

Question Answer System

Shruti Mhatre, Kalyani Pawar

Abstract— The Question Answer System proposes an alternative approach to the process of finding information. Instead of using search engines which provides us with a huge amount of information which is difficult to handle, a system is developed which provides direct answer to the question asked by a user. Semantic Web technology and ontology database have been used in combination to develop an efficient system for answer retrieval. Ontology is used for storing the database while other techniques such as sentence splitting, indexing and keyword extraction are applied so that the system is able to provide the desired result efficiently by mapping extracted keyword to relevant data from the database ontology.

Index Terms— Search engine, Web pages, Semantic web, Ontology, Database, Information retrieval, Query .

1 INTRODUCTION

INFORMATION is the data obtained about a certain subject or event. Before carrying out any task it is always essential to gain information related to that particular area.

In order to search information related to any topic, nowadays it is usually preferred to make use of a search engine. A web search engine is a software system that is designed to search for information on the World Wide Web. The World Wide is a system of interlinked hyperlink documents that are accessed via the internet. With a web browser, one can view web pages and navigate between them via hyperlinks [1]. They search documents for specified keywords and return a list of documents where the keywords were found.

Although the required information is obtained with the help of a search engine, the amount of information thus gathered is huge. Moreover, the information is in the form of Uniform Resource Locators (URL) or documents which the user needs to go through manually so that the relevant knowledge is acquired. URL is the global address of documents and other resources on the World Wide Web (WWW) [1].

Also, the content is searched solely based on keyword extraction without actually understanding the user query. The process of information extraction needs to become more precise while reducing the time and effort required in the process. This can become possible with the use of Semantic Web technology.

The Semantic web is an evolution and extension of the existing web that allows computers to manipulate data and information. It is proposed to have web pages structured and tagged in such a way that it can be directly read by computers [1]. Ontology specifies the concepts, relationships and other distinctions which are relevant for modeling a domain. The specifications are in the form of classes and the relationships among them [1].

2 OBJECTIVE

2.1 Problem Definition

The goal of this project is to develop a system which provides user a direct answer to the input query rather than the user having to manually go through various documents in order to get answer for that particular query. The problem faced while

developing such system is to make sure that the correct answer is extracted from the database by making the system understand the user's query.

The problem can be tackled by applying various pre-processing techniques (such as splitting and tokenizing the query, determining the type of question and keyword extraction) on the input query and then mapping the extracted keywords to the database. The answer will be obtained by mapping keywords to the relevant data in the ontology.

2.2 Relevance

This project is developed to reduce the time consumed in the process of finding answer from huge content available on the web. With the help of this system, the time and effort of the user gets reduced with respect to finding the exact answer to the user query.

A user can get direct answers of his/her query in a single system by firing the required query instead of visiting various links, web pages, documents, etc. Usually a person looking for information about the Engineering colleges in Mumbai would search about each college or visit the website of every college. But using this system, user can get information about all the colleges by just asking a question.

2.3 Scope

The overall scope of the project is to deliver a system which provides a direct answer to the user query regarding the engineering colleges in Mumbai by extracting data from the database ontology. The user will be provided with a web page as graphical user interface to input his/her query. The answer obtained will be displayed on the same webpage.

The database will be stored as an ontology using the protégé tool. The data can be queried using SPARQL which will work at the back end. It will be important to make sure that the answer thus extracted will be correct and obtained in an efficient manner.

3 PREVIOUS WORKS

In [2], S. Garg and S. Kumar have proposed a system involving Semantic web and domain ontology which provides answer to the questions related to a particular domain. Domain ontology plays a vital role in this system. The manual execution result shows the feasibility of this system which can answer 3 types of questions surveyed from different students. The future scope of this system includes developing system based on the algorithm described in the paper and also creating various other ontologies so to include more type of questions in future. Performance of question answering system of getting precise result could be enhanced by using semantic search methodology for fetching answers from ontology model. The system can implement this technique and can be used in intelligent manner for automatic learning method.

In [7], P. Tanwar, Dr T. V. Prasad and Dr. Kamlesh Datta have used an algorithm which is effective in the process of information extraction. Knowledge representation techniques are implemented to form the knowledge base. These techniques form the fundamentals of the process as knowledge base is the core of the system.

Tokenization and tagging of English statement are very important phases of the system. The system is able to provide answer from a 1000-lines input provided in the form of a short story or write-up. The proposed reasoning algorithm that uses semantic net and script will be used for hybrid knowledge representations system and can be used in daily life activities as the system is capable of representing such knowledge. Combination of declarative and procedural techniques makes the system interactive and user friendly. This algorithm will be used for declarative as well as procedural knowledge. The reasoning algorithm can be utilized in many applications of AI and robotics. The proposed reasoning algorithm is used to infer the knowledge from the existing knowledge base. In its advanced stages of development, the proposed system can act as an intelligent system.

In [3], PouryaFarzi, RezaAkbari and Omid Bushehrian have developed a quality of service method for knowledge discovery. The goal of the proposed method is to introduce an extended quality of service framework for ranking semantic web service. In addition, a matchmaking algorithm is introduced based on measuring the similarity between a service query and the published services. For future work, it is planned to create a new context module for viewing the network situation and store the user information history for finding suitable services in the distributed system.

In [4], Shikha Dongre and Swati Singh Lodhi have conducted a survey of different semantic and ontology based question answering system. After studying all the above Question Answering system they have proposed a system which will be able to work in intelligent manner and able to answer the user's query according to his/her requirement. Most of the existing information technology applications use only single source background knowledge. They can improve their performances, if they use huge ontology with knowledge from several sources is available.

4 SYSTEM DESIGN

The system architecture for the working of Question Answer system is given in Fig 1. The major components of the architecture are explained below

- **User:** User is the input end of the system which will enter the query into the system. It will also be the end to receive the output of the query as an answer.
- **Web page:** The web page is a regular xml page which acts as a graphical user interface for the system. It can be made available on the internet and the system may be used by users online.
- **Database Ontology:** The relevant data is stored in the database in form of ontology. It becomes easier to map data stored using such ontology instead of regular databases.
- **Pre-processing:** The various techniques used for analyzing the query are included in the pre-processing phase.

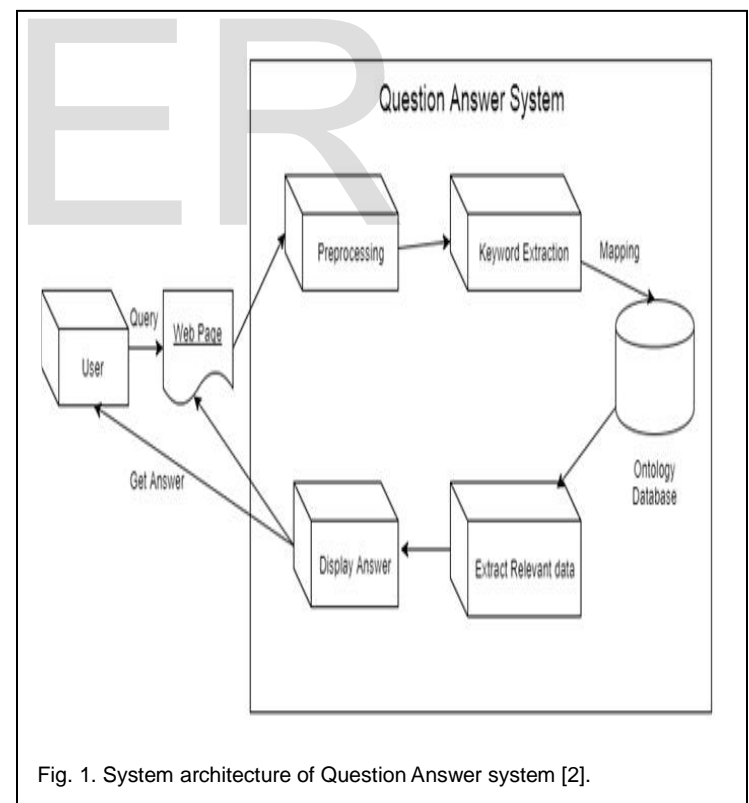


Fig. 1. System architecture of Question Answer system [2].

An information retrieval system consists of some basic components such as an interface to accept user input, the database which contains the information, back end operations

(which perform procedures such as pre-processing of data, mapping data to user query, data connectivity, etc) and output in form of answer display. Each of these components forms the fundamentals of this system. The working of the entire system is dependent on each and every component described above. Out of these components the input and output end are very simple units which can be any kind of user interface. The database and back end is the core of the system which will determine the successful working of the system.

It is therefore essential to select appropriate techniques for data storage, access, processing, etc. A study of work related to such Question Answer system is helpful in deciding upon the methods which are used for implementation of the system.

The data is stored in a system usually done with the help of SQL and the data is in the form of tables. Such method of data storage enables systematic arrangement of data elements which provides simplicity with respect to storage. But it is not enough to have just a simple database because real life knowledge base has a complex structure. Database created using SQL fails to provide the required complexity of data. It is necessary to store the data in ontology which can represent the complicated structure of the database [2].

Ontology is a hierarchical structure of the data elements and thus enables storing data in a tree like structure. Accessing data becomes easier with this technique since mapping the relevant data with respect to the query becomes more convenient. This can be achieved using the various tools available for creating ontology. They not only help for creating ontology but also for storage of data. Also, the query can be mapped to such a database after it is done with other pre-processing operations.

Ontology acts as a service provider in today's semantic web. OWL is the most commonly used language for creating ontology. While SPARQL is the query language used for providing queries to data stored in ontology. Protégé is the most popular and widely used tool for ontology creation and development. It helps us to build RDF/XML or OWL/XML code also from the ontology that is developed. It can also be used for querying the individuals that have been stored in ontology classes and to display this to users in a table format as in SQL [6].

Once user enters his/her query, it cannot be immediately used for answer extraction. It is necessary to apply certain pre-processing techniques in order to make it suitable for further operations. Several techniques are suggested such as indexing, tokenizing, sentence splitting and parsing [5]. It is up to the developer to select from among these techniques according to the needs of the project. It is also wise to classify the question into categories such as yes/no type, wh type, etc and then extract keywords [2].

The pre-processing phase provides with keywords which can be mapped to the database and hence answer gets extracted [2]. This can be achieved by applying reasoning algorithm to the input queries. Also, SPARQL is used for implementing database queries on the ontology class structures which is essential for carrying out extraction of data.

The methodology used in [8] is said to be most efficient according to a study carried out regarding different question

answering systems [4]. The method involving natural language processing techniques is seen to be working more efficiently as compared to other systems carrying out similar tasks.

One of the problems faced by such systems is that user can provide input using any of the synonyms of a single word. The system should be able to determine that all of those words mean one and the same thing. It is practically impossible to have dictionary based function to solve this issue. Words and their relationships to each other are organized in a hierarchical manner similar to the taxonomies which may be found in the natural sciences. Words which are closely related to each other may be found in the same branch of the hierarchy's tree. Each word belongs to a set of synonyms which belongs to the hierarchy [8].

4 CONCLUSION

The concepts of Semantic Web and the technologies related to it have been studied thoroughly. From this study, the design of the system is determined. The system architecture gives the basic working operations of the system. Implementation phase has covered the implementation of user front end. Also, the class structure of database ontology has been created in the protégé tool.

The project can be further extended by expanding the database according to the user requirements. For now, the database of that project is only for engineering colleges in Mumbai. It can be extended to include the colleges of an entire state or country.

Also, such kind of systems can be implemented to develop enquiry systems related to organizations, hospitals, institutions, etc. The amount of data stored in such systems should be increased by expanding the database ontology.

ACKNOWLEDGMENT

The authors our project guide Prof. Nikita Lemos , Head of Department Prof. Chhaya Narvekar , Principal Dr. Y. D. Venkatesh and Director Fr. Francis D'mello for their continuous support and guidance throughout the project work.

REFERENCES

- [1] Dhana, N. Semantic Web and Ontology. 1st edition. Retrieved from <http://bookboon.com/en/semantic-web-and-ontology-ebook>
- [2] Shally Garg and Suresh Kumar, "JOSN: JAVA Oriented Question-Answering System Combining Semantic Web and Natural Language Processing techniques", *1st India International Conference on Information Processing (IICIP)*, pp. 1-6, 2016.
- [3] Pourya Farzi, Reza Akbari and Omid Bushehrian, "Improving Semantic Web Service Discovery Method Based on QoS Ontology", *2nd Conference on Swarm Intelligence and Evolutionary Computation (CSIEC)*, pp.72-76, 2016.

- [4] Shikha Dongre and Swati Singh Lodhi, "A Survey of different Semantic and Ontology based Question Answering System", *International Journal of Advanced Computational Engineering and Networking*, pp.69-74, 2014.
- [5]
- [6] Richard J Cooper and Stephan M Ruger, "A Simple Question Answering System", 2000.
- [7] D Venkatraman and Haritha K C, "Knowledge Representation of University Examination System Ontology for Semantic Web", *International Conference on Advanced Computing and Communication Systems*, 2017.

IJSER