Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 8, August 2014, pg.185 – 189

REVIEW ARTICLE

A Review on Question Generation System from Punjabi Text Contain Historical Information

Parshan Singh[1], Rajbhupinder Kaur[2]

[1]M.Tech Research Scholar, Department of Computer Engineering,
Yadavindra College of Engineering, Talwandi Sabo, Punjab, India
[2]Assistant Professor, Department of Computer Engineering,
Yadavindra College of Engineering, Talwandi Sabo, Punjab, India
[1] parshansranparshan@gmail.com; [2] er.rajbhupinder@gmail.com

Abstract*— Automatic Question generation is a process of generating questions automatically from a text with the help of various NLP techniques. Main challenging area while generating the questions from a text automatically is that it must be correct semantically. 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 In this paper we are presenting the review on question generation from historical documents written in Punjabi language. To generate the questions automatically from a Punjabi text a corpora in Punjabi language which contain various named entities such as names of persons, locations, cities, countries and other entities is required which is not yet available. So 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), Question Generation, Natural Language Processing (NLP), Named Entity Recognition (NER), Rule based Approach**

I. INTRODUCTION

1.1 Question generation

Questions are basic medium to learn something or to check or extract information from the existing contents. Question Generation (QG) is the task of generating reasonable questions from an input, which can be structured (e.g. a database) or unstructured (e.g. a text). Question generation is an interesting challenge in Natural Language Processing(NLP) in the field of Indian Languages such as Punjabi, Hindi Etc. Questions are used in various scenarios for example teacher ask questions from his students to test their skills, Now-a-Days questions are asked to candidates to provides jobs in various government and private sectors. Automatic question generation system helps the examiner in all these area to generate the questions. This paper presents the review to generate the questions from Historical documents written in Punjabi text. The main properties of historical documents is that they contain dates (Ex. Date of war, date of birth, date of death, date of alliance etc), various names (Ex. Akbar , Maharaja Ranjit Singh etc), Location names(Ex. Panipat ,Lahore etc) etc. so historical documents contain a lot of questions which are required to test the knowledge of the person. Examiner can skip some of these questions while preparing a test for the students but an automatic question generation system can generate all the possible question combinations from a given text.

For example consider the following sentence from a historical document.
ਗੁਰੂ ਨਾਨਕ ਦੇਵ ਜੀ ਦਾ ਜਨਮ 15 ਅਪ੍ਰੈਲ 1469 ਨੂੰ ਲਾਹੌਰ ਵਿਖੇ ਹੋਇਆ ।

From the sentence given above the following questions can be generated:
ਕਿਸ ਦਾ ਜਨਮ 15 ਅਪ੍ਰੈਲ 1469 ਨੂੰ ਲਾਹੌਰ ਵਿਖੇ ਹੋਇਆ ?
ਗੁਰੂ ਨਾਨਕ ਦੇਵ ਜੀ ਦਾ ਜਨਮ ਕਦੋ ਹੋਇਆ ?
ਗੁਰੂ ਨਾਨਕ ਦੇਵ ਜੀ ਦਾ ਜਨਮ 15 ਅਪ੍ਰੈਲ 1469 ਨੂੰ ਕਿਥੇ ਵਿਖੇ ਹੋਇਆ ?

As seen above question generation system generate all possible combinations of questions from the given sentence.

1.2 Named Entity Recognition (NER) System

To generate the questions automatically from a given Punjabi text corpora is required in Punjabi which contains all the names related to persons, locations, cities, states, countries, and other entities. But the main problem is that there are no proper corpora available in the Punjabi language which can fulfil 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.

1.3 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 cannot be used further. Another challenge of a question generation system is that is must generate all types of questions from a given sentence.

II. EXISTING WORK

A lot of work is left in the field of question generation for Indian languages. 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. From given text the system is able to generate shallow question. The system generate the question starts with the words "ਕੀ (what)" "ਕਿਥੇ (Where)", "ਕੌਣ (who)", ਕਿੰਨੇ (how many) etc. But system is unable to generate the deep question from given text. The system cannot generate the question starts with "ਕਿਉਂ (why)", "ਕਿਵੇ (how)" etc. Hence a lot of work is left behind the system. And system show good results for some kind of question but for other types of question the system show very low results. So need to add new rules to increase and improve the performance of the system.

III. LITERATURE SURVEY

In 2013 Shikha Garg and Vishal Goyal made 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 cannot generate questions with "ਕਿਉਂ (why)", "ਕਿਵੇ (how)" etc. words. [1]

In 2012 Ming Liu and 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 student 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. [2]

In 2010 Husam Ali, Yllias Chali and Sadid A. Hasan: Automatic 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.[3]

In 2008 Chin-Yew Lin: Automatic Question Generation from Queries. In this paper, author proposes 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 academics and industry, author believe that the time is ripe for such endeavour. The results would change ways that people interact with information and provide new perspectives in natural language generation, information retrieval, and other related fields. [4]

IV. RULE BASED APPROACH

Rule based approach used in question generation system. Some dependent rules for Punjabi language have been developed for the generation of questions. Some of these rules are:

Rule 1: - If we found name of any person entity such as "ਵਿਵੇਕ" we replace it with "ਕੌਣ" and end with a question mark. To find the name we use certain rules such as:-

a) If there is a prefix such as "ਮਿਸਟਰ"
b) If there is a suffix such as "ਸਿੰਘ"
c) Direct names from the database
d) There is a special table to analyse the words which cannot be used as name.

Rule 2:- If we found name of any location, city, country, state entity such as "ਰੋਡ" or "ਜਲੰਧਰ" or "ਪੰਜਾਬ" or "ਪਾਕਿਸਤਾਨ" we replace it with "ਕਿਥੇ" and end with a question mark. To find the location we use certain rules such as:-

a) If there is a prefix such as "ਪੁਰ" for "ਸੁਲਤਾਨਪੁਰ".
b) If there is a suffix such as "ਰੋਡ" for "ਮੁਲਤਾਨ ਰੋਡ".
c) Direct names from the database
d) There is a special table to analyse the words which cannot be used as location.

Rule 3:- If we found date or time such as "22 march 1982" or "1989" we replace it with "ਕਦੋਂ" and end with a question mark. To find the date we use certain rules such as:-

a) If there is a prefix such as "ਈ" or "ਵਿੱਚ" for the year.
b) Special month recognizer.
c) There is a special table to analyse the words which cannot be used as date.

Rule 4:- If we found name of any organization we replace it with "ਕਿਸ" and end with a question mark. To find the organization we use certain rules such as:-

a) If there is a prefix such as "ਮਿਸਲ" or "ਕਮੇਟੀ" and so on.
b) There is a special table to analyse the words which cannot be used as organization.

This can be shown as in the following flowchart:

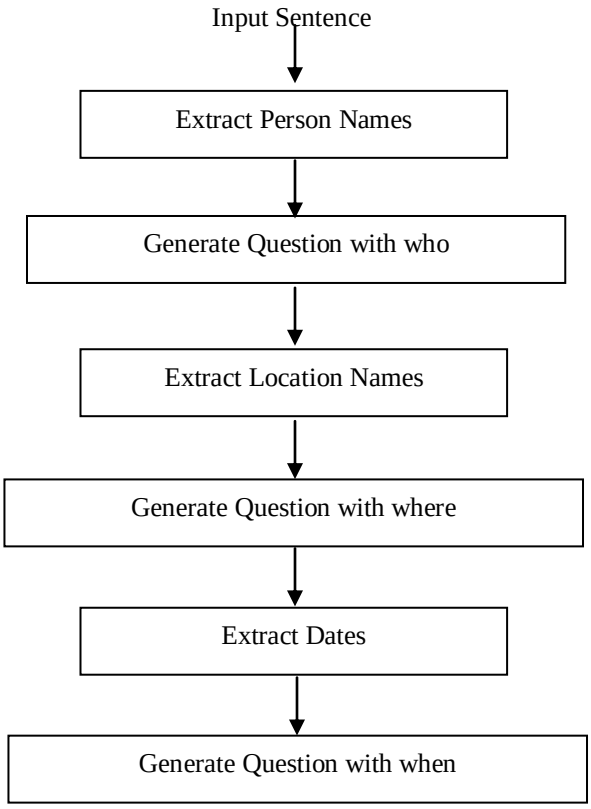

Fig. 1 Flowchart for Question Generation

V. CONCLUSION & FUTURE SCOPE

In this paper we present the review to generate questions automatically from a given Punjabi text contain historical information. As discussed a large corpora is needed to generate questions from given sentences. A lot of work is left to develop a question generation system for Indian languages such as Punjabi, Hindi etc. Present system is based n rule based approach in which a lot of modifications are required in rules to achieve more accuracy. Further, the system can also be improved to generate questions based on "kyo", "kive" etc. A more accurate question generation system is required to be developing so that it can be used by teachers, examiners, and researchers etc to explore or test one's skill.

The method proposed is simple and easy to implement. One of the limitations of the current implementation is its usefulness over limited domain. The method is too specific and lacks necessary framework for other domains. The proposed future work is in the direction of extending the system to a general framework.

REFERENCES

[1] S. Garg and V. Goyal, "System for Generating Questions Automatically from Given Punjabi Text Question Generation", March 23, 2013
[2] M. Liu, R. A. Calvo and V. Rus, "G-Asks: An Intelligent Automatic Question Generation System for Academic Writing Support", ICITF, Dialogue and Discourse 3(2) (2012), pp. 101–124.
[3] H. Ali, Y. Chali and S. A. Hasan "Automatic Question Generation from Sentences" in 2010
[4] Chin-Yew Lin, "Automatic Question Generation from Queries"
[5] V. Gupta, "A Proposed Online Approach of English and Punjabi Question Answering", 2013

[6] S. Pandey and K.C Rajeswari, "Automatic Question generation Using Software Agent for Technical institutions", December 2013
[7] N. Iwane, C. Gao and m. Yoshida, "Question Generation for Learner centered learning" IEEE 13th International Conference on advanced Learning Technologies 2013.
[8] Y. Chali and S. A. Hasan, "Towards Automatic Topical Question Generation", Proceedings of COLING 2012: Technical Papers, pages 475–492, COLING 2012, Mumbai, December 2012.
[9] X. Yao, G. Bouma and Y. Zhang, "Semantics-based Question Generation and Implementation", Dialogue and Discourse, pp 11- 42, 2012
[10] K. Kaur and V. Gupta, "Name and Entity Recognition for Punjabi Language", June 2012.
[11] M. Liu, R. A. Calvo and V. Rus, "Hybrid Question Generation Approach for Critical Review Writing Support", Proceedings of the 20th International Conference on Computers in Education. Singapore: Asia-Pacific Society for Computers in Education 2012.
[12] V. Gupta and G. S. Lehal, "Named Entity Recognition for Punjabi Language Text Summarization", International Journal of Computer Applications (0975 – 8887) Vol.33, No.3, pp. 28-32 in 2011.

Authors Bibliography:

Parshan Singh has received his B. Tech degree in Computer Engineering from Yadavindra College of Engineering, Punjabi University Guru Kashi Campus, Talwandi Sabo (Bathinda) in 2012.
He is pursuing M. Tech (Regular) degree in Computer Engineering from Yadavindra College of Engineering, Punjabi University Guru Kashi Campus, Talwandi Sabo (Bathinda). His research interests include natural language processing.

Rajbhupinder Kaur has received her M. Tech Degree from Punjabi University, Patiala in 2010 and B-Tech degree from Punjab Technical University, Jalandhar in 2006. She is working as Assistant Professor in Yadavindra College of Engineering, Talwandi Sabo, Bathinda Punjab. Her research interests are in the fields of Mobile Ad-Hoc Network, Network Security, Nanotechnology, wireless sensor networks. She has published many national & international papers.

