevant information instead of large list of answers. Here we have used triplet extraction algorithm which extract triplets from sentence. Then Jena API is used to map SPARQL query with RDF database in order to retrieve only the relevant information. This system works well for an

academic library system where we can retrieve book using multiple queries such as book name, author, title and edition of book.

References

[1] XiangbinXu, Gonzalez Moctezuma Luis, Andrei Lobov, Jose L. Martinez Lastra. "Multiple ontology Workspace Management and Performance Assessment" International Conference on Industrial information(INDIA)22015 IEEE 13th.

[2] Yinghui Huang, Guanyu L, Oiang piang Li". Rough Ontology Based Semantic Information Retrieval", international Symposium on Computational intelligence and Design(ISCID),2013 16th.

[3] T.Kanimozhi,A. Christry," Incorporating ontology and SPARQL for semantic image annotation"2013 IEEE Conference on Information and Communication Technologies(ICT).

[4] Domai 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] Rashmi Chauhan, Ryan Goudar, Robin Shama, Atul Chauhan, "Domain ontology based semantic search for information retrieval through automati query expansion ", Dept. of Computer Science and Engineering GEU Deharadu, India,2013 International Conference.

[7] Rachid Ahmed-Ouamer, Arezki Hammache, "Ontology based information retrieval for e-learning of computer science", IEEE conference, 2010.

[8] B. Chandrashekran, John R. Josephson, "What are ontologies and why do we need them? ", IEEE Intelligent System, [J], 1999.PP20-25

[9] Feng Luo and L. Khan, "Ontology construction for information selection", Technical Report, Computer Science Department, University of Texas at Dallas. 2002

[10] Astrova, N. Korda, and A. Kalja, "Storing OWL ontologies in SQL relational databases", international journal ofelectrical, computer, and systems engineering volume 1 number 4 2007 issn 13075179