Discourse Based Advancement On Question Answering System

Soumya Santhosh[1] and Jahfar Ali[2]

[1]Department of Computer Science and Engineering, MES College of Engineering
Kuttipuram, India
sou.santhu@gmail.com

[2]Department of Computer Science and Engineering, MES College of Engineering
Kuttipuram, India
jahfar.ali@gmail.com

ABSTRACT

Question-answering (QA) research emerged as an attempt to tackle the information-overload problem. Question answering systems take users natural language questions and locate answers from large collections of documents. The current existing Question-Answering systems can deal with shallow questions (factual questions) somehow easily and correctly, but asking deep questions (non-factual questions or complex questions) such as Why-questions are more difficult than asking shallow questions. However, these complex questions cannot be neglected as input for a QA system as they comprise about 5 percentage of all why-questions and it shows that this kind of questions do have relevance in QA applications.

The techniques that have proven to be successful in QA for factual or factoid questions that returns answer as a noun phrase are found to be not suitable for questions that expect an explanatory answer. The work made an analysis on different question answering methods and found that discourse structure based method is effective in dealing with non - factual questions. So the work discuss about the role of dscourse structure in dealing with 'why' questions since it is based on Rhetoric Structure Theory, which helps in identifying the relationship between sentences or paragraphs from a given text or document and investigate to what extent discourse structure does indeed enable answering why-questions.

KEYWORDS

Network Protocols, Wireless Network, Mobile Network, Virus, Worms &Trojon Factual, Discourse structure, Rhetorical Structure Theory, Nucleus, Satellite, Mean reciprocal rank, Elementary Discourse Units.

1. INTRODUCTION

Question Answering (QA) is a specialized form of information retrieval. Given a collection of documents, a Question Answering System(QAS) attempts to retrieve the right answers to questions posed in natural language. Generally question answering system has three components such as question classification, information retrieval and answer extraction.

Question classification is the first phase which classifies user questions, derives expected answer types, extracts keywords and reformulates a question into semantically equivalent multiple questions. Information retrieval (IR) system recall is very important for question answering. If no correct answers are present in a document, no further processing could be carried out to find an answer. Precision and ranking of candidate passages can also affect question answering performance in the IR phase. Answer extraction is the final component in question answering system, where these systems are often requiring ranking and validating a candidate answer. Answer extraction technology becomes an influential and decisive factor on question answering system for the final results. The work shows the importance of using discourse structures in improving the quality of question answering system, in particular to the answer extraction part.

The rest of the work is organized as follows: section 1.1. presents the Motivation, 1.2. discuss about the architecture of a typical question answering system and 1.3. its evaluation. Section 1.4. presents the underlying technologies where a comparison is made on different QA methods for factual as well as non factual questions and based on that certain observations are found.. Chapter 2 describes the Literature Survey that analyses the di�erent QA approaches for non factual questions. Chapter 3 presents the problem definition. Chapter 4 presents the Conclusion and Future works.

1.1. Motivation

Question-answering (QA) research emerged as an attempt to tackle the information-overload problem. The amount of information on the web has developed exponentially over the years, with content covering almost any subject. As a result, when user looks for information, he/she is often confused by the vast quantity of results from search engines. Virtually all kinds of information are available on the World Wide Web (WWW) in one or another form. The number of web pages on the Internet increased tremendously and crossed 1 trillion landmark in 2008 which was only 200 billion in 2006 [1]. Therefore, managing such a huge volume of data is not an easy task. Search engines like Google and Yahoo return links along with snippets to the documents for the user query. User's browse the content carefully through a long list of outcome to look for a precise answer.

Recent researches in QA systems shows that the system have been expanded to answer simple questions correctly; but now researches have been focused on methods for answering complex questions truthfulness [2]. A little research has focused on QA models for non-factual or non-factoid questions, such as 'how', 'why', or 'manner' complex questions. One reason for this is that the frequency of non factoid questions posed to QA system is lower than that of other types of questions such as factoid questions. However, these complex questions cannot be neglected: as input for a QA system, they comprise about 5 percent of all why- questions and 4 percent questions asked to QA are 'how' questions, it shows that this kind of questions do have relevance in QA applications [3]. A second reason why this type of question has largely been disregarded until now is because those methods analyze and parse complex question to multi simple questions use existing techniques for answering them. But they are not found to be suitable for questions that expect an explanatory answer.

A few works have been done recently to deal with complex questions based on discourse structure. The relevance of discourse analysis for QA applications has been suggested by Marcu and Echihabi (2001) and Litkowski (2002). So this work discuss about the various methods used in QA system and investigate to what extent discourse structure does indeed enable answering why-questions.

1.2 The architecture of a typical question answering system

For a generic free text-based question answering system, each time a user issues a question as a query in natural language, the system usually parses and analyzes the question before classifying it. The system then retrieves one or more queries generated from the question against a large size corpus. From the returned results, one or multiple clear and precise selected answers, each in the form of a sentence or a piece of a sentence, are responded to the user. Fig 1 shows the three phases of a generic free text-based question answering system [4].

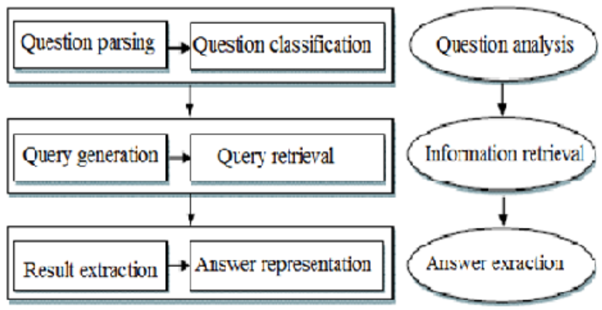

Figure 1. Generic Text-based Question Answering System

1.3. The evaluation of a typical question answering system

When evaluating the results of a question answering system, several performance metrics have to be considered. When processing answers to factual questions, the mean reciprocal rank (MRR) calculates the average over a set of n queries, where different scores are attributed inversely proportional to the rank ($rank_i$) of the first correct answer in the answer list. More formally it is defined as:

$$MRR = 1 \square n \sum_{i=1}^{n} \qquad (1)$$

Since the MRR metric does not allow evaluating questions to which there is no answer in the information source, and does not credit the recall of the system, precision and recall measures are used to evaluate the performance of the system.

For questions in which there is no answer in the document collection, the recall and precision of the NIL answers are computed, respectively as the percentage of NIL answers found and as the percentage of correct NIL answers in the found NIL answers. For factoid questions, accuracy is used as one of the major evaluation metrics, for which the answers are judged to be globally correct. Within TREC, list questions are evaluated with instance precision (IP) and instance recall (IR) which are based on the complete list of known distinct instances of the answers. If S is the number of known answers, D is the number of correct, distinct responses returned by the system, and N is the total number of responses returned by the system, then $IP = D \square N$ and $IR = D \square S$. The F-score is then estimated giving equal weights to recall and precision ($\beta = 1$) using the equation below.

$$F(\beta) = \frac{\beta^2+1\times precision\times recall}{\beta^2 \times precision+recall} \qquad (2)$$

Since different types of questions are evaluated differently and may not reflect the overall performance of the system, evaluating over individual series of questions should provide more accurate scores. Each series is a mixture of different question types and a weighted score can be computed taking into account only the types of questions available in accordance with the below equation. The final score is called pre-series score and is estimated as an average over all weighted scores.

$$\text{WeightedScore} = \mu \times \text{factoid} + v \times \text{list} + (1 - \mu - v) \times \text{other} \qquad (3)$$

1.4. Underlying technologies

When the information content is expressed in the form of text, there are number of valuable information extraction techniques that can structure the information found in the text. The work illustrate the different methods used in QA with the following example question: Who killed Lee Harvey Oswald?

1.4.1. Bag-of-words representations

The simplest approach is to consider the question as a bag-of-words, i.e., a set of words, possibly with stop words or non content bearing words removed (see Fig 2) without taking into account any structural characteristics that signal sentence or discourse grammaticality, or positional information of the words [5].

The most straightforward case when information query and information objects are represented in a bag-of-words manner (Fig 3) uses a Boolean model that requires the Boolean expression built from the query to be matched by the document representation as shown in (Fig 4). This retrieval function is used in the first iteration of a question answering system to find answer candidates to the posed query. Since this model requires a strict deterministic match, other re trieval models have been proposed, including the algebraic vector space model and various probabilistic models. Under the vector space model, an information object and natural language query are represented as a vector of terms in a p-dimensional space, where p is the number of terms in the vocabulary. The document and query vector are compared by computing their similarity (e.g., the cosine of the angle between the vectors) or distance. Compared to Boolean models, this approach returns answers to the questions even if the constraints posed by the question or query are only partially met, but at the cost of precision this model was a failure. also this method cannot identify the structural relations between sentences.

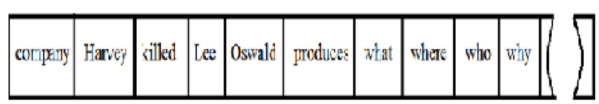

Figure 2. The vocabulary of a bag-of-words representation

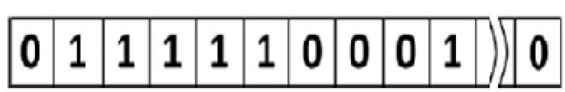

Figure 3. Bag-of-words representation indicating the presence or absence of a word from the vocabulary.

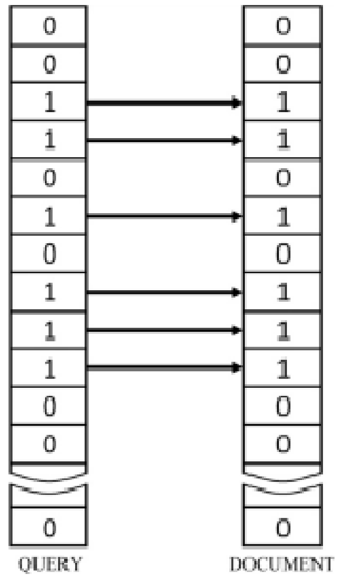

Figure 4. Matching mechanism between a document and a query given a bag-of-words representation.

It is a simple approach to question analysis and likewise to the analysis of documents and sentences, it ignores structural relations between words and the retrieved information may not be very precise.

These models, although very popular in document retrieval and search technology, typically lack the precision needed for question answering, even for answering simple factual questions requiring the retrieval of an entity (e.g., person name, date). They have the advantage of low analysis complexity and low storage overhead. However, in traditional question answering they have proven their usefulness for the initial rough filtering of documents and sentences that obviously have no relevance to the information question at hand.

1.4.2. Semantic classification of the expected answer type

A natural language question or query statement gives us additional information on the type of information that is expected as an answer. For instance,we might be looking for a person, the name of a company of interest, the place, the date etc. It has become common practice in question answering to automatically identify the semantic class of the expected answer in the question and the corresponding semantic class of the answer candidate in the information source [6].

Here question classification make use of 13 basic conceptual semantic question classes including Causal antecedent, Goal orientation, Enablement, Causal consequent, Verification, Request and others. Unfortunately few studies have implemented this classification due to the need for a deep semantic analysis with the exception of the investigation of Why questions in QA research. Motivated by early TREC evaluations and by the task settings, questions were categorized as factoid, list, definition, hypothetical, causal, procedural and confirmation queries [6]. For factoid questions the question type classes correspond with an expected answer type (EAT). The following example illustrates how interrogative words can be used to identify the expected answer type:

what/which Aux (be) NP (city - country - company) ?
CITY - COUNTRY - COMPANY

who/whom? PERSON
when? TIME/DATE
where/what place? LOCATION
what time (of day)? TIME

This method also makes use of a deterministic and a probabilistic approach. Under the deterministic approach, once the expected answer type has been identified, the answer candidates can be filtered by this expected semantic class. Under the probabilistic approach, the confidence score of the EAT identification or the scores of a few better hypotheses may be incorporated into a probabilistic retrieval model. List questions especially require the information from different places in one or several documents which is then linked and merged into an understandable answer. Yes / no questions need assessments whether the events or entity-relationships found in the question hold based on the information obtained from the documents. Another difficulty arises when questions are complex, entailing several question and answer types, possibly combined with conjunctions, disjunctions or conditional operators, often requiring the aggregation of pieces of information found.

The next problem of this method is to identify the expected answer type of respectively the question and the information found in documents. Symbolic approaches use hand-crafted rules possibly combined in a grammar. Features typically included in the rules are lexical elements such as words or phrases and syntactic information obtained through POS tagging and sentence parsing.

With respect to the information source, the analysis of answer candidates is done by employing shallow semantic analysers, such as named-entity recognizers, which provide semantic labels to the tokens in the candidate answer. The set of labels is limited to capabilities of the recognizer and adjusted to the chosen question taxonomy. The importance of correctly identifying the expected answer type was studied based on this method. The author reports that 36.4 percentage of errors in question answering on TREC-8, 9 and 10 data are caused by incorrect EAT estimation of this method. The problem is that many different question and answer type (complex question answering) taxonomies exist in the form of flat lists or hierarchies which need a kind of interlingua between questions and answers to facilitate their matching and it is found to be ineffective through this method [6].

1.4.3. Identifying the necessary discourse relationships

In the document collection, the complete answer to a question is not necessarily located in one sentence. Typically answers to list questions are spread throughout the document collection. Here also it is important to identify how data are connected across documents. This method identifies a number of rhetorical relations in questions and data which is needed for dealing with non factual questions types like ’How’, ’Why’ etc, that needs an answer in a sentence or more. Here, the propositions of a question topic and its answer are both represented by a text span in the source text, and identifies an RST relation holds between these spans. A why-question can then be answered by matching its topic to a span in the RST tree and selecting the related span as answer. Since recent advances in text-based argumentation detection open up new research avenues for answering Why and How questions. On this aspect the other two methods fails since they fail to identify the relations between sentences and can retrive answers only in a single word [7].

1.4.4. Observation And Analysis

A comparison made on the different QA methods is shown in Fig 5. Even though all the methods are efficient in dealing with factual questions that returns a single word or phrase as answer, it has been found that Discourse based approach is found to be effective in dealing with non-factual questions that need answer in a sentence or more since it is good enough in capturing structural relation between sentences. One of the main disadvantage of this approach is that it consumes time and space.

Methods	**Time and space complexity**	**Capturing structural relation**	**Factual**	**Non-factual**
Bag of words	Low	Low	Good	Average
Semantic classifica-tion	Low	Average	Good	Low
Discourse relation-ship	High	High	Good	Good

Figure 5. comparison made on the di erent QA methods

2. LITERATURE SURVEY

This chapter discuss about the work that have been carried out for non factual questions especially why-question answering (why-QA). “ Automatic question answering using the web: Beyond the Factoid [1]” described and evaluated a Question Answering (QA) system that goes beyond answering factoid questions.This approach to QA assumes no restrictions on the type of questions that are handled, and no assumption that the answers to be provided are factoids.

The work presented an unsupervised approach for collecting question and answer pairs from frequently asked question (FAQ) pages, they collected a corpus of 1 million question/answer pairs from FAQ pages available on the Web. This corpus is used to train various statistical models employed by the QA system: a statistical chunker used to transform a natural language-posed question into a phrase-based query to be submitted for exact match to an o -the-shelf search engine; an answer/question translation model, used to assess the likelihood that a proposed answer is indeed an answer to the posed question; and an answer language model, used to assess the likelihood that a proposed answer is a well-formed answer. Even though the work attempts to go beyond the factoid world, in which the authors focus on finding mechanisms for retrieving whole documents relevant for natural language-posed questions and is not restricted to factoid questions. The authors, however, do restrict the questions they use to evaluate their system to 4 general types of questions: Who, How, Where, and What. For each of these types of questions, they learn di erent transformations from question to more e ective queries, which are then sent to a search engine to retrieve relevant documents from the Web.

Although this technique has been shown e ective for factoid as well as non- factoid questions, it still has problems in handling complex questions as they excluded ”why” question types which have a major role in non factual questions.

“Exploring the Use of Linguistic Analysis for Answering Why-Questions [8]” investigated the possibilities of why question answering method based on textual cues. To that goal they analyzed the text fragments related to each question- answer pair in data collection. For each of these pairs, the item in the text is identified that indicates the answer. For 50 percent of the questions, they could identify a word or group of words that in the given context is a cue for the answer. Most of these cues, however, are very frequent words that also occur in many non- cue contexts. For example, the 'subordinator' that occurs 33 times in a document collection, only 3 of which are referred to by one or more why-questions. This means that only in 9 percentage of the cases, the subordinator that is a why-cue. The only two words for which more than 50 percent of the occurrences are why- cues, are 'because' and 'since' . Both are a why-cue in 100 percentage of their occurrences.

For almost half of the question-answer pairs that do not have an explicit cue in the source text, the answer is represented by the sentence that follows or precedes the sentence that represents the question that may not always be correct. These systems acheived a recall of 25 percentage which is better compared to the above method.

Since the simple cue - based method is found to have its own limitations, Rhetorical Structure Theory(RST) based method was introduced by Suzan Verbane.

"Discourse-based answering of why-questions [9]", used a method for why-QA that is based on discourse structure and relations. The main idea of this approach is that the propositions of a question topic and its answer are both represented by a text span in the source text, and that an RST relation holds between these spans. A why-question can then be answered by matching its topic to a span in the RST tree and selecting the related span as answer.

2.3. Discourse structure analysis

Discourse usually refers to a series of segments or sentences which constituted an overall language unit with certain hierarchy. The basic constituent units of all levels in discourse include clauses, sentences, paragraphs, sections. Usually only by analyzing the hierarchical structure and semantic relationship between constituent parts of text, discourse can determine the relations of paragraphs or sentences. Discourse structure includes two aspects: physical structure and logical structure. Discourse of physical structure is the basic elements of text (such as title, paragraphs, sentences, words and punctuation etc). Logical structure of the text refers to the logical relationship formed by the constituted article topic, structural levels, paragraphs, sentences and keywords of text ideological contents in the conceptual sense [3]. The rhetoric structural theory (RST) is originally developed by Mann and Thompson that commonly used in logical structure analysis of discourse. At present the discourse structure theory has been widely used in automatic text generation, automatic summarization, text analysis, and machine translation field, etc.

2.2. Rhetorical Structure Theory (RST)

In RST, the smallest units of discourse are called Elementary Discourse Units (EDUs). In terms of the RST model, a rhetorical relation typically holds between two EDUs, one of which (the nucleus) is more essential for the writers intention than the other (the satellite). If two related EDUs are of equal importance, there is a multinuclear relation between them. Two or more related EDUs can be grouped together in a larger span, which in its turn can participate in another relation. By grouping and relating spans of text, a hierarchical structure of the text is created. Such a hierarchical structure can be called as an RST tree [9]. For example, the following text is described by the schema depicted in Fig 6.

a. Come to the party for the new President.
b. There will be lots of good food.
c. The Fluted Mushroom is doing the catering.
d. The party is in the ballroom at eight o'clock on Friday.

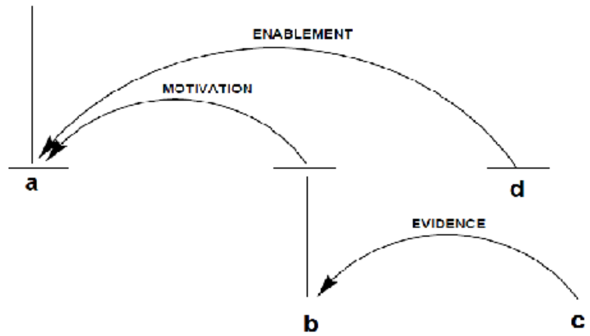

Figure 6. Graphical Representation of an RST Analysis of e.g.

In this example, (a) is the nucleus of the entire text and it presents an action that the speaker wishes the hearer to perform. (b-c) presents information intended to increase the hearer's desire to perform the action, and is therefore a satellite related to (a) by the MOTIVATION relation. (b-c) is further decomposed into a nucleus, (b), and a satellite, (c), which in this case are related by EVIDENCE since (c) is intended to increase the hearer's belief in (b). In (d) the speaker provides information that is intended to increase the hearer's ability to perform the action in the nucleus, and thus (d) is a satellite span related to (a) by the ENABLEMENT relation [9]. The RST analysis of e.g. is shown in Fig 6.

The work considered a why-question-answer pair and the RST structure of the corresponding source text. The following hypothesis was obtained:

1. The question topic corresponds to a span of text in the source document and the answer corresponds to another span of text;
2. In the RST structure of the source text, an RST relation holds between the text span representing the question topic and the text span representing the answer.

If both hypotheses are true, then RST can play an important role in answering why questions . For the purpose of testing hypotheses, a number of RST annotated texts and a set of question-answer pairs that are linked to the texts. The final set of selected relations for the method is shown in Fig 7. For majority of these relations in Fig 7, the span of text that needs explanation (or elaboration, evidence, etc.) is the nucleus of the relation, and the span of text giving this explanation is the satellite. The only exception to this rule is the cause relation, where the cause is given by the nucleus and its result by the satellite. Knowing this the following procedure was used for analyzing the questions and answers [9].

cause	circumstance	condition
elaboration	explanation - argumentative	evidence
interpretation	list	problem- solution
purpose	reason	result
sequence		

Figure 7. Selected relation types

Based on this, the following procedure was used for analyzing the questions and answers:

I. Identify the topic of the question.

II. In the RST tree of the source document, identify the span(s) of text that express(es) the same proposition as the question topic.
III. Is the found span the nucleus of a relation of one of the types listed in Table 1 (or, in case of cause relations, the satellite)? If it is, go to IV. If it is not, go to V.
IV. Select the related satellite (or nucleus in case of a cause relation) of the found span as an answer.
V. Discard the current text span

Even though this method can solve the problem in why question answering and having a recall of 58 percent i.e. it seems to be far better compared to simple cue- based method. But the method failed for:

1. Questions whose topics are not or only implicitly supported by the source text.
2. Questions for which both topic and answer are supported by the source text but there is no RST relation between the span representing the question topic span and the answer span.
3. Questions for which the correct (i.e. user-formulated) answer is not or only implicitly supported by the text.
4. Questions for which the topic can be identified in the text and matched to the nucleus of a relevant RST relation, but the corresponding satellite is not suitable or incomplete as answer.

More over this approach requires manually annotated corpora with RST relations. So improvements are to be done for why-questions where both topic and answer are supported by the source text, but there is no RST relation between the span representing the question topic span and the answer span. Even though the work made a better improvement in dealing with why QA, an investigation is to be done to check whether other types of linguistic analysis, or different exploitation of the RST structure can result a better recall value and hence can improve the QA technology.

Since this work has been done based on manualy annotated corpora as an improvement it has to make use of automatically annotated data.

2.3. Observation and Analysis

On analysisng the above mentioned methods the observations found are shown in Fig 8. Eventhough Method 3 based on discourse structure is found to better compared to other methods, it still have some short coming which have to be improved for a better QA system.

Methods	**Technolo-gy used**	**Factual**	**Non- factual**
Method 1	Based on phrase	Good	All non- factual questions except “why”
Method 2	Based on textual cues	Good	Used for “why” QA, failed for 70%
Method 3	Based on RST	Not necessary	Better method for “why” questions

Figure 8. comparison made on the di erent non-factual QAS.

3. PROBLEM DEFINITION

To enhance the efficiency of Question Answering System for non factual - Why questions based on discourse structure.

On analysing different QA methods it has been found that most of the meth ods failed to deal with non - factual questions like how and why. Reason for this is that the frequency of why-questions posed to QA systems is lower than that of other types of questions such as who and what questions and the techniques that have proven to be successful in QA for factual questions that returns answer as a noun phrase was found to be not suitable for questions that expect an explanatory answer. However, why-questions cannot be neglected as input for a QA system, they comprise about 5 percent of all wh-questions and they do have relevance in QA applications [5] [6] [7].

After analysing different works on non factual questions, it has been found that certain methods even though can deal with non factual question they fail for why question types. Certain approach made use of cue based method which can solve this problem to certain extent but still have limitation because almost half of the QA pair do not have an explicit cue in the source text. So discourse based answering of Why questions based on RST is analysed. Though it is one of the efficient approach for Why QA with a recall of around 60 percent, this method failed for:

1. Questions for which both topic and answer are supported by the source text but there is no RST relation between the span representing the question topic span and the answer span.
2. Questions for which the correct (i.e. user-formulated) answer is not or only implicitly supported by the text.
3. Questions for which the topic can be identied in the text and matched to the nucleus of a relevant RST relation, but the corresponding satellite is not suitable or incomplete as answer.

More over this approach requires manually annotated corpora with RST relations which may consume time [1] [8] [9]. So a better method is needed to improve the current question answering system for non factual questions. An investigation is to be done to check whether other types of linguistic analysis, or different exploitation of the RST structure can result a better recall value than the existing one.

4. CONCLUSION AND FUTURE WORK

The techniques that have proven to be successful in QA for factual questions that returns answer as a noun phrase are not suitable for questions that expect an explanatory answer. So the work made an analysis on different question an-swering methods and found that discourse structure based method is effective in dealing with non - factual questions. Then the work discussed about different QA approaches on non factual questions and analysed the extent discourse structure does indeed enable answering why-questions, since it is based on Rhetoric Structure Theory, which helps in identifying the relationship between sentences or paragraphs from a given text or document.

Even though discourse structure based approach using RST have some short comings. This is one of the best method for non factual questions that can retrive around 60 percent of answers in Why question answering. By improving the existing limitations an advanced question answering system can be made for Why QA as they comprise about a minimum of 5 percent in a question answering system.

REFERENCES

[1] R. Soricut and E. Brill, (2006) in “Automatic question answering using the web: Beyond the factoid,” Information Retrieval, vol. 9, pp 191-206.

[2] M. R. Kangavari, S. Gh, and M. Golpour, (2007) “A new model for question answering systems,”.

[3] H. Zong, Z. Yu, J. Guo, Y. Xian, and J. Li, “An answer extraction method based on discourse structure and rank learning,” in Natural Language Processing and Knowledge Engineering (NLP-KE), 7th International Conference, vol. 3, nov. 2011, pp. 132 -139.

[4] D. Wen, S. Jiang, and Y. He, “A question answering system based on verbnet frames,” in Natural Language Processing and Knowledge Engineering, NLP-KE '08. International Conference, vol. 3, 2008, pp 1 -8.

[5] S. Verberne, L. Boves, N. Oostdijk, and P.-A. Coppen, ‘What is not in the bag of words for why-qa?” Comput. Linguist., vol. 36, pp 229- 245.

[6] N. Schlaefer, J. Ko, J. Betteridge, G. Sautter, M. Pathak, and E. Nyberg, (2007) “Semantic extensions of the ephyra qa system for trec 2007,” in Proceedings of the Sixteenth Text REtrieval Conference (TREC).

[7] O. Kolomiyets and M.-F. Moens, (2011)“A survey on question answering technology from an information retrieval perspective,"”Inf. Sci., vol. 181, pp. 5412- 5434.

[8] S. Verberne, L. Boves, and N. Oostdijk, (2007) “Exploring the use of linguistic analysis for answering why questions,”.

[9] S. Verberne, L. Boves, N. Oostdijk, and P.-A. Coppen, (2007) “Discourse-based answering of why-questions," pp. 21- 41.