Natural Language Processing for Question Answering in the Legal Domain

Er. Inakshi Garg
Computer Science Department
Chandigarh University
Gharuan, Punjab
Inakshi.e12349@cumail.com

Aman Saini
Computer Science Department
Chandigarh University
Gharuan, Punjab
19bcs1733@gmail.com

Rupam Bhandari
Computer Science Department
Chandigarh University
Gharuan, Punjab
19BCS1840@cuchd.in

Abhinav Verma
Computer Science Department
Chandigarh University
Gharuan, Punjab
19bcs1232@gmail.com

***Abstract* — The main objective of the paper is to built a smart question answering (QA) system that takes some input like .txt file and processing it and with the help of the provided .txt file that acquiring knowledge and that can be used in the legal domain. Utilizing the knowledge gained, our system endeavors torespond to user queries by means of a Question Answering System (QAS). Main target of this QASis to inspire the exploration of answer-returning systems, as a substantial number of users prefer direct responses.**

Keywords – QA: Question Answering System, AI: Artificial Intelligence, NLP: Natural Language Processing

I. INTRODUCTION

Question answering system is an interdisciplinary testing field which merges the work in various areas, such as Information Retrieval (IR), Information Extraction (IE), and Natural Language Processing (NLP), to achieve a common goal. We use different kinds of web engines to search they help us to find answers through the information, but those application have some limitation and they could not help us or they could not give us information beyond some limitations. [1]Unlike current search engines that only perform "document retrieval" by returning ranked documents containing the given keywords, but not providing an accurate answer, the main goal of the QAS is help the users by giving the accurate answer for the particular questions with in a limited estate.

Open-domain QA system and closed domain QA system are the two main types of question answering system. Open-domain QA systems handle questions about a wide range of topics, such asthose found on the World Wide Web. Conversely, closed-domain QA systems address questions that fall within a specific domain, such as music or weather forecasting. The latter relies heavily on natural language processing systems.

There are two fundamental differences between QA and search engines. First, in QA, the query is a question rather than a keyword.

The second major difference between QA and search engines is that a QA system provides a particular answerfor the particular question, as opposed to a list of potential resources that may contain the answer.

QA Approaches

1) Linguistic Approach:

The comprehension of natural language texts involves in this approach, with the help of this it can utilize linguistic techniques (like tokenization, POS tagging, and parsing). To transform questions into the exact query format it used this linguistic technique. Basically, to develop a Q&A (question answering system) system requires something like it should be understood of natural language text, linguistics and basic knowledge. [2]This approach addresses factoid-type questions that necessitate a comprehensive semantic understanding.

2) Statistical Approach:

The vast quantity of information accessible on the internet has heightened the significance of utilizing statistical methods. These techniques, along with online text databases, can construct queries using natural language rather than relying on structured query languages. Examples of statistical approaches include Support Vector Machine classifiers and Bayesian classifiers.

3) Pattern-Matching technique:

It involves analyzing the communicative influence of text pattern [3] and with the help of other computational techniques it replaces more complex processing. It is particularly well-suited to smaller websites and can effectively handle factoid and definition-based questions. However, its semantic understanding is relatively limited compared to other approaches. Surface text patterns are typically used in most pattern-matching question-answering (QA) systems.

Components of QAS

As previously noted, by using three kinds of different module built the QA system's architecture: (i) module for how to process a question, (ii) module for how to process a document, (iii) module for how to give the extracting and constructing replies. It categorized by two main parts which questions received by the process: (i) factual and (ii) expert. There are some basic/common words in factual like what, who, when, where, why, how etc.

1) The Document Processing Module:

The system prompts the user to select a specific passage from a displayed list. Once the passage is selected, a POS Tagger is used to tag the tokens in the passage. The system then identifies the verbs in the passage using the tags. Basically, a QA system produce a data structure, usually an array, that carry the identified verbs through with their tenses including ing form based on irregular verbs and regular verbs.

2) Question Processing Module:

A list of a keywords increase the document/information at the query point when the question processing give the result. [4] That system receives a query from user and utilizes the stringTokenizer method to tokenize it. The tokenized question is saved and return inside the data structure and a program for more use.

3) Question-Answer module:

That question answering system firstly detects the verb which is existin the query and compares it and the tokens created in the information processing stage. There are different kinds of factual queries/questions based on one kind of factual question (like when, where, what, how), it extracts and formulates the answer for more distant.

In this system, every client firstly selects those desired paragraph, then secondly selects the which types of question answering module those are, which makes different types of extrications which is obtainedduring document processing part as well as the processed documents that had the tagged format of the actual input documentation. Formulation module than applies the necessary algorithms to generate the desired outputs.

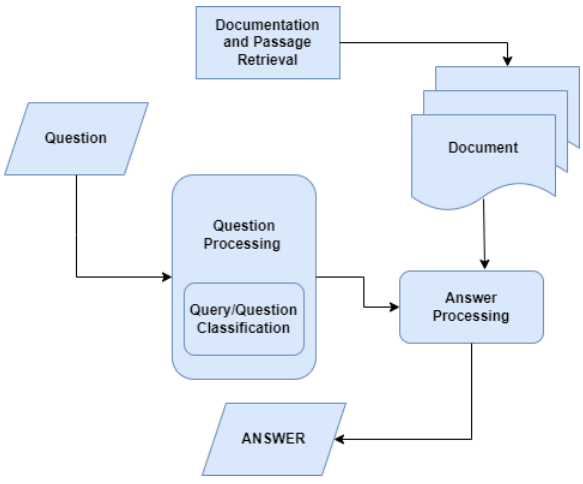

Fig.1 Architecture of QAS

II. LITERATURE SURVEY

As previously explained, with the help of 3 mainly components make a QA system: query categorization, document replenishment, and answer extrication. As a result, each kind of these elements has captured interest of any experimenters in the field of QA.

Classification of question

Classification of question can be categorized by using question classes such as what, how, which, who, where, when, why, whom. These questions can be further answered and will have their answer types.

Questions tend to follow predictable language patterns, which leads to their categorization formed on taxonomies. Flat and hierarchical are the main two varieties of taxonomies's classification. Flat taxonomiesonly have one level of classes and those classes do not have any subclasses, and another side hierarchical taxonomies include multiple levels of classes. Lehnert introduced a flat taxonomy called "qualm", which have 17 categories like people, location, day, time, digit, clarity, concern, explanation/representation, contraction, known for, figure, measure, cash, cause, time scale, determination, titular, and so on.

Different kinds of machine learning classifiers that differentiated by Zhang and Lee, that are Nearest Neighbors (NN), support vector machines (SVM), Decision Trees (DT), Naïve Bayes (NB), using a hierarchical taxonomy for their QA system.

Retrieval of Document

It is main things to evaluate the postulate effectiveness of using noun, verb, named entities and prepositional phrases as authentic match terms information replenishment query is worth evaluating. Basically, there are two men who presented an approach for question answering that involves this type of analysis, they are Gaizauskas and Humphreys [4].

Answer Extraction

By leveraging manually build aspect patterns the users can find answers to use aspect text information. However, due to limited reminder ofmanually crafted patterns and many experimenters have turned to mechanized methods for generating text patterns, as seen in the work of Xu et al.

III. PROPOSED METHODOLOGY

Currently, there exist numerous Question Answering Systems, each with its own unique features and functionalities. To authenticate the beneficial answers in reply to any user's natural language question answering is of practical importance, [5] that is known as developing automated methodology. Below, we describe the main characteristics of some of the most popular QAS available today.

ELIZA :

In 1964, ELIZA was developed as one of the earliest Question Answering Systems. DOCTOR, a successful application of ELIZA, was a computer program that utilized a text chat interface to interact with users. It makes good interactions between users and the interaction should be in simple way. [6] It provided answers to their questions and responded to their dialogue in a manner.

EVI :

If we talk about EVI what is Evi? EVI basically known as TrueKnowledge, which is search engine based on knowledge-based which is designed to help the users find the information the users need by comprehension each user's individual needs and the environment in where they live in.

Quora :

In 2010, QUORA was launched and which is available for only two languages, they are English and Spanish languages. The QUORA community includes several notable individuals, including Dustin Moskovitz and Jimmy Wales.

BING :

BING was introduced by Microsoft in 2009 as a competitor to Google. Bing is the Microsoft's web browser that serves as a default search engine and the BING is available in 40 languages. In addition to web search, it also offers various other services, including image and video search, as well as maps.

STOYANCHEV :

Stoyanchev et al. (StoQA) made a contribution to the field of question answering systems in 2008 with their research on document retrieval. They conducted an experiment using exact phrases as search queries and to extract phrases use stop-word lists, part of speech taggers and employed named-entity recognition.

WOLFRAM|ALPHA :

Basically, the Wolfram|Alpha is a type of unique Computational Knowledge Engine that through dynamic computation gives answers and information. Unlike traditional search engines that rely on web pages, Wolfram|Alpha uses a different approach to deliver results. It returns answers in a tabular format, with relevant information for the query separated into categories. For example, a query about a person would typically include details such as images, timelines, notable facts, and familial relationships, among others.

ANSWERBAG :

Answerbag is a website where users can ask questions and receive answers from a combination of professional researchers and community members. The website offers answers to factual questions, opinions, and entertainment-related inquiries. Essentially, Answerbag can be seen as a community-driven question-answering platform that provides expert-level answers to a wide range of questions.

BLURTIT :

Blurtit is an online community designed to help users find the answers they need. It is also a question answering platform, anyone can use it unlimited because it's not a chargeable platform, and the users can use it anytime. It's comprehension part or documentation part is continuously growing, incorporating reality, document, and user thoughts.

KANGAVARI :

In 2008, some researchers researched on a model to improve question answering (QA) systems and they build a model that is called KANGAVARI. Early, the model only use for asked any questions or queries and it take feedback from the user's end to increase system's performance. If we talk about the system's uses then it gives forecasting weather information in a closed aerologic estate and it gives a positive result.

However, the model had some limitations. It has been assessed in a sparse experimental domain where that environment was very determined, and there were not a lot of queries or questions. Additionally, it is depending on user's feedback for validating answers can be risky as itmay not always be accurate.

ASK.COM :

Originally known as Ask Jeeves, Ask.com and it is basically a web engine to search anything that is focused on answering questions.

By using natural language queries it allows the users to input the queries and it will give the answers, this is the main idea behind to build the Ask Jeeves. This option is still supported by Ask Jeeves and which has additional support for mathematics, dictionary and conversation queries. This system has a work method and the method's target is to "understand" user's questions and the work method give three types of explanation simultaneously: a direct response, it searches websites on relevant types of topics, and it collects the same types of queries an same types of answers from the various question answering webpages and in this way it gives the answers to the users.

IV. FUTURE SCOPE

To handle natural language questions is designed the Question Answering (QA) system and that provide answers within a line or within a statement. Within certain developments, the proposed system can be applied for various types of framework, such as if any people have disabilities they can hear the answer without read the answer by adding speech recognition which would be beneficial for every disabilities, and anyone can using as for query processing in database systems and other information gathering, document corresponding and analyzing documentation, and by restricting lectures the system make online lectures more associative it get going only when the related queries are answered correctly. The performance of the proposed system can be increased by integrating a search engine

CONCLUSION

The article describes in the English Language for the Question answering system (QA), it collects the queries as a input from the users and it decides which answer will be appropriate for a particular queries and then it gives the most correct answer. This paper provides an overview of different QAS approaches and types. The use of a QAsystem can improve system interaction and overall user experience. The paper also addresses some common challenges in QA, such as restricted answer domains, prescribed question formats, and accurate answer extraction. To overcome these challenges, the paper suggests the use of semantic representation for natural language, effective logical operations, and a formalism for answer verification and extraction. TheQA domain offers a lot of potential for exploring these challenges and developing new solutions.

ACKNOWLEDGEMENT

We express our heartfelt gratitude to Professor Inakshi Garg from the Computer Science Department at Chandigarh University, Gharuan for providing us with her valuable guidance and motivation throughout the course of this work.

REFERENCES

a. Bhushan Zope, Article Question Answer System: A State-of-Art Representation of Quantitative and Qualitative Analysis, Big Data Cogn. Comput. 2022

b. Sanjay K Dwivedia, Vaishali Singh, Research and reviews in question answering system, International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013

c. Bolanle Ojakhu, A Review of Question Answering Systems, Article in Journal of Web Engineering (JWE)
· January 2019

d. A Question Answering System Supported by Information Extraction* supported in part by the SBIR grants F30602-98-C-0043 and F30602-99-C-0102 from Air Force Research Laboratory (AFRL)/IFED.

e. AILA: A Question Answering System in the Legal Domain, Weiyi Huang, Proceedings of the Twenty- Ninth International Joint Conference on Artificial Intelligence (IJCAI-20) Demonstrations Track

f. Question Answering System, Approaches and Techniques: A Review, International Journal of Computer Applications (0975 – 8887) Volume 141 – No.3, May 2016