AiXAM - AI assisted Online MCQ Generation Platform using Google T5 and Sense2Vec

*Dr. Dhanya N M
Department of Computer Science & Engineering, Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India
nm_dhanya@cb.amrita.edu

Balaji R K
Department of Computer Science & Engineering, Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India
balajirk2308@gmail.com

Akash S
Department of Computer Science & Engineering, Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India
akashsachithanandham@gmail.com

***Abstract*— Multiple Choice Question (MCQ) tests provide a quick way to assess knowledge objectively. This pandemic scenario has transformed majority of exams as MCQ based. However, the manual creation of quality MCQ tests is a tedious and time-consuming process. This paper proposes a method to automatically build such tests from a given piece of text. Thus, it is expected to reduce the time and effort required to formulate a MCQ test. This will aid schoolteachers, college faculties and students (for self-assessment). This paper first introduces the readers to the components of an MCQ, then presents some background on such systems with relevant examples. Moreover, it discusses the list of techniques and a few approaches adopted from literature, later shows the implementation of the proposed system - aiXAM MCQ generation platform. Towards the end of this paper, survey results and inferences pertaining to the system have been presented.**

Keywords—Natural Language Processing, multiple choice questions, Google T5, Sense2vec

I. INTRODUCTION

MCQ tests are seen as an effective way to assess students' knowledge, being able to cover a wide range of topics in a limited period of time. Each MCQ consists of a question and a group of responses, only one of which is correct. Incorrect answers are called distractors.

A. Advantages

Multiple-choice questions have a number of benefits over other types of questions. These are,

• Quick evaluation. The evaluation of descriptive questions, on the other hand, consumes comparatively longer time.

• Machine evaluation. MCQs may be evaluated electronically or automatically without human involvement.

• Consistent scoring. The scoring of descriptive questions, on the other hand, can vary from examiner to examiner.

• Factors unrelated to the content being evaluated are not taken into account. Handwriting and presentation consistency, for example, have no bearing on the marking.

• Time required to conduct and complete a test is less.

B. Structure

There are three main elements in MCQ. They are,

- Stem
- Key
- Distractors

The stem is otherwise referred to as the question or the question sentence. It can be presented in a variety of ways, including fill-in-the-blank or interrogative sentences.
The key is the only right answer to the presented question.
Distractors are the set of incorrect options supplied along with the right answer to confuse the test-taker.

To make things simpler, the below MCQ is considered as an example.
The capital of Finland is —.

1. Oslo
2. Helsinki
3. Stockholm
4. Dublin

The stem of this MCQ is "The capital of Finland is". It could also be presented in interrogative/question form: "What is the capital of Finland?". There are four options provided with the question. Among these options, 'Helsinki' is the right one. So, 'Helsinki' is the key. The remaining three options are the distractors.

MCQ generation involves the following major processes

1. Sentence selection
2. Key selection
3. Question formation
4. Distractor generation

The section 2 of this paper, looks at the existing MCQ generation systems. The forthcoming sections, explain the components of the implementation and share the results achieved.

II. EXISTING SYSTEMS

There are several mechanisms where text processing can be done [37] such as, deals with the review of the same [10], deals with sentiment analysis on tweets, where the text data is analysed to get the sentiments. Neural networks can be used for text prediction [12], in which a recurrent neural network is used [23] to combine document summarisation and topic modelling, and to evaluate them. Since it is a combination, it produced better accuracy than Latent Dirichlet Allocation (LDA) topic modelling [31] that deals with two level text summarizations. This generates summaries of a document, for a given corpus with multiple topics [35]. An application has been created for question generation using T5 Transformers. This model performs very well on unseen data and generates well-formed and grammatically correct questions. To measure the news information, a quiz style MCQ type of questions was generated for a news article [36]. NewsQuizQA dataset was generated and the model was checked against the dataset and achieved high accuracy. A transformer-based approach was used in [38] for creating robust question generation system. A pre-trained language generation model was used that outperformed more complex models such as RNN and Seq2Seq models with 8.62 and 14.27 increase in meteor and Rougue_L Scores respectively.

This section presents the major steps involved in MCQ generation with their various approaches and examples from existing literature.

A. Sentence selection

Valid questions cannot be generated from any sentence belonging to the text. A candidate sentence should include a fact that can be put forth as a question. The following are the popular approaches for the same with examples from existing literature that use it.

- *Sentence Length:* A sentence of few words cannot provide enough detail to frame questions. Because of various facts and relationships, an exceedingly long sentence adds ambiguity. Several journals, like [21], [2], [3], [4], [5], and [22], used sentence length as a criterion for sentence selection.
- *Parts-of-speech Information:* Along with sentence length, reference [6] looked for occurrences of verb form patterns. In reference [7], adjective-noun pairs were picked as candidate sentences.
- *Machine Learning:* Many papers have described the use of machine learning for this purpose. For instance, Naive Bayes has been used by Hoshino and Nakagawa [8], Neural Network by Banchs, D'Haro and Kumar [9], Support Vector Machine by Correia et al. [2].

B. Key Selection

Selecting the right key (answer word/phrase) is of utmost importance in framing a MCQ.
Some approaches are listed below.

- *Part-of-speech (POS):* In certain domains or implementations, observations suggest that certain POS types have a higher probability of becoming the keyword. For instance, Lee and Seneff [11] considered prepositions to be the keyword. Similarly, Sumita, Sugaya, and Yamamoto [1] considered verbs as keywords. Agarwal and Mannem [13], Mitkov, Ha, and Karamanis [14] and Gates et al. [16] considered particular POS types for selection of keywords.
- *Pattern Matching:* Pattern matching involves finding a matching sentence structure. Chen, Liou, and Chang [15] identified a few common patterns for key selection by observing sentences with common structural features. Keywords were identified with the help of those patterns. Elsewhere, Gates et al. [16] identified the keys using syntactic patterns.
- *Machine Learning:* Machine learning techniques were used by Hoshino and Nakagawa [8] to select verbs, idiom pieces, and adverbs as keywords. A conditional random field classifier was used by Goto et al. [17] & [18] to select the key.

C. Question formation

As a general observation, in many systems, conversion of 'fill in the blanks' to the interrogative form of the sentence has been ignored. However, the work of some other systems that presents questions in interrogative form has been summarized. A majority of the works use rule and pattern-based approaches. However, these rules and patterns differ for each systems.

- *Choosing the right question word:* Various works sought to determine the right question word and construct the query appropriately. For instance, "what" denotes an entity, "when" denotes time, "where" denotes place, "who" denotes persons, and so on. Majumder and Saha [4] examined the target sentence's structure to determine the position of the key. Then, using grammar rules, they identified the appropriate question word and constructed the query.
- *Knowledge in Sentence*: A rule-based methodology was used by Pabitha et al. [19]. The rules were determined by the class of information contained in any particular sentence. A variety of information classes were employed - concept, definition, example, calculation, method, and outcome. For example, the rules were escorted by these knowledge labels. If it's a definition, the question is "What does X mean?" If it's a process, the question is "How do you perform X?" and so on.
- *Syntactic Transformation:* Heilman [20] used syntactic transformation to create questions by identifying the answer phrases. Their main steps were to mark unmovable phrases, use a rule-based approach to generate potential question phrases, splitting of the main verb, inversion of subject and auxiliary verbs, removal of answers and insertion of question words/phrases.

D. Distractor generation

The role of distractors is significant in the generation of MCQs. The quality of a MCQ is primarily determined by the quality of the distractors. If the distractors are unable to adequately confuse the test-taker, he/she can promptly choose the right answer. As a consequence, the overall purpose and usefulness of the MCQ suffer. In literature, many approaches to distractor generation have been used. A few of them are summarized below.

- *Using POS*: A distractor is chosen such that it has semantic similarity with the answer/key. Therefore, the key and the distractors invariably belong to the same part of speech. In [1], [17], [3], [13], [24] this finding is used as an indication.
- *WordNet:* WordNet [25] contains an organized description of the lexemes present in English, which includes an inventory of known morphemes and information about their meaning. WordNet organizes terms into collections of synonyms called "synsets" and tracks the relationships between these synonym sets and their member elements. Concepts that display semantic similarities can be used as a distractor. As a result, many researchers describe the use of WordNet to generate distractors.
- *Distributional Hypothesis:* According to the distributional hypothesis, identical terms occur in

similar contexts. Smith, Sommers, and Kilgariff [29], Mitkov et al. [27], Karamanis, Ha, and Mitkov [26] and Afzal and Mitkov [28] used a distributional similarity-based technique. Lee and Seneff [18], and Agarwal and Mannem [13] explored a collocation-based technique.

In this section, the core processes behind a MCQ generation system is studied. The different approaches to each of the processes from literature has been identified. In Section III, MCQ generation platform – aiXAM has been proposed.

III. AIXAM MCQ GENERATION PLATFORM

aiXAM has been built as an end-to-end platform for creating, hosting, and managing MCQ tests online. The platform provides the following features:

i) *User Authentication* – For users to sign up and sign in, to the platform.

ii) *Test creation* – Test creators can create a MCQ test by entering test details like start time, end time, time limit, subject, topic, test password etc. Once test is created, a test id is generated by the system which he/she can share with the test takers.

iii) *MCQ generation* – During test creation, user can submit a file with text that he/she intends to create MCQs based upon. Once the MCQs are generated, user can review the MCQs and add/modify/remove any question as per his/her wish before publishing the test.

iv) *Test environment* – The test-taker needs to enter the test ID and password to attempt the test. The test environment displays the MCQs for which he/she has to choose the correct answer. The test is submitted, once the test-taker clicks 'Finish Test' or auto-submitted on the completion of time limit.

v) *Results* – After attempting the test, test-taker can view his marks for all the tests he/she has attempted. The test creator also can view the results of each test-takers. He/she can also download the results in CSV format.

Except the above feature iii) all others are standard occurrences on any other online test platforms. Hence, only on feature iii) has been expanded in section IV.

IV. MCQ GENERATION IN AIXAM

Architecture

The following Fig. – 1 shows the architecture diagram for the proposed system.

A. Text extraction using Textract

As discussed in the previous section, the test-creator can submit a text file which he wants the MCQs to be based upon. This file can be of any of the popular text formats like .txt, .doc, .pdf, .odf, .pptx etc. Textract library on python has been used to extract the text from the files.

B. QA generation using Google Text-To-Text-Transfer Transformer

Many of the modern NLP systems are based on the Transformer architecture which is introduced by Vaswani et al. [32]. These days, there is a large variety of different architectures. After reading about several recent architectures, Google's T5 model has been considered, which was introduced by Raffel, C et al. [30].

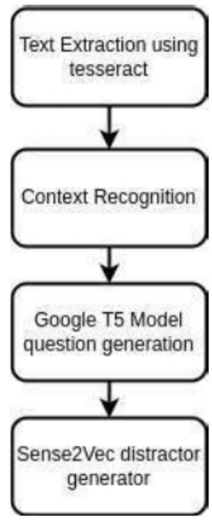

Fig. – 1 Architecture Diagram

Google T5's fundamental principle is to reframe all NLP tasks as sequence-to-sequence tasks. For e.g., in summarization, the model accepts the text to be summarized as an input sequence and returns the summary as a sequence. The model takes the text to be analyzed as an input sequence and outputs a sequence that specifies the sentiment of the text. This is useful because, although the model was not constructed or pretrained with Question-Generation in mind, it can be easily repurposed for Question-Generation. The response and context can simply be used as inputs, and the model can be trained to give a question as the output sequence.

The question generation model fine-tuned over Google T5 by Suraj Patil et al. [34] has been used to generate question-answer pairs from a given text corpus. The article is summarized into few important sentences and keyword is taken from each sentence to form the answers. The sample sentence is given below.

Amrit's favourite sport is **cricket**

And the question generated will be,

"What is Amrit's favourite sport?" and the MCQ can be the following with cricket as the correct answer and 3 distractor words. The distractor word generation is explained in the next section.

FootBall
Cricket
Polo
Hockey

Given the input sentence, Amrit's favourite sport is **cricket** , the keyword (cricket) is extracted and the sentence is given to the T5 transformer algorithm. This will take context, the sentence and keyword, cricket, as input and generate a question. Since context is taken as an input, the following sentence will generate a different question since the context for the work cricket is totally different. For e.g.,

Amrit is annoyed by **cricket** in his room.
What insect is in Amrit's room?

Grasshopper
Cricket
Cat
Dog

C. Distractor generation using Sense2Vec

Andrew Trask, Phil Michalak, and John Liu [33] presented a subtle modification over word2vec that lets us learn more interesting and detailed word vectors. Semantic similarities of a given key with the word embeddings in a pre-trained sense2vec vector like the s2v_reddit_2015_md, trained on Reddit threads of the year 2015 has been found. The ones with the closest semantic similarities are chosen as the distractors. Sense2Vec claims advantage over other distractor generation methods, on the fact that it is able to provide distractors for phrases rather than just a word. This is a neural network algorithm trained on millions of sentences to predict a focus word given other words, or predict surrounding words given a focus word. The second method is used in this paper to generate distractor words. The vector representation is in such a way that similar kind of words in real world will be closer in vector space, thus creating very similar words as distractors.

Once the above steps are performed, the auto-generated MCQs are displayed for the test-creator to make modifications to the questions, answers, and distractors, if necessary, before publishing the test.

It is observed that, despite some MCQs being erratic, it greatly reduces the effort and time taken to create a MCQ test. To verify this, a survey was taken asking for their feedback on this platform. The results of the survey are interpreted in section V.

V. RESULTS AND EVALUATIONS

The system is evaluated by 2 methodologies. One using a metric and another by manual evaluation using a questionnaire. The document is converted into summary sentences and from those summary sentences, the questions are generated. A metric called ROUGE is used for evaluation at this stage. The Recall-Oriented Understudy for Gisting Evaluation (ROUGE) scoring algorithm calculates the similarity between a candidate document and a collection of reference documents. It also evaluates the quality of document translation and summarization models. To be more specific, this paper uses ROUGE-L and ROUGE-N as a measure.

ROUGE–L measures the Longest Common Sub-sequence (LCS) between the original text and the summarized one. It counts the longest shared sequence of words shared by both. The idea here is that a longer shared sequence would indicate more similarity between the two sequences. ROUGE-N measures the number of matching 'n-grams' between the model-generated text and a reference.

TABLE – I ROUGE SCORE

Sample	ROUGE Score			
	Paragrapgh	*Sentences generated*	*ROUGE-N F-Score*	*ROUGE-L F-Score*
Science	An electric charge will experience a force if an electric field is applied.	What will be the experience on an electric charge if an electric field is applied?	0.56	0.56
History	Buddhist missionaries from China introduced hand printing technology to Japan	Who introduced hand printing technology to Japan?	0.53	0.53

User experience is very important for a system like this. So a survey was conducted asking the users for their feedback on this platform. A total of thirty-one users responded to the survey. The survey questions and responses were recorded on Google Forms. The relevant survey questions are listed below, along with the graphical representations and key takeaways from them.

A. Rate the test creation experience on a scale of 1-5 for ease of use. (1 for 'poor' and 5 for 'It was a breeze!')

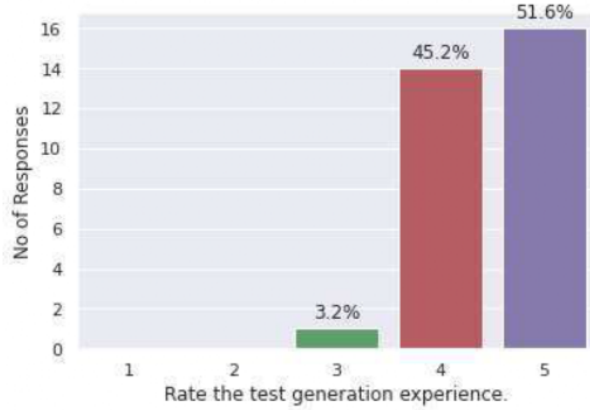

Fig. 2: Test generation Experience

45.2% of users gave a rating of 4 out of 5 and 51.6% gave a rating of 5 out of 5. This response indicates that users prefer assistance and suggestions to build MCQs while creating tests. This makes it easier for them to create MCQ based tests shown in fig. 2.

B. What percentage of the questions (question sentences) did you find relevant?

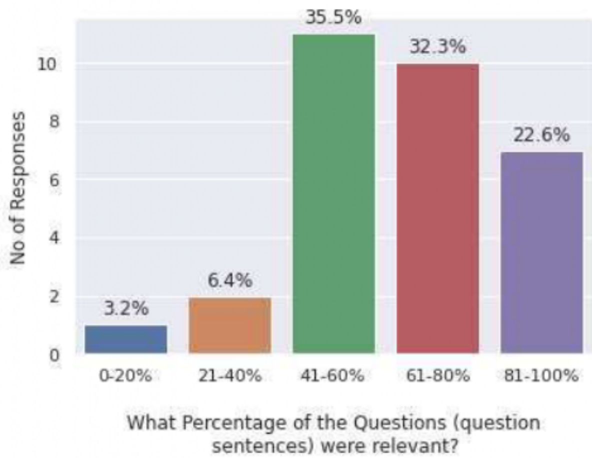

Fig. 3: Relevant Question Analysis

35.5% of users found about 41-60% of the generated questions to be useful/relevant. 32.3% of users said that 61-80% of the generated questions are relevant. About 22.6% of users found 81-100% of the generated questions to be relevant. Also 6.4% of users found 21-40% of the generated questions to be relevant as shown in fig 3.

C. Percentage of MCQs for which incorrect options were auto generated.

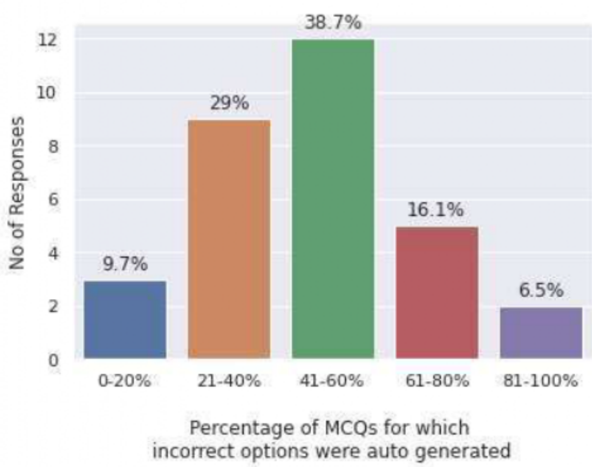

Fig. 4: Incorrect Option Analysis

Text can contain various keywords or phrases belonging to different subject domains. The robustness of any distractor generation model lies in its ability to generate distractors for all kinds of keywords. Here, 22.6% of users claimed that they had distractors are generated for at least 60% of the MCQs. To improve this, training the model with more examples and increased focus on domain-based learning can be explored as shown in figure 4.

D. Percentage of auto generated incorrect options that were relevant.

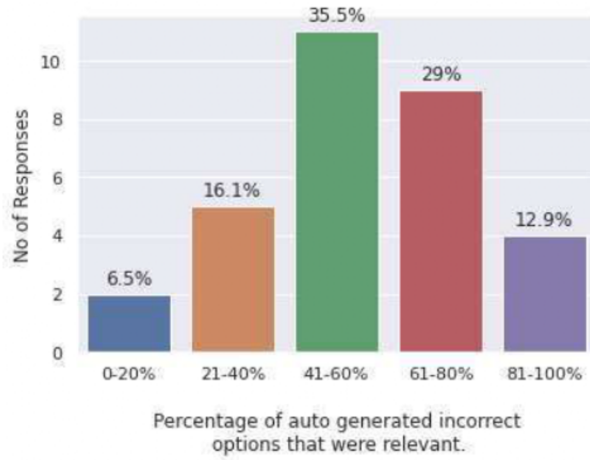

Fig. 5: Relevant Option Analysis

The usefulness of a distractor generation model is determined by its ability to generate the 'right' distractors. Here 'right' distractors mean that they are indeed close to, or share a relationship with the answer keyword. This is the underlying notion behind this survey question. The results are displayed above as shown in fig 5.

E. What extent of time and effort did you save on creating a test using aiXAM platform vs Conventional platforms?

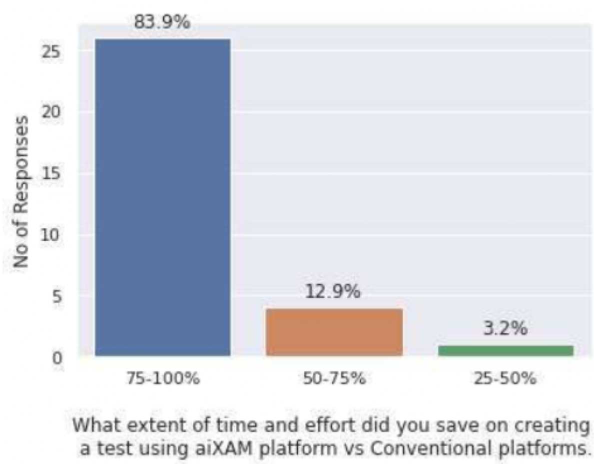

Fig. 6: XAM Platform Vs Conventional Platforms

As opposed to conventional test creation platforms, a whopping 83.9% of users felt that this platform drastically reduced their time and effort by 75-100% in creating a MCQ test. 12.9% felt that 50-75% of their time and effort was saved. Zero users felt that this platform did not save any time and effort. This positive response can be largely attributed to the MCQ generation feature offered as shown in fig 6.

F. Your overall rating for the product.

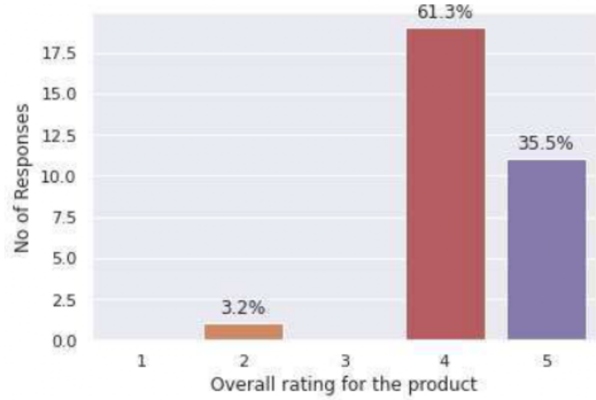

Fig. 7: Overall Rating of the Product

Overall rating is considering all aspects of the platform – UI/UX, test creation, attempting tests etc as shown in fig 7.

G. Would you pay to use a more mature version of this product?

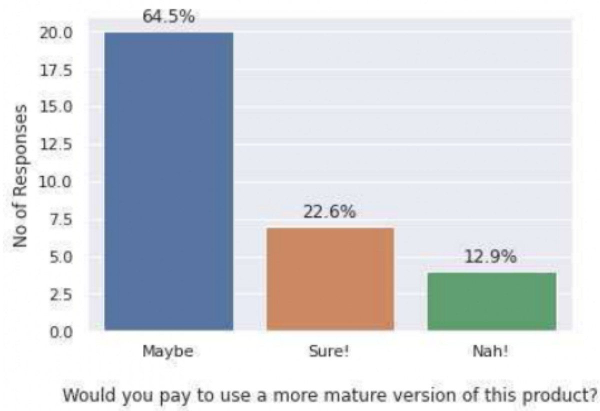

Fig. 8: Mature version

Potential funding options is required to keep the site going and to improve the customer experience in the future. With that in mind, the users were asked if they would be willing to pay for a more mature version of this platform. 22.6% of users responded positively, even though small, it indicates that there might be a potential customer base who find the service worth paying for. Also 64.5% of users were undecided, hinting that some of them might opt in once they see a matured version as shown in fig 8.

VI. CONCLUSION

This paper discusses the important processes and techniques involved in MCQ generation, along with relevant examples from the literature. Although there exists a lot of literature on question generation, not a lot of systems have evolved into products that are available for use. However, the aim was to provide a solution to this by providing a platform - aiXAM, to assist in MCQ test creation. The positive survey results are a testament to the fact that providing such solutions will benefit schoolteachers, college faculty, students (for self-assessment) and other professionals by saving them a lot of time, effort, and potential stress. Looking into the future, teaching assistant robots is another space where this finds an application. With rapid developments in deep-learning based language models, it is expected that the quality of question-answer generation and distractor generation will also improve quickly. This will help to provide better assistance to users, further saving their precious time and effort.

REFERENCES

[1] E. Sumita, F. Sugaya, and S. Yamamoto, "Measuring non-native speakers' proficiency of English by using a test with automatically-generated fill-in-the-blank questions," in Proc. 2nd Workshop Building Edu. Appl. Using NLP, 2005, pp. 61–68.

[2] R. Correia, J. Baptista, M. Eskenazi, and N. Mamede, "Automatic generation of cloze question stems," in Proc. 10th Int. Conf. Comput. Process. Portuguese Lang., 2012, pp. 168–178.

[3] R. Correia, J. Baptista, N. Mamede, I. Trancoso, and M. Eskenazi, "Automatic generation of cloze question distractors," in Proc. Interspeech Satellite Workshop 2nd Lang. Stud.: Acquisition Learn. Edu. Technol., 2010, pp. 1–4.

[4] M. Majumder and S. K. Saha, "A system for generating multiple choice questions: With a novel approach for sentence selection," in Proc. 2nd Workshop Natural Lang. Process. Techn. Edu. Appl., 2015, pp. 64–72.

[5] J. Pino, M. Heilman, and M. Eskenazi, "A selection strategy to improve cloze question quality," in Proc. Workshop Intell. Tutoring Syst. IllDefined Domains 9th Int. Conf. Intell. Tutoring Syst., 2008, pp. 22–32.

[6] I. Aldabe, M. Maritxalar, and R. Mitkov, "A study on the automatic selection of candidate sentences distractors," in Proc. 14th Int. Conf. Artif. Intell. Edu. Workshops, 2009, pp. 656–658.

[7] Y.-C. Lin, L.-C. Sung, and M. C. Chen, "An automatic multiple-choice question generation scheme for English adjective understanding," in Proc. Workshop Model. Manage. Gener. Problems/Questions eLearn. 15th Int. Conf. Comput. Edu., 2007, pp. 137–142.

[8] A. Hoshino and H. Nakagawa, "A real-time multiple-choice question generation for language testing: A preliminary study," in Proc. 2nd Workshop Building Edu. Appl. Using NLP, 2005, pp. 17–20.

[9] G. Kumar, R. E. Banchs, and L. F. D'Haro, "Automatic fill-the-blank question generator for student self-assessment," in Proc. IEEE Frontiers Edu. Conf., 2015, pp. 1–3.

[10] Dhanya, N.M., Harish, U.C., Detection of Rumors in Tweets Using Machine Learning Techniques Lecture Notes in Electrical Engineering, 2021, 700, pp. 3095–3111

[11] J. Lee and S. Seneff, "Automatic generation of cloze items for prepositions," in Proc. 8th Annu. Conf. Int. Speech Commun. Assoc., 2007, pp. 2173–2176.

[12] Bindu K.R.,, Aakash C., Orlando B., Latha Parameswaran .p, "An algorithm for text prediction using neural networks", Lecture Notes in Computational Vision and Biomechanics, Springer Netherlands, Page No.186 - 192, ISSN: 22129391.

[13] M. Agarwal and P. Mannem, "Automatic gap-fill question generation from text books," in Proc. 6th Workshop Innovative Use NLP Building Edu. Appl., 2011, pp. 56–64.

[14] R. Mitkov, L. An Ha, and N. Karamanis, "A computer-aided environment for generating multiple-choice test items," Natural Lang. Eng., vol. 12, no. 2, pp. 177–194, Jun. 2006.

[15] C.-Y. Chen, H.-C. Liou, and J. S. Chang, "FAST: An automatic generation system for grammar tests," in Proc. COLING/ACL Interactive Presentation Sessions, 2006, pp. 1–4.

[16] D. Gates, G. Aist, J. Mostow, M. McKeown, and J. Bey, "How to generate cloze questions from definitions: A syntactic approach," in Proc. AAAI Fall Symp., 2011, pp. 19–22.

[17] T. Goto, T. Kojiri, T. Watanabe, T. Iwata, and T. Yamada, "An automatic generation of multiple-choice cloze questions based on statistical learning," in Proc. 17th Int. Conf. Comput. Edu., 2009, pp. 415–422.

[18] T. Goto, T. Kojiri, T. Watanabe, T. Iwata, and T. Yamada, "Automatic generation system of multiple-choice cloze questions and its

evaluation," Knowl. Manage. E-Learn. Int. J., vol. 2, no. 3, pp. 210–224, 2010.

[19] P. Pabitha, M. Mohana, S. Suganthi, and B. Sivanandhini, "Automatic question generation system," in Proc. Int. Conf. Recent Trends Inf. Technol., 2014, pp. 1–5.

[20] M. Heilman, "Automatic factual question generation from text," Ph.D. dissertation, Language Technol. Inst., School Comput. Sci., Pittsburgh, PA, USA, 2011, Art. no. aAI3528179.

[21] D. Coniam, "A preliminary inquiry into using corpus word frequency data in the automatic generation of english language cloze tests," CALICO J., vol. 14, no. 2–4, pp. 15–33, 1997.

[22] J. C. Brown, G. A. Frishkoff, and M. Eskenazi, "Automatic question generation for vocabulary assessment," in Proc. Conf. Human Lang. Technol. Empirical Methods Natural Lang. Process., 2005, pp. 819–826.

[23] subathra P., Chidambaram A.R.V., Ragapriya D., Pragadeesh C., "Document summarization using topic modeling", Journal of Advanced Research in Dynamical and Control Systems, Volume 10, Issue 5 Special Issue, 2018, Pages 1773-1781,

[24] R. Mitkov and L. A. Ha, "Computer-aided generation of multiplechoice tests," in Proc. HLT-NAACL Workshop Building Edu. Appl. Using Natural Lang. Process., 2003, pp. 17–22.

[25] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, "Introduction to WordNet: An on-line lexical database," Int. J. Lexicography, vol. 3, no. 4, pp. 235–244, 1990.

[26] N. Karamanis, L. A. Ha, and R. Mitkov, "Generating multiple-choice test items from medical text: A pilot study," in Proc. 4th Int. Natural Lang. Gener. Conf., 2006, pp. 111–113.

[27] R. Mitkov, L. A. Ha, A. Varga, and L. Rello, "Semantic similarity of distractors in multiple-choice tests: Extrinsic evaluation," in Proc. Workshop Geometrical Models Natural Lang. Semantics, 2009, pp. 49–56.

[28] N. Afzal and R. Mitkov, "Automatic generation of multiple choice questions using dependency-based semantic relations," Soft Comput., vol. 18, no. 7, pp. 1269–1281, Jul. 2014.

[29] S. Smith, S. Sommers, and A. Kilgarriff, "Learning words right with the sketch engine and WebBootCat: Automatic cloze generation from corpora and the web," in Proc. 25th Int. Conf. English Teaching Learn. Int. Conf. English Instruction Assess., 2008, pp. 1–8.

[30] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv preprint arXiv:1910.10683.

[31] Saikumar D., Subathra P. Two-Level Text Summarization Using Topic Modeling. In: Satapathy S., Bhateja V., Janakiramaiah B., Chen YW. (eds) Intelligent System Design. Advances in Intelligent Systems and Computing, vol 1171. Springer, Singapore.

[32] Vaswani, Ashish & Shazeer, Noam & Parmar, Niki & Uszkoreit, Jakob & Jones, Llion & Gomez, Aidan & Kaiser, Lukasz & Polosukhin, Illia, "Attention is all you need", 2017.

[33] Andrew Trask, Phil Michalak, and John Liu. sense2vec-a fast and accurate method for word sense disambiguation in neural word embeddings. arXiv preprint arXiv:1511.06388, 2015.

[34] S. Patil, "Question Generation using Transformers". [Online]. Available: https://github.com/patil-suraj/question_generation

[35] Grover, K., Kaur, K., Tiwari, K., Rupali, & Kumar, P. (2021). Deep learning based question generation using T5 Transformer. *Communications in Computer and Information Science*, 243–255. https://doi.org/10.1007/978-981-16-0401-0_18

[36] Lelkes, A. D., Tran, V. Q., & Yu, C. (2021). Quiz-style question generation for news stories. *Proceedings of the Web Conference 2021*. https://doi.org/10.1145/3442381.3449892

[37] Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2020). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education*, *30*(1), 121–204. https://doi.org/10.1007/s40593-019-00186-y

[38] Lopez, L.E., Cruz, D.K., Cruz, J.C.B., Cheng, C.: Transformer-based end-to-end question generation (2020)