

“Automatic Question Generation from Punjabi Text :A Review”

Nisha Sharma

M.Tech Final Year Student
GZS Campus, Bathinda
Sharma20neetu@gmail.com

Er.Abhilasha

Associated Professor
Department of Computer Science & Engg.

Abstract

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.

Keywords: NLP (Natural Language Processing), Rule Based Approach, Question Generation (QG), NER (Named Entity Recognition), Information Extraction (IE).

1. Introduction:

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]. 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]. 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) & “ਕਿਉਂ”(kio) (why) . For example for the following statement the questions can be generated as [1]

Input : ਜਵਾਹਰਲਾਲ ਨਹਿਰੂ ਦਾ ਜਨਮ 14 ਨਵੰਬਰ 1889 ਨੂੰ ਇਲਾਹਾਬਾਦ ਵਿਖੇ ਹੋਇਆ।

Output:

1. ਜਵਾਹਰਲਾਲ ਨਹਿਰੂ ਦਾ ਜਨਮ ਕਦੇ ਇਲਾਹਾਬਾਦ ਵਿਖੇ ਹੋਇਆ ?
2. ਕਿਸ ਦਾ ਜਨਮ 14 ਨਵੰਬਰ 1889 ਨੂੰ ਇਲਾਹਾਬਾਦ ਵਿਖੇ ਹੋਇਆ?
3. ਜਵਾਹਰਲਾਲ ਨਹਿਰੂ ਦਾ ਜਨਮ 14 ਨਵੰਬਰ 1889 ਨੂੰ ਕਿੱਥੇ ਹੋਇਆ ?

1.1Types of Question

QG system takes the input text and gives the output as questions. Question generation is two types [1][8] as shown in Figure1.

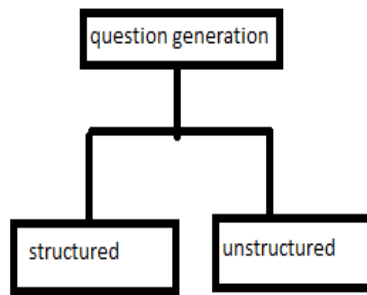


Figure1: question generation types

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 [1][8]. 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 NER (Name Entity Recognition)

NER is a tool for generating the question generation [9]. NER tool is widely used in Natural Language Processing. NER tools are performed to identifying the named entities from text input. QG made the corpus for generating the question which contains all the names related to persons, locations and other entities. Husam Ali also has been used the NER tool for QG [5]. Name entity recognition is applied on sentence to encode necessary information of questions.

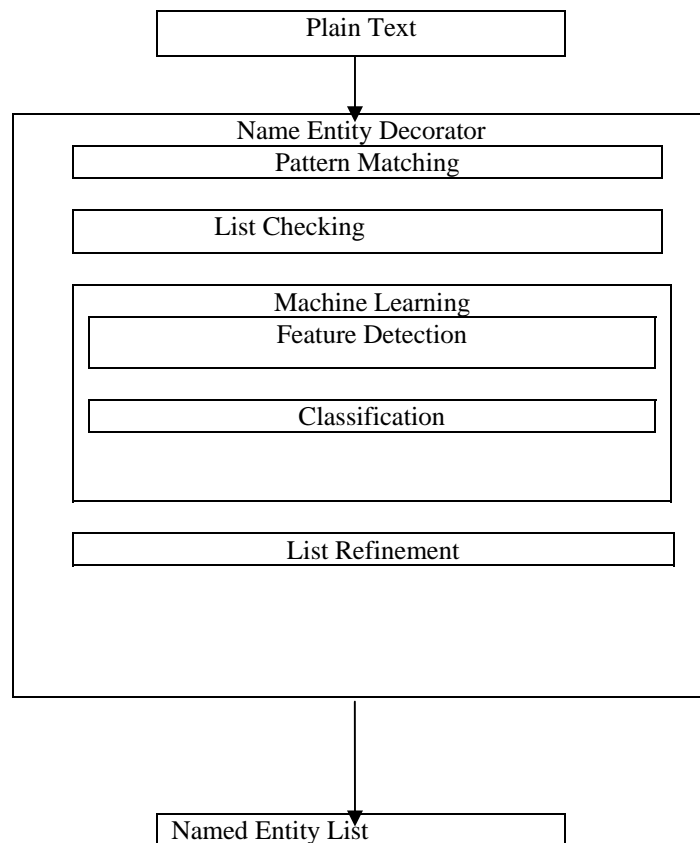


Figure2: Architecture of NER System

2. Existing Approaches for Automatic Question Generation:

2.1 Rule Based Approach is represented by Sikha Garg and Vishal Goyal in 2013 [1]. Rule defines in this paper is as follow:

Rule 1:-if we found name of any location, city such as “ਪੰਜਾਬ” (punjab) replace it with “ik`Qy”(kithe) 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) and end with a question mark.

Rule 3:-if we found of any name we replace it with “ਕਿਸ”(kis) and end with a question mark.

Rule 4:-if we found of integer type and the word is preceded by the word "ਸੰਨ" (sun), "ਸਾਲ"(sal), “ਈਸਵੀ”(esvi) then replace it with “ਕਦੇ”(kado) and with a question mark.

Limitations: It is not contain any named entity. It cannot recognize the difference between name and location entity. Kamaldeep kaur and Vishal Gupta also has been used some rules with the help of NER for QG. As many first names in Punjabi are also common nouns, this limitation lower the performance of the system [9].

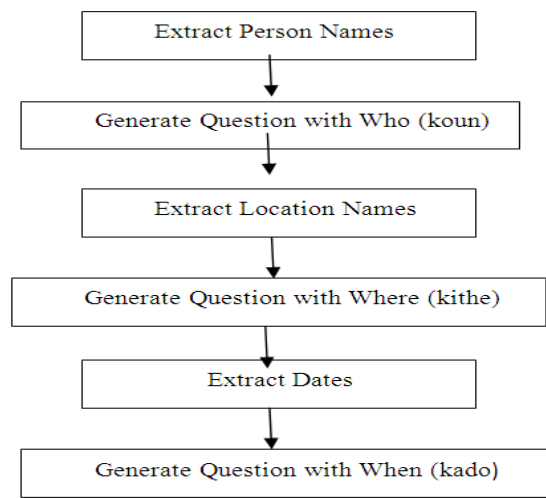


Figure4: Flow chat input sentence

Poonam Gupta and Vishal Gupta also has been used some rule for QG. In hybrid system take the inputs and generate the question answering but these answer extract from full line as well as input such as [10]

Input:- ਮੁੰਬਈ ਦੇ 26/11 ਅੱਤਵਾਦੀ ਹਮਲੇ ਵੇਲੇ ਪੂਰਾ ਦੇਸ਼ ਸੁੰਨ ਹੋ ਗਿਆ ਸੀ ।

Output:- ਮੁੰਬਈ ਤੇ ਅੱਤਵਾਦੀ ਹਮਲਾ ਕਦੇ ਹੋਇਆ ?

Answer: ਮੁੰਬਈ ਦੇ 26/11 ਅੱਤਵਾਦੀ ਹਮਲੇ ਵੇਲੇ ਪੂਰਾ ਦੇਸ਼ ਸੁੰਨ ਹੋ ਗਿਆ ਸੀ ।

By the Poonam Gupta and Vishal Gupta give overall accuracy of the existing system is 76.66% [10]. It works on the three parameters Precision, Recall, F-score. It is not like as a multiple choice question.

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].

3.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. Further, the system can also improve with new rules and using example based and dictionary lookup approach.

4.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