INFORMATION RETRIEVAL FROM ONTOLOGY USING NATURAL LANGUAGE INTERFACE

[1]Sudarshan Awale, [2]S.J.Karale
Dept. of Computer Technology, Yeshwantrao Chavan College of Engg.
Nagpur (Maharashtra), India
Email: [1]sudarshan_awale@yahoo.com, [2]sjkarale@gmail.com

***Abstract*— Information retrieval from triple based ontological database play important role for many organizations. General search engines use keyword based search mechanisms. They retrieve huges amount of data from which sometimes it becomes difficult for users to identify relevant information. Semantic approach can be used to retrieve relevant information from ontological database. Here we have proposed and implemented an information retrieval system based on NL triplet extraction algorithm.**

Index terms — RDF, SPARQL, Ontology, OWL

I. INTRODUCTION

Domain specific question answering systems play significan role in many organizations. General search engines are based on keyword searching mechanisms. It retrieves huge amount of data from which sometimes it become difficult for the users to recognize relevant information. Semantic approach searches information by understanding the intent of user and meaning of words in user query. It uses semantics to produce highly relevant results.

This technique can be used to retrieve information for knowledge bases like ontology. Ontology is a technology used to enable the domain knowledge at a high level and improve the information retrieval time used in question answering system.

User can use ontological concept to search conceptual and semantic information. Ontology plays a significant role to access information, exchange of information, use and reuse of knowledge, sharing of information. Ontology can describe things and their properties and interrelations in a way that computers can process and automate. Ontology provides a knowledge-sharing infrastructure that supports the representation and sharing of domain knowledge. An increasing number of ontologies are being developed, and their reuse and sharing offer several benefits. One important advantaget is that we can subatantially save time and effort by reusing existing ontologies instead of building new ones. Another benefit is that heterogeneous systems and resources can interoperate seamlessly by sharing a common ontology.

II.SEMANTIC APPROACH

Semantic web is an extension of World Wide Web. It allows user to find, share and combine information more easily and efficiently. Ontologies form an important component of

semantic web which is used to improve understanding the intent of user and meaning of words.

To retrieve information related to semantics, ontologies are one of the main approaches used for knowledge management. Ontologies are defined as conceptualization which contain set of concepts, their interrelations and rules that govern these concepts to be interpreted by machines. Most ontologies describe individuals (instances), classes (concepts), attributes, and relations.

For creating ontologies Web Ontology Language (OWL) is used. OWL is based on W3C standards and help in defining ontologies which contain information representation features. OWL us built on XML and allows users to provide machine readable semantic annotations for specific communities of interest. OWL is used to describe classes, properties and individuals.

III. TRIPLES BASED MODEL

To translate NL query to intermediate triple-based representation linguistic components are used. Linguistic components consist of *English tokenizer*, *sentence splitter*, *POS tagger* and *VP chunker*. The annotations returned after the sequential execution of these resources include information about sentences, tokens, nouns and verbs. These annotations are used to query ontology. It is preprocessing step which help in precise classification of query. It is needed to understand particular NL query and also guide NL query in creating equivalent triple based representation.

Tokenizer is used to break a stream of text up into words, phrases, symbols or other meaningful elements called tokens These tokens becomes input for further processing such as POS tagging, parsing etc. Automatically assigning descriptors to given tokens is called tagging. Tag may indicate one of the parts of speech, semantic information and so on. The process of assigning one of the parts of speech to the given word is called Parts Of Speech tagging. Parts of speech include nouns, verbs, adverbs, adjectives, pronouns, conjunction and their sub-categories. Here off the shelf Stanford NL parser can be used to get triples from a natural language query.

IV. PROPOSED WORK

When user enters question in natural language, that question is first processed to get query triples. Triples are in the form of {Subject, Object & Predicates). Linguistic components are used to classify query in triples. The result we get is "Query Triples". There are two main reasons for adopting a triple-based data model. First of all, although not all possible queries can be represented in the binary relational model, in practice these exceptions occur less frequently. Secondly, RDF-based knowledge representation (KR) formalisms for the semantic web, such as OWL also subscribe to this binary relational model and express statements as <subject, predicate, object>. Hence, it makes sense for a query system targeted at the semantic web to adopt a triple-based model that shares the same format as many millions of other triples on the Semantic Web. Here a triplet extraction algorithm is used.

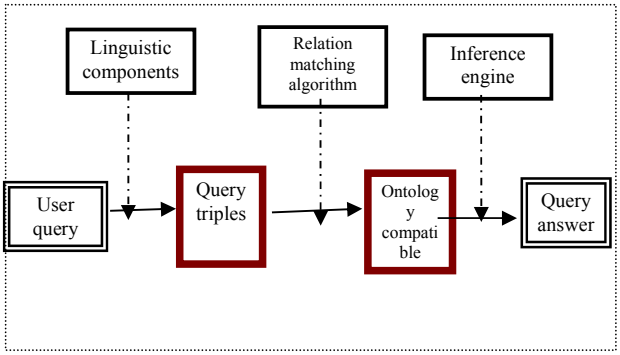

Fig 1: Natural language query processing

After getting Query Triples next step is to map these triples to onto compatible triples using relation matching algorithm. The purpose of this algorithm is to match these query triples against an existing knowledge base, consisting of semantically described words or phrases.

While trying to classify parts of the user input, a comparison between user input and knowledge base must eventually be made. Because of user input being an infinite set, either a reduction of input words or expansion of the knowledge base must be made. Popular approaches that solve this problem include stemming, lemmatization and various distance functions. Stemming is based upon a set of

rules, which determine word morphing, and is therefore limited to weakly inflected languages, where such rule collections exist. Lemmatization is used in conjunction with large language specific dictionaries, which are used to expand the knowledge base dictionary. This knowledge is then used to derive morphed words into their lemma. For measuring the results we have to create a test set, which would allow us to compare sequences against each other and would at the same time contain the information about the closest match.

Fig 1 shows user's NL query gets convetyed into "Query Triples" using linguistic components. Then these triples are mapped to ontology compatible triples using relation matching algorithm to get desired efficient and relevant answer to user query.

V.IMPLEMENTATION

Implementation of this propoed system includes various steps that have been shown in Fig 2. *Normalization, interpretation, query strings and finally the answer finding mechanism.*

Normalization process gives the "would be" condition where the answer can be specific to the question. For example as shown in the example, the question is "who is the hod of computer department?" and the normalization would be like "who be the hod of computer department". Here it is normalization that took us to next step in order to find the answer.

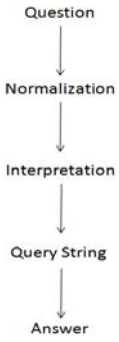

Fig 2: Implementation flow

Interpretation is process of finding objectives of the question, for example Property, target and context. It gives where to focus in order to find the answer. For example the question is "who is the hod of computer department?" and the interpretation will be "Property: Name, Target: hod and Context: computer department". Target shall give the actual attribute which have to be looked up.

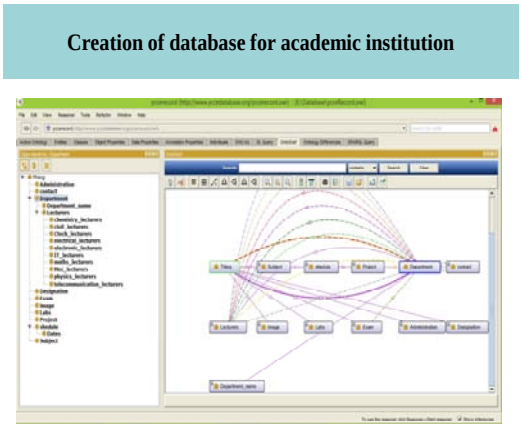

Fig 3: Current screenshot for Question Answering System giving relevant answer

Fig 3 shows current screenshot of the system.

VI.CONCLUSION

Currently our question answering system is under construction. It works for small domain specific ontology only. We are trying to create large ontology which will be capable of answering all possible question for particular domain specific ontology.

VII.REFERENCES

[1] Borut Gorenjak, Marko Ferme, and Milan Ojsteršek, "A Question Answering System on Domain Specific Knowledge with Semantic Web support", *International journal of computers*, vol. 5, no. 2, 2011.

[2] I. Čeh, and M. Ojsteršek, "Developing Question Answering System for the Slovene Language", *WSEAS transactions on information science and applications*, vol. 6, no. 9, Sep. 2009.

[3] N. Shadbolt, W. Hall, and T. Berners-Lee, "The Semantic Web Revisited", *IEEE Intelligent Systems*, vol. 21, no. 3, 2006.

[4] G.Madhu, A.Govardhan, and T.V.Rajinikanth, "Intelligent Semantic Web Search Engines: A Brief Survey",

International journal of Web & Semantic Technology (IJWesT), vol.2, no.1, Jan. 2011.

[5] Tim Berners-Lee, James Hendler, and Ora Lassila, "The Semantic Web," *Scientific American,* May 2001.

[6] Ramprakash et al., "Role of Search Engines in Intelligent Information Retrieval on Web," in *Proc. 2nd National Conference: INDIACom-2008.*

[7] F. F. Ramos, H. Unger, and V. Larios (Eds.), "Semantic search engines", *LNCS 3061 Springer-Verlag Berlin Heidelberg* , pp. 145–157, 2004.

[8] Sara Cohen et al.,"XSEarch: A Semantic Search Engine for XML,"in *Proc. International conference on very large databases*, 2003, pp. 45-56.

[9] D. Bhagwat and N. Polyzotis, "Searching a file system using inferred semantic links,"in *Proc. HYPERTEXT '05* ,Salzburg, 2005, pp. 85-87.

[10] H. L. Wang et al.,"Semantic search on Internet tabular information extraction for answering queries,"in *Proc. CIKM '00,* McLean, 2000, pp. 243-249.

[11] 11. E. Kandogan et al., "Avatar semantic search: a database approach to information retrieval,"in *Proc. SIGMOD '06,* Chicago, 2006, pp. 790-792.

[12] Gauri Rao et al., " Natural language query processing using semantic grammar", *International journal on computer science and engineering*, vol. 2, no. 2, 2010, pp. 219-223

[13] 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.

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

[15] 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.

[16] 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).

[17] 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.

[18] 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.

[19] 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.

[20] *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).

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

[22] 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).

[23] 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.

[24] 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.

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

[26] *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.

[27] 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.

[28] Ayesha Ameen, Khaleel Ur Rahman Khan, B. Padmaja Rani*,"Construction of University Ontology"*, IEEE conference 2012.

[29] 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.

[30] Zeng Dan, *"Research on Semantic Information Retrieval Based on Ontology"*, Library of Wuhan University of Technology, Wuhan, P.R. China, 430070.

[31] LIU Xiaoming, XU Jinzhong,LI Fangfang, *"Domain-specific Ontology Construction from Hierarchy Web Documents"*, School of Computer Science and Technology, Beijing Institute of Technology, China.