SURVEY ON DIFFERENT APPROACHES OF QUESTION ANSWERING SYSTEM

Toral Desai[1], Avani Parmar[2]

ME, Computer engineering department, HGCE, Ahmadabad, Gujarat, India[1]

Assistant professor, Computer engineering department, HGCE, Ahmadabad, Gujarat, India[1]

Abstract: QA is a application or specific type of information Retrieval. Today's world of the web, information is store with large repositories of information. The Big Problem we face is that large amount of information available that allows us to find what is relevant and cannot be managed without automatic Search. We solve this problem by using Question answering System(QAS). The main goal of QA system is able to answer users Question in Natural Language. The QA system can fulfil the needs of user as they provide with faster and appropriate answers to user question.

*Keywords***: Question answering System, Natural Language Processing, Information retrieval, closed domain, open Domain**

INTRODUCTION

The first question answering system introduce in the 1960s, Baseball was able to answer domain-specific natural language questions which was about the baseball games played in American league over one season. There are various types of QA systems are developed so far such as, closed –domain QAS, open domain QAS ,web based QAS, information retrieval(IR) QAS , Information Extraction (IE) based QAS, rule based QAS. QA systems can be classified in following categories:

1.1 Classification of Question Answering Systems

The primary Classification of QAS are classified on the basis of content and on the basis of language paradigm [3]. Following are the different categories of Question Answering System

A.   Classification based on Data Content

The most important classes of techniques based on data content are open-domain and domain specific.

  i. Open domain question answering :

Open domain question answering deals with unrestricted topics. Hence, questions may concern any subject. The corpus may consist of unstructured or structured texts. On the other hand, these systems tackle huge amount of data to extract the most relevant answer.

  ii. Closed-domain question answering (domain specific):

Closed-domain question answering deals with questions under a specific domain, which means that the topics of the questions are restricted. In this QAS only a limited type of questions are accepted, such as questions asking for descriptive rather than procedural information. This type of QA is easier, for the vocabulary is more predictable, and

ontologies describing the domain are easier to construct. In medical or railways ,the closed domain question answering system is used.

B. Classification based on Language

The another way of classifying the field of QA deals with language. There are three type of language base QA. In monolingual QA both the questions and the answer are in the same language. It is useful for people speaking one of the popular languages. cross-language QA the language of the questions (source language) is different from the language of the documents (target language). Multilingual systems deal with multiple target languages. User asks questions in one language and gets answers different from the source language or same as source language in Multilingual QA system.

1.2 Classification of Answer Classes

There are two type of Answer Classes,That are Factoid and non- Factoid.

   i. **Factoid:**
      Factoid answer means short fact based answers like names, dates.

   ii. **Non- Factoid:**
      Non – factoid means description and Definition.

ARCHITECTURE OF QUESTION ANSWERING SYSTEMS

In Figure 1.1., a Basic Architecture of QA system consists of three Common modules, each of which has a core component beside other supplementary components: "Question Processing Module" whose heart is the question classification, the "Document Processing Module" whose heart is the information retrieval, and the "Answer Processing Module" whose heart is the answer extraction[6].

The Question processing module include three parts: Query Interface, question analyzer and Question classification, The information retrieval component is used to retrieve the relevant documents based upon important keywords appearing in the question. The answer processing module is responsible for identifying, extracting and validating answers from the set of ordered paragraphs passed to it from the information retrieval module.

The detailed description on the working of QA system is given in section 2.1.with the study of various approaches used to develop different modules of QA system.
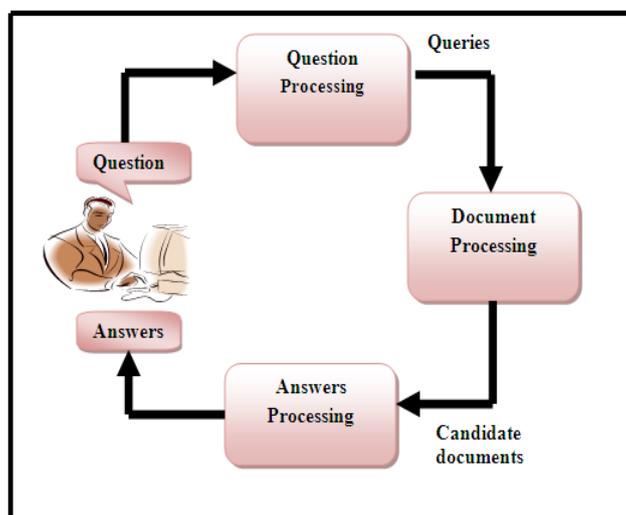

Fig 1. Basic Architecture Question Answering System.

LITERATURE REVIEW

In [1], investigates the role of distributional Semantic Models in Question Answering System(QAS).The Method to integrate DSM into QAS ,Called QuestionCube. The QuestionCube is framework for QA that merge several techniques to retrieve passages Containing the exact answers for Natural Language Question. The authors propose several kinds of DSMs based on classical Term- Term co-occurrence Matrix (TTM), latent semantic analysis (LSA), Random Indexing (RI) and combination of last two Techniques. Authors idea is that DSMs approaches can help to Semantic relatedness between users 'questions and candidate answers by exploiting paradigmatic relations between words.

In [2], shows the all implementation approaches for different categories of QAS. First approaches is Closed-domain QAS. It is introduce in 1961(e.g., BASEBALL) and 1973 (e.g., LUNAR). In Open Domain based QAS, The most important challenge of an open domain system is its database The efficiency of any system depends on how well the database is arranged and maintained, A vector space model is a kind of model which can be used for classifying the candidate answers. In WEB BASED QAS, the most important property is "snippet − tolerant" property which allows it to provide correct responses to the QAS while searching answer through search engines like Google, yahoo etc. In Information Retrieval or Information Extraction (IR/IE) based QAS, IR system works on the interaction between human and computer when used to search the answer for posed question. IE systems are used for extracting the correct answer from the retrieved documents. Developing Rule based QAS is bit challenging task as the developer needs to consider virtually all the possible topics on which the system may get tested.

In [3], discussion regarding different Question Answering types. In addition they describe Mean Reciprocal Rank (MRR) used to evaluate the performance of different question answering systems. They also discuss the recent question answering systems developed and their corresponding techniques.
Semantic relatedness measures quantify the degree in which some words or concepts are related, considering not only similarity but any possible semantic relationship among them. Relatedness computation is of great interest in different areas, such as Natural Language Processing, Information Retrieval, or the Semantic Web. In [4], we explore the use of a semantic relatedness measure between words, that uses the Web as knowledge source. This measure exploits the information about frequencies of use provided by existing search engines. Semantic measures can also be defined between lexically expressed word senses, or between whole texts. Three main kind of measures are defined in this paper: semantic similarity, semantic relatedness and semantic distance.

QA system consists of three Common modules, each of which has a core component beside other supplementary components: "Question Processing Module" whose heart is the question classification, the "Document Processing Module" whose heart is the information retrieval, and the "Answer Processing Module" whose heart is the answer extraction. In this survey[6] ,all question answering system approaches is defined by author and all described the minor lamination of all QA researchers. In Question Processing Module, different approach such as Machine learning based approach, rule based approach, hierarchical taxonomy, flat taxonomy

are mainly used in different systems. In Document Processing Module, used the web corpus and knowledge-based corpus approach. In Answer Processing Module, used the text pattern and named entity approach.

It summarized and organized recent research results in a novel way that integrated and added understanding to work in the question-answering system field. It is impossible for a survey to include all or even most of previous research, this survey included only the work of the top-publishing and top-cited authors in the QAS field. This survey [6] also included research containing minor limitations to show how these limitations were discovered, faced and treated by other researchers.

The goal of a question answering system is to retrieving answers to questions rather than full documents or best-matching passages, as most information retrieval systems. In [5], Author discussed some of the approaches used in the existing QA system and proposed a new architecture for QA system retrieve the exact answer. Answering system has become an important component of the online education platform. Question answering system for Indian languages like hindi, telugu, bengali is discussed. No Punjabi QAS is discovered. The focus of the system has been mainly on four kind of questions of type What, Where, How many, and what time. On analysis of the system the overall efficiency of the system was found to be significant. The next generation of question answering systems will have to take into consideration presently available multimedia data. There exists a mixture of natural language text, images, video, audio, user added tags, and metadata. On the question side, users may express their queries using a variety of modalities.

In [7], research different semantic relatedness functions called "Measure of Semantic Relatedness (MSR)" are discussed and compared. They found that the quality and accuracy of MSRs are different when applied in various contexts. In this paper they compared several MSR algorithms using different corpuses and have analyzed the results.

In [8], they discussed some of the approaches like Web Based QAs , IR / IE Based QAS , Restricted Domain QAS , Rule Based QAs used in the existing QA system and proposed a new architecture for QA system retrieve the exact answer. In [8] discussed all basis component of Question Answering system, Answering system has become an important component of the online education platform.

In [9], presents a survey of various types of QA systems. These QA systems are classified as Text based QA systems, Factoid QA systems, Web based QA systems, Information Retrieval or Information Extraction based QA systems, Restricted Domain QA systems and Rule based QA systems. The paper further investigates a comparative study of these models for different type of questioners which led to a breakthrough for new directions of research in this area.

In [10], They present a logic-based semantic approach for the recognizing textual entailment task. The system participating in the RTE competition used a set of world-knowledge, NLP, and lexical chain-based axioms and an in-house logic prover which received as input the logic forms of the two texts enhanced with semantic relation instances. Because the state-of-

the-art semantic parsers cannot extract the complete semantic information encoded in text, the need for semantic calculus in NLP became evident. They introduce semantic axioms that either combine two semantic instances or label relations between the frame elements of a given frame.

CONCLUSIONS

This paper include Different Classification of QAS based on Data Content and Language. Also show the answer classes. There are two answer classes factoid and Non- factoid.

In this paper, show many approaches for the Question Answering System. Different approaches are Closed Domain QAS(e.g. BASEBALL), Open Domain based QAS , Web Based QAS and Rule based QAS and also Defined the Implementation of this Different Approaches.

ACKNOWLEDGEMENT

REFERENCES

[01] Piero Molino, Pierpaolo Basile, Annalina Caputo, Pasquale Lops, Giovanni Semeraro, "Exploiting Distributional Semantic Models in Question Answering", IEEE Sixth International Conference on Semantic Computing, pp. 146-153,19-21 September, 2012

[02] Walke P.P, Karale, S. "Implementation Approaches for Various Categories of Question Answering System", In Proceedings of 2013 IEEE Conference on Information and

[03] Communication Technologies (ICT 2013),11-12April, 2013.

[04] Jaspreet Kaur, Vishal Gupta, "Effective Question Answering Techniques and their Evaluation Metrics", International Journal of Computer Applications, vol.65, no. 12, pp. 30-37, March 2013

[05] Emadzadeh, E., Nikfarjam, A. , Muthaiyah, S.,"A Comparative Study on Measure of Semantic Relatedness Function", In Proceeding of the 2nd International Conference on Computer and Automation Engineering (ICCAE),pp. 94 - 97, 26-28 Feb. 2010.

[06] Loni, Babak., "A survey of state-of-the-art methods on question classification.", Delft University of Technology , Technical Report, pp. 1-40, Aug, 2011.

[07] Ali Mohamed Nabil Allam, Mohamed Hassan Haggag, "The Question Answering Systems: A Survey", International Journal of Research and Reviews in Information Sciences (IJRRIS) , Vol. 2, No. 3, pp. 211-221, September 2012.

[08] Ehsan Emadzadeh, Azadeh Nikfarjam, Saravanan Muthaiyah, "A Comparative Study on Measure of Semantic Relatedness Function",vol 1 , 2010 IEEE,

[09] Poonam Gupta, Vishal Gupta, " A Survey of Text Question Answering Techniques", International Journal of Computer Applications (0975 – 8887) Volume 53– No.4, September 2012

[10] R.Mervin, " An Overview of Question Answering System", International Journal Of Research In Advance Technology In Engineering (IJRATE) Volume 1, Special Issue, October 2013

[11] Marta Tatu and Dan Moldovan, "A Semantic Approach to Recognizing Textual Entailment", Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), pages 371–378, Vancouver, October 2005.