The Question Answering System Using NLP and AI

Shivani Singh
Student, SCS,
Lingaya's University,Faridabad

Nishtha Das
Student, SCS,
Lingaya's University,Faridabad

Rachel Michael
Student, SCS,
Lingaya's University,Faridabad

Dr. Poonam Tanwar
Associate Prof. , SCS
Lingaya's University, Faridabad

Abstract:

The Paper aims at an intelligent learning system that will take a text file as an input and gain knowledge from the given text. Thus using this knowledge our system will try to answer questions queried to it by the user. The main goal of the Question Answering system (QAS) is to encourage research into systems that return answers because ample number of users prefer direct answers, and bring benefits of large-scale evaluation to QA task.

Keywords:

Question Answering System (QAS), Artificial Intelligence (AI), Natural Language Processing (NLP)

1. INTRODUCTION

Question Answering (QA) is a research area that combines research from different fields, with a common subject, which are Information Retrieval (IR), Information Extraction (IE) and Natural Language Processing (NLP). Actually, current search engine just do "document retrieval", i.e. given some keywords it only returns the relevant ranked documents that contain these keywords. They do not provide a precise answer to that. Hence QAS is designed to help people find specific answers to specific questions in restricted domain.

QA systems are classified into two main categories, namely open-domain QA systems and closed- domain QA systems. Open-domain question answering deals with questions about nearly everything such as the World Wide Web. On the other hand, closed-domain question answering deals with questions under a specific domain (music, weather forecasting etc.) The domain specific QA system considers heavy use of natural language processing systems.

QA is different from the search engines in two aspects:
(i) In QA, query is the question not a keyword
(ii) QA response with specific answer to a specific question instead of a list of documents.

1.1. APPROCHES in QA

There are three major approaches to Question Answering Systems: Linguistic Approach, Statistical Approach and Pattern Matching Approach.

A. Linguistic Approach

This approach understands natural language texts, linguistic techniques such as tokenization, POS tagging and parsing.[1] These are applied to reconstruct questions into a correct query that extracts the relevant answers from a structured database. The questions handled by this approach are of Factoid type and have a deep semantic understanding.

B. Statistical Approach

Availability of huge amount of data on the World Wide Web increased the importance of Statistical Approach. Statistical approaches and online text repositories are not dependent on structured query languages and can formulate queries in natural language.

Statistical techniques: Support Vector Machine classifier, Bayesian classifiers, etc.

C. Pattern Matching Approach

This approach deals with expressive power of text pattern, it replaces the sophisticated processing involved in other computing approaches. This approach uses the communicative power of text patterns. This approach best suits to small and medium sized websites. The type of questions handled by this approach are mainly factoid based,

definition based and it has a less semantic understanding as compared to other approaches. Most of the pattern matching QA systems uses the surface text pattern.

1.1. QAS COMPONENTS

The architecture of the QA system as mentioned earlier, would consist of following three modules: 1. Question processing module. 2. Document processing module. 3. Answer extraction and formulation module.

The questions that the system receives can be divided into two major categories: FACTUAL & EXPERT. Factual questions are those which contain words like what, where, when, who, etc. Expert questions are those which contain words like how, why etc.

- **TheDocument Processing Module**

It takes in the choice of the user for a particular passage from the displayed list .Then using POS Tagger tags the tokens .With the help of tags finds the verbs in the passage. Using the list of irregular verbs and the logic for regular verbs a data structure (array) is created which contains the verbs along with their tenses and ing form.

- **TheQuestion Processing Module**:

It takes in a question from the user Using StringTokenizer tokenizes it and stores it in another data structure returns it for further use in the program

- **The Question-Answering Module**

First finds the verb in the question. Matches the verb just found with the tokens created in the document processing stage. According to the selected case for type of factual question (what, when, etc) it further tries to extract and formulate the answer.

The user is first asked to select the passage of his choice and then the type of question. The Question processing module will process the question and pass it to the Question Answering module which will make use of the various extractions received from the Document Processing phase, along with the Processed Documents containing the tagged format of the original input document. By applying required algorithms this module will pass it to the Formulation module for getting the desired answer.

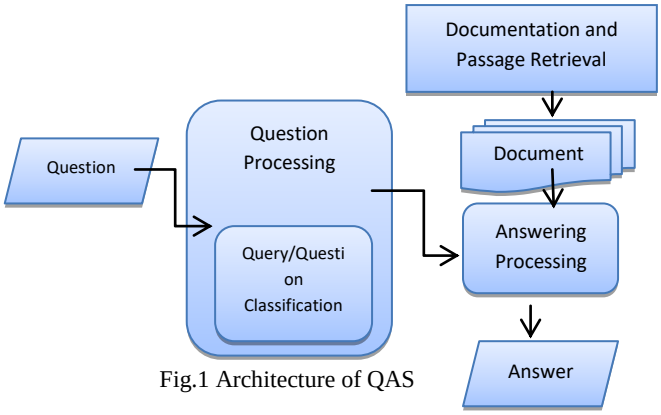

Fig.1 Architecture of QAS

2. LITERATURE SURVEY

QA systems, as explained before, have a backbone consisting of three main parts: categorization of question, information retrieval, and answer extraction. Therefore, each of these three components attracted the attention of QA researchers.

Question Classification

Question class	Question	Answer Type
WHAT	basic-what/ what-who /what-when/ what-where	Money/ No./ Definition/ Title/ NNP/ Undefined
WHO		Person /
HOW	basic-how how-manyhow-long how-much how-muchhow-far how-tall how-rich how-large	Manner Number Time/Distance Money / Price how-much Undefined Distance Number Undefined
WHERE		Location
WHEN		Date
WHICH	which-who which-where which-when which-what	Person Location Date NNP
NAME	name-who name-where name-what	Person/ORG. Location Title / NNP
WHY		Reason
WHOM		Person

Questions usually gurantee to predictable language patterns, and therefore are categorized based on taxonomies. Taxonomies are distinguished into two main types: flat and

hierarchical taxonomies. Flat taxonomies have only one level of classes without having sub-classes, on t,he other hand hierarchical taxonomies have multi-level classes. Lehnert [2] proposed "QUALM".

QA system used a flat taxonomy with seventeen classes e.g. PERSON, PLACE, DATE, NUMBER, DEFINITION, ORGANIZATION, DESCRIPTION, ABBREVIATION, KNOWNFOR, RATE, LENGTH, MONEY, REASON, DURATION, PURPOSE, NOMINAL,OTHER.

Zhang and Lee [3] compared various choices for machine learning classifiers using the hierarchical taxonomy propose such as: Support Vector Machines (SVM), Nearest Neighbors (NN), Naïve Bayes (NB), Decision Trees (DT).

Information Retrieval

Evaluate the use of named entities and of noun, verb, and prepositional phrases as exact match phrases in a document retrieval query. Gaizauskas, and Humphreys [4] described an approach to question answering

Answer Extraction

Finding the answers by exploiting surface text information using manually constructed surface patterns. In order to enhance the poor recall of the manual hand-crafting patterns, many researchers gained text patterns automatically such as Xu et al [5].

3. PROPOSED WORK

Nowadays there are many different Question Answering Systems. Most popular QAS and their main characteristics are described below.

ELIZA

One of the earliest QA systems was ELIZA, developed in 1964. One successful ELIZA application was DOCTOR, a computer program that basically interacted with users through a text chat interface, answering questions and responding to the users dialog in a way that mimicked the client-centered psychotherapy between a client (the user) and their therapist.

Evi (originally known as TrueKnowledge) is a knowledge Web search engine that assists people gaining what they desire and require through Evi understanding of each user and the world they live in. [6].

QUORA is available in English and Spanish launched in 2010. The QUORA community includes some well know people, such as Dustin Moskovitz, Jimmy Wales.

BING is Microsoft's answer to google and it was launched in 2009. Bing is the default search engine in Microsoft's web Browser.it is available in 40 languages. It provides different services including image, web and vedio search along with maps.

Stoyanchev et al (StoQA) , 2008: Contribution In their research, they presented a document retrieval experiment on a question answering system. They used exact phrases, as constituents to search queries. The process of extracting phrases was performed with the aid of named-entity (NE) recognition, stop-word lists, and parts-of-speech taggers.

Wolfram|Alpha is a Computational Knowledge Engine that introduces a fundamentally new way to get knowledge and answers, not by searching the Web sites, but by dynamic computations based on a vast collection of built-in data, algorithms, and methods. [7]. Wolfram|Alpha returns an answer in a form of a table where information, which is relevant to a query, is separated by categories (e.g. an answer to a query about some person usually contains such se image, timeline, notable facts, familial relationships and others).

Answerbag is question answering website where users can get answers to their questions, whether they're looking for facts, opinions or simply entertainment. Questions are answered by Answerbag professional researchers and community members.[8]so Answerbag can also be considered as an expert community question answering website..

Blurtit is an online question answering community that provides the answers users are looking for and gives them free, 24/7 access to a whole world of information, and to millions of knowledgeable friends. Blurtit knowledge base contains facts, information and users' opinions, enlarging with each new answer.

Kangavaril, 2008: Contribution The research presented a model for improving QA systems by query reformulation and answer validation. The model depends on previously asked questions along with the user feedback.
Experimental environment and results: The system worked on a closed aerologic domain for forecasting weather information.
Limitations:the model was tested in a sparse experimental environment in which the domain was very particular and the number of questions relatively small. Also, depending only on the users as a single source for validating answers is a two-edged weapon.

Ask.com (originally known as Ask Jeeves) is a question answering-focused web search engine. The original idea behind Ask Jeeves was to allow users to obtain answers to questions posed on daily basis, natural language, as well as

by traditional keyword searching. The current Ask.com still supports this, with added support for math, dictionary, and conversion questions. This system tries to "understand" any users query and gives three forms of answer at once: a direct answer, a list of links to webpages on related topics and a list of similar questions with answers from other question answering websites.

The general characteristics of described QAS are summarized in Table 1.

Table 1

QA System	Domain	Description	Year
ELIZA	Closed	Attempt to mimic basic human interaction Q&A exchanges.	1964
EVI	Open	Specializes in knowledge base & semantic search	2007
Quora	Open	Knowledge based answering ability	2009
Bing	Open	Provide diff. services like image, web and video seach	2009
Stoyanchev	Closed	Extract phases	2008
Wolfram/Alpha	Closed	Computation search engine	2009
Answerbag	Open	Web based QA system	2003
Blurtit	Open	People ask query of regular user provide answer based on their opinion.	2006
Kangavari	Closed	Depend on Previously asked Question.	2008
ASK.com	Open	Web based QA	1996

4. APPLICATION AND FUTURE SCOPE

QA system handles a question in natural language and tries to provide its answer in a single statement. With certain improvements; the proposed system can be used for the following applications:

- Adding speech recognition abilities to the current system will enable people with reading disabilities to take advantage of this system.
- Query processing in database systems and many analytical tasks that involve collecting, correlating and analyzing information; can naturally be produced as QA queries.
- Can be used to make online lectures more old-school type by allowing lectures to proceed only when questions related to the previous lecture are answered correctly.

The existing system can be integrated with a search engine to enhance the performance.

CONCLUSION

This paper describe about the Question Answering System for an English Language i.e. it receives query from the user and selects most appropriate answer. QAS is approach to find the correct answer to the question asked from user. This paper also describes different QAS approaches, different types of QAS.QA system help in improving system interaction. In this paper we also concentrated on finding the solution of some problem: Answer is restricted to a precise domain, user has to follow a particular path while entering a question and Extracting correct answer. The solution consists: semantic representation for Natural Language, effective logic is to be performed on them and developing a formalism to represent the answer verification and specific answer extraction. Thus there is great potential for exploring the challenges in QA domain.

ACKNOWLEDGMENT

We sincerely thank prof. PoonamTanwar, Computer department, Lingaya's University, Faridabad for her valuable guidance and motivation for this work.

REFERENCES

[1]. Saranya R, Christopher Augustine, " Schemes and Approaches in Question Answering System", International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) ,Vol. II, Special Issue X, March 2015.
[2]. Lehnert, W., 1977. The Process of Question Answering - A Computer Simulation of Cognition. Yale University, ISBN 0- 470-26485-3.
[3]. Zhang, D. & Lee, W., 2003. Question Classification using Support Vector Machines.In Proceedings of the 26th Annual International ACM SIGIR Conference.
[4]. Gaizauskas, R. & Humphreys, K., 2000. A Combined IR/NLP Approach to Question Answering Against Large Text Collections.In Proceedings of the 6th Content-based Multimedia Information Access (RIAO-2000).
[5]. Xu, J., Licuanan, A. &Weischedel R., 2003. TREC2003 QA at BBN: Answering Definitional Questions. In Proceedings of the 12th Text Retrieval Conference.
[6]. The story of Evi .http://www.evi.com/about/.
[7]. http://www.wolframalpha.com/about.html.
[8]. About us, Answerbag2014.
http://www.answerbag.com/about-us/