

International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IJCSMC, Vol. 3, Issue. 10, October 2014, pg.86 – 91

RESEARCH ARTICLE

Rule Based Question Generation System from Punjabi Text Contain Historical Information

Parshan Singh¹, Rajbhupinder Kaur²

¹M.Tech Research Scholar, Department of Computer Engineering, Yadavindra College of Engineering, Talwandi Sabo, Punjab, India

²Assistant Professor, Department of Computer Engineering, Yadavindra College of Engineering, Talwandi Sabo, Punjab, India

¹parshansranparshan@gmail.com; ²er.rajbhupinder@gmail.com

Abstract— Automatic Question generation is a computer program of generating questions automatically from a text. This is possible with the help of NLP (Natural Language Processing). Main challenging of this is generating the questions from a text that are must be syntax and semantically correct. Rule based approach used to generating the questions automatically from a text. In this paper we are presenting rule based approach to generate question from Punjabi historical documents or text. Punjabi corpus used to generate the questions automatically from a Punjabi text. This corpora contain various named entities such as names of persons, locations name, cities name, countries name 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, Rule based Approach, Natural Language Processing (NLP), Named Entity Recognition (NER)

I. INTRODUCTION

1.1 Question generation

Questions are used to extract usefully information from the text. This is essential element of learning. These are used in many applications such as student learning, in medicine, in security contexts, in FAQ, etc. Now-a-Days questions are asked to candidates to provide jobs in various governments and private sectors Question generation is task of generating the question from the text. Question Generation (QG) is the task of generating reasonable questions from an input or text, input 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. Question generation is not easy. It is very complicated to generate the question from text. But the generation of question from sentence is easy as compare generation of question from the paragraph. Question Answering system are not useful until question generate automatically. On the basis of complexity question are divided into two categories. Firstly Shallow questions (what, where when etc) and secondly deeply

questions (why, how). This paper presents the rule based approach to generate the questions from Punjabi Historical documents written. 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. Banda Singh bahadur, bhagat Singh etc), Location names (Ex. Patna, Jhansi etc) etc. so historical documents contain a lot of questions which are required to test the knowledge of the person.

For example consider the following sentence from a historical document.

ਭਗਤ ਸਿੰਘ ਦਾ ਜਨਮ 28 ਸਤੰਬਰ 1907 ਨੂੰ ਪਿੰਡ ਬੰਗਾ ਵਿਖੇ ਹੋਇਆ।

From the sentence given above the following questions can be generated:

1. ਭਗਤ ਸਿੰਘ ਦਾ ਜਨਮ ਕਦੇ ਹੋਇਆ (based on date format)
2. ਕਿਸਦਾ ਜਨਮ 28 ਸਤੰਬਰ 1907 ਨੂੰ ਪਿੰਡ ਬੰਗਾ ਵਿਖੇ ਹੋਇਆ (Based on person name)
3. ਭਗਤ ਸਿੰਘ ਦਾ ਜਨਮ 28 ਸਤੰਬਰ 1907 ਨੂੰ ਕਿਥੇ ਹੋਇਆ (based on location name)

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

1.2 Named Entity Recognition (NER) System

It is task to identity and classifies the element from the text into predetermine classes. This tool is very necessary to generate the question from the text. Question generation system generate question based on the Punjabi corpus. There is no corpus for the Punjabi language. So name entity recognition tool is used to create the Punjabi corpus. This corpus contains all the names related to persons, locations, cities, states, countries, and other entities. So a tool is also needed to create which extract the named entities from a given Punjabi text and classifies element into predefined classes like location names, person names etc.

II. RELATED WORK

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 converts a simple sentence to its interrogative sentence. It accepts sentences as an input which can be structured or unstructured. And generate 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)”, “ਕਿਵੇਂ (how)” etc. words. The recall value of the system is 46%. That is very low so need to improve it. [1]

In 2012 Itziar Aldabe, Itziar Gonzalez Dios and Montse Maritxalar: Two Approaches to Generate Question in Basque. In this paper author developed two systems to generate the question for Basque language. Both systems generate at question at the sentence level. Both systems use name entity recognition tool to generate the question from the sentence. These experiments were carried out based on the educational text. Then a linguist analysed the grammatically and appropriateness of the question generated from the sentence as well as their interrogative pronoun. [2]

In 2010 Prashanth Mannemx, Rashmi Prasady, and Aravind Joshiy: - Question Generation from Paragraphs at UPenn: QGSTEC System Description. In this paper author explained the question generation system developed at UPenn for QGSTEC, 2010. The system analysis structures of sentences to generate the question from this. System takes paragraph as input and generates a set of six question from the paragraph. The generated questions are then ranked to pick final six best questions Semantic roles labeling (SLR) is used to analysis the structure of the system. Question generation system has three stages to generate the question. Content selection, question formation and ranking. The performance of the system is 85 percentages. [3]

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. [4]

In 2006 Itziar Aldabe, Maddalen Lopez de Lacalle, Montse Maritxalar, Edurne Martinez, and Larraitx Urria: - Arikiturri: an Automatic Question Generator Based on Corpora and NLP Techniques. In this paper authors designed system Arikiturri that generate question automatically fro that text. Input of the system represent by XML mark up language. And output of the system also represent in XML mark up language. This system generate four types of question fill-in-blank, word formation,

multiple choice, and error correction question types. Author obtained positive results regards the system. And performance of the system is more than 80%. [5]

III. PROPOSED SYSTEM

The proposed system use rule based approach to generate the possible questions from a text paragraph. Rules are developed by considering the features of Punjabi language. Proposed system is also integrated with the NER system for Punjabi language which generates the possible named entities from a Punjabi text paragraph. For system handcraft rules are to be made from reference of input the given text. This system also generate question of types “ਕਿਉ” and “ਕਿਵੇਂ”.

Proposed system uses the following model to generate the question from the text paragraph. For Example if question is to be generated from the line containing person name the following steps will be used:

Question Rule to generate the question from named entity:

Step 1: Read the input string

Step 2: Extract the name from the input string. If name does not found then run NER tool to extract the name from the input.

Step 3: if name is found in the step 2 then replace the name with “ਕੋਣ” word.

Step 4: perform the post processing on the input string and generate the final question.

Step 5: Concatenate the sign of question mark (?) at the end of question.

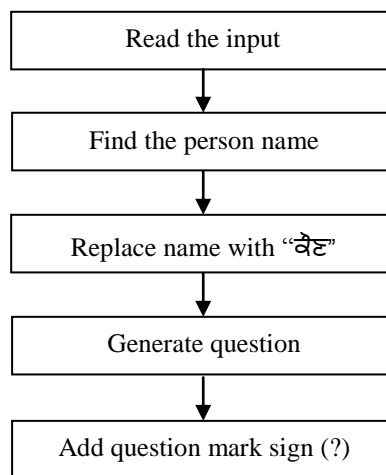


Fig. 1:- Architecture of Question Generation for Punjabi Language.

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. The following rules are developed to generate the questions from the text paragraph of Punjabi language:

Rule 1: If person name is found followed by words such as “ਨੇ”, “ਨੂੰ” etc. then question with “ਕਿਸ” keyword is generated.

Rule 2: If location name is found followed by words such as “ਦਿੱਚ”, “ਚ” etc. then question with “ਕਿੱਥੇ” keyword is generated.

Rule 3: If direction name is found followed by words such as “ਦਿਸ਼ਾ” etc. then question with “ਕਿਸ ਦਿਸ਼ਾ” keyword is generated.

Rule 4: If one of date formats dd.mm.yyyy, mm.dd.yyyy, dd-String-YYYY is found followed by words such as “ਦਿੱਚ”, “ਚ” etc. then question with “ਕਦੇ” keyword is generated.

Rule 5: if the strings “ਇਸਲਈ” and “ਕਿਉਂਕੀ” are found in the line from which question is to be generated then extract the string from the ending index of ਇਸਲਈ and starting index of ਕਿਉਂਕੀ. Add word “ਕਿਉ” in the starting of the extracted string to make it question.

Rule 6: if “ਈਸਵੀ” or “ਸੰਨ” words are found followed by the numeric value then question with “ਕਦੇ” keyword is generated.

Rule 7: If monetary expression is found followed by words such as ਰੁਪਈਆ, "ਪੈਸਾ" etc. then question with “ਕਿੰਨੇ” keyword is generated.

V. EVALUATION & RESULTS

Proposed system tested on 400 different sentences to evaluate the results. Results obtained by our system are very good. Proposed system uses rule based approach along the NER tool developed to generate the questions. The overall precision of the system for the test data has been calculated to 85.50% which shows very good improvement over existing question generation systems.

Various types of questions can be generated by the system. The following table will compare the types of questions generated from the existing system and by proposed system.

Question Type	Existing Systems	Proposed System
ਕਿੱਥੇ (Kithe)	Yes	Yes
ਕਿਸ (Kis)	Yes	Yes
ਕੀ (Ki)	Yes	Yes
ਕਦੇ (Kado)	Yes	Yes
ਕਿੰਨੇ (Kinee)	Yes	Yes
ਕਿਉਂ (Kyo)	No	Yes
ਕਿਵੇਂ (Kive)	No	Yes
ਦਿਸ਼ਾ(Direction based)	No	Yes
Monetary Expressions	No	Yes

Table 1:- Comparison between Existing and Proposed system.

Parameters to evaluate the results of the system:

- I. **Recall Value:** this is value to total number of questions that are generated by the system to the total number of question that can be generated manually by human being.
- II. **Precision:** This is the total number of accurate questions from all the questions generated by the system.
- III. **F-Measure:** This is defined as the harmonic mean of the recall and precision.

Question Type	Recall	Precision	F-Measure
ਕਿੱਥੇ (Kithe)	92.89%	91.56%	91.25%
ਕਿਸ (Kis)	62.86%	74.52%	65.67%
ਕੀ (Ki)	85.50%	70.97%	77.56%
ਕਦੇ (Kado)	89.79%	94.79%	91.75%
ਕਿੰਨੇ (Kinee)	91.24%	98.30%	94.64%
ਕਿਉਂ (Kyo)	82.01%	95.09%	88.06%

ਕਿਵੇਂ (Kive)	70.81%	84.79%	77.17%
ਦਿਸ਼ਾ(Direction based)	90.68%	88.22%	89.43%
Monetary Expressions	89.22%	99.27%	93.97%
Total	83.88%	88.61%	85.50%

Table 2:- Result table of proposed system

VI. 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] I. Aldabe, I. Dios and M. Maritxalar (2013), “Two Approach to Generate Question in Basque”, pp 101 – 108 September 2013
- [3] P. Mannem, R. Prasad, and A. Joshi, “Question Generation from Paragraphs at UPenn: QGSTEC System Description”, March 22, 2013
- [4] H. Ali, Y. Chali and S. A. Hasan “Automatic Question Generation from Sentences” in 2010
- [5] I. Aldabe, M. Lacalle, M. Maritxalar, E. Martinez, and L. Uria (2006), “ArikIturri: An Automatic Question Generator Based on Corpora and NLP Techniques”, ITS 2006, LNCS 4053, pp. 584 – 594.
- [6] Chin-Yew Lin, “Automatic Question Generation from Queries”
- [7] V. Gupta, “A Proposed Online Approach of English and Punjabi Question Answering”, 2013
- [8] S. Pandey and K.C Rajeswari, “Automatic Question generation Using Software Agent for Technical institutions”, December 2013
- [9] N. Iwane, C. Gao and m. Yoshida, “Question Generation for Learner centered learning” IEEE 13th International Conference on advanced Learning Technologies 2013.
- [10] 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.
- [11] X. Yao, G. Bouma and Y. Zhang, “Semantics-based Question Generation and Implementation”, Dialogue and Discourse, pp 11- 42, 2012
- [12] K. Kaur and V. Gupta, “Name and Entity Recognition for Punjabi Language”, June 2012.
- [13] 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.
- [14] 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.