A Review on Novel Scoring System for Identify Accurate Answers for Factoid Questions

[1]Harpreet Kaur, [2]Rimpi

[1, 2]Swami Vivekanand Institute of Engineering and Technology
Banur, Punjab, India

Abstract: *In this research work we have develop a new scoring mathematical model that works on the five types of questions. The question text failures are first extracted and a score is found based on its structure with respect to its template structure and then answer score is calculated again the question as well as paragraph. Text to finally reach at the index of the most probable answer with respect to question.*

Keywords: Natural language processing, Question answering System, Information retrieval.

1. Introduction

NLP focuses on communications between computers and natural languages in terms of theoretical results and practical applications, and on information sharing now that information is exchange as it never has been before and sharing information becoming the leading theme in the domain of NLP systems. Question Answering (QA) systems go beyond the usual Information Retrieval (IR) systems which underlay popular Internet search engines. QA systems have the aim of responding to natural language questions whereas IR systems take up keywords from users and deploy some intelligent search mechanisms on a document collection to get back to the user with a ranked list of documents rather than an exact answer.

There is a need for tools that would reduce the amount of text in order to obtain the desired information [1] [2] [3]. People have questions and they need answers, not documents and they always prefer to express the questions in their native language without being restricted to a particular query language, query formation rules, or even a particular knowledge domain. Automatic question answering system will help for the above technology. In this Question Answering System consists of three distinct phases: Question classification, information retrieval or document processing and answer extraction.

The design of a standard QA system assumes that the language in which the question is asked and the text collection available to be processed are all in the same language. However, there might be a need for cross-lingual QA system which takes in questions in one language and searches through a document collection in a different language to get to the answer. English QA system research attempts to deal with a wide range of question types like WHEN, WHERE, WHAT, HOW, WHOM, WHY & WHOSE. Thus the aim of a QA system is to localize the exact answer to a question from a structured or a non-structured collection of texts. The QA system states that a QA task can be decomposed into three main sub problems. The sub problems are:- Question processing, information retrieval or document Processing and answer processing and. The question processing stage is responsible for taking a question in a natural language and producing some kind of intelligent representation of the raw question string so that it

becomes more useful for finding answers. The document processing stage is used to reduce the search space of the document collection where the answer to the question can be expected. This stage is basically a complete Information Retrieval system where the idea is to take in some keywords and produce a ranked list of documents related to those keywords. The final stage of a QA system is the answer processing stage where the system does some intelligent matching with the output of the previous two stages to produce an answer to the given question. Any QA system should have these four basic components and may have a number of other components to make the system more useful and robust.

1.1 What is Question and Answer System?

Questions answering (QA) systems look for the answer of a question in a large collection of documents. The question is in natural language. QA systems select text passages. Then, after that the answer is extracted from these passages, according to criteria issued from the question analysis. To facilitate the question generation task, we build text from the input complex text using a syntactic parser. We classify the text based on their subject, verb, object and preposition for determining the possible type of questions to be generated. The ability of QA systems to recognize a great amount of answer types is related to their powerfulness for extracting right answers [4] [5] [6].

1.2 Why do we need QSA System?

Question Answering (QA) Systems allow the user to ask questions in a natural language and obtain an exact answer. In this, we tried to learn the important issues in the field of Question Answering (QA) systems. We peeked into the internals of many established QA systems. we do not only consider simple questions but text problems consisting of several sentences. Our approach to translating the natural language question uses an underlying corpus and the knowledge base to derive meaningful and relevant patterns which can then be used to process the questions and capture their meaning with respect to the underlying knowledge base.

2. Types of QA systems

Different types of QA systems which are divided into two major groups based on the methods used by them. First group of QA system belongs to simple natural language processing and information retrieval methods, while another group of QA systems are dependent upon the reasoning with natural language.

2.1 Web Based Question Answering System

This web bases QA system submits the question to the search engine like Google, Yahoo etc and grabs its top 100 search results..Given a user's natural language question, the system will submit the question to a search engine, then extract all possible answers from the search results according to the question type identified by the question classification module, finally select the most possible answers to return. The Web Based QA systems mostly handles wh-type of questions such as "Who was the first American in space?" **Or "Which of the following is correct".** 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 [7]:

- **Question Classification:-** In order to correctly answer a question, usually one needs to understand what type of information the question asks for, e.g., the sample question "Who was the first American in space?" asks for a person name. The question classification is made to provide better accuracy in the results.
- **Answer Extraction: -** In this, extracts the correct possible answers for different classification of questions.
- **Answer Selection:-** Among the possible answers obtained, ranking approaches are used to find out the best accurate answers based on its weight age factor. Answer classes generally is of factoid and non - factoid types. The factoid is getting short fact based answers like names, dates, and non-factoid is getting descriptions or definitions [9]. Given a user's natural language question, the system will submit the question to a search engine, then extract all possible answers from the search results according to the question type identified by the question classification module, finally select the most similar answers to return. The architecture of web based question answering system is shown in figure 2[8].
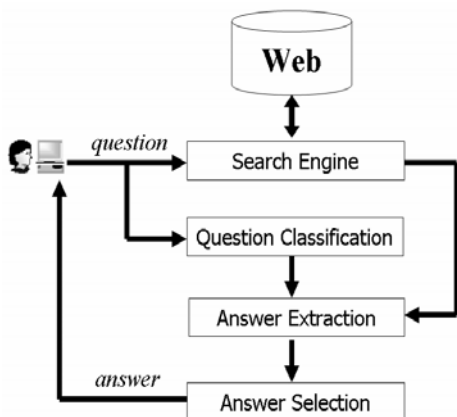

Figure 1: Architecture of Web based question answering system

2.2 IR / IE Based Question Answering Systems

Question Answering, the process of extracting answers to natural language questions is profoundly different from Information Retrieval (IR) or Information Extraction (IE). IR systems allow us to locate relevant documents that relate to a query, but do not specify exactly where the answers are. In IR, the documents of interest are fetched by matching query keywords to the index of the document collection. 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 build into distinct levels:

- **L**evel 1 NE tagger is use 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 the correlated entities by using the most important entity in the question and prepares General Element (GE) which consists of asking point of view. For *Eg: "How did John pass the exam?"* The ASKING POINT is clearly defined i.e. Person (Noun) if we by passing this question into the separate levels which was mentioned above.
- The Architecture of IR/IE based question answering system is given in figure3
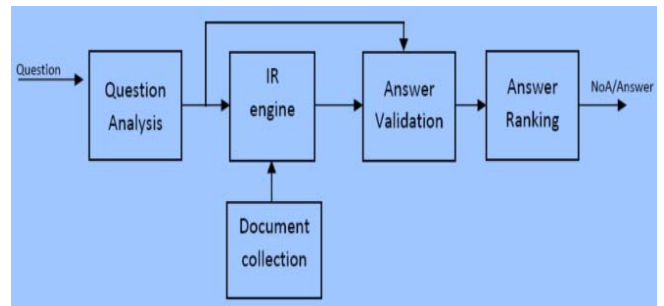

Figure 2: Architecture of IR/IE based question answering

2.3 Restricted Domain Question Answering systems

QA systems for restricted domains may be designed to retrieve answers from so-called unstructured data (free texts), semi-structured data (such as XML-annotated texts), or structured data (databases). Question answering applied to restricted domains is interesting and challenging in two important respects. Question answering on restricted domains requiring the processing of complex questions and offering the opportunity to carry out complex analysis of the text sources and the questions [3] [10]. The main difference between open-domain question answering and restricted-domain question answering is the existence of domain-dependent information that can be used to improve the accuracy of the system.

2.4 Rule Based Question Answering Systems

The rule-based system uses lexical and semantic heuristics to look for evidence that a sentence contains the answer to a question. Each type of WH question looks for different types of answers, so Quarc uses a separate set of rules for each question type (WHO, WHAT, WHEN, WHERE, WHY,WHOSE,WHOM). Given a question and a story,

Quarc parses the question and all of the sentences in the story using our partial parser Sundance. Much of the syntactic analysis is not used, but Quarc does use the morphological analysis, part-of- speech tagging, semantic class tagging, and entity recognition. The rules are applied to each sentence in the story, as well as the title of the story, with the exception that the title is not considered for WHY questions. **"Wh**o" rules looks for Names that are mostly Nouns of persons or things. The "**Wh**at" questions were the most difficult to handle because they sought an amazing variety of answers and it consists of DATE expression or nouns. "**Wh**en**"** questions almost always require a TIME expression, so sentences that do not contain a TIME expression are only considered in special cases. The "**Wh**ere" questions almost always look for specific locations, so the WHERE rules are much focused. "**Wh**y" questions are handled differently than other questions. The WHY rules are based on the observation that the answer to a WHY question often appears immediately before or immediately after the sentence that most closely matches the question. We believe that this is due to the causal nature of WHY questions. The "**Wh**ose/**Wh**om" these questions usually ask about an individual or an organization. The 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.

3. Application of QAS

Question answering has many applications. We can subdividing these applications based upon the source of the answers: structured data (databases), semi-structured data (for example, comment fields in databases) or free text. We can further distinguish among search over a fixed set of collections, as used in TREC (particularly useful for evaluation); search over the Web, search over a collection or book, e.g. an encyclopaedia or search over a single text, as done for reading comprehension evaluations[11]. Another application is in education can also find uses for Question Answering in fields where there are frequently asked questions that people wants to search.

4. A Review on Methodology

First, we collect the articles from encyclopaedia. These articles bank created for extraction of text. When the full extraction system generated multiple outputs from an input sentence and text, we randomly sampled one of them.
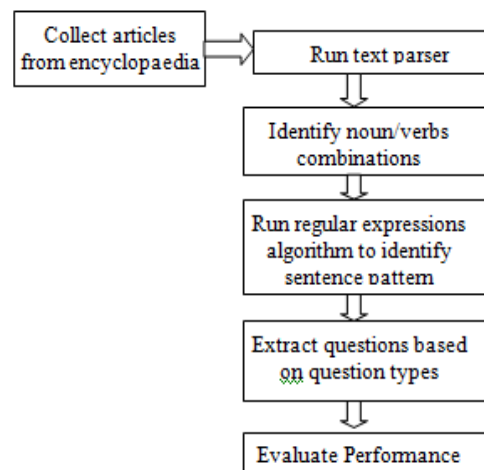

Figure 3: Flow of Question Answer

The run text parser is that in which we can take text as a input and break it up into meaning components in some way. After that using text parsing algorithm identify noun/verbs combination to develop questions. The last but not least regular expressions algorithm is used to specify the pattern that provides concise and flexible means to match. Strings of text such as, particular character, words or patterns of character. Our main purpose of a Question Answers System (QAS) is to find out who did what to whom, where, when, how why and whose? And after that the answers are extracted for the questions of types (What), (Where), (When), (Who), (Why) & (Whose). At the end based on these questions the performance is evaluated.

5. Conclusion

In this paper we discussed some of the approaches used in the existing QA system and proposed a new architecture for QA system retrieves the exact answer. It presents a method checking that an answer is of the specific type expected by the question. It can be used to improve question answering system by checking all returned answers. However, it cannot be used alone to select the good answer. Answering system has become an important component of the online education platform. 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.

References

[1] Haque, Nafid. and Rosner, Mike. A prototype framework for a Bangla question answering system using translation based on transliteration and table look-up as an interface for the medical domain. University of Malta Gertjan Van Noord, University of Groningen

[2] Ashish Kumar Saxena, Ganesh Viswanath Sambhu, L. Venkata Subramaniam*, Saroj Kaushik, "IITD-IBMIRL System for Question Answering using Pattern Matching, Semantic Type and Semantic Category Recognition" OCT 2007.

[3] Poonam Gupta & Vishal Gupta(2012)" A Survey of Text Question Answering Techniques" International Journal of Computer Applications (0975 – 8887)Vol No. 2012

[4] Arnaud Grappy, Brigitte Grau"Answer type validation in question answering systems"Le centre de hautes etudes internationals dtnnformatique documentaire Paris, France, France ©2010

[5] S. M. Harabagiu, M. A. Pa_sca, and S. J. Maiorano. Experiments with open-domain textual question answering. In Proceedings of the 18th conference on Computational linguistics, Morristown, NJ, USA, 2000. Association for Computational Linguistics

[6] E. Hovy, L. Gerber, U. Hermjakob, C.-Y. Lin, and D. Ravichandran. Toward semantics-based answer pinpointing. In HLT '01: Proceedings of the _rst international conference on Human language technology research, Morristown, NJ, USA, 2001. Association for Computational Linguistics

[7] Guda, Vanitha., Sanampudi, Suresh. Kumar. And Manikyamba, I.Lalkshmi ,"Approaches For Question Answering Systems" , Vanitha Guda et al. / International Journal of Engineering Science and Technology (IJEST) ISSN : 0975-5462 Vol. 3 No. 2011. 990-995

[8] Rodrigo, Alvaro. Perez-Iglesias, joaqum., Penas, Anselmo., Garrido, Guillermo and Araujo,Lourdes. A Question Answering System based on Information Retrieval and Validation

[9] Quarteroni, S. and Manandhar S. "Designing an Interactive Open-Domain Question Answering System". Journal of Natural Language Engineering 1. 1-23.

[10] Anette Frank, Hans-Ulrich Krieger, Feiyu Xu, Hans Uszkoreit, Berthold Crysmann, Brigitte Jörg, Ulrich Schäfe" Question answering from structured knowledge sources" German Research Center for Artificial Intelligence, DFKI, Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany, Journal of Applied Logic 5 (2007) 20–48

[11] Gupta, Vishal and Lehal, Gurpreet S. "A Survey of Text Mining Techniques and Applications". Journal of Emerging Technologies in web Intelligence, VOL. 1, No. 1.

Author Profile

Harpreet Kaur is currently perusing M. Tech in Computer Science and Engineering from Swami Vivekanand Institute of Engineering & Technology, Banur, Punjab. She holds the degree of B. Tech in Computer Science and Technology from Baba Banda Singh Bahadur Engineering and Technology, Fathegarh sahib, Punjab.

Er. Rimpi is currently working as Assistant Professor in Computer Science and Engineering Department at Swami Vivekanand Institute of Engineering and Technology, Banur. She has completed her M. Tech in Computer Engineering from Guru Nanak Dev University, Amritsar, Punjab in 2011. She holds the degree of B. Tech in Computer Science and Technology from Guru Nanak Dev University, Amritsar, Punjab in 2009.