

A SURVEY OF DIFFERENT SEMANTIC AND ONTOLOGY BASED QUESTION ANSWERING SYSTEM

¹SHIKHA DONGRE, ²SWATI SINGH LODHI

^{1,2}Computer Science and Engineering, Sanghvi Innovative Academy, Indore, India
E-mail: shikha.dongre30@gmail.com¹, swati.singh0710@gmail.com²

Abstract- The question answering system is a system, which answers the user's question in the relevant way according to user's requirement. It is based on semantic web provides the meaningful search on the web. This is a challenging task to provide the accurate answer to the user. When a user asks any question than user find many of irrelevant and unuseful answers. From these answers user get irritate in searching require answer. In this paper we will study different question answer systems, which use semantic web and ontology. We will get a extraction from those papers to solve the problems, which comes in any question answering system, while user search the answer for his/her question.

Keywords- Knowledge Base, Ontology, Question Answering (QA), Semantic Web.

I. INTRODUCTION

Now a day internet is using by every person. Users prefer the internet to search the answer of their questions meaningfully. During searching user require meaningful search. This meaningful search is known as "Semantic Web". The term Semantic Web is first used by Tim Berners Lee for web data. Semantic Web provides a framework which allows sharing and reusing data across enterprise. On Internet, Semantic web plays an important role in searching the data.

The aim of Semantic web is to converting the structured and unstructured data into "Web of data". Structured data refers to the data, which is proper arranged and stored in the file like Database. Database is a structured data. Unstructured data is data that do not organise in any pre-defined manner or in file like dates or numbers.

There are too many search engines, which are used by the users to get meaningful answer. Swoogle, SenseBot, Sindice, START are some search engines and Question Answering System that, uses semantic web. Researchers are doing too much research on semantic question answering system, and to provide an exact and suitable answer to the user according to the user's requirement in a perfect way. The biggest problem of question answering system is that, the user cannot find the accurate answer. Because old system provide answer to the user, but they provide enormous data to the user, and user have to find the answer. Even user is not sure about the answer, because they do not guarantee that the answers, those systems are providing are correct. User search on the web, because they do not know the answer, but those traditional systems provide bulk of links and data, and let that upon the user to find their answer. Another problem occurs in question answering system is that, they use keyword based search. Keyword based search means, the search based on the

exact word matching. If the exact word not found, then no answer will be show.

As we have studied these problems in the question answering system, we found the solution for those problems. The solution we found is that, we should make a system which will able to search the answer according to users answer. Instead of providing a list of data we should provide exact answer to the user. The other solution for the problem is we can use content based search spite of using keyword based search.

In this paper first, we will study about the different question answering systems and the problems occur in those systems. This paper is mainly literature survey paper, so the main focus of our paper is on the old question answering systems and their result. This paper has seven sections: II, III and IV section provide the terminologies and concept we will be going to use in our system implementation. V section contain the literature review, in which, we will introduce different question answering system and the problems in those system. After analysing different question answering system, what should be the proposed system will be define in section VI. In section VII we will describe conclusion.

II. QUESTION ANSWERING SYSTEM

When we talk about the term semantic search, one more term occurs, which is Question Answering system. We require QA systems because; users prefer web search engines to search the answer according his questions. Question Answering (QA) system is a system, which takes user's question in natural language (like English), process the question, search for the question in database, Find the appropriate result and provide the answer to the user according to his/her question. The QA systems provide a semantic search on the web, which helps the user to get exact result instead of the ambiguous and irrelevant data.

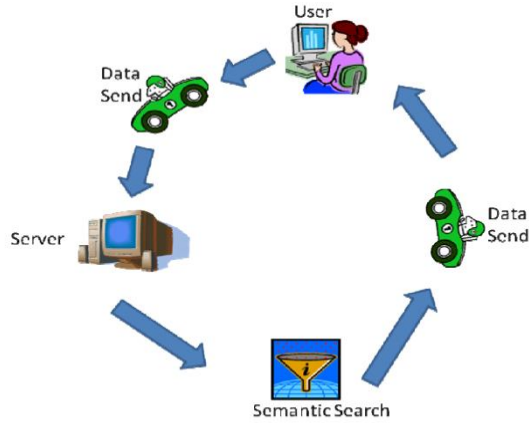


Fig 1: Question Answering System uses Semantic Search

The question answering system, search the answer from the database using its knowledge, which is known as knowledge base. Knowledge base QA system helps to give situation specific knowledge from multiple sets of information. Main aim of this paper is to introduce and survey what is semantic question answering system and what researches have been done on this. Analyze all semantic question answering and find the appropriate solution to the problems, which are found in the old semantic question answering systems.

III. RDF

Another terminology, which is used in Semantic search engines, is Resource Description Framework (RDF). Resource Description Framework (RDF) is a directed graph, which represent the information on the web. It defines the term metadata, which is “Data about the Data”. The RDF is similar to class diagrams. It uses RDF triples as subject-predicates-objects. Means, it divides the question in subject, predicates and object form to search the answer from the database.

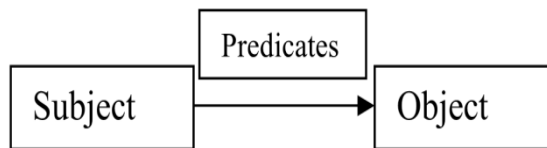


Fig 2: RDF Triple

When user asks the question in natural language, that question is converted into the SPARQL. SPARQL generates query in triple pattern with the help of RDF triples. Question-answering system uses OWL. Web Ontology Language (OWL) is a family of knowledge representation languages or ontology languages for authoring ontologies or knowledge bases. The languages are characterized by formal semantics and RDF/XML-based serializations for the Semantic Web. OWL is endorsed by the World Wide Web Consortium (W3C) and has attracted academic, medical and commercial interest.

IV. ONTOLOGY

Ontology is very important factor in Semantic Web to search more appropriate answer. It makes communication easy with the database. It uses the Knowledge Base (KB) for communicating with the system. Ontology formally defines a conceptual representation of concepts and their relationships within a specific domain. Ontology play an essential role in the semantic web by enabling knowledge sharing and exchange, ontological knowledge structures play an important role in the utility of background knowledge for question answering and information retrieval. QASYO. Knowledge base in Ontology provides a convenient way to obtain knowledge from Database. Ontology uses classes, attributes, and relationships. These three are the basic components of any ontology.

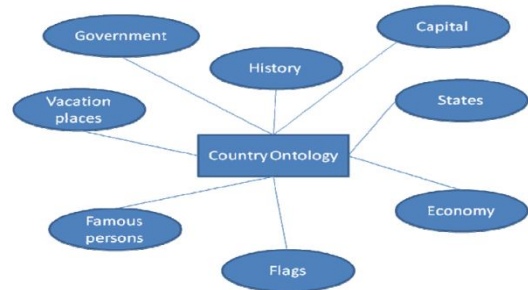


Fig 3: Ontology for Country

V. LITERATURE REVIEW

A. AQUA: Ontology Based Question Answering System (Maria Vargas-Vera and Enrico Motta 2003); This paper was an experimental question answering system that combines ontology, NLP and retrieving of information and logics in the uniform framework.

This was basically designed to answer the questions related to academic people from its own database. AQUA used RDF for framework for expressing metadata. AQUA reduces the chance of failure of any question related to academic peoples and authors by using Ontology.

The architecture for AQUA is shown in Figure 4. The disadvantage of AQUA was that they uses nly Academic domain. They do not able to answer other domain questions.

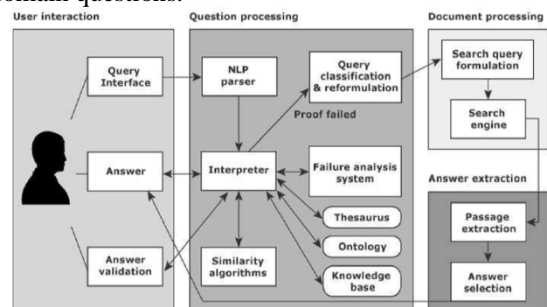


Fig 4: AQUA Architecture

B. Querix: Natural Language Interface to Query Ontologies

Based on Clarification Dialogs (Esther Kaufmann, Abraham Bernstein, and Renato Zumstein 2006); This system was designed to solve the problem of ambiguity which occur in Natural Language. Querix is a domain independent Natural language interface which is used in semantic web. Querix is simple and portable.

The Querix system is consists of 7 components which are: User Interface (Generate SPARQL Query), Ontology Manager (Find Synonyms from WordNet), A Query Analyzer (uses Noun, Verb, Prepositions, Question Word, Compositions), A Matching Center, A Query Generator, Dialog Component (used for Ambiguity), Ontology Access Layer.

Matching Center is very important part of the Querix System. It tries to find the match with query.

Querix approach is depends on the quality of ontology with high knowledge of English grammar. It is able to translate 78.6% queries.

C. SMART: A web-based, ontology-driven, semantic web query answering application (Battista, A.D.L., N. Villanueva-Rosales, M. Palenychka and M. Dumontier, 2007); SMART stands for Semantic web information Management with Automated Reasoning Tool. The aim of SMART is to provide tool to life Scientist for represent, integrate, manage and query heterogeneous and distributed biological knowledge. SMART uses Ajax, SVG, JSF, RDF, OWL and SPARQL. For implementing its system SMART uses Java Programming Language and operates with open source and semantic web including OWL, API, and Protégé.

System Architecture consist of: Application Mediator, Web based user interface, Ontology repository, ontology index, Hybrid reasoned. Future work for SMART was to include exploratory content browsing, interactive graphical query formulation knowledge composition and truth maintenance.

D. PowerAQUA: A Multi-Ontology Based Question Answering System – v1 (Vanessa Lopez, Enrico Motta, Marta Sabou, Miriam Fernandez 2007); This paper presents the PowerAQUA, which was a multi-ontology-based Question Answering (QA) system that takes input queries in natural language and able to return answers from suitable distributed resources on the Semantic Web.

As compared to other natural language, PowerAqua is not restricted to a single ontology. PowerAQUA uses 3 components which are Linguistic Component: This analyzes the natural language query and translates it into linguistic triples.

Power Map: This identifies the type of ontology form the database.

Triple Similarity Services: Analyzes the ontologies and generate the appropriate result.

The first PowerAqua prototype was available on <http://kmi.open.ac.uk/technologies/aqualog/okdeliverable>.

The ranges of queries are limited in PowerAQUA. E. QASYO: A Question Answering System for YAGO Ontology (Abdullah M. Moussa and Rehab F. Abdel-Kader 2011); the old system uses the keyword based search but the question answering system uses natural language queries. In ontology based Question Answering system the Knowledge base is used, which provides a suitable way to obtain knowledge for use.

QASYO joins different techniques in a simple way to make sense of NL queries and map them in semantic markup. QASYO uses YAGO Ontology. YAGO was developed by using the oracle database and visual C#.

QASYO system consists of 4 phases: Question Classifier, Linguistic component, Query generator, Query processor.

The architecture for QASYO is shown in Figure 5. Linguistic component generates a parse tree which divides in grammatical form of English like question words, noun, object, verb.

QASYO simplifies natural languages questions into simple query which term as Query Simplification Technique. QASYO system uses the WordNet 2.1 to implement their system. QASYO was able to answer 91% of question from which 40% was based on Knowledge.

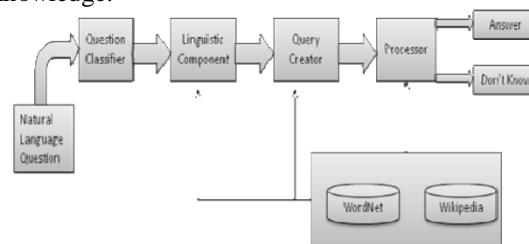


Fig 5: QASYO System Architecture

F. FREyA: an Interactive Way of Querying Linked Data Using Natural Language (Danica Damljanovic, Milan Agatonovic, and Hamish Cunningham): It uses syntactic parsing with the ontology-based lookup to interpret the question, and involves the user according to requirement. FREyA is an interactive Natural Language Interface to querying ontologies, which joins advantages enhancement methods like feedback and clarification dialogs attempt to improve recall and improve precision by resolving ambiguities.

FREyA starts its working with a Natural Language (NL) question, and end when the answer searched

from database. This results generation of syntactic parsing and the analysis generates a parse tree, which uses Stanford Parser, after that it uses several rules to identify Potential Ontology Concepts (POCs). POCs are a question terms/phrases, which not compulsory have to be linked with Ontology Concepts (OCs). POCs are selected on based on the analysis of syntactic parse tree. This analysis does not require strict way to syntax and works on ill-formed questions and question fragments and on the grammatically correct ones. For example, nouns, verbs, or WH phrases(Where, Who, When, How many) are expected to be found by POC identification algorithm. FREyA algorithm was based on identification of prepreterminals and preterminals in the parsed tree.

FREyA can be used with different ontologies, and querying Linked open data (LOD). FREyA tested DBpedia and MusicBrainz data sets.

G. QAAL: Comparison of Question Answering Systems Based on Ontology and Semantic Web in Different Environment (Kalaivani and K. Duraiswamy 2012): The found the problem in old Question Answering (QA) system is that for finding the answer of user's question systems take several steps that were followed for conversion of questions to query form to get an exact answer. QAAL system surveys different types of question answering system, which were based on ontology and semantic web and uses different query format. QAAL question answering system is an automatic learning system architecture, which is used to overcome the difficulties raised from the different QA models. The QAAL system is shown in figure 6. In this system they used semantic search methodology; which is implemented by RDF graph. The applications of data structure domain and the performance is analyzed. For retrieving the answer from the question this system uses ontology, Semantic Search approach and question-to-query algorithm. This process evaluates system for analyzing performance and evaluation of system.

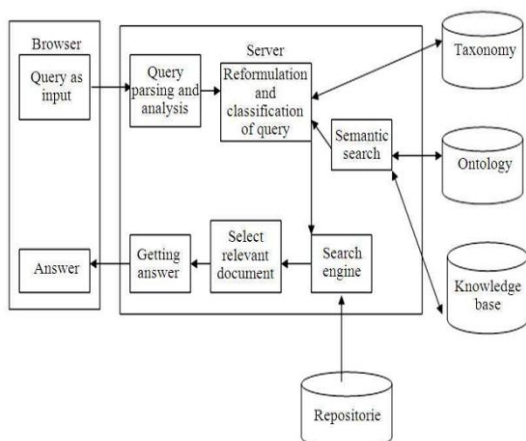


Fig 6: QAAL System

This system successfully implements the technique of semantic search and ontology to retrieve the exact answer and behave like an intelligent system.

H. Intelligent Question answering System (V.S.Babanne, Dr.S.T.Patil, D.J.Joshi 2013): The old systems were facing the problem of irrelevant huge amount of data, which was bombarded to user while he puts any question to get the answer. Intelligent Question Answering System minimizes the problem of these irrelevant output data. This paper was a rule-based system, which aims to delivering brief information that contains answers to user's questions. It was a domain specific system, which was able to answer the question related to java. This Question Answering system combination of two parts: the knowledge base and the reasoning/inference, engine. For searching result, this system uses IF THEN rule. The Intelligent Question Answering System is shown in figure 7. The execution of this question answering system for getting exact result can be improved by using Semantic search. The answer can be retrieved from knowledge base.

This system was able to give specific answer to user's question on the place of document retrieval.

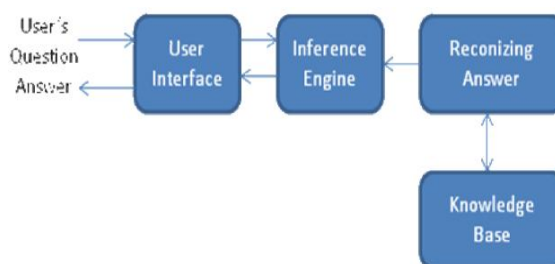


Fig 7: Intelligent Question Answering System

I. Intelligent Semantic Question Answering System (Erfan Najmi, Khayyam Hashmi, Fayez Khazalah and Zaki Malik 2013): Huge information is available on the Web and it requires new techniques to handle and organize that huge data. Ontology is focal methodology to represent manor-specific concept based knowledge and help to provide solution for Question Answering (QA) systems. QA systems can be facilitating by using semantic technologies to better serve knowledge to the users.

QA systems are using "search engine approach", in which user get multiple pages as the result of their output and user need to select the answer from this list according to his knowledge.

Intelligent Semantic Question Answering system introduces a paper, which is an approach for improving the strength of QA systems using semantic technologies. This system implement by taking the input in natural language and converting this natural language query into the RDF triples to find the relevant answer.

This system uses question classification techniques which are: Factoid classification, Keyword classification.

In factoid classification the classification is done as: people, organizations, dates, location, quantities, linear measures etc.

In keyword classification the classification is done by using natural language parser.

After the question classification the query is translated and then finds the answer.

This system requires less computation power as compared to other. This system was only able to answer one word answer but it was not able to answer list questions.

J. Ontology Based Meaningful Search Using Semantic Web and Natural Language Processing Techniques (K Palaniammal and S. Vijayalakshmi 2013): This system is ontology based meaningful search system, which uses semantic web and Natural Language Processing (NLP) techniques in the educational domain. The search model of this system is consists of three parts: embedding spell-check; search for synonyms using WordNet API; querying ontology using SPARQL.

This system is made to develop a reliable and efficient search engine to access the accurate results for the user according to his query. The aim of this system is to retrieve the same result for synonyms words. This prevents the visualization of irrelevant search results. This system provides the details related to education domain, correct URL and metadata to search, which takes more time in the syntactic search engine. In this details are generated by using ontologies and relationships between classes, entities, individuals. This system implemented a reliable and efficient system, which suggests the user about all the effective details to know about an educational domain.

This system provides very accurate results and details for the selecting field in a single page as compared to other systems.

VI. LITERATURE EXTRACTION

After studying all the above Question Answering system we are proposing a system which will be able to work in intelligent manner and able to answer the user's query according to his/her requirement. The objective of our system is, to make a system, which will able to answer the list question, effective and efficient answer by ignoring un-useful and irrelevant answer. The aim of our proposed system is to make a system which will overcome the problems comes in traditional question answering system.

The purpose of a Question Answering system is to find the correct and exact answers to user's questions in both non-structured and structured collection of data. Most of the existing information technology applications use only single source background knowledge mostly WordNet or Wikipedia. They can improve their performances, if they use huge ontology with knowledge from several sources was available.

CONCLUSION

In this paper we have discuss about semantic and ontology based question answering systems and the work, which have been done in this field. We studied that, Semantic web and ontology are very important part of question answering system because semantic web makes this system meaningful and ontology makes the system easy to search.

Different systems used many different ways for answering user's query, like some system uses natural language processing, some uses RDF triples and some of the system uses SPARQL.

ACKNOWLEDGEMENT

I would like to thank my project guide Miss Swati Singh Lodhi and Mr. Sumit Jain for helping me, in the creating this research paper.

REFERENCES

- [1]. Maria Vargas-Vera and Enrico Motta 2003: AQUA-Ontology Based Question Answering System.
- [2]. Esther Kaufmann, Abraham Bernstein, and Renato Zumstein Querix: A Natural Language Interface to Query Ontologies Based on Clarification Dialogs In proceedings of the 5th International Semantic Web Conference (ISWC 2006), Athens, GA, November 2006.
- [3]. Battista, A.D.L., N. Villanueva-Rosales, M. Palenychka and M. Dumontier, 2007. SMART: A web-based, ontology-driven, semantic web query answering application.
- [4]. Vanessa Lopez, Enrico Motta, Marta Sabou, Miriam Fernandez 2007 PowerAQUA: A Multi-Ontology Based Question Answering System – v1
- [5]. Abdullah M. Moussa and Rehab F. Abdel-Kader QASYO: A Question Answering System for YAGO Ontology International Journal of Database Theory and Application Vol. 4, No. 2, June, 2011.
- [6]. Danica Damljanovic, Milan Agatonovic, and Hamish Cunningham FREyA: an Interactive Way of Querying Linked Data Using Natural Language 2011
- [7]. S. Kalaivani and K. Duraiswamy 2012. QAAL: Comparison of Question Answering Systems Based on Ontology and Semantic Web in Different Environment. Journal of Computer Science 8 (9): 1407-1413, 2012 ISSN 1549-3636.
- [8]. V.S.Babanne, Dr.S.T.Patil, D.J.Joshi Intelligent Question answering System International Journal of Scientific & Engineering Research, Volume 4, Issue 5, May-2013.

- [9]. Erfan Najmi, Khayyam Hashmi, Fayez Khazalah and Zaki Malik Intelligent Semantic Question Answering System 2013 IEEE International Conference on Cybernetics.
- [10]. K Palaniammal and S. Vijayalakshmi: Ontology Based Meaningful Search Using Semantic Web and Natural Language Processing Techniques.
- [11]. Chong Wang, Miao Xiong, Qi Zhou and Yong Yu 2007: PANTO: A Portable Natural Language Interface to Ontologies.
- [12]. Nicolas Kuchmann-Beauger, Falk Brauer, Marie-Aude Aufaure: QUASL: A Framework for Question Answering and its Application to Business Intelligence, Proceedings of the 7th International Conference on Research Challenges in Information Science (2013) 143

