APPROCHES FOR QUESTION ANSWERING SYSTEMS

Article *in* International Journal of Engineering Science and Technology · February 2011

Source: DOAJ

3 authors, including:

Guda Vanitha
Chaitanya Bharathi Institute of Technology
10 PUBLICATIONS **30** CITATIONS

SEE PROFILE

Suresh Kumar Sanampudi
Jawaharlal Nehru Technological University, Hyderabad
11 PUBLICATIONS **70** CITATIONS

SEE PROFILE

APPROCHES FOR QUESTION ANSWERING SYSTEMS

VANITHA GUDA

Computer Science, Chaitanya Bharathi Institute of Technology,Gandipet,
Hyderabad, AndhraPradesh, India.

SURESH KUMAR SANAMPUDI

Information Technology, JNTUH College of Engineering, Nachupally, Karimnagar,
AndhraPradesh, India.

I.LAKSHMI MANIKYAMBA

Computer Science, JNTUH College of Engineering,
Hyderabad, AndhraPradesh, India

Abstract :

Question Answering (QA) system is a man machine communication device. The basic idea of QA systems in Natural Language Processing (NLP) is to provide correct responses to the questions in a human like manner giving short and accurate answers. This paper presents a survey of various types of QA systems. These QA systems are classified as WEBBASED QA system, Information Retrieval or Information Extraction(IR/IE)BASED QA system, RESTRICTED DOMAIN QA (RDQA) system and RULE BASED QA system. 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.

Keywords: Question Answering System; Temporal Information; Questioners.

1. Introduction

With the wide spread use of information in the internet exploration era there is a recently renewed interest for retrieving answers to the questions which are short and accurate. **Q**A systems aim to retrieve point-to-point answers rather than flooding with documents or even matching passages as most of the information retrieval systems do. For E.g. ***"who is the first prime minister of India?"*** the exact answer expected by the user for this question is (Pandit Jawaharlal Nehru),but not intends to read through the passages or documents that match with the words like first, prime minister, India etc.,.

The major challenging issues in Question answering system is to provide accurate answers from tremendous data available on the web. It also aims to recognize the cross linguistic questions which allow the users to ask the questions and obtain the answers in their native language. The processing of time based information to answer temporal queries still remains as a challenge.

In this paper we focus on different types of QA systems which are categorized into two major groups based on the methods used by them. First category of QA system belongs to simple natural language processing and information retrieval methods, while another category of QA systems are depended on reasoning with natural language.

In [Hermjakob.U,(2001)] the first category popular QA systems such as Web based QA systems [Zheng.Z,(2002)], IR/IE based QA systems [Rohini Srihari,(2002)] make use of various retrieval methods to answer the questions. Some of the present online QA systems which falls under this retrieval based methods are Webclopedia [Brin,(1998)],AnswerBus [ZhipingZheng,(2002)], Mulder [C.Kwok,(2001)], Lamp [DellZhang,(2002)] etc.,.The next category of QA systems such as Domain oriented QA systems and Rule Based QA systems make use of reasoning methods inorder to answer the questions, the popular online QA systems that are comes under this category are Geographic[Chung.H(2004)], Biomedicine[Zweigenbaum.P,(2003)], WEBCOOP[Benamara.F,(2004)] and [Saint-Dizier,(2004)],START[Katz.B,(2002) and Katz. B,(1997)] etc.,.

The two QA systems are compared with characteristics of different dimensions such as techniques used, question that deals with and so on. In **table-1** provides the details of the comparisons of these QA systems.

Table-I characterization of QA systems

DIMENSIONS	QA system based on NLP and IR	QA systems Reasoning with NLP
Technique	Syntax processing, Named Entity tagging and Information Retrieval	Semantic Analysis or high reasoning
Data Resource	Free text documents	Knowledge Base
Domain	Domain Independent	Domain Oriented
Responses	Extracted Snippets	Synthesized Responses
Questions Deals with	Mostly wh- type of questions	Beyond of wh- type of questions
Evaluations	Uses exisisting Information Retrieval	N/A

The remaining part of the paper is organized as follows, **Section 2** describes Web Based QA Systems, **Section 3** discusses IR/IE based QA Systems, **Section 4** deals with Restricted domain QA Systems, **Section 5** deals with Rule Based QA Systems, **Section 6** deals with the classification of various levels in the questioners, **Section 7** focuses on a Comparative Study of QA Systems with various levels of questioners. Finally **Section 8** concludes the paper by giving a brief glimpses into the future directions of research in this area.

2. Web Based Question Answering Systems

With the wide spread usage of internet a tremendous use of data is available, web is one of the ideal source to obtain the information. Web based question answering systems uses search engines (Like Google, Yahoo, Alto Vista etc.,) to retrieve webpage's that potentially contains answers to the questions. Most of these Web based QA systems works for open domain while some of them works for domain oriented also. The web based QA systems such as MULDER [C.Kwok,(2001)], NSIR [D.R.Radev,(2002)] ANSWERBUS [ZhipingZheng,(2002)] falls into the category of domain independent QA systems, while START [Katz,B,(2002) and Katz,B,(1997)] is referred as domain specific QA systems.

The Web Based QA systems mostly handles wh-type of questions such as ***"who killed Mahatma Gandhi"? Or "Which is the longest river in the World".*** This QA system provides answers in various forms like text documents, Xml documents or Wikipedia. The common levels that are used by different web based Question Answering systems architectures are as follows:

- Question Classification: This level provides correct answers by classifying the user query into one of the question type to which it belongs to. The question classification is made to provide better accuracy in the results.
- Answer Extraction : This level extracts the correct plausible answers for different classification of questions.
- Answer Selection:Among the plausible answers obtained, ranking approaches are used to mine the best accurate answers based on its weightage factor.

Different types of web based QA systems uses different properties like snippet tolerant property, world knowledge axioms that are supplied by word net glossaries, as well as rewriting rules representing equivalent classes of linguistic patterns. One of the popular open domain web based QA systems such as LAMP [DellZhang,(2002)] make use of snippet tolerant property to calculate ranking percentages based on which responses are produced.

Dialogue based QA systems [RamiReddyNandiReddy,(2004)],WEBCOOP[Benamara.F,(2004)]and[Saint-Dizier,(2004)], is a cooperative type of web based QA systems which integrates knowledge representation and advanced reasoning procedures to generate cooperative responses. The comparative study of these web based QA systems for its availability of different features is shown in **table-2.**

Most of the Web based QA systems performs a better effort to produce correct answers. Though these systems are mostly capable to handle wh-type of questions, but lacks to produce accurate answers, instead these systems retrieves the relevant passages that contain keywords to extract answer from the knowledge base which

require an additional process to obtain exact answer. This QA systems fails to provide answers to temporal based queries like for instance “When did the X died?” etc.,.

Table-2 comparative study for some web based QA

Features	Answer Bus	Start	Lamp	NSIR
Sentence Parsing	√	√	√	√
Open domain	√	×	√	√
Relation Extraction	×	√	√	√
Explanation of Failure	×	×	√	√
Dynamic Answer generation	×	√	√	√

3. IR / IE Based Question Answering Systems

Most of the IR based QA systems returns a set of top ranked documents or passages as responses to the query. Information Extraction (IE) system uses natural language processing (NLP) systems to parse the question or documents returned by IR systems, yielding the “meaning of each word”. IE systems need several resources like Named Entity Tagging (NE), Template Element (TE), Template relation (TR), Correlated Element (CE), and General Element (GE). IE systems architecture is built into different levels like

- **L**evel 1 NE tagger is used to handle named entity elements in the text(who, when, where, what etc..,) .
- **L**evel 2 handles NE tagging +adj like(how far, how long ,how often etc..,),
- **L**evel 3 builds correlated entities by using the major entity in the question and prepares General Element(GE)which consists of asking point of view. For ***Eg: “Who won the first Nobel Prize in India?***”. By passing this question onto the levels mentioned above ASKING POINT is Person (Noun) KEY WORDS such as won, noble, prize etc.., are retrieved.

The architecture of IE systems consists of two common modules, they are

- ***Question processor*** which takes the question as input and generates asking point for the question which in turn helps to match for the answer in the text.
- ***Text Processor*** retrieves named entities keywords from the text to generate accurate answers. Some of the IR Based systems like AskJeeves, LaSiE system performs text analysis which uses some basic modules like Tokenizer, Sentence splitter, Parse process, Name matcher, Discourse Interpreter [Robert Gaizauskas,(1998)].

The IR/IE based QA systems depends on knowledge base which requires an extension to CE and GE components to handle yes/no types of questions in the text. This systems can answer only wh-type of questions but other than wh-type of questions such as ***“How can I assemble a computer?”,*** remains unanswered.

4. Restricted Domain Question Answering systems

This type of Question answering system requires a linguistic support to understand the natural language text in order to answer the questions correctly. An efficient approach of improving the accuracy of QA system is done by restricting the domain of questions and the size of knowledge base which resulted in the development of restricted domain question answering system (RDQA). RDQA have specific characteristics like “System must be Accurate” and “Reducing the level of Redundancy”. RDQA over comes the difficulties occurred in open domain by achieving better accuracy. Early RDQA systems like LUNAR[Woods.W,(1972)] allows to ask geologist questions about rocks. BASEBALL[Green W,(1961)] is another restricted domain QA system, which can only answer about one season’s Baseball data. These early systems has encoded large amount of domain knowledge in data bases.

RDQAS doesn’t focus on language understanding it focuses on specific set of domain rules. Current RDQA systems are restricted to specific domains like Railways, Medicine, Weather Forecast[Diekema A,(2004)] and Geographic Systems [Chung,H,(2004)] etc,. The RDQA systems uses IE engines which consist of web crawlers and Wrappers, WebCrawler is used to select set of extraction rules that can extract domain information, while Wrappers used to retrieve relevant domain oriented WebPages which contains answers. RDQA first analyzes questions and translates into Structured Query Language (SQL) statements which are further processed to obtain results by retrieving data from database.

Domain Oriented systems make use of domain oriented knowledge bases, domain servers, information systems, parsers etc. Knowledge Acquisition and Access Systems such as KAAS[Anne R,(1999)] uses IR systems to retrieve relevant passages that are processed by NLP. RDQA mostly handles specific questions like for instance Eg**: *"what is the recently invented medicine for cancer ?".*** Domain oriented QA systems contribute great effort to achieve accuracy from the data which is retrieved by the information retrieval. The system requires "Situated evaluations" while comparing with that of open domain QA systems. This type of QA systems needs a domain classifier to handle different domains which is one of the difficult tasks to achieve.

5. Rule Based Question Answering Systems

The rule based QA system is an extension for IR based QA system. Rule Based QA doesn't use deep language understanding or specific sophisticated approaches. A broad coverage of NLP techniques are used in order to achieve accuracy of the answers retrieved. Some popular rule based QA systems such as Quarc [Ellen Riloff,(2003)] and Noisy channel [Abdessamad Echihabi,(2000)] generates heuristic rules with the help of lexical and semantic features in the questions. For each type of questions it generates rules for semantic classes like ***who, when, what, where*** and ***Why*** type questions. **"W**ho" rules looks for Names that are mostly Nouns of persons or things. **"W**hat" rules focuses on generic word matching function shared by all question types it consists of DATE expression or nouns. "**W**hen" rules mostly consists of time expressions only."**W**here" rules are mostly consists of matching locations such as "in", "at', "near" and inside. "**W**hy" rules are based on observations, that are closely matched to the question.

These Rule Based QA systems first establish parse notations and generate training cases and test cases through the semantic model. This system consists of some common modules like IR module and Answer identifier or Ranker Module.

- ***IR module*** **:** It retrieves set of documents or set of sentences that contain answers to a given question and returns the results ranker module..
- ***Ranker Module*** **:** Assigns ranks or scores to the sentences which are retrieved from IR module.
- ***Answer identifier*** **:** This module identifies the answer substrings from sentences depending on the score or rank.

Rule Based QA system approach is a wonderful test-bed for NLP to provide accurate answers. The co-reference resolution requires automatic extraction of semantic knowledge [AbdessamadEchihabi,(2000)] which is a difficult task to achieve.

6. Classification of Questioners Levels

In Question Answering system the questions are classified into different levels based on its context. The questions may be assertive, informative, interrogative or interactive in normal context. The perspectives of these types of questions may vary, but the common goal is to obtain accurate answer from the system. This section presents a classification of different levels of Questioners.

- **CASUAL QUESTIONERS:** In this type of questioners normal questions are posed to the system. Majorly it focuses in normal "context" to handle the questions like Eg:" When did Abraham Lincoln Died?" and "who invented computer?" . All these type of questions are having normal context.
- **TEMPLATE QUESTIONERS:** In this type of questioners, templates are generated for the given question, which focuses on "linguistic" knowledge of the question. For Eg: "How was Abraham Lincoln Died?" and "**Does any specific reason to invent computer?"** all these questions are having a specific reason to be answered for the question.
- **CUBE REPORTER:** In this type of questioners the complex questions are breaked into small set of questions. It majorly consists of context and specific relations to answer the questions of this type. The QA system needs to search answers from multiple sources which lies beyond the database search. It can answer the questions like Eg**: "Does any specific actions occurred by US government after Lincoln's death?". c**ube reporter generates small set of questions which are related to the main question, that are Eg: "**When did Abraham Lincoln died?" "What was his citizen ship?" and "what was released by US government after Lincoln's death?".**
- **PROFESSIONAL INFORMATION ANALYST:** The questions of this type are having future perspectives. It identifies different taxonomies and multiple facts which are involved in the questions, but it requires much reasoning techniques for answering, the questions like Eg**: "What are the actions done by US government after Lincoln's death to honor their president?" .**

7. Comparative study of Questioner Levels with QA Systems

A Comparative study of different types of Question Answering systems are mapped on to the various levels of Questioners, which are listed in **Table-3(mentioned after references).**

Table-3 Comparative study of questioner levels with QA systems

"Casual Questioner"	**Focuses**: Normal context. **Property:** Snippet tolerant or keyword based etc.., **Responses**: all possible answers in different forms from Knowledge Base.	**Focuses:** Normal context. **Property**: IR retrieval from documents or passages **Responses:** all possible answers from document pool.	**Focuses:** Normal context. **Property**: specific domain servers **Responses:** named entity tagging with corresponding domain	**Focuses:** Normal context. **Property**: specific rules structures. **Responses:** in different forms from IR
"Template Questioner"	**Focuses**: linguistic concern in generated templates. **Property:** Snippet tolerant or keyword based etc.., **Responses**: all possible answers from different Knowledge Bases	**Focuses**: linguistic concern and correlated entities **Property:** NLP techniques to parse the questions **Responses**: large passages as answers which are top ranked from IR.	**Focuses:** linking of generated templates with that domain **Property**: NE tagging with the domain corpus **Responses:** answers in the asked domain	**Focuses:** rules for each template **Property**: parse notations **Responses:** in different forms from IR
"Cube Reporter"	**Focuses**: resolves the ambiguities in set of question. **Property:** Snippet tolerant or keyword based etc.., **Responses**: all possible answers from different "ontologies".	**Focuses**: ambiguity resolutions. **Property:** Named Entity tagging and CE to relate entities. **Responses**: large no.of documents or passages	**Focuses:** linking of generated templates with that domain **Responses:** doesn't relates with different domains only provides from single source	**Focuses:** rules for each semantic clauses **Property**: heuristic and semantic rules **Responses:** in different forms from IR
"Professional Information analyst"	**Focuses**: future constraint type of questions. **Responses:** Fails to handle this type of questions. Need temporal knowledge .	**Focuses**: Future perspectives **Property:** Named Entity tagging and CE and GT **Responses**: very less responses some times fails to give response.	This type of questions are not handled because high level of semantic interpretation is required.	Very difficult to handle these type of questions because high level of knowledge base is required.
"Error rate"	Proportional to the no.of responses produced for the question.	More, because answers are scattered from different sources.	Very less because small set of data sources	Depends on generated rules structure
"Precision and Recall"	High	Low	High	N/A

8. Conclusion

Question answering system is one of the emerging areas of research in natural language processing applications of Artificial Intelligence. QA systems aim to produce accurate answers, but the current QA systems are succeeded only to some extent. The survey of different QA systems has shown that there are different aspects in answering the questions, while temporal aspect has got little attention. As the data changes in the dynamic world it requires accessing the information and facts which are true at current point of time. Hence forth, it shows that answering temporal based questions is essential in the present QA systems, answering the questions like for Eg "*Who is the president of India during 1950-52?" as* remained unsolved in the present QA systems. The survey of various QA system models has given a novel idea to develop temporal Based Question Answering system that can answer different types of temporal queries.

References

[1] Anne R. Diekema, Ozgur Yilmazel, and Elizabeth D. Liddy,"Evaluation of Restricted Domain Question-Answering Systems".(1999) Center for Natural Language Processing.

[2] Abdessamad Echihabi and Daniel Marcu "A Noisy-Channel Approach to Question Answering", (2000).

[3] Benamara.F.,"Cooperative Question Answering in Restricted Domains: the WEBCOOP Experiment",2004. In Proceedings of the ACL Workshop on Question Answering in Restricted Domains.

[4] Benamara,F.and Saint-Dizier,P.,Advanced Relaxation for CooperativeQuestion Answering,(2004).In New Directions in Question Answering. MIT Press.

[5] C. Kwok, O.Etzioni, D. Weld."Scaling Question Answering to the Web". In Proceedings of WWW10, Hong Kong, 2001.

[6] Chung,H.,Han, K., Rim, H., Kim, S., Lee, J., Song, Y. & Yoon, D. "A Practical QA System in Restricted Domains" (2004). In Proceedings of the ACL Workshop on Question Answering in Restricted Domains.

[7] Dell Zhang and Wee Sun Lee."A Web-based Question Answering System" (2002) ,October 31.

[8] D.R.Radev,W.Fan,H. Qi, H. Wu and A. Grewal."Probabilistic Question Answering from the Web",(2002), In Proceedings of the 11th World Wide Web Conference,Hawaii.

[9] Diekema A.R, Yilmazel Ozgur, and Liddy E.D. " Evaluation of Restricted Domain Question-Answering Systems" (2004).In Proceedings of the ACL2004 Workshop on Question Answering in Restricted Domain ,p.p 2-7,.

[10] Ellen Riloff and MichaelThelen."A Rule Based Question Answering System for Reading Comprehension Tests",(2003){rioloff,thelen}.

[11] Green W, Chomsky C, and Laugherty K..BASEBALL: An automatic question answerer.(1961). Proceedings of the Western Joint ComputerConference, p.p. 219-224.

[12] Hermjakob.U."Parsing and Question Classification for Question Answering" (2001). In Proceedings of the ACL Workshop on Open-Domain Question Answering.

[13] Katz, B. & Lin.J., START and Beyond.(2002). In Proceedings of the 6th World Multi conference Systemics, Cybernetics and Informatics,.

[14] Katz.B.,"Annotating the World Wide Web using Natural Language" (1997).In Proceedings of the 5th Conference on Computer Assisted Information Searching on the Internet..

[15] Rohini SrihariCymfony and WeiLi Cymfony "A Question Answering System Supported By Information Extraction" (2002) Inc.wei@cymfony.com.Inc.rohini@cymfony.com..

[16] Rami Reddy Nandi Reddy, Sivaji Bandyopadhyay "Dialogue based Question Answering System in Telugu", (2004). Dept. of Comp. Sc. & Engg, Jadavpur University, Kolkata..

[17] Robert Gaizauskas and Kevin Humphreys "A Combined IR/NLP Approach to Question Answering Against LargeText Collections". (1998),University of Sheffield UK .

[18] S. Brin and L. Page."The anatomy of a large-scale hypertextual (Web) search engine" (1998).In proceedings of the 7th InternationalWorldWideWeb conference (WWW7)/Computer Networks, 30(1-7),p.p. 107—117.

[19] Woods W.A, Kaplan R.M, and Webber B.N. "The Lunar Sciences Natural Language Information System", (1972). Final Report. BBN Report 2378, Bolt Beranek and Newman Inc., Cambridge, Massachusetts.

[20] Zhiping Zheng "AnswerBus Question Answering System" (2002). In Proceeding of HLT Human Language Technology Conference.San Diego,CA. HLT March,.pp.24 –27.

[21] Zheng, Z., "Developing a Web-based Question Answering System" (2002). In Proceedings of the 11th International Conference on World Wide Web.

[22] Zweigenbaum, P., "Question Answering in Biomedicine" (2003). In Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics.