

A Review on Hindi Question Answering System

Amreen Kaur | Nishi | Beant Kaur

Assistant Professor, Department of Computer Science Engineering, BGIET, Sangrur, Punjab, India

To Cite this Article

Amreen Kaur, Nishi and Beant Kaur, "A Review on Hindi Question Answering System", *International Journal for Modern Trends in Science and Technology*, Vol. 05, Issue 11, November 2019, pp.-181-185.

DOI: <https://doi.org/10.46501/IJMTST051189>

Article Info

Received on 01-November-2019, Revised on 19-November-2019, Accepted on 21-November-2019, Published on 26-November-2019.

ABSTRACT

With advancement in technology, the process of question answering is the main field of research of text mining. In India, the natural language of people is mainly Hindi. In automatic question generation the system generate multiple choice questions automatically from Hindi text using question generation techniques. There are various approaches that can be used to generate the questions from a given text. This paper presents the review to various question generation techniques.

KEYWORDS : NLP, Hindi, Question, Answer.

Copyright © 2019 International Journal for Modern Trends in Science and Technology
All rights reserved.

1. INTRODUCTION

NLP is a field of computer science and linguistics concerned with the interactions between computer and human (natural) languages. In theory, NLP is a very attractive method of human-computer interaction. Natural language understanding is sometimes referred to as an AI-complete problem because it seems to require an extensive knowledge about the outside world and the ability to manipulate it. NLP has significantly overlapped with the field of computational linguistics, and is often considered a sub-field of artificial intelligence.

The foundation of NLP lies in a number of disciplines like computer and information sciences, linguistics, mathematics, electrical and electronic engineering, artificial intelligence and robotics, psychology, agriculture, weather forecasting, etc.

Applications of NLP include a number of fields of studies, such as machine translation, natural

language interface to databases, natural language text processing and summarization, user interfaces, multilingual and Cross Language Information Retrieval (CLIR), speech recognition, artificial intelligence and expert systems, and so on.

Natural language processing (NLP) [10] is a theory-motivated range of computational techniques for the automatic analysis and representation of human language. NLP research has evolved from the era of punch cards and batch processing, in which the analysis of a sentence could take up to 7 minutes, to the era of Google and the likes of it, in which millions of webpages can be processed in less than a second. NLP[11] enables computers to perform a wide range of natural language related tasks at all levels, ranging from parsing and part-of-speech (POS) tagging, to machine translation and dialogue systems.

Question generation [12] is a task of generating reasonable questions from given input. The task of generating question containing multiple subareas.

The approach taken for question generation depends on the purpose of the question generation application. The purpose of asking question is not limited to such tasks but it may serve much more than this. In a class teacher questions her student not because she is unaware but to know the intellect of her. Such questions provide a good hint to the students which help them to solve the problem. This paper presents the review to generate the questions from Historical documents written in Hindi text. The main properties of any documents are required to be find from which questions are to be generated.

Database Management System[2] is a collection of Interconnected Data and set of a programs to get those Data. There is a requirement to design and Develop an Interface in the Native Language so User can easily use System without Knowledge of English as well as Query Language. With the help of interface, the end user can query the system in a Query Language like English, Hindi, Marathi, etc., and can see the result in the same language. NLIDB system is proposed as a solution to the problem for accessing information in an easy way, allowing any type of users, mainly non-technical people to retrieved Information from a Database using Natural Language(NL). We are going to develop a system for people who know Hindi Language. User can access Database using in the Hindi Language and get the result in the same Language.

II. RELATED WORK

Poonam Gupta et al. [3] discussed that with advancement in technology, the process of question answering is the main field of research of text mining. In this process user is assigned a particular answers in place of number of text documents or paragraphs. It provides a way for obtaining appropriate and useful answers for questions of user put in native natural language instead of formal query in that language. Hindi, Punjabi, Bengali, Kannada, Telugu and Marathi etc are main spoken languages in India. For these Indian languages, very less number of language resources is available and much of research is going on for developing basic language resources in these languages. This paper discusses a survey of different question answering techniques for Indian languages.

Kumar et al. (2005) [4] implemented the Hindi search engine for the retrieval of the relevant passages from the collection of the passages. The architecture for the question answering was

proposed. The architecture involved the various modules. Automatic Entity Generator module which identified domain related entities to which the user wanted to ask questions. Then the question was classified based upon the categories in the question classification module. Question parsing module was searching for the domain entities from the question after removing stop words. Query formulation module which converted the question into a query which then input to the retrieval engine for the retrieval of the answers. Query expansion module expanded the query which increased the search process by including the terms whose meanings were same and thus retrieved the text in which query terms do not particularly appear. Answer was extracted by using the answer extraction module. Answers were selected among the candidate answers by using ranking in the answer selection module.

Sahu et al. (2012) [5] discussed an approach for finding out the answers for the questions in the Hindi language. The answers were extracted from Hindi text and the text was completely analyzed to understand the meaning of each sentence. In this paper the main focus was on four types of questions when, where, how many and what time. The architecture for Hindi question answering system has been given. The architecture consists of five stages. The first stage was used to classify the question based upon its type. The second stage used the query logic language (QLL) for the input question. In the next stage the answers were searched in the stored database. The answer produced by that stage was stored as a document. Then the answer was converted into Hindi and then it was presented to the user. In this paper, Query logic language (QLL) which was a subset of Prolog which was used to represent the questions. Hindi shallow parser was used for the identification of the verb, noun, and question word. These words were very helpful for the extraction of the answers. The answers were extracted using set of developed rules.

Reddy et al. (2006) [6] described the dialogue based question answering system in Telugu language for railway specific domain. The main part of this system was dialogue manager which was responsible for the handling of the dialogues between user and the system. The system architecture for railway domain had been given. This architecture was based on the keyword approach in which input query was analyzed by the query analyzer. The query analyzer was responsible for the generation of the tokens and

keywords with the use of knowledge base. Based upon the keywords and tokens which were presented in the knowledge base, an appropriate frame was selected. The words that have some semantic information were needed to be present in the knowledge base. SQL statements were generated from the tokens. There were two main issues in the design of the railway information system, how to design railway database and knowledge base. The railway database had been constructed which contained the information about the arrival / departure time of each train, information regarding their fares. For these purposes, relational model had been used. Railway database contained the tables like schedule tables, route tables and temporal tables while knowledge base contained the tables like train name, station name and also contained alias tables for station name and train name. For each input statement, root words were identified by the query analyzer during query analysis. Based upon the detection of the keywords and tokens, a query frame was identified during query frame decision. The basic responsibility of the dialogue manager was to manage the flow of dialogues. Dialogue manager was also responsible for the coordination of the other components in the system. After the generation of the query frame, SQL query was generated. Then the answer was retrieved from the database using SQL query.

Stalin et al. (2012) [7] discussed the web based application for the extraction of answers for a question posed in Hindi language from Hindi text. If the answer was not present in the Hindi text then the answers were searched on Google. This paper proposed a question answering architecture that used words of sentence (question). This acts as a source. This architecture searched the answer for a question posed in Hindi language in Hindi text. Then the results were displayed to the user. The architecture of the system involved various modules. Query interface was used for the retrieval of the question from the user. Question classification was used for the recognition of the question type. Query formulation was used for the retrieval of the correct answer. Database was used for the retrieval of the document based upon the keyword present in the question. It required the knowledge of the pattern of the question. Database sent all the candidate answers to the next module which was responsible for the extraction of the answers from the retrieved documents. Then all the candidate answers were displayed on the screen.

Banerjee et al. (2012) [8] discussed how to classify

a question. Since classification step was necessary towards the construction of the question answering system. In this paper discussion of the lexical, semantic and syntactic characteristics was given for the classification of the Bengali question. The proposed classification work was based upon the machine learning techniques. The question classification work for Bengali Language used the Bengali shallow parser. This parser was used to analyze the sentence in terms of chunking, POS tagging etc. This paper also discussed the various interrogatives present in the Bengali language. The question types for Bengali had been discussed. For the classification of the questions, there were three types of classifiers which were used to train the classifiers. These are lexical features, semantic features and syntactic features. Lexical features of the question were extracted based upon the words were present in the question. These features include wh-word, wh-word position, wh-type, length of the question, end marker and the shape of the word. Syntactical features include parts of speech (POS) tags and head words. POS tags were like nouns, adjectives, noun phrases and verb phrases. Head words were those words that were used to identify the objects from the question. Semantic features can be extracted from the question based upon meaning of the words that were present in the question. These features include related words and named entities. For the retrieval of related words Bengali dictionary had been used and for named entities, Bengali NER system had been used.

Sekine et al. (2003) [9] developed a question answering system for Hindi and English. The questions were created in Hindi Language and the answers were found in Hindi Language through Hindi newspapers and then these answers were converted into English Language back. With the help of the tagger, person names, location names, organization names were identified for the English paragraphs. First of all, the examiner examined the questions and searched their answers from Hindi newspapers. An English Hindi bilingual dictionary was used to find out the top 20 Hindi articles. These articles were used to find out the candidate answers. In the end, Hindi answers were returned back to the English language.

Gauri Rao et al. [13] discussed that the field of natural language processing (NLP) has seen a dramatic shift in both research direction and methodology in the past several years. In the past, most work in computational linguistics tended to focus on purely symbolic methods. Recently, more

and more work is shifting toward hybrid methods that combine new empirical corpus-based methods, including the use of probabilistic and information theoretic techniques, with traditional symbolic methods. The main purpose of Natural Language Query Processing is for an English sentence to be interpreted by the computer and appropriate action taken. Asking questions to databases in natural language is a very convenient and easy method of data access, especially for casual users who do not understand complicated database query languages such as SQL.

This paper proposes the architecture of a new NLDBI system including its probabilistic context free grammar, the inside and outside probabilities which can be used to construct the parse tree and the usage of dependency structures and verb sub categorization in analyzing the parse tree.

Jasmeen Kaur et al.[14] discussed that the field of query processing has recently been coupled with natural language processing (NLP) that has shown dramatic shift in both research direction and methodology in the past few years. In past, most of the work was done on computational linguistics which drew focus on purely symbolic methods. Recently, more prominence is given to hybrid methods that combine new empirical corpus-based methods, including the use of probabilistic and information theoretic techniques, with traditional symbolic methods. The main purpose of Natural Language Query Processing is to interpret an English sentence and hence a complementary action is taken. Querying to databases in natural language is a convenient method for data access, especially for newbie's who have less knowledge about complicated database query languages such as SQL. This paper emphasise on the structural designing methods for translating English Query into SQL using automata.

Payal Garg et al.[12] discussed that Question Generation is an important area of text processing in Natural Language Processing. Automatic Question generation is a process of generating questions automatically from a text with the help of various NLP techniques. There are various approaches that can be used to generate the questions from a given text. Rule based approach is most common approach to generate the questions automatically from a text. In this paper we are presenting the review on question generation from documents written in Punjabi language. A NER (Named Entity Recognition) Tool is also need to be created which recognizes the names from a given sentence and generate the appropriate questions

from it.

Nisha Sharma et al.[15] discussed that question generation is an application of the NLP (Natural Language Processing). In automatic question generation the system generate multiple choice questions automatically from Punjabi text using question generation techniques. Question generation systems use the rule based approach to generate the questions from given text. Generation of multiple choice questions is very important because this helps anyone to test their knowledge in specific field. One can give the answer easily by choosing one option from a given set of options provided by the system and then system evaluate the given answer and generate the report for all the answers given. This paper presents the review to various question generation techniques are the rule based approach, pattern matching and information extraction are discuss in this paper from used various researchers.

III. CONCLUSION

In this paper the review of work done by various researchers in the field of automatic question generation from a given Hindi text has been reviewed. The work illustrated here in context of Hindi Question Answering System is first of its kind. For future scope we suggest more work must be done to add more types of questions for question classification and more types of answer patterns must be incorporated with scoring system that balances the probability between different natures of answers.

REFERENCES

- [1] Mr. Mahesh Singh and Ms. Nikita Bhati(2014), "A Noval Hindi Language Interface for Databases", International Journal of Computer Science and Mobile Computing, Vol.3 Issue.4, pg. 1179-1189.
- [2] Prof. K.S. Chandwani, Monika Bramahankar, Dhanashri Gowardhan, Pallavi Mohadikar, Poonam Majalgaonkar and Shrutika Jumde(2018), "Hindi Language Graphical User Interface to DBMS for Transport System", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 6 Issue I.
- [3] Poonam Gupta and Vishal Gupta(2014), "A Survey of Existing Question Answering Techniques for Indian Languages", Journal Of Emerging Technologies In Web Intelligence, Vol. 6, No. 2.
- [4] P. Kumar, S. Kashyap, A. Mittal and S. Gupta, "A Hindi Question Answering System for E-learning documents," In Proceedings of IEEE International Conference on Intelligent sensing and Information processing, Bangalore, India, 2005, pp. 80-85.
- [5] S. Sahu, N. Vashnik and D. Roy, " Prashnottar: A Hindi Question Answering System", International Journal of Computer Science and Information Technology (IJCSIT), vol.4, no.2, 2012, pp.149-158.

- [6] R. Reddy, N. Reddy and S. Bandyopadhyay, "Dialogue based Question Answering System in Teulgu," In Proceedings of EACL Workshop on Multilingual Question Answering, 2006, pp. 53-60.
- [7] S. Stalin, R. Pandey and R. Barskar, "Web based Application for Hindi Question Answering System," International Journal of Electronics and Computer Science Engineering, vol. 2, 2012, pp. 72-78.
- [8] S. Banerjee and S. Bandyopadhyay, "Bengali Question Classification: Towards Developing QA System," In Proceedings of the 3rd Workshop on South and Sotheast Asian Language Processing (SANLP), COLING, 2012, pp. 25-40.
- [9] S. Sekine and R. Grishman, "Hindi-English Cross-Lingual Question Answering System," ACM Transaction on Asian Language Information Processing, vol. 2, 2003, pp.181-192.
- [10] E. Cambria and B. White(2014), "Jumping NLP curves: A review of natural language processing research," IEEE Computational Intelligence Magazine, vol. 9, no. 2, pp. 48-57.
- [11] Tom Young, Devamanyu Hazarika, Soujanya Poria and Erik Cambria(2018), "Recent Trends in Deep Learning Based Natural Language Processing".
- [12] Payal Garg and Er.Charandeep Singh Bedi(2014), "A Review on Question Generation System From Punjabi Text", International Journal of Emerging Trends & Technology in Computer Science, Volume 3, Issue 4.
- [13] Gauri Rao and S.H.Patil(2015), "Natural Language Query Processing Based On Probabilistic Context Free Grammar".
- [14] Jasmeen Kaur, Bhawna chauhan and Jatinder Kaur Korepal (2013), "Implementation of Query Processor Using Automata and Natural Language Processing, International Journal of Scientific and Research Publications, Volume 3, Issue 5.
- [15] Nisha Sharma and Er.Abhilasha(2015), "Automatic Question Generation from Punjabi Text :A Review", International Journal of Computer Science & Engineering Technology (IJCSET), Vol. 6 No. 07.