

# Automatic Question Generation from Language Independent System

Meenakshi Sharma

Student OPJS University, Churu, Rajasthan, India

## ABSTRACT

Natural Language processing is a basic component of Artificial Intelligence. NLP means to train the computer/system as it behaves like a human being. Question production is an application of the NLP (Natural Language Processing). In automatic question generation, the system generates multiple choice questions automatically from Punjabi text, English text and Hindi text, using question generation algorithms. The text will be in the form of paragraph. As we know that Indian Languages basically English, Punjabi and Hindi have their own grammar rules. So Question generation system use different approaches according to the rules of grammar to generate the questions from given text. Generation of multiple choice questions is imperative because it 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 score for all the answers given. This paper presents about the language Independent system having integration of Punjabi, English and Hindi languages. Hybrid Approach is used that is a combination Dictionary look up. Rule based approach, pattern matching and information extraction, Knowledge Based approach, Example based and SMT approach.

**Keywords:** NLP (Natural Language Processing), Rule Based Approach, Question Generation (QG), Information Extraction (IE), SMT (Statistical Machine Translation).

## I. INTRODUCTION

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 [5]. Natural Language Processing (NLP) 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[9]. Good readers ask themselves

questions during reading [4]. Students learn to formulate and respond to question about situations facts and ideas while engaged in understanding a text. Teacher ask the question (“what is your Aim”) to students but some time not reminder frequently that time Automatic Question generation system will help for question generation. In this review paper categorize the question into various types like “ਕੀ”(ki)(what) “ਕਦੇ” (kado) (when) “ਕਿੱਥੇ”(kithe) (where) “ਕਿਹੜੇ / ਕਿਹੜਾ / ਕਿਹੜੀ” (which), “ਕੌਣ / ਕਿਸ / ਕਿਸੇ”(koon)(kis) (who) & “ਕਿਉ”(kio) (why) . For example for the following statement the questions can be generated as [1]

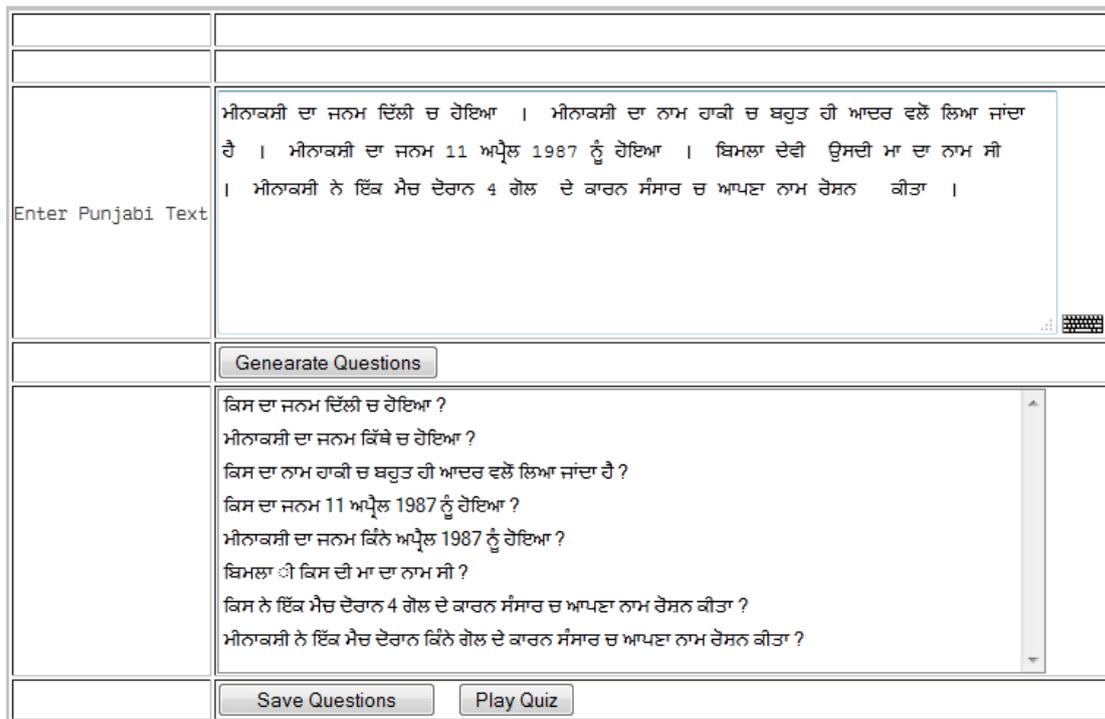


Figure 1

## II. Literature Survey

### Deana L. Pennell and Yang Liu, Normalization Of Text Messages For Text-To-Speech

This paper describes a normalization system for text messages to allow them to be read by a TTS engine. To address the large number of texting abbreviations, author use a statistical classifier to learn when to delete a character. The features we use are based on character context, function, and position in the word and containing syllable. To ensure that our system is robust to different abbreviations for a word, system generate multiple abbreviation hypotheses for each word based on the classifier's prediction. System then reverse the mappings to enable prediction of English words from the abbreviations. Results show that this approach is feasible and warrants further exploration. Author evaluate classifier accuracy by performing 10-fold cross validation on the training data. Always choosing the positive class System yields a baseline accuracy of 74.7%. [13]

### Shikha Garg, Vishal Goyal “ System for Generating Question Automatically from given Punjabi Text” (2013)

This paper introduces a system for generating questions automatically for Punjabi. The System transforms a declarative sentence into its interrogative counterpart. It accepts sentences as an input and produces a possible set of questions for the given input. Not much work has been done in the field of Question Generation for Indian Languages. The paper represents the Question Generation System for Punjabi language to generate questions for the given input in Gurmukhi script. For Punjabi, adequate annotated corpora, Part of speech taggers and other Natural Language Processing tools are not yet available in the required measure. Thus, this system relies on the Named Entity Recognition tool. Also, various Punjabi Language dependent rules have been developed to generate output based on the named entity found in the given input sentence [1].

### Poonam Gupta ,Vishal Gupta “ Algorithm for Punjabi Question Answering System”, (2013)

In this paper a hybrid algorithm for Punjabi Question Answering system has been implemented. A hybrid system that works on various kinds of question types using the concepts of pattern matching as well as mathematical expression for developing a scoring system that can help differentiate best answer among

available set of multiple answers found by the algorithm and is also domain specific like sports. This system is designed and built in such a way that it increases the accuracy of question answering system in terms of recall and precision and is working for factoid questions and answers text in Punjabi. The system constructs a novel mathematical scoring system to identify most accurate probable answer out of the multiple answer patterns. The answers are extracted for various types of Punjabi questions. The experimental results are evaluated on the basis of Precision, Recall, F-score and Mean Reciprocal Rank (MRR). The average value of precision, recall, f-score and Mean Reciprocal Rank is 85.66%, 65.28%, 74.06%, 0.43 (normalised value) respectively. MRR values are Optimal. These values are act as discrimination factor values between one relevant answer to the other relevant answer [2].

### III. Approaches for Automatic Question Generation in Integrated system:

#### 2.1 Rule Based Approach

Rule 1:-if we found name of any location, city such as “ਪੰਜਾਬ” (punjab) replace it with “kithe” in Punjabi language, “Where” in English language and “kahan” in Hindi language and end with a question mark.

Rule 2:-if we found date or time such as “15 ਅਪਰੈਲ (April) 1469” we replace it with “ਕਦੇ”(kado) in Punjabi language , When in English language, Kab in Hindi language and end with a question mark.

Rule 3:-if we found of any name we replace it with “ਕਿਸ” (kis) in Punjabi language, whose in English language, kis in Hindi language and end with a question mark.

#### 2.2 Pattern Matching Approach

In this approach a set of patterns along with all their possible questions are stored into the database [2]. When an input text in Punjabi will be given to the system, system compares the pattern of the input to the patterns stored into the database. If the pattern

matches with the pattern stored into the database then set of questions with the help of available stored questions are get generated. A hybrid system that works on various kinds of question types using the concepts of pattern matching as well as mathematical expression for developing a scoring system[10]. Limitation in this approach, If pattern does not found then system will generate the questions on the basis of Rule Based approach [10].

#### 2.3 Information extraction Approach (IE)

Information extraction is one of the Approach for question generation [3][6]. Information extraction system is using the natural language processing system to parse the question or documents returned by information retrieval [6].It is solves the problems with extraction of documents. The aim of IE technique is to search online documents collection [3]. Limitation in this technique, it is use for online text document. IE addresses the problem of transforming a corpus of textual documents into a more structured database [7].

#### 2.4 Corpus based machine translation

Corpus based machine translation systems have advantages that they are fully automatic and require significantly less labor then traditional rule based approaches. They require sentence aligned parallel sentences of paragraphs for each language pair and can not be used for language pairs for which such corpus does not exist.

#### 2.5 Statistical Machine Translation approach

In this approach translation of input text is done with the help of existing translated text. In this approach a large corpus is created which contain input text along with their translated text and output is generated according to the given translated text. This approach works in two phases which are (i) training Phase (ii) Translation phase. In the training phase, various combinations are generated and stored in the system which is used in the second phase. These combinations contain the input text along with the translated text. In the second phase actual

translation is done with the help of the combinations generated in the first phase. This approach uses a further approach N-Gram approach to generate the combinations from the input text. [12]

As we use hybrid approach, so we will integrate all these above techniques.

#### IV. CONCLUSION & FUTURE SCOPE

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. A Rule based approach, Pattern matching and Information Extraction techniques for QG has been discussed. In these approaches a lot of modifications are required to obtain more accurate results. So Hybrid approach will be used. Further, the system can also improve with new rules and can also be work on other language rather than Indian languages.

#### V. REFERENCES

- [1]. Shikha Garg, Vishal Goyal " System for Generating Question Automatically from given Punjabi Text", International Journal of Computer Science and Mobile Computing, pp.324-327 ,2013
- [2]. Poonam Gupta ,Vishal Gupta " Algorithm for Punjabi Question Answering System", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no.7, Pp.902-909 July 2013
- [3]. Vishal Gupta " A Proposed Online Approach of English and Punjabi Question Answering", International journal of Engineering Trends and Technology(IJETT)-vol. 6, no. 5, pp 292-295,dec 2013
- [4]. Wei CHEN, Gregory AIST, and jack MOSTOW "Generating Question Automatically from information Text", International Journal of Artificial Intelligence in Education,vol.21,no.4, pp 255-269,oct2013
- [5]. Husam Ali Yllias Chali Sadid A. Hasan " Automatic Question Generation From Sentences" , TALN 2010, Montreal,19-23 juillet 2010
- [6]. Poonam Gupta.,Vishal Gupta "A Survey of Text Question Answering Techniques" International Journal of Computer Application(0975-8887)Volume 53-No.4,September 2012
- [7]. Vishal Gupta"A Survey of Text Mining Techniques and Applications" J.Emerging Technologies in Web Excellence Vol.1,No.1,August 2009
- [8]. Xuchen Yao,G., "Semantic-based Question Generation and Implementation"Dialogue and Discourse3(2) (2012) 11-42,doi:0.5087/dad.2012.202
- [9]. Kamaldeep Kaur,V., "Name Entity Recognition for Punjabi Language" International Journal of Computer Science and Information Technology & Security(IJCSITS),ISSN:2249-9555 Vol.2,No.3,June 2012
- [10]. Poonam Gupta and Vishal Gupta "Hybrid Approach for Punjabi Question Answering System" springer international publishing Switzerland ,pp 133-149,2014
- [11]. Nisha Sharma"Automatic Question generation from Punjabi text"IJCEA , ISSN : 2229-3345 Vol. 6 No. 07 Jul 2015
- [12]. Meenakshi Sharma " A review paper on SMS text to plain English translation" International Journal of Computer Science & Engineering Technology (IJCSET), ISSN : 2229-3345 Vol. 5 No. 07 Jul 2014
- [13]. Deana L. Pennell and Yang Liu, Normalization Of Text Messages For Text-To-Speech , 978-1-4244-4296-6/10/\$25.00 2010 IEEE