

ANALYSIS OF NATURAL QUERY LANGUAGE METHODS

Harmanjot Kaur¹, Kailash Bahl²

Abstract- Question generation is an application of the Natural Language Processing. In automatic question generation the system generate multiple choice questions automatically from Punjabi 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, Punjabi, Question, Answer.

1. INTRODUCTION

Natural Language Processing (NLP) [2] deals with the human understandable language. NLP is an important field of computer science. It is an area of research that explores how computer can be used to understand and manipulate natural language text. The various Applications of NLP such as Automatic Question Generation, Automatic Summarization, Machine Translation, Optical Character Recognition, Part-of-speech tagging etc. Automatic question generation is one of the most important Applications of the NLP. It is a way to generate the question from Punjabi text sentence. Ideal learners are often curious question generations who actively self regulate their learning. Good readers ask themselves questions during reading. Students learn to formulate and respond to question about situations facts and ideas while engaged in understanding a text.

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 [3] 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 Punjabi text. The main properties of any documents are required to be find from which questions are to be generated.

1.1Types of Question

QG [2] system takes the input text and gives the output as questions. Question generation is two types as shown in Figure1.

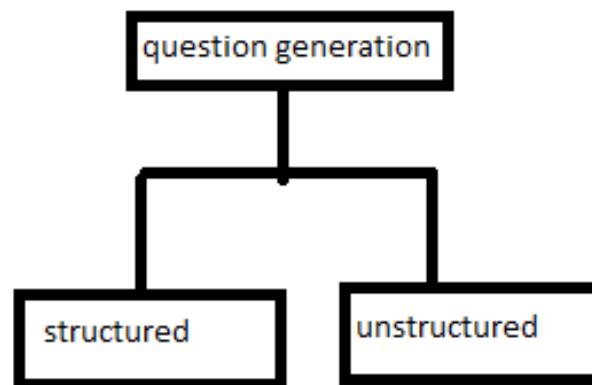


Figure1: question generation types

¹ Department of Computer Science Engineering, PIET, Punjab, India

² Department of Computer Science Engineering

Structured based approach is a data base look up and unstructured technique is a rule based approach for text and documents. Question generation can be splitting into deep QG and shallow QG[12]. Deep QG generates deep questions that involve more logical thinking such as why and how questions whereas shallow Question generation that focus more on facts such as who, what, when, where, which, how questions.

1.2. Question generation system helpful in following areas[3]

Intelligent tutoring system: QA(question answering) can ask questions based on learning materials in order to check learners accomplishment or help them focus on the keystones in study. QA can also helpful for tutors to generate questions.

Closed domain question answering systems: some closed domain QA systems use some hand written question-answer pairs to provide QA services. By employing a QG approach such systems should be ported to other domains without any efforts.

Natural language generation system: QA can help to generate, for instance, frequently asked questions from the provided information source in order to provide a list of FAQ candidates.

1.3 Requirement of Natural Language Interface to Database Systems

Language[13] is the primary means of communication used by humans. It is the tool we use to express the greater part of our ideas and emotions. Learning new concepts and expressing ideas through them is so natural that we hardly realize how we process natural language. Natural Language Processing (NLP) is concerned with the development of Computational models of aspects of human language processing. There are two main reasons of such development:

1.4 To develop automated tools for language processing.

1.4.1 To gain a better understanding of human communication.

Building computational models with human language processing abilities requires Knowledge of how humans acquire store and process language. It also requires the knowledge of world and of language. Companies have related the problem of extracting data from a Data Base Management System (DBMS) by using the tools like MS Access, Oracle and others. A person with no knowledge of Structured Query Language (SQL) may find himself or herself handicapped in corresponding with these tools. Thus it creates the need to develop products for people to interact with the database in their own native language. In Punjab (Pb.) Punjabi is used as a source of language for communication. We have proposed a model for Punjabi language interface to enable the people of Punjab to use the databases in their own native language.

Database applications play an important role in today's commercial system[14]. Most of the businesses need these types of applications by using the SQL language. Natural language processing (NLP) is becoming one of the most active techniques used in Human-computer Interaction. It is a branch of AI which is used for Information Retrieval, Machine Translation and Language Analysis. The main goal of NLP is to enable communication between human and computers without memorization of complex Commands and procedures. In other words, NLP is the techniques that can make the computer to understand the natural languages used by humans. Today's requirement of commercial system is to extracting data from a Database Management System such as MS Access, Oracle and others. A person without knowledge of SQL may find himself/herself handicapped in corresponding with the database. Therefore in this work the development of system for people to interact with the database in simple English language is implemented. This enables a user to input their queries in simple English and get the answer in same language. This is known as a Natural Language Interface to a Database (NLIDB).

2. RELATED WORK

The very first attempts at NLP database interfaces are just as old as any other NLP research. In fact database NLP may be one of the most important successes in NLP since it began. 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. The success in this area is partly because of the real-world benefits that can come from database NLP systems, and partly because NLP works very well in a single database domain. Databases usually provide small enough domains that ambiguity problems in natural language can be resolved successfully.

Fazal Mithani et al. [1] discussed that the world's latest scenario suggests that the database size will pilling-up epidemically. Due to this, requirement of larger database is eminent. Furthermore, these data needs to be scalable enough to support future up gradation, keeping in mind that performance is not compromised. The structure of Structured Query Language (SQL) query also plays an enormous role in gauging the performance of execution. In this paper we have suggested a model that allows checking of the query performance with respect to predefined rules, SQL benchmark, strategy of SQL tuning and data freezing. In this paper we systemized existing query formation rules which need to be followed by every query that needs to be optimized. Our approach also suggests the user an optimized query for normalization if required. If filtration is used in given query our approach will also be stating any requirement for indexing on particular column. By using any of the above approach along with avoiding unnecessary usage of data, columns, our proposed model converts the input user SQL query into an optimized SQL query and main aims to give assurance of reduced query execution time.

Poonam Gupta et al. [6] implemented an algorithm for Punjabi Question Answering system in this paper. The proposed system is designed and built in such a way that it increases the accuracy of Question Answering Systems in terms of Recall and Precision and is working for factoid questions and answers text in Punjabi. The system constructs a novel pattern finding

and matching system to identify most accurate probable answer out of multiple answers. This algorithm is based on understanding the meaning of the given Punjabi question and expresses them in query logic language.

Gauri Rao et al. [5] 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.[4] 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.

Saravjeet Kaur et al. (2012) [7] proposed a method of querying with the databases by means of a natural language interface. This is hot issue in the area of database management is to provide a high level interface for nontechnical users. Normal users are not aware with the formal language like SQL. Then the problem is how they interact with the database system. A normal user may find him/her self handicapped to deal with the database system. The paper presents an interface module that converts user's query given in natural language into a corresponding SQL command. Asking questions to databases in natural language like English is a very convenient and easy method of data access from database system, especially for normal users who do not understand complicated database query languages such as SQL. This paper proposed the architecture for translating English Query into SQL.

Wiqas Ghai et al., (2012) [9] discussed that Punjabi, Hindi, Marathi, Gujarati, Sindhi, Bengali, Nepali, Sinhala, Oriya, Assamese, Urdu are prominent members of the family of Indo-Aryan languages. These languages are mainly spoken in India, Pakistan, Bangladesh, Nepal, Sri Lanka and Maldives Islands. All these languages contain huge diversity of phonetic content. In the last two decades, few researchers have worked for the development of Automatic Speech Recognition Systems for most of these languages in such a way that development of this technology can reach at par with the research work which has been done and is being done for the different languages in the rest of the world. Punjabi is the 10th most widely spoken language in the world for which no considerable work has been done in this area of automatic speech recognition. Being a member of Indo-Aryan languages family and a language rich in literature, Punjabi language deserves attention in this highly growing field of Automatic speech recognition. In this paper, the efforts made by various researchers to develop automatic speech recognition systems for most of the Indo-Aryan languages, have been analysed and then their applicability to Punjabi language has been discussed so that a concrete work can be initiated for Punjabi language.

Payal Garg et al.[3] 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.[2] 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.

Preeti Verma et al. [8] discussed that unlike most user-computer interfaces, a natural language interface allows users to communicate fluently with a computer system with very little preparation. Databases are often hard to use in cooperating with the users because of their rigid interface. A good NLIDB allows a user to enter commands and ask questions in native language and then after interpreting respond to the user in native language. For a large number of applications requiring interaction between humans and the computer systems, it would be convenient to provide the end-user friendly interface.

Punjabi language interface to database would prove fruitful to native people of Punjab, as it provides ease to them to use various e-governance applications like Punjab Sewa, Suwidha, Online Public Utility Forms, Online Grievance Cell, Land Records Management System, legacy matters, e-District, agriculture, etc. Punjabi is the mother tongue of more than 110 million people all around the world. According to available information, Punjabi ranks 10th from top out of a total of 6,900 languages recognized internationally by the United Nations. This paper covers a brief overview of the Natural language interface to database, its different components, its advantages, disadvantages, approaches and techniques used. The paper ends with the work done on Punjabi language interface to database and future enhancements that can be done.

3. CONCLUSION

In this paper the review of work done by various researchers in the field of automatic question generation from a given Punjabi text has been reviewed. The work illustrated here in context of Punjabi 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.

4. REFERENCE

- [1] Fazal Mithani, Sahista Machchhar and Fernaz Jasdanwala(2016), "A Novel Approach for SQL Query Optimization", IEEE International Conference on Computational Intelligence and Computing Research.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] Gauri Rao and S.H.Patil(2015), "NATURAL LANGUAGE QUERY PROCESSING BASED ON PROBABILISTIC CONTEXT FREE GRAMMAR".
- [6] Poonam Gupta and Vishal Gupta(2013), "Algorithm for Punjabi Question Answering System", International Journal of Advanced Research in Computer Science and Software Engineering,, Volume 3, Issue 7.
- [7] Saravjeet Kaur, Rashmeet Singh Bali, "SQL Generation and Execution from Natural Language Processing", International Journal of Computing & Business Research, Proceedings of 'I-Society 2012' at GKU, Talwandi Sabo Bathinda (Punjab).
- [8] Preeti Verma, Suket Arora, Kamaljit Batra, "Punjabi Language Interface to Database: A Brief Review".
- [9] Wiqas Ghai, Navdeep Singh, "Analysis of Automatic Speech Recognition Systems for Indo-Aryan Languages: Punjabi A Case Study", International Journal of Soft Computing and Engineering (IJSCE),ISSN: 2231-2307, Volume-2, Issue-1, March 2012.
- [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] Xuchen Yao,G.(2012), "Semantic-based Question Generation and Implementation" Dialogue and Discourse3(2), 11-42,doi:0.5087/dad.2012.202.
- [13] Priyanka Arora ,Punet Goswami(2014), "An Efficient Hindi Language Interface using Relational Databases", International Journal of Research in Computer and Communication Technology, Vol 3, Issue 6.
- [14] Deepjot Kaur, Navjot Kaur(2013), "Implementation of DJ Rule Based Algorithm for DhuniVishleshan of Compound Punjabi Words", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 7.