A Review on Question Generation System From Punjabi Text

Payal Garg[1], Er.Charandeep Singh Bedi[2]

[1] M.Tech Final Year Student
Department of Computer Science &Engg. BFGI, Bathinda

[2]Assistant Professor
Department of Computer Science &Engg. BFGI, Bathinda

Abstract
Question Genearation 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.

Keywords: Question Generation System (QGS), Natural Language Processing (NLP), Named Entity Recognition (NER)

1. INTRODUCTION

Question generation 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 This paper present the review to generate the questions from Historical documents written in Punjabi text.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. These properties of documents includes names of persons, locations, organizations , dates , monentary expressions etc. With the help of these properties questions from the text can be generated easily. For example consider the following sentence from a punjabi document `rwm cOQI jmwq iv``c hY[`From the sentence given above the following questions can be generated: `kOx cOQI jmwq iv``c hY?` As seen above question generation system generate all possible combinations of questions from the given sentence.

2. Question generation system helpful in following areas

- 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.

3. Named Entity Recognition(NER) System

To generate the questions automatically from a given Punjabi text a corpa is required in Punjabi which contain all the names related to persons, locations, cities, states, countries, and other entities. But the main problem is that there is no proper corpa available in the Punjabi language which can fulfill the requirement of our system. So a tool is also needed to create which extract the named entities from a given Punjabi text and classifies them in proper categories line location names, person names etc. These names can be used to generate the questions from the given text. Here NER(Named entity Recognition) is a tool which can generate the named entities from a given Punjabi text. NER system can be implemented by using Rule based approach in which rules are to created to extract the named entities from a given text. Accuracy of the NER system depends mainly on the rules created for the system. An accurate NER system also tends to increase the accuracy in the Question Generation (QG) system.

4. Main challenge of a Question Generation System

The main challenge in generating a question from a given text is that the question must be correct semantically. If generated question does not provide a meaning then it can not be used further. Another challenge of a question generation system is that is must generate all types of questions from a given sentence.

5. Existing work

To generate the questions automatically from a Punjabi text only rule based approach is used to generate the questions from a given Punjabi text. In rule based

International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 3, Issue 4, July-August 2014   **ISSN 2278-6856**

approach hand crafted rules are created according to the rules of the Punjabi language to generate the questions from text. Questions can only be generated if input text matched with the rule base. In the existing system for question generation from Punjabi language developed by shikha garg(2013) a rule based approach is used to generate the question automatically. The system generate the question starts with the words "kithe (Where)" , "kisne (who)", kinne (how) etc. The sysem can not generate the question starts with "kuo (why)" ,"kive (how)" etc. Hence a lot of work is left behind the system.

6. Literature Survey

Shikha garg, vishal goyal, System for Generating Questions Automatically From Given Punjabi Text, 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 current paper represents the Question Generation System for Punjabi language to generate questions for the given input in Gurmukhi script. Proposed system can not generate questions with "kuon" , "kive" etc. words.

Chin-Yew Lin

Automatic Question Generation from Queries In this paper, author propose automatic generation of questions from queries as a shared task. With large amount of cQA data available online, together with real world query logs, and interests from both aca-demics and industry, author believe that the time is ripe for such endeavor. The results would change ways that people interact with information and provide new perspectives in natural language gen-eration, information retrieval, and other related fields. [2]

Ming Liu, Vasile Rus

G-Asks: An Intelligent Automatic Question Generation System for Academic Writing Support This paper presents a novel Automatic Question Generation (AQG) system, called G-Asks, which generates specific trigger questions as a form of support for students' learning through writing. We conducted a large-scale case study, including 24 human supervisors and 33 research students, in an Engineering Research Method course and compared questions generated by G-Asks with human generated questions. The results indicate that G-Asks can generate questions as useful as human supervisors ('useful' is one of five question quality measures) while significantly outperforming Human Peer and Generic Questions in most quality measures after filtering out questions with

grammatical and semantic errors. Furthermore, authors identified the most frequent question types, derived from the human supervisors' questions and discussed how the human supervisors generate such questions from the source text. [3]

Husam Ali, Yllias Chali, Sadid A

HasanAutomatic Question Generation from Sentences

In this paper, authors proposed an approach to automatically generate questions given sentences. They used the dataset provided by the TREC 2007 Question Answering Track and evaluated the performance of their system using Recall and Precision. We filtered out important sentences from the dataset by following a target-driven method. They simplified the process by extracting elementary sentences from the complex sentences using syntactic information. After classifying the sentences based on their subject, verb, object and preposition, They generated the questions automatically from them using a predefined set of interaction rules.

7. Rule Based Approach

Rule based system contain conditional statements which are used to generate the questions from text. In rule based approach rules are created to generate the questions from the existing text written in Punjabi language. For example if a Person name is found in the sentence then question with "kisne (who)" word can be generated, if location name is found in the sentence then question with "kithe(where)" word can be generated. Accuracy of the rule based system mainly depends upon the rules that are created according to the language to generate the questions. If created rules can not apply to the input data then questions cannot be generated from this input data.

For example

This can be shown as in the following flowchart   Input Sentence
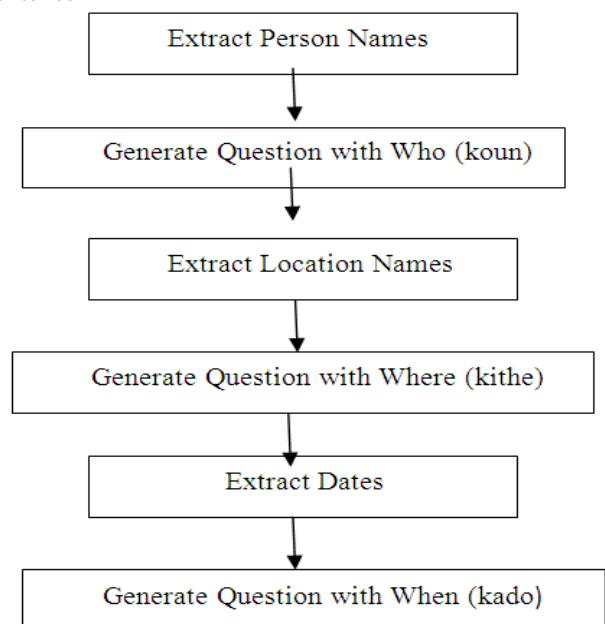
8. Conclusion

In this paper we present the review to generate questions automatically from a given Punjabi text. As discussed only rule based approach is used to generate the questions from a text. In this approach a lot of modifications are required to obtain more accurate results. Further , the system can also be improved to generate questions based on "kuo", "kive" etc. A hybrid model is required to generate the questions from a given text. Example based approach can be merged along with the rule based approach to develop hybrid approach. With the help of hybrid approach and by increasing the size of corpa more accuracy in the field of question generation can be obtained.

References

[1.] Itziar Aldabe, Maddalen Lopez de Lacalle, Montse Maritxalar,Edurne Martinez, and Larraitz Uria ArikIturri: An Automatic Question Generator Based on Corpora and NLP Techniques

[2.] Chin-Yew Lin, Automatic Question Generation from Queries

[3.] Ming Liu,Vasile Rus :G-Asks: An Intelligent Automatic Question Generation System for Academic Writing Support

[4.] Husam Ali, Yllias Chali, Sadid A. : HasanAutomatic Question Generation from Sentences

[5.] Xuchen Yao, Gosse Bouma,Yi Zhang : Semantics-based Question Generation and Implementation

[6.] Berger, A. and J. Lafferty. Information Retrieval as Sta-tistical Translation. In Proceedings of SIGIR 1999.

[7.] Cao, Y., C.-Y. Lin, Y. Yu and H.-W. Hon. Recommending Questions Using the MDL-based Tree Cut Model. In Proceedings of WWW2008.

[8.] Hovy, E., G. Laurie, U. Hermjakob, C.-Y. Lin, and D. Ravichandran Toward Semantics-Based Answer Pinpointing. In Proceedings of HLT 2001.

[9.] Xuchen Yao,Gosse Bouma, Yi Zhang , Semantics-based Question Generation and Implementation

[10.] Yushi Xu, Anna Goldie, Stephanie Seneff ,Automatic Question Generation and Answer Judging: A Q&A Game for Language Learning.

[11.] M. Lemos, R. Muralidharan, V. V. Kamat, Automatic Generation of Questions for On-line Evaluation.

[12.] Yllias Chali, Sadi d A. Hasan ,Towards Automatic Topical Question Generation ,Proceedings of COLING 2012: Technical Papers, pages 475–492, COLING 2012, Mumbai, December 2012.