International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/

AUTOMATIC QUESTION GENERATION MODEL BASED ON DEEP LEARNING APPROACH

Hala Abdel-Galil,	Mai Mokhtar*,	Salma Doma,
Associate Professor and Head of Computer Science Department, Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt	Computer Science Department, Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt	Computer Science Department, Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt
hala.nagy@fci.helwan.edu.eg	maimokhtar@fci.helwan.edu.eg	salmadoma@fci.helwan.edu.eg

Abstract: *Nowadays, students face many difficulties to practice for exams. Professors and teachers spend a lot of time and effort to make exams. Automatic Question Generation Model proposes a solution to save time, effort, and student's learning process which helps in educational purposes. AQGM is user-friendly which is implemented asa GUI-based system that generates Wh- questions which mean" WH" (" What"," Who", and" Where") and formatted into two types of templates, Question Bank template, and Exam template. Exams have different difficulty levels (Easy, Medium, and Hard). Therefore, students can measure their level and teachers will know to what extent the students understand the course. Researches have shown that this method is helpful and successful for educational purposes. AQGM generates questions automatically by using its model that generated by using sequence-to-sequence approach specially encoder-decoder technique with copy mechanism and attention decoder. AQGM model uses SQuAD as a training dataset which helps to get more accurate results. This model gets a BLEU-4 score 11.3 which is good according to it generated automatically using deep learning approaches.*

Keywords: *Automatic Question Generation, SQuAD Dataset, Seq2seq, Feature-Rich Encoder, Attention-Based Decoder, Copy Mechanism*

* Corresponding author: Mai Mokhtar
Computer Science Department, Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt
E-mail address: maimokhtar@fci.helwan.edu.eg

1. Introduction

Recently, As shown when coronavirus disease (COVID-19) outbreak spectacularly. The lockdown has happened to force people to stay at home because it was the only way to reduce the spread of the virus. Education one of the most sectors affected by the lockdown over the whole world. As many countries tended to online learning to cover what was lost in the pandemic. therefore, the impact of online learning has been observed around the world. Because without online learning the education was stopped as it was a disaster when the effect of that might be shown in next generation when they show the lack of education. Automatic Question Generation Model can be used in online learning and offline learning.

Automatic Question Generation Model helps students, teachers, and professors for educational purposes. Also, it helps students to practice well and evaluate their performance by answering more questions without being confused by misleading or huge error-prone questions. Also, it helps teachers and professors because they face some problems when generating questions to make exams that have some misleading questions, lower-level questions, or imprecise questions. Moreover, they spend a lot of time also effort to construct these exams. Also, it helps students to practice for exams by answering more questions because studies have reported that deploying questions in teaching encourages students to self-explain, which is highly beneficial for learning. Also, the human effort consumed in making exams so, the professors and teachers can direct their effort in more helpful researches and invent new methods in teaching.

The questioning enhances learning has been confirmed by research [1, 15] and modern studies reconfirm this important pedagogical principle [12, 14]. Questioning is an essential part of the teaching and learning processes and is rooted in the quality of education and strategic thinking. Questions are used to teach as well as to evaluate student understanding, and thus questioning plays a crucial part in the overall achievement of a classroom.

Automatic Question Generation Model generates questions, it is an interactive system with a graphical user interface (GUI) that provides the user with ease of use. Simply, the user specifies the document needed then, the document will go through four phases before generating questions. The first one is the Data Cleaning phase, the second one is the Context Selection phase, the third one is the Features Extraction phase, and last but not least the Questions Generator phase. Data Cleaning phase will clean text from mathematical equations, special symbols and convert contractions (convert words like it's to it is for better understanding). Next is the Context selection phase that will extract important paragraphs from the document then it will simplify by extracting important and meaningful sentences in the paragraphs. After that the feature extractor part will extract features required by the model to generate questions and lastly question generation part takes sentences and generate Wh questions with different difficulty levels, just Wh questions no other type is supported for now. Also, it produces two types of templates, a question bank template, and an exam template. The question bank template will take the questions as input. Exam template can be generated from question bank using a shuffling algorithm. Exams will have different difficulty levels, so, students can practice well. The rest of the paper is organized as follows. In the next section, Literature Review and Related Work will be presented, followed by a description of the SQuAD Dataset in section 3. Then, the details and technical aspects of AQGM will be presented in section 4. Section 5 explains the experiments and shows evaluation results. Section 6 discusses the analysis. Finally, conclusions and future work will

be presented in section 6.

2. Literature Review and Related Work

Some systems granted choose the informative sentence and the keyword to ask based on Named Entity Recognizer (NER) and the semantic labels that exist in the sentence [4]. Semantic Role Labeling and Named EntityRecognizer techniques are used with some systems to convert the entered sentence to a semantic pattern and developed an Artificial method that will be able to classify according to the question type. The most question types recognized are set of Wh-questions like where, when, what, who, how, and why. Some systems use feature extraction, storage memory, learning, and associative retrieval to solve classification tasks. First, the entered sentence will parse using NER and SRL techniques then from NER and SRL recognizes whether, sentences includea person name, date, location, will be recognized based on this classification question pattern. As an example, ifa location then the question pattern would be WHERE [4].

Some web-based automatic question generation system to generate reading comprehension questions and multiple-choice questions on grammar from a given English text. they can automatically generate Wh-questionsof selected sentences. To generate Wh-questions uses the pre-trained natural language understanding model, Text-To-Text Transfer Transformer, and a modified version of the SQuAD machine reading comprehension dataset [5]. But some different approaches were detected as in [2], all questions are generated by parameterizedconcepts from a set of pre-defined templates. Traditional methods for question generation mainly use definite heuristic rules to convert a sentence into relevant questions. Automatic question generation proposes to generatequestions from paragraphs where the generated questions can be answered by certain sentences of the given paragraph. Zhou proposes a system that uses the neural encoder-decoder model to generate meaningful and several questions from natural language sentences [16]. According to this paper [10], different approaches were developed to help in educational purpose. Awad proposes approach which is based on domain-specific ontologies which differ from other approaches. It is free of lexicons such as WordNet or other linguistic resources [3].

3. SQuAD Dataset

SQuAD Stanford Question Answering Dataset (SQuAD) is a reading comprehension dataset, consisting of questions posed by crowd workers on a set of Wikipedia articles, where the answer to every question is a segmentof text, or span, from the corresponding reading passage, or the question might be unanswerable. SQuAD 1.1, the previous version of the SQuAD dataset, contains 100,000+ question-answer pairs on 500+ articles. It was usedin question generator model. Also, it used in more researches that related to question generation but its mainly prepared and generated for answering question as used many papers which are Vocabulary Matters: A Simple yet Effective Approach to Paragraph-level Question Generation [9], Focused Questions and Answer Generationby Key Content Selection [6], and Focused Questions and Answer Generation by Key Content Selection [7]

```
⟨|Context →
   Beyoncé Giselle Knowles-Carter (/biːˈjɒnseɪ/ bee-YON-say) (born September 4, 1981) is
      an American singer, songwriter, record producer and actress. Born and raised in
      Houston, Texas, she performed in various singing and dancing competitions as
      a child, and rose to fame in the late 1990s as lead singer of R&B girl-group
      Destiny's Child. Managed by her father, Mathew Knowles, the group became one
      of the world's best-selling girl groups of all time. Their hiatus saw the
      release of Beyoncé's debut album, Dangerously in Love (2003), which established
      her as a solo artist worldwide, earned five Grammy Awards and featured
      the Billboard Hot 100 number-one singles "Crazy in Love" and "Baby Boy".,
   Question → When did Beyonce start becoming popular?,
   Answer →
      in the late 1990s,
   AnswerPosition → 270, Answerable →
      True,
   QuestionID → 56be85543aeaaa14008c9063,
   Title → Beyoncé |⟩
```

Figure 1. SQuAD Dataset Sample

4.   Details and technical aspects

4.1   System Overview

The automatic Question Generation Model allows students, teachers, and professors to upload their PDF documents to generate questions. It will take these documents to read as text by converting PDF into text, apply the data cleaning to clean any equations, figures, and tables. Then, the context selection phase simply selects the appropriate text to ask questions on, but not all paragraphs are suitable for asking questions, some paragraphs are redundant, have no special keywords related to the field, talk about the author or any external topic, etc. All those paragraphs are not needed in the QG phase. Therefore, the context selection simply eliminates them. Also, it is simply the removal of nonimportant keywords, punctuation, and other special symbols not needed in the QG phase, by extracting a set of important keywords, identifying the top n important keywords, and based on the keywords a top m list of sentences is defined, allowing the QG model to ask more accurate questions by simplifying and eliminating some redundant or unwanted sentences. Then, the question generation function takes the simplified sentences to generate Wh-questions with different difficulty levels by extending the sequence-to-sequence models. The extended model generates answer-focused questions by incorporating the answer position indicator with the lexical features. The answer position feature denotes the position of the answer span in the input sentence, which is crucial to make the model aware of the target answer. The lexical features contain wordcase, Part Of Speech (POS), and Named Entity Recognizer (NER) tags to encode more linguistic information. Finally, the attention-based decoder generates an answer-relevant question from the sentence. It will store gener-ated questions in the question bank template. The user specifies the number of questions and difficulty levels then AQGM classifies the questions as (Easy, Medium, and Hard) and specifies the question using random selection,and writes in exam template. Also, it generates different exam models.

4.2 Preprocessing

Three phases of preprocessing are involved: data cleaning phase, context selection phase, and features extraction phase.
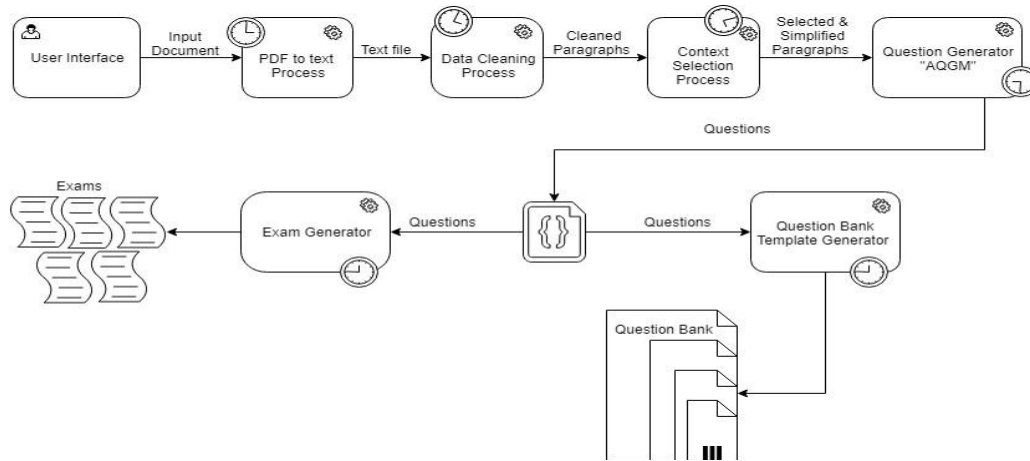

Figure 2. Overview of the whole system Automatic Question Generation.

– Data Cleaning: This is employed by cleaning text from mathematical equations, special symbols, and convert contractions.

– Context selection: In this phase, important paragraphs are extracted from the document then important and meaningful sentences are extracted from the filtered paragraphs.

– Feature extraction: The answer position indicator and the lexical features are required by the model to generate questions. The answer position feature denotes the location of the target answer, which is crucial to generate a question concerning a specific answer in the input sentence. The lexical features contain word case, Part Of Speech (POS), and Named Entity Recognizer (NER) tags to encode more linguistic information.

4.3 Model Approach

This subsection introduces the Question Generator Model, which consists of a feature-rich encoder and an attention-based decoder, as shown in Figure 2.

4.3.1 Feature-Rich Encoder

The AQGM uses Gated Recurrent Unit (GRU) to build the encoder. Considering that bidirectional GRU (BiGRU) read the inputs in both forward and backward orders, which helps to capture more useful contextual features. AQGM concatenates the word vector, lexical feature embedding vectors and answer position indicator embedding vector as the input of BiGRU encoder to produce two sequences of hidden vectors, i.e., the forward
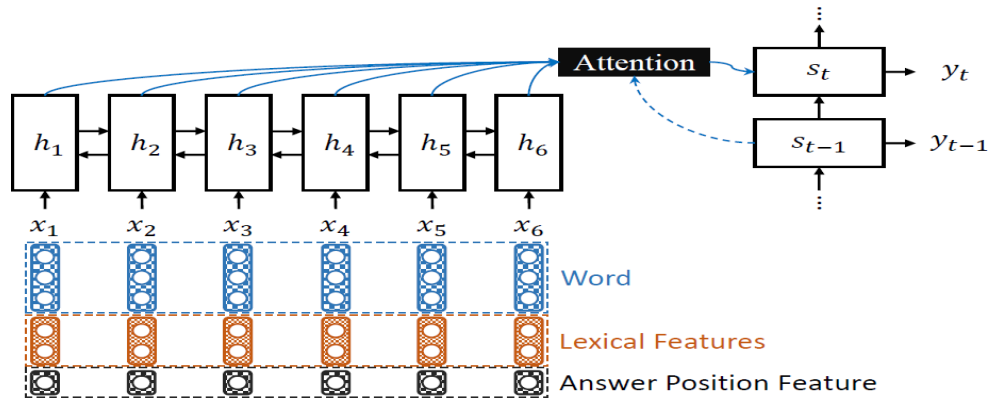
Figure 3. Overview of Question Generator Model" AQGM"[13].

sequence $(\vec{h1}, \vec{h2}, ..., \vec{hn})$ and the backward sequence $(\overleftarrow{h1}, \overleftarrow{h2}, ..., \overleftarrow{hn})$. Concretely, the BiGRU encoder reads the concatenated sentence word vector, custom-built lexical features, and answer position feature to produce a sequence of word-and-feature vectors [13].

Answer Position Feature the proposed model uses the answer position feature to locate the target answer, which generates a question concerning a specific answer in the input sentence. In this model, the BIO tagging scheme is used to label the position of a target answer. Table 1 illustrates the BIO tagging scheme.

Table 1. BIO tagging scheme

Tag	Meaning
B	denotes the start of an answer
I	continues the answer
O	marks words that do not form part of the answer

The BIO tags of answer position are represented as real-valued vectors in the embedding layer of the feature-rich encoder. With the BIO tagging feature, answer-focused questions are generated by encoding the answer position to the hidden vectors [11].

Lexical Features Along with the sentence words, AQGM feeds other lexical features to the encoder. Word case, POS, and NER tags are selected as the lexical features to capture and encode more linguistic information.POS and NER tags features are very significant pre-processing steps for many NLP tasks, such as information retrieval, parsing, Text to Speech (TTS) applications, and question generation, they are used to find patterns in sentences and also to guide the selection of the

question word. Taking into account that SQuAD is constructed using Wikipedia articles, which contain lots of named entities, therefore incorporating NER in the lexical features helps for identifying them.

4.3.2 Attention-Based Decoder

To generate the questions AQGM utilizes an attention-based GRU to summarize the sentence and answer information. The GRU decoder produces the new hidden state that depends on the previous word embedding together with the context vector that arises from RNN hidden states. To initialize the decoder GRU hidden state, the first backward RNN hidden state is used through a linear layer. To calculate the readout state, the previous word embedding is combined with the current context vector and the current decoder state, then the readout state is passed to a maxout hidden layer and a softmax layer to predict the next word in the target question based on the given words over the decoder vocabulary[13].

Copy Mechanism deals with the uncommon and unknown words problem using a pointing mechanism to dynamically copy the words from the input sequence while decoding. When decoding a word, the copying mechanism provides a switch for the model, whose input is a concatenation of the current decoder state and context vector to generate the probability of copying a word from the input sequence [8].

5. Results and Evaluation

The question generation model gets BELU-4 score 11.3 which is good in" Automatic Question Generation" topic by Machine Learning.

5.1 Examples of generated questions

Table 2. Sample of perfect output 1

Sentence	the rhine is the longest river in germany.
The gold question	the rhine is the longest river in what country?
Generated question	the rhine is the longest river in what country?

Table 3. Sample of perfect output 2

Sentence	the v&a is in discussion with the university of dundee, university of abertay, dundee city council and the scottish government with a view to opening a new # 43 mil-lion gallery in dundee that would use the v&a brand although it would be funded through and operated independently.

The gold question	when did the methodist protestant church split from the methodist episcopal church?
Generated question	when did the methodist protestant church split from the methodist episcopal church?

Table 4. Sample of perfect output 3

Sentence	in 1830, the methodist protestant church split from the methodist episcopal church over the issue of laity having a voice and vote in the administration of the church, insisting that clergy should not be the only ones to have any determination in how the church was to be operated.
The gold question	why did the methodist protestant church split from the methodist episcopal church?
Generated question	why did the methodist protestant church split from the methodist episcopal church?

Table 5. Sample of good output 1

Sentence	private education in north america covers the whole gamut of educational activity, ranging from pre-school to tertiary level institutions.
The gold question	what region does use the term 'private schools' to refer to universities?
Generated question	where is private education located?

Table 6. Sample of good output 2

Sentence	in july 1888, brown and peck negotiated a licensing deal with george westing house for tesla's polyphase induction motor and transformer designs for $ 60,000 in cash and stock and a royalty of $ 2.50 per ac horsepower produced by each motor.
The gold question	when did the deal between peck, brown and westing-house take place?
Generated question	when did brown and peck negotiate a license?

Table 7. Sample of bad output

Sentence	in scotland , anyone wishing to teach must be registered with the general teaching council for scotland -lrb- gtcs-rrb-.

The gold question	what is considered an all graduate profession in scotland?
Generated question	what is the school of scotland that anyone to teach school for scotland?

5.2   Examples from GUI

The GUI is made to allow the user to interact with the system through an easy to use and well-designed userinterface. The GUI consists of the following screens:

1. Generate Question

2. Displaying Generated Questions

3. Export Question Bank

4. Generate Exam

The user is represented with two tabs in the first screen displayed. The user can switch between these tabs,each tab opens a different screen.

Figure 4. Generate Question" Start Screen."
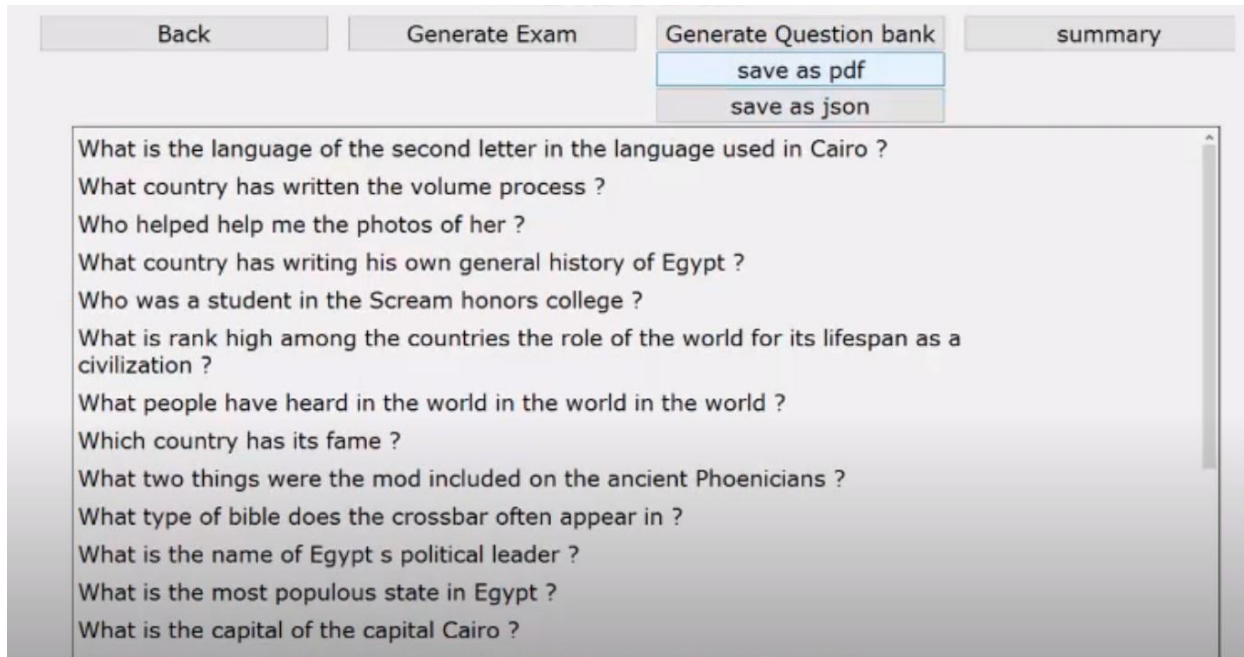
Figure 5. Displaying Generated Questions" Second Screen."
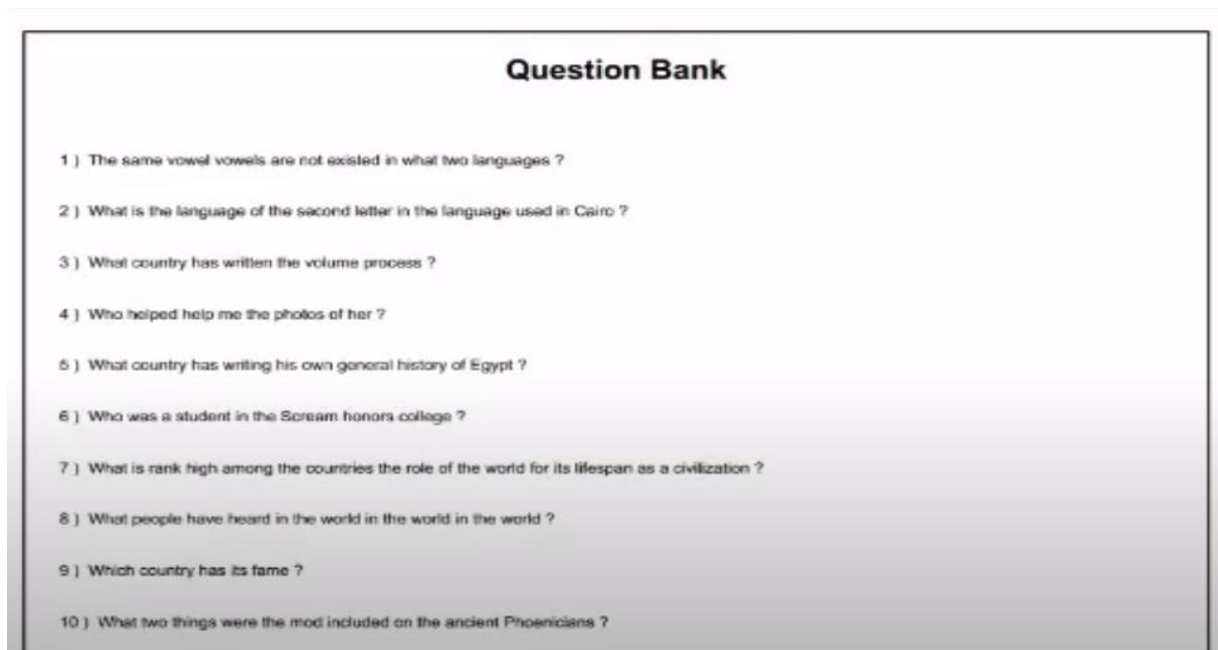
Figure 6. Question Bank PDF

Figure 7. Generate Exam Page.

Figure 8. Exam PDF

5.3 Types of Generated Questions

As shown in Figure 9, The number of gold questions (for each Wh-question type) is close to the generatedones.
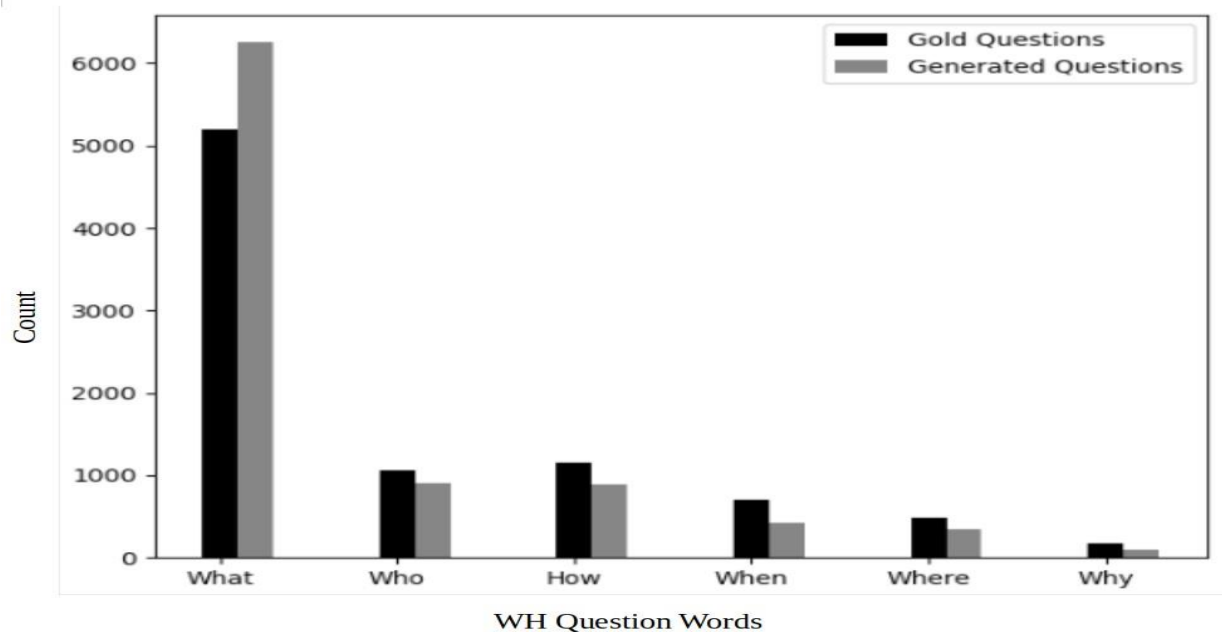
Figure 9.
Types of
Generated
Questions

6.   Analysis

As shown in the results section, the generated questions can be classified into three categories which are perfect, good, and bad that depend on question quality. In the first case, the generated question is the same question as humans wrote before in SQuAD that will be classified as a Perfect question. In the second case, if the generated question is not the same question as human wrote before in SQuAD but the generated question has a meaning and asked right question from the sentence that will be classified as a good question. In the third case, the generated question is a question that hasn't a meaning that is related to the sentence. The generated questions have more types of the most common type is a what question which makes sense because SQuAD is biased to what question and in real life most questions are asked by what.

7.   Conclusions and Future Work

AQGM generates automated questions using seq2seq technique. The question generation model takes the output sentences of preprocessing phase to generate Wh-questions with different difficulty levels by extending the sequence-to-sequence models. The model evaluation gets 11.3 using BLEU-4 score metric which shows an overall good performance regarding automatic question generation using deep learning approach. The generated questions are stored in the question bank template. AQGM facilitates user interface to specify the number of questions and difficulty levels, then classifies the questions as (Easy, Medium, and Hard), specifies the question using random selection, and writes in exam template with also different exam models. More question types would like to be considered like MCQ, fill in the blank, complete, and so on. Therefore, in the future, this work will be extended to generate different question types not only this feature will be added but also, AQGM itself will be converted from supervised to be reinforcement learning. Reinforcement learning technique will improve the BLEU score.

References

1. ANDERSON, R. C., AND BIDDLE, W. B. On asking people questions about what they are reading. vol. 9 of *Psychology of Learning and Motivation*. Academic Press, 1975, pp. 89–132.

2. ANDREAS PAPASALOUROS, K. K., AND KOTIS, K. Automatic generation of multiple choice questions from domain ontologies. In *IADISMulti Conference on Computer Science and Information Sys- tems,Proceedings of e-Learning 2008 ,July 22-25*. (Amsterdam, The Netherlands, Jan. 2008), pp. 427–434.

3. AWAD, A. E., AND DAHAB, M. Y. Automatic generation of question bank based on pre-defined templates. In *IJIACS Academic Science* (April 2014), vol. 3, International Journal of Innovations & Advancement inComputer Science.

4. FATTOH, I. E. Semantic based automatic question generation using artificial immune system. Computer Engineering and Intelligent Systems www.iiste.org ,ISSN 22221719.

5. FUNG, Y.-C., KWOK, J. C.-W., LEE, L.-K., CHUI, K. T., ET AL. Automatic question generation system for english reading comprehension. In *International Conference on Technology in Education* (2020),Springer, pp. 136–146.

6. GANGOPADHYAY, S., AND RAVIKIRAN, S. M. Focused questions and answer generation by key content selection. In *2020 IEEE Sixth International Conference on Multimedia Big Data*

(BigMM) (2020), pp. 45–53.

7. GANGOPADHYAY, S., AND RAVIKIRAN, S. M. Focused questions and answer generation by key content selection. In *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)* (2020), pp. 45–53.

8. HE, S., LIU, C., LIU, K., AND ZHAO, J. Generating natural answers by incorporating copying and retrieving mechanisms in sequence-to-sequence learning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (2017), pp. 199–208.

9. KUMAR, V., JOSHI, M., RAMAKRISHNAN, G., AND LI, Y.-F. Vocabulary matters: A simple yet effective approach to paragraph-level question generation. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference onNatural Language Processing* (Suzhou, China, Dec. 2020), Association for Computational Linguistics, pp. 781–785.

10. LE, N.-T., KOJIRI, T., AND PINKWART, N. Automatic question generation for educational applications– the state of art. In *Advanced Computational Methods for Knowledge Engineering* (Cham, 2014), T. vanDo, H. A. L. Thi, and N. T. Nguyen, Eds., Springer International Publishing, pp. 325–338.

11. LOVENIA, H., LIMANTA, F., AND GUNAWAN, A. Automatic question-answer pairs generation from text.

12. PEVERLY, S. T., AND WOOD, R. The effects of adjunct questions and feedback on improving the read- ing comprehension skills of learning-disabled adolescents. *Contemporary Educational Psychology 26*, 1 (2001), 25–43.

13. QINGYU ZHOU, NAN YANG, F. W. C. T. H. B., AND ZHOU, M. Neural question generation from text: A preliminary study. In *Natural Language Processing and Chinese Computing - 6th CCF International Conference, NLPCC 2017, Dalian, China, November 8-12, 2017, Proceedings* (Harbin,China, Apr. 2017),pp. 662–671.

14. ROEDIGER III, H. L., AND KARPICKE, J. D. The power of testing memory: Basic research and implications for educational practice. *Perspectives on psychological science 1*, 3 (2006), 181–210.

15. ROTHKOPF, E. Z. Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal 3*, 4 (1966), 241–249.

16. ZHOU, Q., YANG, N., WEI, F., TAN, C., BAO, H., AND ZHOU, M. Neural question generation from text: A preliminary study. In *Natural Language Processing and Chinese Computing* (Cham, 2018), X. Huang, J. Jiang, D. Zhao, Y. Feng, and Y. Hong, Eds., Springer International Publishing, pp. 662–671.