International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 9
September 2017

A conceptual review of Automatic Question Generation from a given Punjabi Text

Ashminder Singh Gill
M.Tech. Scholar
Department of Computer Engineering
Punjabi University, Patiala

Gagandeep kaur Virk
Assistant Professor
Bahra Group of Institutes
Patiala Campus

Abhinav Bhandari
Assistant Professor
Department of Computer Engineering
Punjabi University Patiala

Abstract— *Questions simulate the mind to think and probably arrive at some answer. It is a reasonable thing to make questions a part of the learning process. Questions, asked by anyone, can also address our common queries about the subject in question. The current scenario involves vast amount of digital information in the form of websites, blogs, forums, online course, tutorials etc. Hence, questions in the form of quiz, FAQ's, queries etc can form a significant part of the digital media. Using the tools and techniques available in NLP, algorithms for automatic question generation (AQG) have been designed and implemented. In this paper, different concepts related to Automatic Question Generation from given text and work done by various researchers on present topic of research using different methodologies have been reviewed.*

Keywords: Automatic Question Generation (AQG) , FAQ, NLP etc.

1. INTRODUCTION TO NATURAL LANGUAGE PROCESSING

Natural Language Processing came into existence in 1950s. NLP research focuses on various research areas like machine translation, information retrieval, text summarization, question answering, information extraction, topic modeling, and opinion mining or Sentiment analysis. Natural Language Processing is an Artificial Intelligence method used by humans for communicating with computer systems using natural languages such as English, Hindi, Punjabi etc. It is also concerned with programming computers to process large natural language corpora for different applications like text summarization, question generation, translation, sentiment analysis etc. The input and output for an NLP system can be:

- Speech
- Written Text

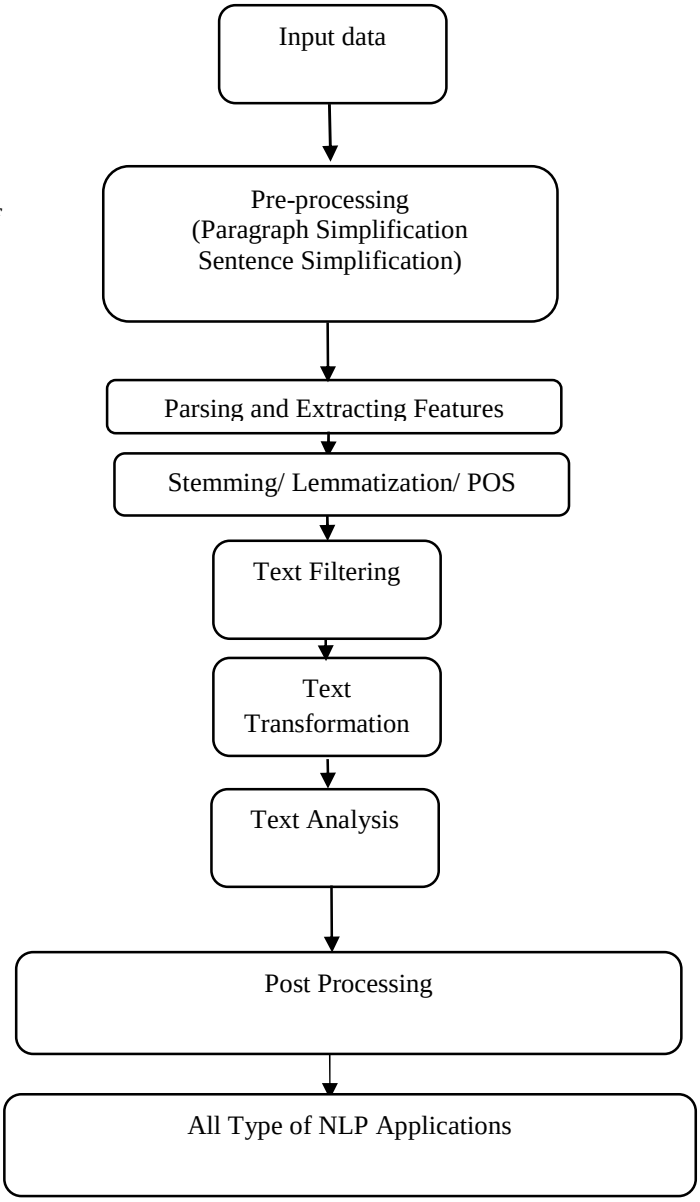

Fig.1.1. NLP workflow

,

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 9
September 2017

1. NLP TASKS

- **Morphological segmentation**

This task is used to separate words into individual morphemes and identify the class of the morphemes (small units of text that carry some meaning). Morphology means the structure of words of the language being considered. English has fairly simple morphology.

- **Sentence breaking**

It finds the sentence boundaries by identifying periods or other punctuation marks and then breaks the paragraphs into sentences.

- **Stemming/Word segmentation**

It divides the continuous text into separate words. For a language like English, this task is easier because words are usually separated by spaces. But in languages like Chinese, Japanese and Thai, no word boundaries are present in form of spaces, so it becomes difficult in these languages to perform stemming.

- **Part Of Speech Tagging (POS)**

POS tagging is one of the most important task performed in Natural Language Processing. It basically determines all the parts of speech- Noun, Pronoun, Verb, Adjective, Adverb, Cardinals, Punctuation, Echo-Words etc. in the given text.

- **Named Entity Recognition (NER)**

NER helps to determine the items from the text that map to proper names, such as people or places, and what type of each such name is- person, location, organization. Capitalization can also help in recognizing named entities in English like languages but it cannot determine the type of named entity, and is often inaccurate. There are many other languages in non-Western scripts like Chinese or Arabic that do not have any capitalization at all, and even languages with capitalization may not consistently use it to distinguish names. For example, in German language all nouns are capitalized, regardless of whether they refer to names, and in French and Spanish, the names that serve as adjectives are capitalized.

- **Sentence Simplification**

This is used to reduce the complexity of question generation.

1.2 NLG AND CONCEPTS

1.2.1 Natural Language Generation (NLG)

David Lindberg, 2010 explains NLG as part of Natural Language Processing (NLP), which can be more broadly described as the study of computers interacting with human language.NLP also includes Natural Language Understanding (NLU), which can be thought of as the inverse of NLG. An NLU system seeks to convert human language into a machine-understandable form.

Question Generation (QG) from text is an NLG task concerned with generating questions from unstructured text. It uses both NLU and NLG because in order to generate questions from text, we first have to understand it and that understanding may be deeper or shallow [13]. Natural Language Generation (NLG) is one of the grand challenges of natural language processing and artificial intelligence that is concerned with construction of a system that can generate human understandable text.

1.2.2 QUESTION GENERATION SYSTEM

Questioning has been shown to improve learning outcomes, and automatic question generation can greatly facilitate the inclusion of questions in learning technologies such as intelligent tutoring systems.

According to **Lauer et. al., 1992; Piwek et al., 2007**, Question Generation is the most required component of learning systems. It helps the information seeking systems, conversations between virtual agents, and tutoring systems to help the learners have deep insight into the topic by generating questions from the given text input.

Rus & Graesser, 2009 defines Question Generation as the task of automatically generating questions from given input text. The input text can be either in form of database or given in form of raw text.

Question Generation is a three-step process [7]:

- *Content selection*: It selects the target of the question about which (or whom), the question is to be asked.
- *Selection of question type*: It selects the appropriate question-word to be used for the target i.e.- Question type.

,

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 9
September 2017

- *Question construction*: It generates the question(s).

V. Goyal et al., 2016 developed Question Generation System for Punjabi text using Named Entity Recognition (NER) tool because the authors claim that annotated corpora, name dictionaries, good morphological analyzers, POS taggers and other NLP tools are not available for Punjabi Language. The various tags of NER used by their system to generate questions include Person Name, Location, Designation, Number, Date/Time, Abbreviations, Title Person, and Measure. Based on these NER tags, they developed some Punjabi language dependent rules to generate questions. This system generated shallow questions.

From this paper, it can be concluded that without POS tagged data, larger number of comparisons has to be done to find out entities in a given text. The authors state that "POS tagged data and stemmer is very essential for QG system" [7] but they have not used any POS tagger for their QG. So, they leave it as a future scope to improve the performance of the system.

1.2.2.1 QUESTION GENERATION TASKS

According to the first Shared Task Evaluation Challenge on Question Generation (QG-STEC), QG tasks were classified into two categories namely [22]:

- Question Generation from Paragraphs (Task A)
- Question Generation from Sentences (Task B).

1.2.2.2 TYPES OF QUESTIONS

Questions are very helpful in learning systems as well as basic informative systems. For instance, suppose we do not know something, the quite obvious thing to do is to ask about it. Depending on the purpose of the information being gathered from given text, two types of questions can be generated [7]:

DEEP QUESTIONS: The questions that involve more logical thinking (such as why, why not, what-if, what-if-not and how questions) are generated by Deep QG. Deep questions allow multiple perspectives and reflective answers. Past work has shown the use of deep-reasoning questions to be significantly correlated with student learning.

SHALLOW QUESTIONS: The questions that focus more on facts (such as who, what, when, where, which, how many/much and yes/no questions) are generated by shallow QG.

1.2.2.3 APPLICATIONS OF QG

Some applications of QG are as follows [20]:

- QG System can suggest good questions that can help learners to read and understand documents and other media.
- Questions can promote and assess deeper learning.
- QG system can suggest questions for patients and caretakers in medical field.
- QG system can suggest questions can be asked in legal contexts by litigants or in security contexts by interrogators.
- Questions automatically generated from information repositories can help to generate Frequently Asked Question (FAQ) facilities.
- QG system can be helpful for creating hint and prompts in an intelligent tutoring system.

2. QUESTION GENERATION IN PUNJABI LANGUAGE

The research work on Question Generation for Punjabi language is in its initial stage, including methodology examination, evaluation criteria selection and dataset preparation, etc. It is a very challenging task to generate questions in Punjabi Language. Some of the challenges are:

- There is no concept of capitalization in Punjabi. But in English like languages, most of the Named Entities are represented by writing first letter capital (e.g. 'Punjab').
- **Syntax of the language:** As the nouns referring to place and names are context specific, hence, place and person names can be the same words in Punjabi.
- **Structure of Punjabi language:** As Punjabi is a free form of language, which means that the order of the preposition, verb and noun does not count in contrast to English language

,

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 9
September 2017

- **Lack of recourses:** The field of AQG from Punjabi Text is comparatively much newer, hence, concrete resources, such as, adequate annotated Punjabi corpora, POS tagging tools, NER tools, semantic labeling tools and other NLP tools are not available in required extent
- As different approaches have their specific advantages and disadvantages which are dependent upon the type of questions and language used, deciding upon the approach and technique to use posses a great challenge in Punjabi language.
- **Less coverage of Named Entities in gazetteer**: For Punjabi Language, large gazetteer are not available.

2.1 APPROACHES USED FOR QUESTION GENERATION

2.1.1 Semantic based methods

Woo et al., 2016, focused on four semantic roles given in Table 1, and these roles can produce more concrete and specific information. A0 is agent or experiencer, and A1 is usually theme or result. Location and time are specified by AM-LOC and AM-TMP respectively. Questions are constructed by replacing a question type with an argument. Then Q Type is moved to the left and adjust the verb tense and subject position, and keep the rest of the words in a sentence to generate a question. In this paper, QG is explained with the help of following example [27]:

Table 1: Question type mapping from a semantic role

Role	Question Type (QType)
A0, A1	who (a person), what (not a person)
AM LOC	where
AM TMP	when

2.1.2 Syntax based methods

Syntax-based methods are based on simple phenomenon- parse the sentence to determine syntactic structure, simplify the sentence if required, identify the key phrases, and apply syntactic transformation rules and replace with question words. Most of the work in existing literature is done using these methods. **Kalady et al.[10], Varga and Ha [25], Wolfe [26], and Ali et al.[2]** used these methods.

2.1.3 Template based methods

Mostow and Chen used template-based approach for building a question generation system designed for a highly-constrained interaction scenario [17]. This system helped the children to improve their capabilities of reading comprehension.

2.1.4 Rule based method

Rule based method makes use of conditional statements (called as rules) to generate the questions from text. Rules are created by the system developer according to the context and structure of the language. Accuracy of the rule based system is directly dependent on the rules created for the system

If the rules are not in accordance with the given data then questions cannot be generated from this input data. The efficiency of the system highly depends on the way rules are developed.

4. RELATED WORK

In [27], the authors developed a new Question Generation (QG) system for authentication use, where questions are widely used to verify user identity for online accounts. They prompted users to provide a few sentences about their personal life events and transformed those user-provided input sentences into a set of simple fact-based authentication questions. They compared their approach with previous QG systems, and evaluation results show that their approach yielded better performance and promised future personalized authentication question generation.

Most of the previous work in question generation relies on only one view of the sentence provided by a parser. But in [16], authors talk about using multiple views from different parsers to create a tree structure which helps for question generation. This approach has shown 17% reduction in the error rate. It is found in literature that most of the question generation systems transform declarative sentences into questions by using syntactic

,

manipulation or template based techniques. Le et al. [12] observed from their earlier work that template based systems perform better than syntactic-transformation based systems. So, the authors of this paper used a template-based technique to build new question generation system based on dependency parsing and semantic role based methods.

The work in [15] talks of SENA as a tool for parsing the sentence and semantic labelling as follows: The system consists of a straightforward technique. First, the input text is divided into sentences which are processed by SENNA software, described in [4]. SENNA performs tokenization, POS tagging, syntactic constituency parsing and semantic role labeling used in the system. SENNA produces separate semantic role labels for each predicate in the sentence. For each predicate and its associated semantic arguments, a matcher function is called which returns a list of patterns that match the source sentence's predicate-argument structure. Then questions are generated and stored by question type in a question hash table.

Ming Liu et al. 2016, developed a novel Chinese question generation system using sentence simplification, question generation and ranking, to address the challenge of automatically generating factual questions in Chinese. The proposed system was evaluated on sentences from the New Practical Chinese Reader corpus. The experimental results show that ranking technique used in this system improved the acceptability of the top 25% questions by more than 20%in the dataset.

5. CONCLUSION

Recently, some work has been done in the field of Automatic Question generation (AQG) for the regional languages, including the Punjabi language. The work can be done in the field of automatic question generation (AQG) from a given text and focus can be made on improvement upon the quality and variety of questions generated from a given Punjabi text, of any general domain, by improving upon the syntactic approach and rule based techniques already being used.

6. REFERENCES

1. Ali, H. D. A. D. (2012). *Automatic question generation: a syntactical approach to the sentence-to-question generation case* (Doctoral dissertation, Lethbridge, Alta.: University of Lethbridge, Dept. of Mathematics and Computer Science, c2012).
2. Ali, H., Chali, Y., & Hasan, S. A. (2010, June). Automation of question generation from sentences. In *Proceedings of QG2010: The Third Workshop on Question Generation* (pp. 58-67).
3. Bharati, A., & Mannem, P. R. (2007, January). Introduction to shallow parsing contest on south asian languages. In *Proceedings of the IJCAI and the Workshop On Shallow Parsing for South Asian Languages (SPSAL)* (pp. 1-8).
4. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, *12*(Aug), 2493-2537.
5. Garg, P., & Bedi, E. C. S. (2014, August). A Review on Question Generation System from Punjabi Text. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, Issue 4, pp. 285-87.*
6. Gill, M. S., Lehal, G. S., & Joshi, S. S. (2009). Part of speech tagging for grammar checking of Punjabi. *The Linguistic Journal*, *4*(1), 6-21.
7. Goyal, V., Garg, S., & Singh, U. (2016, December). System for Generating Questions Automatically from Given Punjabi Text. In *Language and Technology Conference* (pp. 115-125). Springer International Publishing.
8. Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American educational research journal*, *31*(1), 104-137.
9. Graesser, A., Rus, V., & Cai, Z. (2008). Question classification schemes. In Proc. of the *Workshop on Question Generation.*
10. Kalady, S., Elikkottil, A., & Das, R. (2010, June). Natural language question generation using syntax and keywords. In *Proceedings of QG2010: The Third Workshop on Question Generation* (pp. 1-10). questiongeneration. org.
11. Lauer, T. W., Peacock, E., & Graesser, A. C. (Eds.). (2013). *Questions and information systems*. Psychology Press.
12. Le, N. T., Kojiri, T., & Pinkwart, N. (2014, April). Automatic Question Generation for Educational Applications-The State of Art. In *ICCSAMA* (pp. 325-338).
13. Lindberg, D. L. (2013). Automatic question generation from text for self-directed

,

learning *(Doctoral dissertation, Applied Sciences: School of Computing Science).*

14. Liu, M., Rus, V., & Liu, L. (2017). Automatic chinese factual question generation. *IEEE Transactions on Learning Technologies*, *10*(2), 194-204.
15. Mazidi, K., & Nielsen, R. D. (2014). Linguistic Considerations in Automatic Question Generation. In *ACL (2)* (pp. 321-326).
16. Mazidi, K., & Nielsen, R. D. (2015, June). Leveraging multiple views of text for automatic question generation. In *International Conference on Artificial Intelligence in Education* (pp. 257-266). Springer, Cham.
17. Mostow, J., & Chen, W. (2009, July). Generating Instruction Automatically for the Reading Strategy of Self-Questioning. In *AIED* (pp. 465-472).
18. Piwek, P., Hernault, H., Prendinger, H., & Ishizuka, M. (2007). T2D: Generating dialogues between virtual agents automatically from text. In *Intelligent Virtual Agents* (pp. 161-174). Springer Berlin/Heidelberg.
19. Reiter, E., & Dale, R. (1997). Building applied natural language generation systems. *Natural Language Engineering*, *3*(1), 57-87.
20. Rus, V., & Graesser, A. C. (2009). The question generation task and evaluation challenge. In *Workshop Report, ISBN* (pp. 978-0).
21. Rus, V., Cai, Z., & Graesser, A. C. (2007, February). Experiments on generating questions about facts. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 444-455). Springer, Berlin, Heidelberg.
22. Rus, V., Wyse, B., Piwek, P., Lintean, M., Stoyanchev, S., & Moldovan, C. (2010, July). The first question generation shared task evaluation challenge. In *Proceedings of the 6th International Natural Language Generation Conference* (pp. 251-257). Association for Computational Linguistics.
23. Sharma, N., & Abhilasha, E (2015, July). Automatic Question Generation from Punjabi Text: A Review. *International Journal of Computer Science & Engineering Technology (IJCSET), Vol. 6 No. 07, pp.480-483.*
24. Singh, P., & Kaur, R. (2014). Rule Based Question Generation System from Punjabi Text Contain Historical Information. *International journal of computer science and mobile computing*, 86-91.
25. Varga, A. (2010). Le an ha 2010 wlv: A question generation system for the qgstec 2010 task b. In *Proceedings of QG2010: The Third Workshop on Question Generation* (pp. 80-83).
26. Wolfe, J. H. (1976). Automatic question generation from text-an aid to independent study. *ACM SIGCSE Bulletin*, *8*(1), 104-112.
27. Woo, S. S., Li, Z., & Mirkovic, J. (2016, September). Good Automatic Authentication Question Generation. In *INLG* (pp. 203-206).
28. Y. Wu, J. Shi, and F. Wan (2013). Automatic Identification of Chinese Coordination Discourse Relation. *Acta Scientiarum Naturalium Universitatis Pekinensis, vol.49, no. 1, pp. 1-6.*

,