# Automatic Question Answering System Based On Ontology  Using NLP

**Reena M[1], Santosh Kumar Singh[2], Prof .Krishna Chandaran[3]**
[1,2] Dept of Master Of Computer Applications
[3]Assistant Professor, Dept of Master Of Computer Applications
[1,2,3] New Horizon College Of Engineering

**Abstract-** *automatic question answering system, such as inadequate knowledge expression and weakness of indicating the inherent relations among knowledge, a solution with using Ontology to build curriculum domain knowledge base was put forward in this paper. In this paper, the course of "Natural Language Processing" was took as an example to construct an automatic question answering system based on Ontology. The system extracts keywords by analyzing users question and transforms the intention of the question into the query of basic elements in Ontology. Finally, the system extracts the answer via Jena reasoning. Meanwhile, the system provides related knowledge to help students study systematically. This paper has achieved an automatic question answering system presented in the course of "Natural Language Processing". This automatic question answering system supports asking questions in natural language and has some reference value in other areas.*

*Keywords*- Automatic question-answering, domain Ontology, questions analysis, Jena reasoning

## I. INTRODUCTION

With automatic question answering system, users can raise questions in natural language. The system requests the knowledge base automatically then returns answers to users. The automatic question answering system can solve problems without the limitation of time and space. Furthermore, the system can improve the students' learning efficiency as well as reducing teachers' work load. Therefore, automatic question answering system became the hotspot of current research . The traditional knowledge base of automatic question answering system is Frequently Asked Questions (FAQ). When answering questions, the system calculates the similarity between users' questions and FAQ, then it lists one or more the highest similarity questions for the user to choose .
. This kind of automatic question answering system have high accuracy when answering existing questions in FAQ ,however, if there are no matched questions between FAQ and users' questions, the system cannot answer question. and users' questions, the system cannot answer question automatically. Moreover, because keyword matching is

regarded as priority in calculating questions' similarity, a lot of FAQ need to be prepared in advance.

At the same time, this kind of automatic question answering system can only simply answer the question without reflecting the connections between the knowledge. Considering Ontology has the good concept hierarchy and was applied widely in the knowledge expression, knowledge sharing and reuse, Ontology was used in this paper to implement the automatic question answering system.

After fusing Ontology into the automatic question-answering, the system not only can analyze the users' questions in the semantic level but also combine with the users' questions for semantic reasoning. Therefore, the users can get a better understanding and a more accurate result.

Seven steps which developed by Stanford university hospital are mainly used to the construction of domain ontology [7].

This paper adopts seven steps to construct Ontology for domain knowledge. Ontology can offer a professional knowledge base for the automatic question answering system. The system extract key words through word segmentation and part-of-speech tagging users question. Then, the system reason out the answer and related knowledge from Ontology.

The system consists of four mainly parts: the knowledge base of Ontology, question analysis module, answer extraction module, and the standard answers' extension module.

NLP provides an automatic annotation of WebPages, community mining, extracting hyponyms, question answering, paragraph extraction and key word extraction for inter entity relation representation.

The main goal of NLP based mechanism used to receives user query and provide an accurate answer for the posted query with the help of machine learning algorithms.

The structure of the paper is as follows:

## II. QUESTION ANSWERING PROCESS

Question Answering (QA) system is a process of Information Retrieval (IR). The automatic QA system is the ability to provide the answer with most relevant technical information in human language in which questions are in either simple or complex form. The QA system [1] is targeted to furnish the correct answers for natural human language queries. Most of the QA systems use Natural language processing (NLP) with the IR component for searching the answer, as they depend on open-domain documents. The NLP performs two operations such as question analysis in front-end and answer generation in back-end of the QA system.

### A. Web-based Question-Answering systems

The QA system utilizes the rich web sources to answer the users' natural language questions. The web source is only the dynamic information source to provide the accurate result of the questions in a timely fashion. While missing entity relationship in the large-scale knowledge bases, the web-based searching is used to generate the appropriate answers. Google search engine [4] identifies the complex mappings between the surface form of users' questions and entities through the Google API. The work in presents the framework for a question- answering system that includes question analysis, search, hypothesis generation, and hypothesis scoring. An automatic QA system in exploits the search engine results and n-gram co-occurrence statistics to answer the complex and non-factoid questions by implementing the dynamic programming algorithm. To predict the appropriate answer type of each question, the existing QA systems contemplate on the question analysis regarding question classification, and context based answer type validation [1].

### B. Ontology-based Question-Answering systems

Ontology a common vocabulary, it extends service to expand the queries before submitting them to the web search engine. Conceptual semantic space is used to expand the users' questions by exploiting the WordNet ontology-based

semantic relations of the keywords appearing in the questions. Most of the conventional web-based QA systems present the candidate answers by exploring named entity recognition and N-grams measurement. Hence, this work focuses on linking the entities of search engine results with the ontology which induces the accurate candidate answer prediction. The WAD approach provides the appropriate answer including the adequate, and significant information on the given question by exploiting both the web corpus and ontology sources. This approach determines the answer type of each query based on the probability of the query and candidate answer sentences or context in the ontology structure .Then, it validates the answer type with the list of answer sentences to predict the precise answer to a given question.

## III. GENERAL ARCHITECTURE OF QA SYSTEM

In QA system, User posed a query as a input in natural language. After that this query is going to search the document to extract all the possible answers for the user query. [2]

The basic architecture of Question-Answering system is as shown in Figure
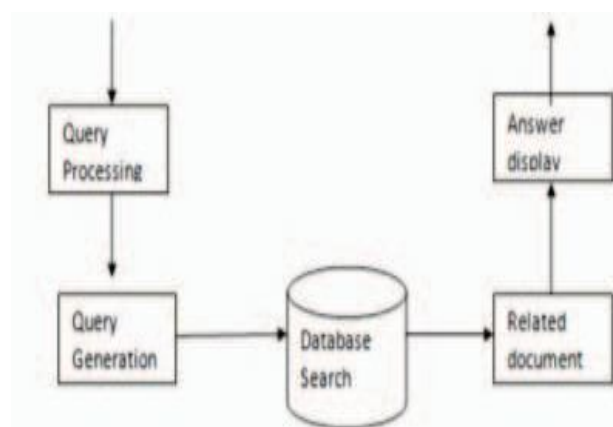


*Query Pre-processing: -* This is the first step in QA system in which input to the system is question ask by user in natural language, the overall function of this module is to process and analyze the input question.

*Query Generation*: - In query generation, Query Logic Language (QLL) is used for expressing the input question.

*Database Search*: - Here the search of the possible results is done in stored database, the related data that satisfy the given query with selected keyword are sent to the next process

*Related Document:* - The result which was generated by the previous stage is stored as a document.

*Answer Display:* - The result is stored as a document. Then the result is converted into accurate text for that the user is looking for and that answer displayed to the user.

## IV. RELATED WORK

At present, many domestic research institutions and remote teaching site have launched in-depth research on question answering system. According to the characteristics of question answering system in remote teaching system, it can be divided into three types [8]: The first type, there is no specific answering system, the communication between teachers and students are limited in e-mail or message board. The second type which has the function of preliminary answer, these systems provide a BBS discussion and a Virtual class which under the leadership of the teacher. The third type which has the function of preliminary automatic question-answering, these systems mostly use keyword query, then they return a lot of related or unrelated materials according to the keywords.

Ontology can be used as Knowledge base. The definition of Ontology in computer is "Ontology is the formal specification of conceptual model", the goal of Ontology is to obtain, describe, and express related fields of knowledge. Document used Ontology to represent domain knowledge, but Ontology was only used to calculate the semantic similarity of questions. Document ontology was used in the automatic question answering system. Document . designed a question answering system based on Ontology, which categorized users' problems. The above documents can't indicate the inherent relations among knowledge. In this paper, Ontology was used to build curriculum domain knowledge base. Through semantic analysis of users' problem, the system transforms problem model into Ontology retrieval, finally, the system extracts the answer from the query to realize answering automatically. The advantages of this automatic question answering system are as follows:
Refining concept, instance, and the relationships among their properties. Using seven steps to construct Ontology without a lot of Frequently Asked Questions.

Extending the standard answer by using the extended Ontology, which can reflects the inherent relations between the knowledge more clearly.

## THE CONSTRUCTION OF DOMAIN ONTOLOGY

### 4.1 Ontology Construction Method

he essence of building domain Ontology is to study the characteristics of related field objects and the relationships between the objects, and then to formulize them and store them in the Ontology. This article has refined concept, instance, and the relationships among their properties through summarizing the knowledge in the course of "natural language processing". In this paper, protégé was used as a template tool for the description of Ontology to create a comprehensive Ontology knowledge base of natural language processing. The Stanford University's seven steps was adopted in this paper to construct Ontology knowledge base.

### 4.2 The Realization of the Ontology and the Consistency Test

Portege can transform Ontology model of natural language processing into corresponding entity for storage. In this paper, Ontology model was edited into Protégé through handling the label of classes, data properties and object properties.For a better searching and reasoning, Ontology which has been built already should pass through the consistency test. The protégé possesses FaCT++ reasoning machine which is based on description logic reasoning and references logic description to depict the Ontology knowledge. Tableau algorithm is used in the reasoning. In this paper, that kind of reasoning machine was used to check whether the Ontology construction is consistent. The results showed there is not conflict in the Ontology construction. Clicking "Reasoner" options in protégé, selecting "FaCT++ reasoning machine", and then clicking "Start Reasoner" to test the consistency of Ontology.

## V. THE DESIGN AND IMPLEMENTATION OF AUTOMATIC QUESTION ANSWERING SYSTEM

### 5.1 System Framework Design

The main functions of system are: Using participle apparatus which combine with user dictionary to handle users' question including word segmentation and part-of-speech tagging; Extracting keyword; Extracting answer via Jena reasoning; If the system have corresponding rules, then it will give an answer, otherwise, the system will give knowledge navigation and the most relevant answer from an Ontology knowledge base; Extending the answer.

### 5.2 Question Analysis Module

Question analysis module mainly includes word processing and keyword extraction. The steps of question analysis are: Chinese word segmentation. Part-of-speech tagging which combine with user dictionary. Getting the

ingredients of a sentence by Stanford Parser syntax analysis. Getting keyword sequence. To be more specific, through the analysis of the question, this module convert the user's intention into the Jena query statement of Ontology. Question analysis process is as follows:
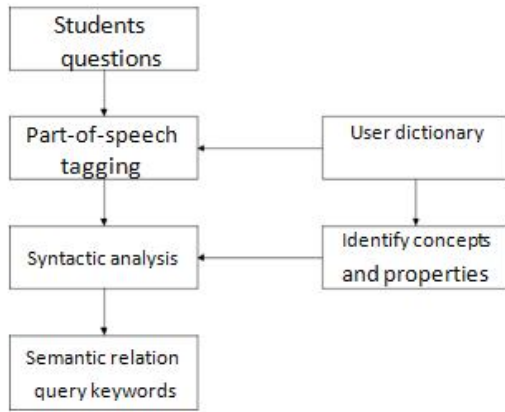


Fig. 1 Question analysis process

## 5.3 Answer Extraction Module

After question analysis, the next step is information retrieval on Ontology. In other words, the answer extraction module transforms the problem into the query of the basic elements in the Ontology (concepts, attributes, relationships, and instance). The SPARQL language and Jena was adopted in this paper for retrieving information on Ontology and extracting the answer. Key words and answer extraction process are shown in the figure below:

Ontology reasoning plays a vital role in answer extraction.

The Ontology should add object properties in order to satisfy the Ontology reasoning base on rules.

For example, in order to reasoning precedence relationship between technologies, we add object properties "hasPremise". For example, [rull1: ?xfa:hasPremise? Y ?y fa:hasPremise? Z ?x fa:hasPremise? z ]. In this rule, x, y, z present technology. Therefore, Jena can reason out the direct and indirect technology before dealing with a technology.
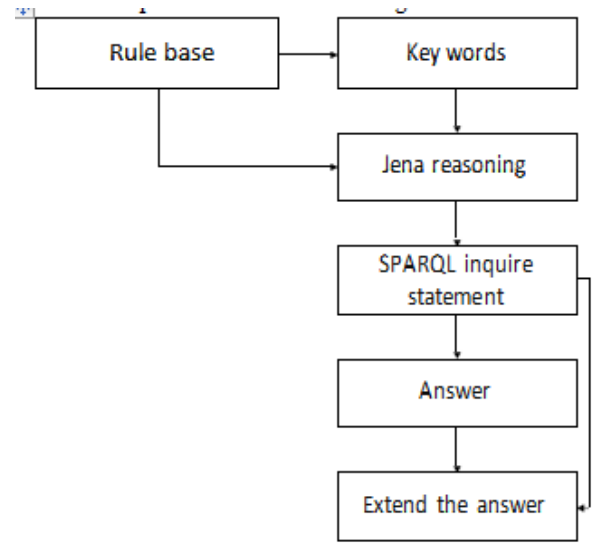


Fig. 2 Key words and answer extraction process

For example, word segmentation is needed before syntax analysis, part-of-speech tagging is needed before word segmentation. Therefore, Jena infers that part-of-speech tagging is needed before syntax analysis.

## 5.4 The Standard Answer Expansion Module

Ontology extension mainly includes synonymous extension, properties extension, the upper extension, the under extension, Instance extension etc.

The father-son relationship of node and the tree node are generally adopted to extend the query concepts. Because father-son relationship is the most common and accessible relationship in the hierarchy of Ontology.

To help students study systematically, we not only use Ontology to give the answer, but also give standard answer extension by the Ontology extension. In the process of giving standard answer, there are two kinds of situations: key words which come from the analysis of questions in the Ontology, or does not exist.

*1) The Ontology concept*

    a) *Using synonyms library to find synonyms.*
    b) *Extending the concept of parent/child in Ontology.*
    c) *Extending the brother concept of Ontology.*
    d) *Extending the properties of Ontology. Properties include data properties and object properties. Some results are new individuals, we give hyperlink to them. Click on the links to continue learn.*

*2) Other words*

If there are no corresponding words in the Ontology, then the system store them in unmatched dictionary to help builder updating Ontology. At the same time, the system provides knowledge navigation according to the classes of Ontology and provides the reference answer.

## VI. CONCLUSIONS

Ontology is a concept collection in special domain,Ontology is a concept collection in special domain, which describes the relationship between the concepts .Ontology has a good concept hierarchy and supports logical reasoning. It is better in reflecting the logical and hierarchical relationships between knowledge than ordinary database. This paper sited "Natural language processing" as an example and set up a knowledge Ontology of "natural language processing" which was used to help [7] realizing the Automatic question answering system. Through the experiment, we found that kind of question answering system is not only able to answer users' questions but also can provide relevant recommendations, which can help students study systematically. This article has realized automatic question answering system, but there are still a lot of works remaining to be done in the future. Because the author's understanding of natural language processing course is limited, and there is almost no ready-made relevant Ontology can be reference, the analysis of the concept and the relationship between knowledge is limited to author's understanding, which leading the Ontology to be not perfect enough. .In the future, we will further improve the domain Ontology and its conceptual structure level.

### REFERENCES

[1] S.Jayalakshmi  Dr.Ananthi Sheshasaayee (2017) Automated Question Answering System Using Ontology & Semantic Role.
[2] Sweta P.Lende, Dr.M.M.Raghuwanshi (2016) Question Answering System on Education Acts Using NLP.
[3] www.google.com
[4] www.ieee explorer.com
[5] sci-hub.tw
[6] www.wikipedia.com
[7] Xingbo Xie, Wei Song, Lizhen Liu,Chao Du, Hanshi Wang(2015) Research & Implementation of Automatic Question Answering System based on Ontology