

Automatic Question Generation System from Punjabi Text using Hybrid approach

Payal Garg¹, Er.Charandeep Singh Bedi²

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

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

Abstract: This paper represents a system for automatically question generation from Punjabi text. Asking questions is a fundamental cognitive process that underlies higher-level cognitive abilities such as comprehension and reasoning. Questions are used from the most elementary stage of learning to original research. Question Generation (QG) is the task of automatically generating questions from various inputs such as raw text, database, or semantic representation. In this system various Punjabi language dependent rules and examples have been developed to generate output based on the given input. Questions are generated by the proposed system on the basis of these rules and examples. System use corpus for Punjabi Named Entities and examples to generate the desired questions.

Keywords: Automatic Question generation system, Rule Based approach, Example based approach, Hybrid approach.

INTRODUCTION

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). In this paper, we narrow the task of QG down to taking a natural language text as input (thus textual QG), as it is a more interesting challenge that involves a joint effort between Natural Language Understanding (NLU) and Natural Language generation (NLG).. The task of question generation contains multiple subareas. Question generation is not so easy. It is very difficult to generate the questions from text. But the generation of question from text is easy as compared to paragraph. In terms of **target complexity**, QG can be divided into deep QG and shallow QG (Grasser et al., 2009). Deep QG generates deep questions that involve more logical thinking (such as why, why not, what-if, what-if-not and how questions) whereas shallow QG generates shallow questions that focus more on facts (such as who, what, when, where, which, how many/much and yes/no questions).

This paper presents the Hybrid approach to generate the questions from given input text. Hybrid approach is a combination of Rule Based Approach, Dictionary Lookup

Approach and Example Based approach. The main properties that any Punjabi text is that it contains are names of persons, locations, cities, state, country, date format , measurement values, directions etc. which are called as Named Entites(NEs). Automatic question generation system generate the questions on the basis of these named entites.

For example from the Named Entites we generate question like

1. ਬਾਨੂ ਬੁਢਾ ਹੋ ਗਿਆ ਸੀ(based on name)

Question : □□□ □□□□ □□ □□□ □□??

2. ਨਾਨਕ ਦਾ ਜਨਮ 15 ਅਪਰੈਲ 1469 ਨੂੰ ਹੋਇਆ (based on date)

ਨਾਨਕ ਦਾ ਜਨਮ kdo ਹੋਇਆ ?

3. ਜਵਾਹਰਲਾਲ ਦਾ ਜਨਮ ਇਲਾਹਾਬਾਦ ਵਿਖੇ ਹੋਇਆ (based on location)

ਜਵਾਹਰਲਾਲ ਦਾ ਜਨਮ □□□□□ ਹੋਇਆ ?

As seen above question generation system generate all possible combination of question from given sentences.

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 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 endeavor. The results would change ways that people interact with information and provide new perspectives in natural language generation, information retrieval, and other related fields.

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 source 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.

Husam Ali, Yllias Chali, 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. They 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. They plan to extend the number of interaction rules in the future. They will also focus on the sentence classification module to make it more robust. Since human generated questions always tend to have words with different meanings and senses, the system can be improved with the inclusion of semantic information and word sense disambiguation.

PROPOSED SYSTEM

The proposed system use the hybrid approach (Rule based and example based approach) to generate the possible set of

questions from given input. Rules are developed by considering the features of Punjabi language. Proposed system is also integrate the NER system for Punjabi language. This system also generate the question like (kyun , kiho jiha).

Proposed system uses the following model. For example if question is to be generated from the line containing person name the following steps to be used.

Step1: Read the input string.

Step2: Extract the name from input string, if the name does not found then run NER tool to extract the name from input.

Step3: if the name is found in 2 step then replace the name with "kiss" or "kon" word.

Step4: perform the post processing and generate final input.

Rule based approach: Rule based approach used in question generation. Some rules for Punjabi language have been developed. The following rules are developed to generate the questions from the text of Punjabi language.

Rule 1: if person name is found in the sentence then "ਕਿਸ" or "ਕੌਣ" type of question can be generated.

Rule 2: if location name is found in the sentence then "ਕਿੱਥੇ" type of question can be generated.

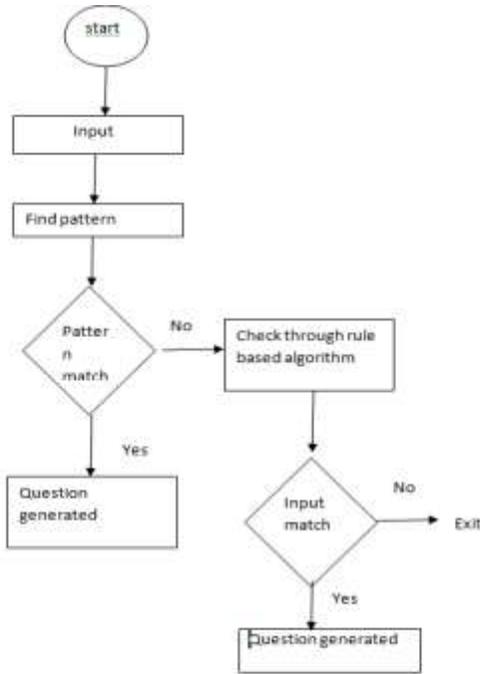
Rule 3: if any type of date format is found in the sentence then "ਕਦੇ" type of question can be generated.

Rule 4: if the strings "ਇਸੀਲਈ" and "ਕਿਉਂਕੀ" is found in the sentence then "ਕਿਉਂ" type of question can be generated.

Rule 5: if "ਸ਼ੀਸਵੀ" or "" word is found followed by the numeric value then "ਕਦੇ" type of question can be generated.

Rule 6: if monetary expression is found followed by words like "rupees" or "paise" then "ਕਿਨੇ" type of question can be generated.

Example based approach:



किने (kinne)	yes	yes
किउ (kyun)	No	yes
Direction based	No	yes
Monetary expression	No	yes
किचे जिहा	No	yes

Results and evaluation:

Proposed system tested on 500 different sentences to evaluated the results. Results obtained by our system are good. The overall precision of the system is 89.5%.which shows good improvement over existing system.

Various types of questions can be generated by the system. The following table will compare the types of question generated between the existing system and purposed system.

Question Type	Existing System	Purposed system
किथे (kithe)	yes	yes
किस (kiss)	yes	yes
कोन (kon)	yes	yes

Parameters to evaluate the results of the system:

- 1. Recall value:** This is the value to total no of question that are generated by the system to the total number of question that can be generated manually.
- 2. Precision value:** this is the total number of accurate questions generated by the system.
- 3. F- measure:** This defines the mean of recall and precision value.

Question type	recall		precision		f-measure	
	old	new	old	new	old	new
किथे	56.6	92.90	74.5	91.56	64.32	91.26
किस	50	62.86	62.1	74.52	55.39	65.67
कोन	48.23	87	73	71	58.13	77.56
किने	52.33	89.79	51.7	94.79	52.01	91.75
किउ	0	91.24	0	98.50		94.64

Conclusion and future work

In this paper we present to generate questions automatically from given Punjabi text. As discussed a large Punjabi corpora is needed to generate questions. A lot of work is left to develop question generation system for Indian languages such as Punjabi, Hindi etc. Present system is based on rule based and example based in which lots of modifications are required.

REFERENCES

- [1] **Xuchen Yao.** Question Generation with Minimal Recursion Semantics. Master's thesis, Saarland University & University of Groningen, 2012[8]
Yi Zhang, Valia Kordoni, and Erin Fitzgerald. Partial Parse Selection for Robust Deep Processing. In Proceedings of ACL 2007 Workshop on Deep Linguistic Processing, 2007
- [2] Y. Xu and S. Seneff, "Mandarin Learning Using Speech and Language Technologies: A Translation Game in the Travel Domain," in *ISCSLP*, Kunming, China, 2008, pp.[9]
- [3] N. Swapna , N.Hareen Kumar, B. Padmaja Rani "INFORMATION RETRIEVAL IN INDIAN LANGUAGES: A CASE STUDY ON CROSS-LINGUAL AND MULTI-LINGUAL" , International Journal of Research in Computer and Communication technology,,2012[3]
- [4] Manish Agarwal_, Rakshit Shah_ and Prashanth Mannem Automatic Question Generation using Disc Theirse Cues[4]
- [5] *Y Ilias Chali Sadid A. Hasan Towards Automatic Topical Question Generation*[5]
- [6] Itziar Aldabe1, Maddalen Lopez de Lacalle1, Montse Maritxalar1, Edurne Martinez2, and Larraitz Uria1 **ArikIturri: An Automatic Question Generator Based on Corpora and NLP Techniques**[6]
- [7] Prashanth Mannemx, Rashmi Prasad c , and Aravind Joshiy Question Generation from Paragraphs at UPenn: QGSTEC System Description [7]
- [8] Itziar Aldabe, Maddalen Lopez de Lacalle, Montse Maritxalar, Edurne Martinez, and Larraitz Uria ArikIturri : An Automatic Question Generator Based on Corpora and NLP Techniques
- [9] **Chin-Yew Lin,** Automatic Question Generation from Queries
- [10] **Ming Liu, Vasile Rus** :G-Asks: An Intelligent Automatic Question Generation System for Academic Writing Support
- [11] **Husam Ali, Yllias Chali, Sadid A.** : Hasan Automatic Question Generation from Sentences
- [12] **Xuchen Yao, Gosse Bouma, Yi Zhang** : Semantics-based Question Generation and Implementation
- [13] **Berger, A. and J. Lafferty.** Information Retrieval as Statistical Translation. In Proceedings of SIGIR 1999.
- [14] **Cao, Y., C.-Y. Lin, Y. Yu and H.-W. Hon.** Recommending Questions Using the MDL-based Tree Cut Model. In Proceedings of WWW2008.
- [15] **Hovy, E., G. Laurie, U. Hermjakob, C.-Y. Lin, and D. Ravichandran** toward Semantics-Based Answer Pinpointing. In Proceedings of HLT 2001.
- [16] **Xuchen Yao ,Gosse Bouma, Yi Zhang** , Semantics-based Question Generation and Implementation
- [17] **Yushi Xu, Anna Goldie, Stephanie Seneff** ,Automatic Question Generation and Answer Judging: A Q&A Game for Language Learning.
- [18] **M. Lemos, R. Muralidharan, V. V. Kamat,** Automatic Generation of Questions for On-line Evaluation.