Journal of Network Communications and Emerging Technologies (JNCET) www.jncet.org
Volume 5, Special Issue 2, December (2015)

A Comparison Study of Question Answering Systems

Poonam Bhatia
Department of CE, YMCAUST, Faridabad.

Rosy Madaan
Department of CSE, GD Goenka University, Sohna-Gurgaon

A.K. Sharma
Department of CSE, BSAITM, Faridabad

Ashutosh Dixit
Department of CE, YMCAUST, Faridabad

Abstract – Question answering system aims to develop techniques that can go beyond the retrieval of relevant documents in order to return exact answers to natural language questions posed by the user. Answering natural language questions requires more complex processing of text than employed by current information retrieval systems. A number of question answering systems have been developed which are capable of carrying out the processing required to achieve high levels of accuracy. Existing Question Answering system are unable to handle variety of questions and reasoning based question. Also, because of the lack of the data sources, the QA system fails to answer the question. In this paper, a comparative analysis of several Question Answering Systems is presented.

Index Terms – Search Engines, Question Answering System, and information retrieval

This paper is presented at International Conference on Recent Trends in Computer and information Technology Research on 25th & 26th September (2015) conducted by B. S. Anangpuria Institute of Technology & Management, Village-Alampur, Ballabgarh-Sohna Road, Faridabad.

1. INTRODUCTION

Search Engines answers the query of a user is in the form of linked pages but if the user enters a question, he/she wants the exact answer instead of pages. Answer should be a text containing the most accurate result to the Query. In quickly changing world Search Engines need to be modified to give the exact answer. This criterion is limited by the assumption that the answer can be found because it uses the question words. It uses existing information retrieval system or search engine can search few keywords, i.e. document retrieval; which gives only relevant documents that contain the keyword. Question answering system aims to develop techniques that can go beyond the retrieval of relevant documents in order to return exact answers to natural language questions. Answering natural language questions requires more complex processing of text than employed by current information retrieval systems. A number of question answering systems have been developed which are capable of carrying out the processing required to achieve high levels of accuracy.

Question Answering (QA) systems extract answers from large text collections by a) Classifying the answer type they expect. b) Using question keywords or patterns associated with questions to identify candidate answer. c) Ranking the candidate answers to decide which passage contains the exact answer. Traditional Question Answering system assumes that the answer can be found because it uses the question words. It uses existing information retrieval system or search engine can search few keywords, i.e. document retrieval; which gives only relevant documents that contain the keyword .Search engines cannot differentiate between the variable documents and spams.Some search engine crawler retrieve only document title not the entire text in the document.

2. GENERAL ARCHITECTURE OF QUESTION ANSWERING SYSTEM

In general a Question Answering system comprises of the following modules, see Figure 2.1:

2.1. Query Interface

2.2. Question processing
2.3. Question Analysis
2.4. Question Classification
2.5. Question Reformulation
2.6. Document processing
2.7. Answer Reformulation

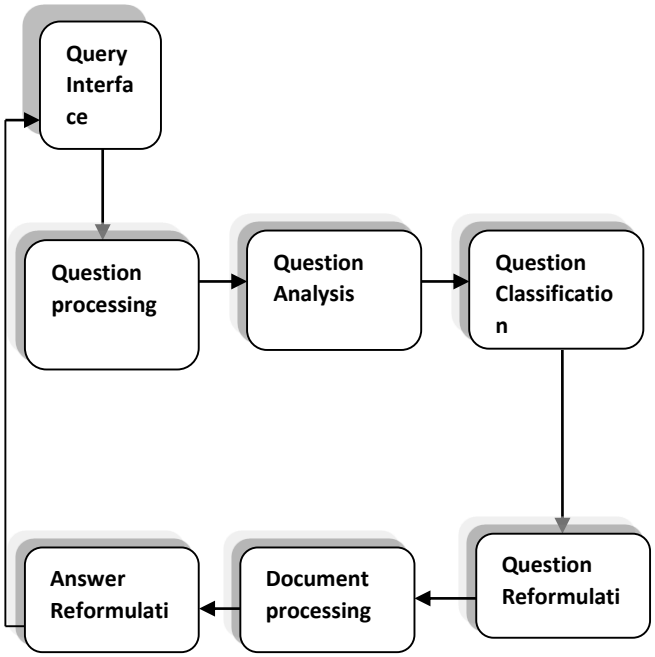

FIGURE 2.1 GENERAL ARCHITECTURE OF A QA SYSTEM

This Question answering system can automatically answer questions about programming technology information that are asked by a user using natural language of English. The fundamental architecture of this system is shown in Figure 2.1 and consists of the Query interface, Query processing, Document retrieval processing, Result processing. The Query interface allows providing a question into the Question Answering system that the user's natural language can be entered into and the output is given to the user. All these modules constitute the whole process and they are most important building blocks of the whole system. All are dependent on each other and none of them is less useful as compared to the other. Query interface is use to accept the query from the user and returns the desired output to the user. When the user enters the query the Query is given to Query processing. Analyze the question, in order to represent the main information that is required to answer the user's question. Classify the question type, usually based on category of questions already coded into the system. Reformulate the question, in order to enhance the question phrasing and to transform the question into queries for the information retrieval. Then Question analysis is done also referred to as "Question Focus". Classifying the question and knowing its type is not enough for finding answers to all questions. After analyzing the Question classification is done in order to correctly answer a question, it is required to understand what type of information the question asks for, because knowing the type of a question can provide constraints on what constitutes relevant data (the answer), which helps other modules to correctly locate and verify an answer.

The document processing module in QA systems is also commonly referred to as paragraph indexing module, where the reformulated question is submitted to the information retrieval system, which in turn retrieves a ranked list of relevant documents. The document processing module usually relies on one or more information retrieval systems to gather information from a collection of document. Finally the answer processing module is responsible for identifying, extracting and validating answers from the set of ordered paragraphs passed to it from the document processing module.

3. LIMITATIONS OF VARIOUS QUESTIONS ANSWERING SYSTEM

The performance of a Question Answering system is firmly tied with the complexity of questions asked and the difficulty of answer extraction. For example, in TREC many systems were quite successful at providing correct answers to simpler, fact seeking questions, but failed to answer questions that required reasoning or linguistic analysis. Q1:"Where is University of Texas?" On the other hand, none could find a correct answer to complex questions such as Q2: "What is the difference between AM radio stations and FM radio stations?" Since performance is affected by the complexity of question processing, we first provide a broad taxonomy of systems.

Question category (Where, What, Who, etc.) provide one of the simplest and yet useful classifications of factual questions. First, they differentiate between declarative ("In 1966, you could rent a car for $2 a day") and interrogative statements ("How much could you rent a car for in 1966?"). Second, they are used in many QA system implementations as (coarse) clues in the identification of the expected answer type. Thus Who questions often (but not always!) ask about Person names, Where questions refer to location. The large majority of TREC questions have the What, Who, Where, and When question stems, with What questions constituting more than half of the evaluation questions. In turn, What questions can be often subcategorized into one of the other question stems. For instance "What was the name of the First Prime Minister?"

corresponds to Who; similarly "What country is the leading manufacturer of copper?" corresponds to Where. The most common What subcategories are Where (17%), modified How (How far, How many, How much etc.) (8%), and Who (5%).

The answer accuracy has a wide variation across the question stems. At the higher end, Where questions are answered with a precision score of 0.657.With the exception of two questions, the system finds exactly one question stem for each question. The system fails to properly identify the question stem for Q3:"Can you give me the name of a jewellery maker in London, England?" Q4:"The U.S. Department first issued paper currency for the U.S. during which war?"

4. QUESTION ANSWERING SYSTEMS:

A.SHRDLU QUESTION ANSWERING SYSTEM

It was a highly successful question-answering program. It was developed by Terry Winograd .It program the operation in a robot and it offered the feasibility to ask the robot questions about the state of the world. SHRLDU was one of the first AI system to perform a realistic task. SHRLDU like dialogs are very plausible when querying database. SHRLDU allows mapping from lexical categories to meaning elements and from syntactic structure to rules for combining meaning elements. For example "destruction" is a noun but refers to an activity.

The stability of this system was the choice of a very specific domain and a very simple world with rules of physics that were easy to encode in a computer program. In the 1970s, knowledge bases were developed that targeted limited domains of knowledge. The QA systems developed to interface with these expert systems produced more repeatable and valid responses to questions within an area of knowledge. Expert systems depend largely on expert-constructed and organized knowledge bases, whereas many modern QA systems rely on statistical processing of a large, unstructured, natural language text corpus. The system answered questions related to the UNIX operating system. It had a knowledge base of its domain, and its target is to phrase the answer to accommodate various types of users.

B.LILOG (TEXT UNDERSTANDING SYSTEM)

It is a close domain Question Answering System It is basically a text understanding system. This system gives tourism information in a German city. Other system also helps the system in linguistic and computational processing.

C.QUALM (STORY UNDERSTANDING SYSTEM)

It works through asking questions about simple, paragraph length stories .QUALM system includes a question analysis module that links each question with a question type. This question type guides all further processing and retrieval of information.

Table 4.1 QUALM Question Categories

Question type	*Example question*
causal antecedent	*How did the glass break?*
Goal orientation	*Mary left for what reason?*
causal consequent	*What happened after John left?*
Disjunctive	*Was John or Mary here?*
Verification	*Did john leave*
instrumental/procedural	*How did John go to New York?*

D.KUPIEC (A SIMPLE WH QUESTION MODEL)

This Question Answering System performs similar function but it rather solves simpler WH-question models to build a QA system. This QA used the interrogative words for informing the kinds of information required by the system.

Table 4.2 Kupiec Question Categories

Question type	*Answer type*
Who/Whose	*Person*
What/Which	*Thing, Person, location*
Where	*Location*

A. ANSWERBUS QUESTION ANSWERING SYSTEM

AnswerBus is an open-domain question answering system based on sentence level Web information retrieval. It accepts users' natural-language questions in English, German, French, Spanish, Italian and Portuguese and provides answers in English. It can respond to users' questions within several

seconds. Five search engines and directories (Google, Yahoo, WiseNut, AltaVista, and Yahoo News) are used to retrieve Web pages that contain answers.

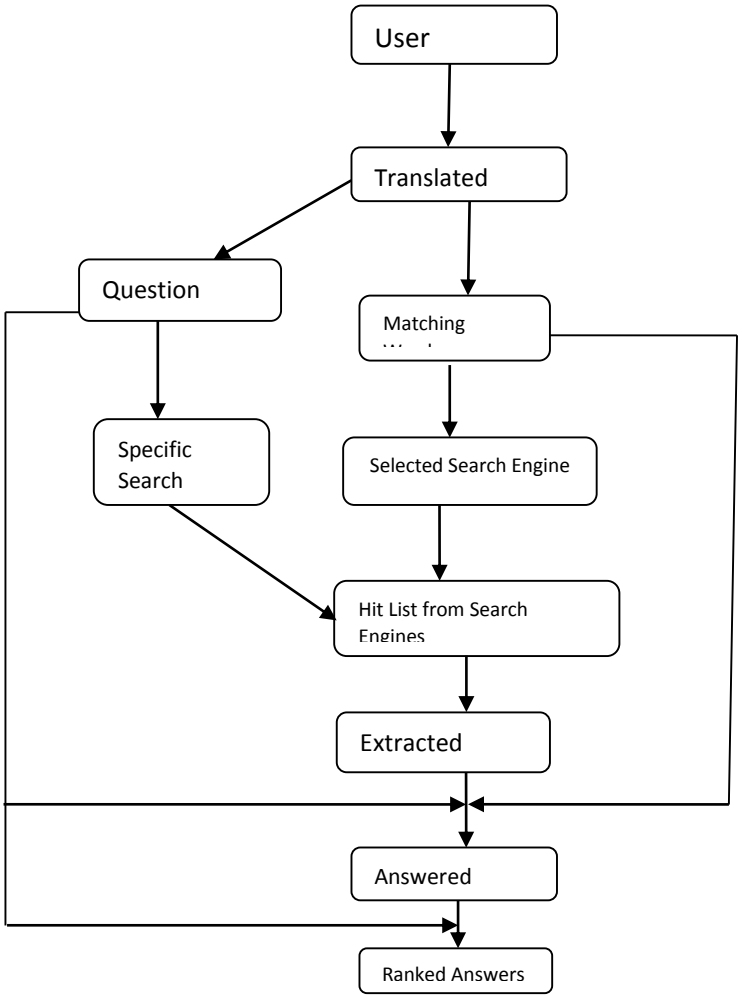

FIGURE 2.2 ANSWERBUS QUESTION ANSWERING SYSTEM

B. ASKMSR QUESTION-ANSWERING SYSTEM

The approach of AskMSR is to take the query as input and rewrites the Query in to the form the system can support. The rewrites generated by the system are simple string-based manipulations. AskMSR do not use a parser or part-of-speech tagger for query reformulation, but do use a lexicon for a small percentage of rewrites, in order to determine the possible parts-of-speech of a word as well as its semantic variation. It created the rewrite rules and associated weights manually for the current system, it may be possible to learn query to answer reformulations and their weight.

5. COMPARITIVE ANALYSIS OF DIFFERENT QUESTION ANSWERING SYSTEM

Question Answering System	Domain	Technique used	Description
START Question Answering	Close Domain	Natural Language Annotation	Start stands for syntactic analysis using Reversible Transformation is the first world web based QA. The Query entered by the user is in natural language and then the Query is matched with the parse tree against it knowledge base and presents the Appropriate answer. Technique used natural language Annotation to extract information from different sources.
MUDLER Question Answering	Open Domain	Natural language parsing and a novel voting procedure	It is the first general-purpose, fully-automated question-answering system available on the web which relies on multiple search-engine queries, natural-language parsing, and a novel voting procedure to yield reliable answers couple.MULDER's recall is more than a factor of three higher than that of AskJeeves. In addition, we find that Google requires 6.6 times as much user effort to achieve the same level of recall as MULDER.
AskJeeves Question Answering System	Open Domain	Direct user to the relevant Web pages.	It uses the natural language parser to solve the question of a Query. It does not provide direct and exact answer to the user .It direct the User to the relevant WebPages just as the Search Engine does. It has done only half of the job for QA.
TEXT RETRIEVAL CONFERENCE Question Answering System	Domain Independent	Effort to find Snippets of text for exact answers	The goal of this QA is to retrieve all the Snippets of text that contains the exact answers instead of the documents returned by text retrieval system.
Answer Bus Question Answering System	Open Domain	Based on Sentence level Web information retrieval	It accepts users' natural-language questions in English, German, French, Spanish, Italian and Portuguese and provides answers in English. Five search engines and directories are used to retrieve Web pages that are relevant to user questions. From the Web pages, AnswerBus extracts sentences that are determined to contain answers.

SHRLDU Question Answering System	Close Domain	Rule of physics encoded in computer program	Expert systems depend largely on expert-constructed and organized knowledge bases, whereas many modern QA systems rely on statistical processing of a large, unstructured, natural language text corpus. The system answered questions related to the UNIX operating system. It had a knowledge base of its domain, and its target is to phrase the answer to accommodate various types of users.
LILOG Question Answering System	Close Domain	Linguistic and computational processing	It is basically a text understanding system. This system gives tourism information in a German city. Other system also helps the system in linguistic and computational processing.
QUALM Question Answering System	Story understanding System	Based on Information Retrieval System	It works through asking questions about simple, paragraph length stories .QUALM system includes a question analysis module that links each question with a question type. This question type guides all further processing and retrieval of information
KUPIEC Question Answering System	Who Question Solver	Based on simple IR technique	This Question Answering System performs similar function but it rather solves simpler WH-question models to build a QA system. This QA used the interrogative words for informing the kinds of information required by the system
AQUALOG Question Answering System	Ontology based	Based on Ontology concept	Aqualog is an ontology - based question answering system that process input queries and classifies them in to 23 categories defined by the system. It takes ontology as an input and return answers drawn from one or more knowledge bases which illustrate the input ontology with domain specific information It uses Word Net to make use of user queries with respect to knowledge base.
TEXTRETRIEVAL CONFERENCE TREC-8	Open Domain	Based on fact based short question concept	TREC 8 has had a question answering track since 1999; in each track the task was defined such that the systems were to retrieve small snippets of text that contained an answer for open-domain, closed-class questions (i.e., fact-based, short-answer questions that can be drawn from any domain)

TREC 2006	Close domain	Based on ontology.	TREC 2006 track focused on retrieval of short passages that answered a biological question, while providing users the textual context of the retrieved answers. The main idea is to save user's time usually spent in manually locating the answer of interest in the full text and user decide whether or not a document is relevant by providing the user with supporting evidence to the "answer".
MORPHEUS(Question Answering System	Open Domain	Based on hidden web concept	Morpheus answers questions by using methods from prior successful searches. The system ranks stored methods based on a similarity defined on assigned classes of queries. Morpheus focuses on the deep (or hidden) web to answer questions because of the large stores of quality information provided by the databases that support it Web forms act as an interface to this information. Morpheus employs user exploration through these web forms to learn the types of data each deep web location provides.

5. CONCLUSIONS

On submitting a query to a search engine, it provides a list of result pages in ranked order. If the user provides a question instead of a query, then he/she is interested in a precise answer and not a complete web page to go through. This calls for the need of a Question answering system. In past, several Question answering systems have been developed. The paper presents details of some existing Question answering systems and provides a comparative analysis of them in terms of domain, technique used and description.

REFERENCES

[1] E. Brill, J. Lin, M. Banko, S. Dumais, and A. Ng. 2001. Data Intensive Question Answering. Proceedings of the TREC 10 Conference NIST, Gaithersburg, MD, 183–189.
[2] R. Gaizauskas, K. Humphreys, 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).
[3] N.J. Belkin, A. Keller, D. Kelly, J. Perez-Carballo, C. Sikora*. Support for Question-Answering in Interactive Information Retrieval: Rutgers. TREC-9 Interactive Track Experience. Y. Sun School of Communication, Information & Library Studies.
[4] A. Lampert, 2004. A Quick Introduction to Question Answering. CSIRO ICT Centre.
[5] Ali Mohamed Nabil Allam1 and Mohamed Hassan Haggag2. Scaling Question Answering to the Web.
[6] Zhiping Zheng . AnswerBus Question Answering System.
[7] Johannes Leveling. A Comparative Analysis: QA Evaluation Questions versus Real-world Queries.
[8] Rohini Srihari,wei li. A Question Answering System Supported by Information Extraction.
[9] Bassam Hammo Hani Abu-Salem Steven Lytinen. QARAB: A Question Answering System to Support the Arabic Language.