INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

AUTOMATIC QUESTION GENERATION TECHNIQUES: A COMPARATIVE STUDY

RASHMI SHRIVASTAVA[1], JEETENDRA KUMAR[2], SHRIYA SAHU[3]

1. Assistant Professor, Bilaspur University.
2. Assistant Professor, Bilaspur University.
3. Assistant Professor, Bilaspur University.

Abstract: Automatic question generation is very important part of Natural Language Processing. Recently question generation has got an immense attention from the researchers. Generating questions automatically and reading assessment is a very complex and tedious task. Questions can be of many types like "wh", descriptive etc. In manual process of generating questions there may be chances of duplicate question generation. Hence there is a need to develop an automatic question generator which automatically generates questions from input text material using Natural Language Processing techniques. The main motivation of this paper is to provide a comparative analysis of various question generators on the basis of types of question generated, accuracy, used tools and nature of over-generation.

Keywords: Question generation, natural language processing, automatic

Corresponding Author: MS. RASHMI SHRIVASTAVA

Access Online On:

www.ijpret.com

How to Cite This Article:

Rashmi Shrivastava, IJPRET, 2015; Volume 3 (9): 732-737

PAPER-QR CODE

INTRODUCTION

A question is a sentence in an interrogative form, addressed to someone in order to get information in reply. Question Generation is the task of automatically generating questions from various inputs such as raw text, database, or semantic representation. Question Generation is regarded as a discourse task involving the following four steps: (1) when to ask the question, (2) what the question is about, i.e. content selection, (3) question type identification, and (4) question construction [1].

Question Generation is an important component in dialogue systems, virtual environments, and learning technologies such as Intelligent Tutoring Systems, inquiry-based environments, and instructional games. For generating a question a person should have subjective knowledge. People often spend hours by themselves contemplating ideas and working through issues raised by what they have read. These ideas and issues are often articulated in the form of questions. Questions can be of many forms like objective type, fill in the blanks, true/false, Wh type question, Match the sentences, Short answer, Long answer, answer in one word etc. A general human procedure for generating questions is to read article thoroughly then generate question accordingly. Sometimes if a written material is very long then it will be tedious task to generate various questions of different types in many sets without any repetition. Hence there is a need to make an automated system which at first read the text material then generates all types of question from that text material. In this paper various question generation techniques are analyzed

RELATED WORK

In 2009 Michael Heilman and Noah A. Smith presented a concept paper on Question Generation via Over generating Transformations and Ranking [2]. They develop a three stage framework for question generation named NLP transformation, Question transducer, Question Ranker. In First phase i.e. NLP transformation selected text is transformed into declarative sentence by altering or transforming lexical items, syntactic structure, and semantics. Many NLP transformations can be used in this stage including extractive summarization, sentence Compression, sentence splitting, sentence fusion, paraphrase, textual entailment, lexical semantics for word substitution. In second phase declarative sentences are converted to question by using syntactic transformation like WH-movement, subject auxiliary inversion, etc. This phase is also known as question transducers. In third phase i.e. Question ranker questions are scores and ranked according to the source content. According to the manual evaluation the

system achieves 43.3% precision-at-10, generating approximately 6.8 acceptable questions per 250 words of source text.

Michael Heilman Noah A. Smith (2010) presented a paper Good Question! Statistical Ranking for Question Generation [3]. Their approach of question generation was based on rule based on over generation, then all generated question was ranked statistically. There question generation approach was two-step process, in first step each sentence of source text is converted to set of declarative sentence. In second step, the declarative sentences derived in step 1 are transformed into sets of questions by a sequence of well-defined syntactic and lexical transformations (subject-auxiliary inversion, WH-movement, etc.). After that question were ranked according to question deficiencies like ungrammatical, does not make sense, vague, missing answer, wrong WH word etc. According to their conclusion 27.3% of test set questions were labeled acceptable (i.e., having no deficiencies) by a majority of raters.

Aquino J.F. Chua D.D. Kabiling R.K., Pingko J.N., Sagum R. (2011) gave the idea of Text2Test [4] This system generate question from text in three steps i.e. text processing , scoring and question over-generation. The text processing step includes parsing resolution and chunking of text that is taken as input. Scoring step consist of sorting, ranking and removal of unwanted sentences. The score of a sentence is the average of the scores of its tokens, and the score of each token is based on how much it persists within the texts, particularly its normalized term frequency. In Over-generation step many questions are generated and ranked. Text2Test generated questions which on average had an accuracy of 3.0558 out of 4. In this paper only short answer type question are generated like wh type question.

Pascal Kuyten, Timothy Bickmore, Svetlana Stoyanche, Paul Piwek, Helmut Prendinger, Mi,tsuru Ishizuka(2011) presented a paper[2] to create virtual instruction from textual input[5]. This system at first analyze the input text then generate various question answer pairs using visual patterns. In First step, text is translated into rhetorical structure theory (RST) trees using disclosure analysis. Then question-answer pairs are generated, by translating the RST tree into coherent dialogue. Question-answer pairs are then translated into an agent scripting language. In the final step, scripts are compiled into a run-time agent system then RST trees are converted into Question answer pairs.

In 2011 Manish Agarwal_, Rakshit Shah_ and Prashanth Mannem gave the idea of generating question using discourse cues[6]. There suggested system was able to generate the questions from written text instead of single sentence. In this paper it was emphasized that some discourse relations are important such as casual, temporal and result than other discourse

relation. They make question type identification based on many discourse cues like since, although, because, when etc. They make the content selection task for question generation by finding the target argument. After finding the target argument syntactic head is located. Then target argument is extracted after identifying syntactic head. Then by applying syntactic transformation content is converted to questions. But the system was not able to handle the conference resolution. As the number of pronoun increased, semantic rating of the question was decreased. Sometimes parser is failed to give correct parse of the sentence with complex structure. For inter sentential connectives system, consider only those sentence in which connective occurs in the beginning of the sentence. The system does not handle the removal of predicative adjuncts.

Yao X., Bouma G., Zhang Y. (2012) proposed a system MrsQG [7] to generate questions on semantics basis. This paper proposes three methods to address these issues: mrs decomposition for complex sentences to simplify sentences, mrs transformation for simple sentences to convert the semantic form of declarative Sentences into that of interrogative sentences, and hybrid ranking to select the best Questions. This paper also generates question from one sentence.

In 2013 Lindberg D. Popowich F., Nesbit J. Winne P., presented a paper on Generating Natural Language Questions to Support Learning OnLine [8]. They developed a template based framework for question generation which is not tightly constrained based and having surface form flexibility. This system uses pronominal anaphora resolution for text pre processing. Then source sentence is tokenized and annotated with POS tags, named entities, lemmata, and its SRL parse.SRL parse were used to extract set of predicate frames. Questions are generated from individual predicate frame rather than complete sentence. Then generated predicate frame is matched with a single template. Then by algorithm template slots are filled with role text to produce a question.

Table1: Comparative chart of Question generation techniques

Question Generation Techniques	Type of Question generated	NLP tool used	Over-generation	Performance
Question Generation via Overgenerating Transformations and Ranking[2]	Factual	Tregex, Tsurgeon	Yes	43.3%
Good Question! Statistical Ranking for Question	Factual	Tregex, Tsurgeon	Yes	27.3%

Generation[3]				
Test2Text[4]	Factual	Tregex, Tsurgeon	Yes	76.55
Fully Automated Generation of Question-Answer Pairs for Scripted Virtual Instruction[5]	Factual	HLDA CODA	No	-
Automatic Question Generation using Discourse Cues[6]	Factual Why When Give an example Yes/No	-	No	78.75 on QGSTEC-2010 dataset 72.5 on Wikipedia dataset
Semantics-based Question Generation and Implementation[7]	Factual Wh How many Yes/no	-	Yes	-
Generating Natural Language Questions to Support Learning On-Line[9]	Factual and descriptive	Stanford NER	No	66%

CONCLUSION

In this paper analysis of various question generation techniques have been done and conclusion is that most of the question generation techniques are able to generate only factual questions like wh type question. Some question generation techniques like "Generating Natural Language Questions to Support Learning On-Line" are also able to generate descriptive type of questions. Some of the techniques "Question Generation via Over-generating Transformations and Ranking", "Good Question! Statistical Ranking for Question Generation", "Test2Text" and "Semantics-based Question Generation and Implementation" use method of over-generation means over generate questions and rank them. In comparison performance [6] and [4] was found most effective. In the near future there is a need to develop a question generator which will be able to generate all type of questions along with high precision and accuracy.

REFERENCES

1. http://www.questiongeneration.org/

2. Heilman M. and Noah A. Smith. "Question generation via overgenerating transformations and rankink" Technical Report CMU-LTI-09-013, Language Technologies Institute, Carnegie Mellon University, 2009.

3. Heilman M. Noah A. Smith "Good Question! Statistical Ranking for Question Generation, Human Language Technologies" The 2010 Annual Conference of the North American Chapter of the ACL, pages 609–617, Los Angeles, California, June 2010.

4. Aquino J.F. Chua D.D. Kabiling R.K., Pingko J.N., Sagum R. "Text2Test: Question Generator Utilizing Information Abstraction Techniques and Question Generation Methods for Narrative and Declarative Text " Proceedings of the 8th National Natural Language Processing Research Symposium, pages 29-34 De La Salle University, Manila, 24-25 November 2011.

5. Kuyten P., Bickmore T., Stoyanche S., Piwek P., Prendinger P., Ishizuka M. "Fully Automated Generation of Question-Answer Pairs for Scripted Virtual Instruction" © Springer-Verlag Berlin Heidelberg 2011.

6. Agarwal M., Shah R. and Mannem P. "Automatic Question Generation using Discourse Cues" Proceedings of the 6th Workshop on Innovative Use of NLP for Building Educational Applications ISBN: 9781937284039 Pages 1-9, 2011.

7. Yao X., Bouma G., Zhang Y. "Semantics-based Question Generation and Implementation, Dialogue and Discourse" 3(2) 11-42 (2012).

8. Lindberg D. Popowich F., Nesbit J. Winne P., "Generating Natural Language Questions to Support Learning On-Line" Proceedings of the 14th European Workshop on Natural Language Generation, pages 105–114, Sofia, Bulgaria, August 8-9 2013. 2013 Association for Computational Linguistics.