International
Journal
Of Advanced Research in Engineering& Management (IJAREM)
ISSN: 2456-2033 || PP. 15-21

Modified Question Generation Technique for e-Learning-Aid

Dushyant Sharma[1], Dhanroop Mal Nagar[2] and Vinay Baghela[3]
[1,3] Dept. Of Comp. Sc. Engg., Govt. Engg. College, Bikaner.
[2] Dept. Of Information Technology, Govt. Engg. College, Bikaner

Abstract: Nowadays readers, researchers, and intellectual people are going online to read a text, information, academic material. As technology has advanced, you have reached an age where almost everybody from a teenage kid to an adult has a supercomputer in the form of mobile in his/her pocket. Thus, Every piece of information is at the tip of your hands. So, to help students, researchers to learn better we need to ask the right set of questions. To design a question test for every material available on the internet is a tedious task, we need to automate this process. This can be done using Rhetorical Structure Theory(RST) which is one of the techniques to define text spans into Nucleus and Satellite pairs, we propose a modified set of Nucleus-Satellite Pairs to cover a large set of text-domain.
Keywords: NLP, Question Generation, E-Learning.

1. Introduction

In the present times when information is available online most of the students are taking the online way of accessing information and gaining knowledge whether it be academic, personal or daily updates. To understand and use information in the future we know we need to comprehend it to use it for the future. To do this we need to learn it first but how can we learn information? In the past, educators, teachers, professors, and learners have gained and learned knowledge by asking questions after reading the material. Thus, we need to ask questions about the recently read topic to gain knowledge. But, we can't manually form questions about all the available material on the internet. Thus, we need some automated way to question generation from reading comprehension. An automated question generation system can help here. Previously, various research work was done on question answering systems which were based on factoid questions and template-based questions. Our work is based on Rhetorical Structure Theory and we present modified categories of Nucleus-Satellite pairs that cover a wider range of text-domain. We proposed modified Nucleus-Satellite Categories based on which looking into the question template questions can be formed.

2. Literature Survey

In our research work about question generation for reading comprehension and various methods to do it, we will start from the very beginning of the research done on improving the quality of questions asked in Classroom settings and Tutor settings in the paper titled "Question Asking During Tutoring" by Arthur C. Graesser[1] et. al. published in 1994. In this paper, the study looks into comparing which is a better classroom setting or tutor setting. In a classroom setting, where a group of students sits together and the professor teaches them whereas in tutor settings there is a one-to-one interaction between student and teacher. It was found that students hesitate in asking questions in classroom settings due to feeling guilty or might ask a question that shows ignorance in class to peers, this is eliminated in the tutoring session. It was found that there are many barriers in classroom settings like many teachers don't know how to ask good questions, questions asked by teachers are shallow which expects explicit knowledge of the material. Only a few teachers know how to ask deep sequential Socratic questions to challenge students into a deep level of thinking. It is also hard to address every student's knowledge deficits in classroom settings. But we are in a better position in one-to-one tutoring sessions.

In the research paper "Bloom Taxonomy" by Mary Forehand [2] it is described that three parts: [i] The cognitive – knowledge-based domain.[ii] The affective - attitudinal-based domain and [iii] The psychomotor – skills-based domain. Through the years, the stages have many times been described as a path leading many teachers to motivate their students to "climb to a higher(level of) thought." The lowest three stages are knowledge, comprehension, and application. The highest three stages are analysis, synthesis, and evaluation. In the old version of Bloom's Taxonomy from bottom to top the levels are Knowledge, Analysis, Synthesis, Comprehension, Application, and Evaluation. While in the new version of Bloom's Taxonomy, the levels are

Remembering, Understanding, Applying, Analysing, Evaluating, and Creating. Though the original Bloom's Taxonomy was an excellent tool for an educator, the revised Bloom's Taxonomy is new and improved.

In the Rhetorical Structure Theory: Toward a functional theory of text organization by W. C. Mann [3], it is stated that Rhetorical Structure Theory is mainly the arrangement of the natural text. It is a useful method of describing natural text in terms of a bond between parts of texts and the meaning of the natural text. The author establishes a new definition of RST. The research also explores three main terms of RST that are the nucleus/satellite structural patterns, their hierarchy, and the role of text structure. It includes relations, schemas, and structures. Relations consist of four fields namely constraints on the Nucleus, constraints on the Satellite, Constraints on the combination of Nucleus and Satellite, and The Effect. Schemas consist of structural constituency arrangements of text. Another is Structural analyses and structure diagrams. In Structural analysis of text in Schema applications constraints holds completeness, connectedness, uniqueness, and adjacency.

In the research paper "Automatic Question Generation for Vocabulary Assessment" by J. C. Brown[4], the author states that in REAP Systems users are automatically rated on their reading levels. In this paper, an approach is proposed to generate a vocabulary assessment. Six types of vocabulary assessment questions are generated namely cloze, hypernym, hyponym, synonym, antonym, and definition questions. The author has shown that the computer-generated questions give an assessment of vocabulary skill for every single word which relates well with written questions by humans and standardized measure of vocabulary skill.

In the research paper "Adaptive E-Learning Content Generation based on Semantic Web Technology " by E. Holohan[5], it is stated that the good and fruitful authoring of educational e-content is the main puzzle in the course of the process. Courseware writers will praise if some tool will automate various authoring tasks. In this paper, a tool is presented, OntAWare, which is a group of e-tools that helps developing e-content by content authoring, and its management and up to delivery. This makes use of new notable knowledge demonstration standards and knowledge-processing techniques. This OntAWare comprises of features like The Generation and Export of Static Courseware i.e. The Modularisation of Knowledge and Courseware and The Learning Objects Generation, The Generation of Courseware for Flexible Delivery i.e. Constrained, Sequenced E-Learning versus Free Navigation, Free Ontology-based Navigation, Option of Guided Navigation, Option of Adaptive Guided Navigation, Combining Flexible Delivery with Interoperability and other types of Flexibility. First, the author verified that the author could automate the process of generation and export of useful learning objects which belong to domain ontologies of the Semantic Web. Second, the author builds a free content navigation system that is user-friendly based on ontologies – this then becomes the basis for further developments. Third, by being the first two steps, leads this navigation guidance which is based on ontologies, for the student use of the system. Fourth, the author attached real-time monitoring of users and corresponding adaptive navigation guidance. Fifth, the author takes comfortable transparent ontology modification for the teacher.

3. Proposed method

Question Generation approaches from the previous methods were not appropriate and were only either factoid-based or template-based. Thus, we were only able to test the understanding level of Bloom's Taxonomy. In our proposed method, we present modified Nucleus Satellite pairs that cover a broad range of text-domains and allow the user to generate questions on a broad range of text-domains.
For this, first, we have to understand the Rhetorical Structure Theory (RST). The elements of RST are defined as follows: Relations, Schemas, Schema Applications, and Structures. There are two independent text spans - Nucleus and Satellite. And there exists a Relation between them. In the previous research, there exist Nucleus-Satellite pair relations that do not cover a full range of text. We propose modified Nucleus-Satellite pairs as follows:

Modified Nucleus-Satellite Pair Relations

Relation (N,S)	Obtained from
Explaination (N,S)	Evidence, Reason, Explaination
Background (N,S)	Background, Circumstance
Cause (N,S)	Cause, Purpose
Result (N,S)	Result, Consequence
Solutionhood (N,S)	Problem-Solution
Condition (N,S)	Condition, Hypothetical
Evaluation (N,S)	Evaluation, Conclusion
Actioness (N,S)	Goal, Action
Understanding(N,S)	(Data)Pattern, Concept(Algo)
Achievement (N,S)	Plan(agen),Accomplishment(Goal)
Ablement(N,S)	Device,Action(Perform)

Fig. 1. Figure shows modified Nucleus-Satellite Pair Relations.

We have added four new categories namely Actioness, Understanding, Achievement, and Ablement. The Question Generation System takes text as input. It then categorizes it into text span types as follows:

Text Span Types

Span Type	Characterstic of span	Example
Type 0	A group of many sentences	A bomb exploded in the building. It destroyed its installations.
Type 1	One sentence, or a phrase or clause not begining with a verb, but containing one	The bomb destroyed the building
Type 2	Phrase or clause begining with a verb	destroyed the buildings
Type 3	Phrase or clause that does not contains a verb	destruction of the building

Fig. 2. Figure shows Text span types with examples.

Then the question is generated from classified categories based on the following table.

International
Journal
Of Advanced Research in Engineering& Management (IJAREM)
ISSN: 2456-2033 || PP. 15-21

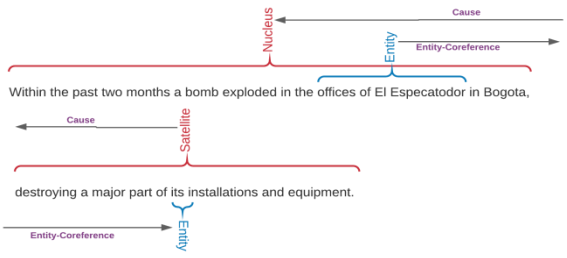
Template for Existing Question Generation Categories

Relation	Template for type 0	Template for type 1	Template for type 2	Template for type 3
Explanation	[Nucleus]. What evidence can be provided to support this claim?	Why [Nucleus]?	What [Nucleus]?	What caused [Nucleus]?
Background	[Nucleus]. Under what circumstances does this happen?	Under what circumstances [Nucleus]?	What circumstances [Nucleus]?	What circumstances led to [Nucleus]?
Solutionhood	[Nucleus]. What is the solution to this problem?	What is the solution to [Nucleus]?	What solution [Nucleus]?	What is the solution to the problem of [Nucleus]?
Cause	[Satellite]. Explain the reason for this statement.	Why [Satellite]?	What [Satellite]?	Explain the reason for [Satellite]?
Result	[Nucleus]. Explain the reason for this statement.	Why [Nucleus]?	What [Nucleus]?	Explain the reason for [Nucleus]?
Condition	[Nucleus]. Under what conditions did this happen?	Under what conditions [Nucleus]?	What conditions [Nucleus]?	What conditions led to [Nucleus]?
Evaluation	[Nucleus]. What lets you assess this fact?	What lets you assess [Nucleus]?	What assessment [Nucleus]?	What assessment can be given for [Nucleus]?
Actioness	**[Nucleus] What led to start of action?**	**Why [Nucleus]?**	**What [Nucleus]?**	**What caused [Nucleus]?**
Understanding	**[Nucleus] What led to formation of concept?**	**What [Satellite]?**	**What [Nucleus]?**	**What caused [Nucleus]?**
Achievement	**[Nucleus] What led to the [accomplishment]?**	**Why [Satellite]?**	**What [Nucleus]?**	**What caused [Satellite]?**
Ablement	**[Nucleus] What led to action?**	**Why [Nucleus]?**	**What [Nucleus]?**	**What caused [Satellite]?**

Existing Categories Added Categories

Fig. 3. Template of Question Generation with Modified Categories.

Let us understand it with the help of the example below.



Example of Cause Relation.

Fig. 4. Figure shows a sentence with an example of cause relation in nucleus satellite pair.

Here, the nucleus is "Within the past two months a bomb exploded in the offices of El Especatodor in Bogota" and satellite is "destroying a major part of its installations and equipment." Now, Questions were classified by (a) degree of specification, text span is classified into type 2 category, We will make the question as "What satellite?" i.e. " What destroyed a major part of installations and equipment of the offices of El Especatodor in Bogota ? "

4. Experimental setup

For the research work, we build a separate system that uses the i7 processor 8GB RAM and 2 GB Nvidia Graphics Card, and 1 TB Hard Disk Memory. Ubuntu 20.01 OS was installed in Oracle Virtual Box with Anaconda Framework containing Python Language version 3.8.3. Oracle version 6.1.16 was used while Anaconda Framework version 2020.07 was installed. Spacy, Nltk libraries were used for natural language processing. Pycharm was used as IDE.

5. Results

We generated questions from our self-generated corpus. We compared our question generation approach from the previous question generation system presented in the research work from T. Desai and P. Dakle[6] on several parameters like Grammatical correctness, Relevance to the topic, Average no. Of Inference Steps, Average no. of inference steps, Coverage of text-domain, etc...and found that coverage of text-domain has improved and our system was as par in other parameters. These parameters improve because of the introduction of new categories. Graphs and table are as follow:

Evaluation Criteria	No. of Inference Steps		Gram. Correctness		Cover. of text domain	Releva. to topic
System	PQG	PA	PQG	PA	PA	PA
R1	0.43	0.4	0.95	0.85	0.72	0.66
R2	0.46	0.49	0.92	0.88	0.86	0.78
R3	0.42	0.45	0.91	0.9	0.73	0.81
R4	0.56	0.51	0.98	0.9	0.8	0.66
R5	0.39	0.4	0.97	0.85	0.77	0.75
R6	0.33	0.37	0.87	0.9	0.72	0.72
R7	0.27	0.29	0.8	0.78	0.84	0.9
R8	-	0.33	-	0.84	0.81	0.84
R9	-	0.3	-	0.88	0.83	0.85
R10	-	0.32	-	0.83	0.7	0.73
R11	-	0.28	-	0.81	0.74	0.91

Table 1: results table

R1- Explanation, R2- Background, R3- Solutionhood, R4- Cause, R5- Result, R6- Condition, R7- Evaluation, R8- Actioness, R9- Understanding, R10-Achievement, R11- Ablement

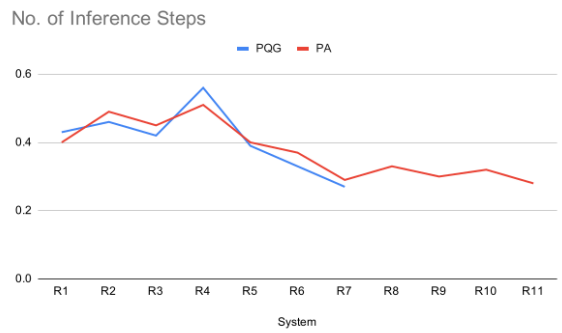


Fig 5: grammar correctness result plot

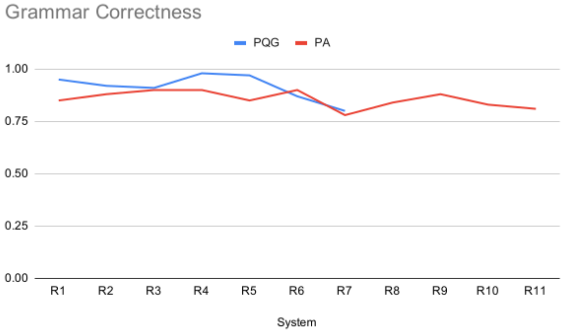


Fig 6: No. of Inference Steps Result Plot

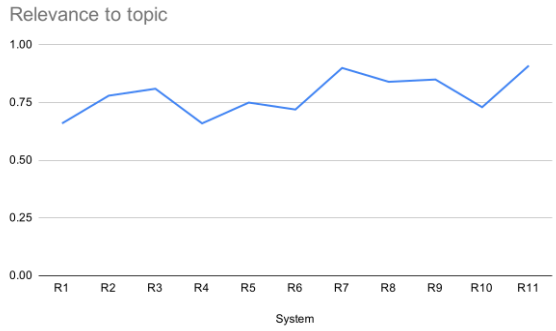


Fig 7: Relevance to topic Result Plot

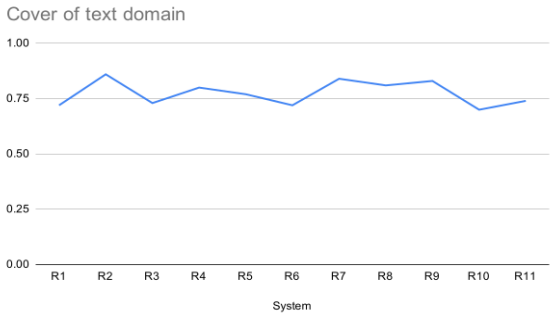


Fig 8: Cover of text domain Result Plot

6. Conclusion

In this paper, we proposed our modified new techniques for question generation which improves on previous techniques which were factoid and template-based. We introduce new categories in the existing Nucleus Satellite pair categories given in the paper by Takashak Desai[6], thus improving on Coverage of text-domain, No. Of Inference Steps etc. We found future scope and areas of improvement in Parsing errors, Incorrect identification in complex sentences, Clause rearrangement. The future scope also includes working on complex sentences generating questions on higher levels of Bloom's Taxonomy.

References

[1]. Arthur C. Graesser and Natalie K. Person, "Question asking during tutoring"(1994) American Educational Research Journal, Vol. 31, No. 1, pp. 104-137

[2]. Mary Forehand,"Bloom's Taxonomy"(2011), Emerging Perspectives on Learning, Teaching and Technology.

[3]. William C. Mann and Sandra A. Thompson, Rhetorical Structure Theory: Toward a functional theory of text organization (1988) Walter de Gruyter, Berlin / New York | 2009

[4]. Jonathan C. Brown, Gwen A. Frishkoff, Maxine Eskenazi, Automatic Question Generation for Vocabulary Assessment (2005) Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing(HLT/EMNLP), pages 819-826, Vancouver, October 2005. Association for Computational Linguistics.

[5]. Edmond Holohan, Mark Melia, Adaptive E-Learning Content Generation based on Semantic Web Technology (2005) Dublin City University, School of Computing, Dublin 9, Ireland.

[6]. Takshak Desai and Parag Dakle, Generating Questions for Reading Comprehension using Coherence Relations (2018). Proceedings of the 5th Workshop on Natural Language Processing Techniques for Educational Applications, pages 1-10 Melbourne, Australia, July 19,2018

[7]. Jerry R. Hobbs, Coherence and Coreference (1979). Cognitive Science 3, 67-90

[8]. Mengqiu Wang, A Survey of Answer Extraction Techniques in Factoid Question Answering (2006) Association for Computational Linguistics

[9]. Vasin Punyakanok, Dan Roth, Wen-tau Yih, The Importance of Syntactic Parsing and Inference in Semantic Role Labeling (2008) Computational Linguistics Volume 34, Number 2

[10]. Silviu Cucerzan and Eugene Agichtein, Factoid Question Answering over Unstructured and Structured Web Content (2005). Microsoft Research

[11]. Ruslan Mitkov, Leanha and Nikiforos Karamanis, A Computer-aided environment for generating multiple-choice test items (2006) Natural Language Engineering 12(2) 177-194 Cambridge University Press, UK

[12]. Rhetorical Structure Theory: A Theory of Text Organization(1987). Information Science Institute. June, 1987

[13]. Andrew M. Olney, Arthur C. Graesser and Natalie K. Person, Question Generation from Concept Maps (2012). Dialogue and Discourse 3(2) 75-99

[14]. Lee Becker, Sumit Basu and Lucy Vanderwende, Mind the Gap: Learning to Choose Gaps for Question Generation (2012). Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 742-751

[15]. J. Araki, D. Rajagopal, S. Sankaranarayanan, S. Holm, Y. Yamakawa, and T. Mitamura, "Generating Questions and Multiple-Choice Answers using Semantic Analysis of Texts."