iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 11 **Issue:** III **Month of publication:** March 2023

**DOI:** https://doi.org/10.22214/ijraset.2023.49390

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 11 Issue III Mar 2023- Available at www.ijraset.com*

# Automated Question Generator using NLP

Tejas Chakankar[1], Tejas Shinkar[2], Shreyash Waghdhare[3], Srushti Waichal[4], Asst Prof. Mrs. M. M. Phadatare[5]
[1, 2, 3, 4, 5]*Department of Computer Engineering AISSMS COE, Pune - SPPU*

***Abstract:*** ***The purpose of the Automated Paper Generator is to automatize the winning guide machine with the aid of the help of computerized gadgets and full-fledged laptop code, pleasurable their requirements, simply so their precious facts/data can be maintained for an extended quantity with clean gaining access to and manipulation of the desired software program and hardware are honestly available on the market and easy to discern with. automatic query Paper Generator, as represented above, will result in blunders loss, at ease, reliable, and brief control system. it can help the person to not forget their opportunity sports but rather to concentrate on file preserving. So, it's going to assist an organization in the higher utilization of sources. The enterprise will keep computerized records while not redundant entries. which means that one needn't be distracted by info it is now not applicable while having the ability to acquire the statistics. The goal is to automatize its existing guide gadget by way of the assistance of picked equipment and complete-fledged laptop software, pleasurable their requirements, just so their precious facts/statistics can be preserved for an extended quantity with sincere access to and manipulation of the equal. on the whole, the venture describes the manner to manipulate permanently overall performance and better offerings for the customers. Texts with capacity academic well-worth have ended up on the market through the net. however, the victimization of those new texts in lecture rooms introduces several challenges, one in every of which is that they typically lack compliance with physical activities and assessments. right here, we will be inclined to address part of this task by way of automating the creation of a specific form of assessment object and the use of various questions developing with taxonomies. specially, we concentrate on mechanically generating actual WH queries. Our purpose is to make an automatic gadget if you want to take as enter a text and flip out as output questions for assessing a reader's understanding of the records within the text. The questions would possibly then be supplied to a trainer, who would possibly pick and revise people who she or he judges to be beneficial.***
***Keywords: NLTK, POS Tagging, AQG, NLP, Word Net Lemmatizer.***

## I. INTRODUCTION

The "Automated Question Generator" has been developed This software program is supported to take away and, in a few cases, reduce the hardships confronted by way of this present system. to override the issues winning in the training guide system.

The utility is reduced as an awful lot as viable to avoid mistakes at the same time as coming into the facts. No formal information is needed for the user to use this system. accordingly, utilizing this all proves its miles are person friendly. Computerized question Paper Generator, as defined above, can result in error-free, relaxed, dependable, and rapid management structures. it can help the consumer to pay attention to their other sports instead of concentrating on record maintenance.

Every corporation, whether large or small, has challenges to conquer and manage the data of route, branch, query, challenge, and Semester. each computerized query Paper Generator has unique department wishes; therefore, we design distinctive worker management structures which are tailored to your managerial necessities. this is designed to assist in strategic making plans and will assist you to ensure that your business enterprise is geared up with the right degree of statistics and information on your destiny desires. also, for those busy executives who are continually on the pass, our structures include remote get-right of entry to features, with a view to can help you manage your body of workers whenever, at all times. those systems will in the end let you better control resources. Those automatic systems help us with many price and time-efficient answers. within the training area, the academicians are majorly dependent on their personnel for producing questions for numerous examinations. however, numerous successful tries were made for the development of automated evaluation structures. The paintings executed within the field of AQG, focus basically on the technology of easy conceptual questions, like who is the president of India? whilst becoming the first plane invented? Or what is supposed by way of the term 'cosmology"? this may no longer show to be very green for judging the scholars' learning. So, on way to efficiently investigate the students, step one is to design a question paper that covers all of the necessary elements to check his/her understanding.

Generally, the three major components of Question Generation are input pre-processing, sentence selection, and question formation. The input text is filtered by removing unnecessary words and punctuations that do not contribute to the meaning of the sentence. The sentences or phrases from which questions can be formed are segregated from the remaining text.

©IJRASET: All Rights are Reserved 

These are mapped to the type of question (what, where, when, etc.) that can be formulated with the selected sentence, followed by the final step of framing a grammatically sound question.

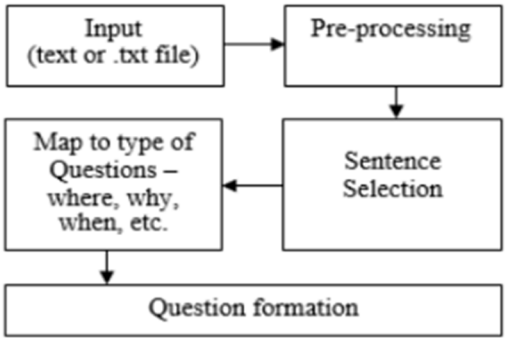



Fig. 1. Generic Approach of Question Generation

There are many changes being made now in various fields that tend to move from manual systems to automated systems. These automatic systems help us with less cost and time-efficient solutions. In the education field, the academicians are majorly dependent on their own for generating questions for various examinations.

## II. PROCESS OF QUESTION GENERATION AND CLASSIFICATION

The question generation method turned into carried out based on sample identification of every sentence to be at a loss for words. using pattern matching became chosen because it became taken into consideration clean to be carried out, had immoderate accuracy, and did not require any extra resources or devices. The scheme of producing questions and answers that have been classified in step with the degree of difficulty in the new taxonomy of bloom became proven in the figure. Paragraphs entered with the resource of the consumer have been broken down into consistent sentences. every one of those sentences was identified. The way checking query generation pattern comes to be performed is by checking the prevailing key terms in the sentence. key phrases located had been then matched towards a listing of gift styles. The gadget of classifying questions becomes finished via key-phrase identity and pattern identification of every question. each question that has been raised then recognized the volume of trouble primarily based completely on the key phrases and patterns that the question has. After the identification system is completed, the original questions have been categorized based mostly on the level of trouble.

## III. LITERATURE SURVEY

Automated Question Generation (AQG) is a rapidly growing field in Natural Language Processing (NLP) that aims to generate questions from a given text with the purpose of enhancing the understanding and retention of the information contained in it. AQG systems have a wide range of applications in education, e-learning, knowledge assessment, and other areas.

AQG can be approached in several ways, including rule-based methods, information retrieval methods, and machine learning methods. Rule-based methods generate questions based on predefined templates and grammar rules. Information retrieval methods, on the other hand, generate questions based on existing questions in a database. Machine learning methods, including deep learning techniques, generate questions by training statistical models on annotated data.

Several studies have explored the use of machine learning for AQG. In [1], a deep neural network was used to generate questions from a given text, achieving an accuracy of over 80% in question classification.

In [2], a reinforcement learning approach was used to optimize the trade-off between question difficulty and coverage of the text, showing promising results in terms of question quality and text coverage.

In [3], a multi-task learning framework was proposed to generate questions and answer choices simultaneously, resulting in more coherent and relevant questions.

Evaluation of AQG systems is a crucial aspect of the research. Common evaluation metrics include question quality, question relevance, and text coverage. In [4], a framework was proposed for the evaluation of AQG systems based on human annotations, demonstrating its effectiveness in evaluating the quality and coherence of generated questions. In [5], a user study was conducted to assess the impact of AQG on learning, showing a significant improvement in the retention of information compared to traditional methods.

Despite the progress made in AQG research, there are still several challenges that need to be addressed. These include generating questions with appropriate levels of difficulty, handling questions with multiple correct answers, and generating questions in different domains and languages. In [6], a survey was conducted to identify the current challenges and future directions in AQG research, providing valuable insights for future work in this field.

In conclusion, AQG using NLP is a rapidly growing field with a wide range of applications. Machine learning methods, including deep learning techniques, have shown promising results in AQG. Further research is needed to address the challenges and limitations of AQG systems, as well as to explore new applications and domains

### *A. NLTK*

Natural Language Processing is the manipulation of data, textual content, or speech through any software or device. An analogy is that people interact, understand each different attitude, and reply with the perfect solution. In NLP, this interplay, information, and reaction are made through a laptop rather than a human. NLTK stands for herbal Language Toolkit. This toolkit is one of the maximum outstanding NLP libraries which incorporate programs to make machines understand human language and respond to it with an appropriate response. It presents particular libraries for textual content processing, category, tokenization, stemming, tagging, labelling, parsing, and semantic reasoning. Tokenization in NLP is greater robust. It consists of breaking the given sentence into tokens and punctuations earlier than processing the facts. Tokenization is done as tested:

```
>>>textual content=" I used to be absent the day before this!"
>>>tokens=nltk.word_tokenize (text)
>>>tokens
['I', 'was', 'absent', 'yesterday', '!'][34]
```

### *B. POS Tagging*

POS tagging is step one in any NLP-based software. Tagging is a sort of elegance that can be defined due to the fact the automated challenge of description to the tokens. here the descriptor is called tag, which also can constitute one of the detail-of-speeches, semantic statistics, and so forth. Now, if we speak about aspect-of-Speech (POS) tagging, then it can be interpreted as a manner of assigning factors of speech to the given phrase. it's far generally called POS tagging. In smooth phrases, we are in a function to mention that POS tagging is an assignment of labelling every word in a sentence with its appropriate label of speech. We already remember that elements of speech embody nouns, verbs, adverbs, adjectives, pronouns, conjunction, and their sub-instructions. Taggers use several facts of the form: dictionaries, lexicons, policies, and so forth. most of the POS tagging falls underneath Rule Base, Stochastic, and Transformation-based tagging. A tag set is a group of tags used for a specific undertaking. each tagger is probably given a famous tag set. The tag set may be coarse which includes NN (Noun), VB (Verb), JJ (Adjective), RB (Adverb), IN (Preposition), CC (Conjunction), and so on.[34]

### *C. Spacy*

Whenever we are working with a huge amount of text, we will eventually want to know more about the text. Questions like:What does the words mean in the sentence, How do they act together to give a meaningful sentence? , Which texts are similar to each other and so on. Spacy is specifically built to process and help us understand large volumes of text. Spacy framework is written in Cython, and is a quite fast library. It provides access to its techniques and functions which are instructed by AI/machine learning models. In its package contains different models which contain the information about vocabularies, trained vectors, syntaxes an entities. It provides features for many natural language tasks, can be used to build information extraction systems. These models are to be loaded into our code to access them.Following is an example of loading the default package“english-core-web”:[35]

### *D. Word Net Lemmatizer*

Lemmatization companies collectively different types of words for reading them as an unmarried entity for you to pick out the dictionary root word, ideally called 'lemma'. Lemmatization is similar to stemming to a point. the important thing difference is that Lemmatization aims to remove the inflectional endings that might occur even as stemming. The output, after the procedure of lemmatization, has some context to it, and importantly, the phrase holds a means in contrast to stemming. WordNet is a massive, unfastened, and publicly available lexical database of the English language. it can be viewed as a thesaurus wherein comparable words are grouped into sets (synsets), each one personally expressing a distinct idea. the main aim is to broaden dependent semantic dating among words.

NLTK gives an interface to get entry to this dictionary- WordNet corpus reader. After downloading and setting up, an example of WordNetLemmatizer() is needed to lemmatize phrases, much like the stemming instance.

```
>>>lemmatizer = WordNetLemmatizer()
>>>print ("hassle: ", lemmatizer.lemmatize("trouble"))
>>>print ("rocks: ", lemmatizer.lemmatize("rocks"))
>>>print ("corpora: ", lemmatizer.lemmatize("corpora"))
```

Output:
problem: hassle
rocks: rock
corpora: corpus[10]

## IV. ANALYSIS AND RESULTS

The process of making a question paper was done manually which included steps like processing the paper and allocating the marks to the questions. This was a tedious task, too much time was consumed in this process.

### *A. Current System Drawbacks*

*1)* The current system is extremely time-consuming.
*2)* An excessive amount of time is consumed within the process of creating question papers on more subjects.
*3)* Numerous questions are evaluated before finalizing them for the question paper.
*4)* The probability of paper leakage is more in the current system as compared ed to the proposed system.
*5)* Processing of the paper takes longer because it is done manually.

### *B. Proposed System*

*1)* Large portion is covered by the system which helps to generate paper skilfully.
*2)* Generating question papers paper will be faster as it is automated.
*3)* Probability of paper leakage will drastically decrease be the cause admin will have all the control over the system.
*4)* This system is totally unbiased and generates random questions with a click of a button.

## V. EVALUATION

An evaluation process was needed to see the success rate of a developed method. The success rate of the developed method can be seen from the achievement of the accuracy value owned, the greater the value of the accuracy the better the method developed. The process of calculating accuracy was shown in Equation 1.

$$Accuracy = The\ amount\ of\ correct\ data\ /\ The\ amount\ of\ all\ data$$

Evaluation of Automated Question Generator using Natural Language Processing (NLP)

Automated question generation (AQG) systems aim to generate questions from a given text in order to help learners better understand the material. The performance of AQG systems can be evaluated in several ways, including accuracy, relevance, fluency, and diversity.

Accuracy refers to the degree to which the generated questions correctly reflect the content of the text. This can be evaluated using metrics such as precision, recall, and F1-score, which measure the ability of the AQG system to generate questions that are both correct and relevant to the text.

Relevance refers to the degree to which the generated questions are relevant to the content of the text and aligned with the intended learning objectives. Relevance can be evaluated by comparing the generated questions to a set of ground truth questions, or by assessing the quality of the questions based on criteria such as grammatical correctness, coherence, and topical relevance.

Fluency refers to the degree to which the generated questions are grammatically correct and natural sounding. This can be evaluated by comparing the fluency of the generated questions to a set of ground truth questions, or by subjective assessments from human evaluators.

Diversity refers to the degree to which the generated questions are diverse and cover a wide range of topics and perspectives. This can be evaluated by measuring the number and variety of questions generated by the AQG system, or by assessing the uniqueness of the generated questions based on criteria such as novelty and originality.

©IJRASET: All Rights are Reserved 

In addition to these metrics, it is also important to evaluate the usability and user experience of AQG systems. This can be done by conducting user studies to assess the effectiveness and ease of use of the AQG system, or by evaluating the user satisfaction and engagement with the generated questions.

In conclusion, evaluating AQG systems requires a multi-faceted approach that considers both the technical performance of the system and the user experience. A combination of objective metrics and subjective assessments can provide a comprehensive picture of the performance of AQG systems and help to identify areas for improvement.

## VI. CONCLUSION

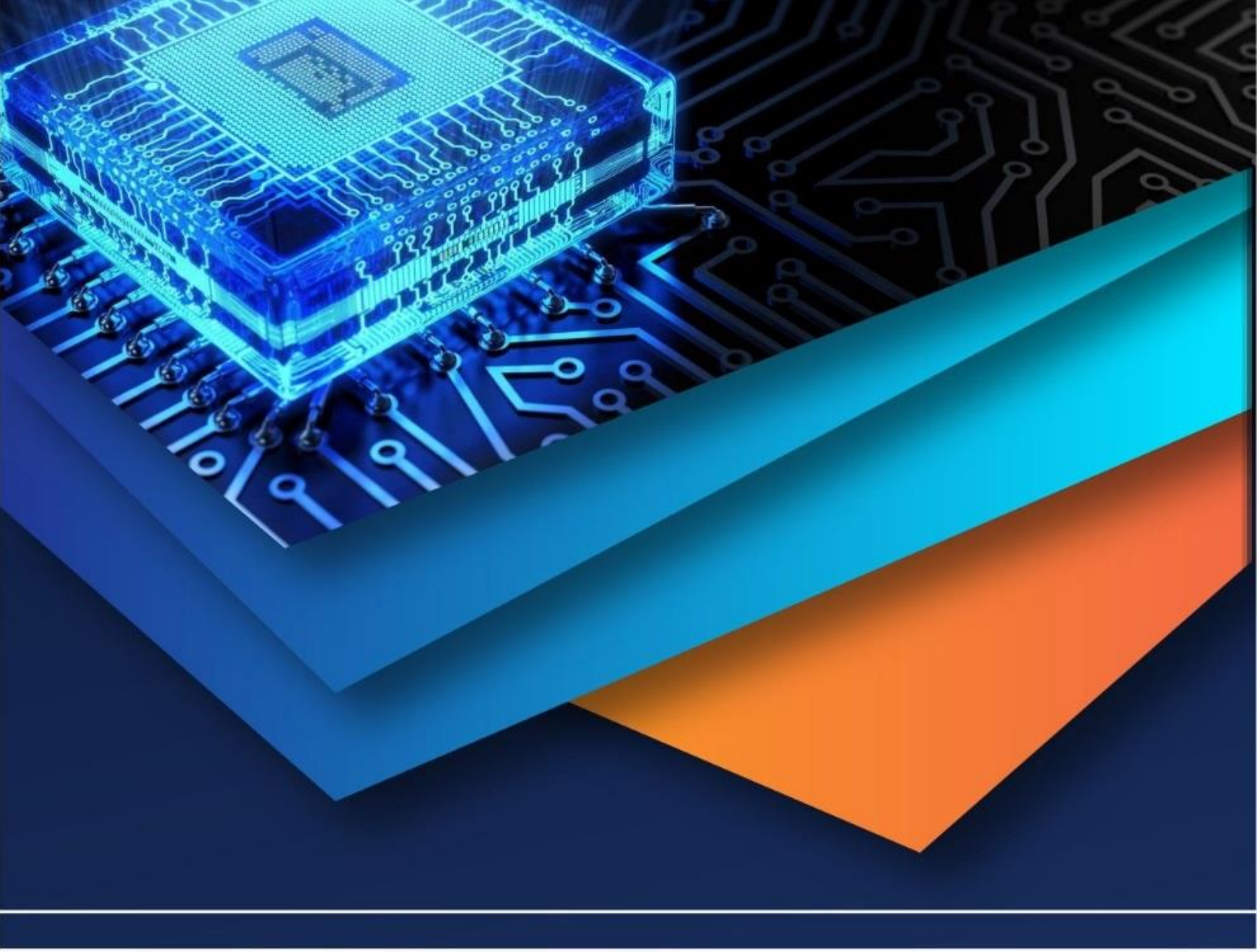
This studies the usage of a dataset of 60 samples of paragraphs derived from nine publications of study guides in Informatics Engineering. The 60 paragraphs produced 278 sentences and 654 questions with various tiers of difficulty in line with the degree of the problem within bloom's taxonomy. The templates used to automate query era routinely quantity to 64 units of Paragraph enter: Li-ion stands for Lithium-Ion, which can be typically used for the gadget. it is distinct from the kinds of battery predecessors, Li-ion battery doesn't have memory impact and can be recharged before it's far empty. some other things that can cause a lowering of the battery performance is overcharging. The result of the paragraph extraction process: 1. it's miles one of a kind from two forms of battery predecessor, Li-ion battery doesn't have reminiscence effect and can be recharged before it is empty. 3. every other issue that may trigger reducing of the battery's overall performance is overcharge templates. The complete query generated became then tested by a unique professional. Validation changed into performed to make sure the generated questions are comprehensible. A question is said to be valid if the query may be understood that means properly, so it may be replied to. The 534 questions were declared to be legitimate, and 120 questions were declared invalid, so the accuracy received for 81.65%. The accuracy of the outcomes received changed too much less than the maximum because of the shape of questions that had been not according to the structure of the query in widespread that become used so that there has been a discrepancy while placed into the template question. This look additionally suggests that this method generates extra questions at the level of remembering and expertise.

## REFERENCES

[1] Vijay Krishan Purohit', Abhijeet Kumar', Asma Jabeen' ,Saurabh Srivastava', R H Gouda' , Shinagawa, “Design of Adaptive Question Bank Development and Management System”, 2nd IEEE International Conference on Parallel, Distributed and Grid Computing, 2012.
[2] G-Asks: An Intelligent Automatic Question Generation System for Academic Writing Support by Ming Liu and Rafael A. Calv
[3] D. R. CH and S. K. Saha , "Automatic Multiple Choice Question Generation From Text: A Survey," in IEEE Transactions on Learning Technologies, vol. 13, no. 1, pp. 14-25, 1 Jan.-March 2020, Doi: 10.1109/TLT.2018.2889100.
[4] Anderson LW, Krathwohl DR. A taxonomy for learning, teaching, and assessing: a revision of Bloom’s taxonomy of educational objectives. New York NY: Longmans; 2001.
[5] Stephen A. Zahorian, Vishnu K. Lakdawala Oscar, R. Gonzalez, Scot Starsman, and James FLeathrum, Jr., "Question Model for Intelligent Questiong Systems in Engineering Education", 31st ASEE/IEEE Frontiers in Education Conferencee, October 10 - 13, 2001 Reno, NY, © 2001 IEEE.
[6] Noor Hasimah Ibrahim Teo, Nordin Abu Bakar and Moamed RezduanAbd Rashid, “Representing Examination Question Knowledge into Genetic Algorithm”, IEEE Global Engineering Education Conference (EDUCON), 2014.
[7] Onur KEKL –“Automatic Question Generation Using Natural Language Processing Techniques.”,July 2018. https://pdfs.semant icscholar.org /ec5e/fc7435 1f0339e34b91b965f99624aedf9200.pdf
[8] Aleena, Vidya – “Implementation of Automatic Question Paper Generator System”, International Research Journal of Engineering and Technology (IRJET), Feb 2019.
[9] Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Lawrence Zitnick, C., Parikh, D. (2015). Vqa: Visual question answering, In Proceedings of the IEEE International Conference on Computer Vision (pp. 2425–2433).
[10] Onur KEKL –“Automatic Question Generation Using Natural Language Processing Techniques.”, July 2018. https://pdfs.semanticscholar.org /ec5e/ fc743 51f0339e34b91b965f99624aedf9200.pdf
[11] Amruta Umardand, Ashwini – “A survey on Automatic Question Paper Generation System”, International Advanced Research Journal in Science, Engineering and Technology (IARJSET), Jan 2017.
[12] Aleena, Vidya - "Implementation of Automatic Question Paper Generator System", International Research Journal of Engineering and Technology (IRJET), Feb 2019.
[13] Kalpana B. Khandale1, Ajitkumar Pundage, C. Namrata Mahender - "Similarities In Words Using Different Pos Taggers.", 10SR Journal of Computer Engineering (IOSR- JCE),(PP 51-55).
[14] Edward Loper and Steven Bird "Nitk: The Natural Language Toolkit.", July 2002.
[15] Ankita, K. A. Abdul Nazeer - "Part-Of-Speech Tagging And Named Entity Recognition Using Improved Hidden Markov Model And Bloom Filter, International Conference on Computing, Power and Communication Technologies (GUCON), 2018.
[16] Surbhi Choudhary, Abdul Rais Abdul Waheed, Shrutika Gawandi and Kavita Joshi, “Question Paper Generator System,” International Journal of Computer Science Trends and Technology, vol. 3, issue 5, Sept–Oct 2015.
[17] Prita Patil and Kavita Shirsat, “An Integrated Automated Paperless Academic Module for Education Institutes,” International Journal of Engineering Science Invention Research and Development, vol. I, issue IX, March 2015.

[18] Ashok Immanuel and Tulasi.B, “Framework for Automatic Examination Paper Generation System,” International Journal of Computer Science Trends and Technology, vol. 6, issue 1, Jan - March 2015.
[19] Kapil Naik, Shreyas Sule, Shruti Jadhav and Surya Pandey, “Automatic Question Paper Generation using Randomization Algorithm,” International Journal of Engineering and Technical Research, vol. 2, issue 12, December 2014.
[20] Dan Liu, Jianmin Wang and Lijuan Zheng, “Automatic Test Paper Generation Based on Ant Colony Algorithm,” Journal of Software, vol. 8, no. 10, October 2013.
[21] D. R. CH and S. K. Saha, "Automatic Multiple Choice Question Generation From Text: A Survey," in IEEE Transactions on Learning Technologies, vol. 13, no. 1, pp. 14-25, 1 Jan.-March 2020, doi: 10.1109/TLT.2018.2889100.
[22] Narendra, A., Manish Agarwal and Rakshit shah, "Automatic Generation." RANLP, 2013. Cloze-Questions Agarwal, Manish & Shah, Rakshit & Mannem.
[23] Agarwal, Manish & Shah, Rakshit & Mannem, Prashanth, Automatic question generation using discourse cues, 2011, pp. 1-9.
[24] Kalpana B. Khandale, Ajitkumar Pundage C. Namrata Mahender, Similarities in words Using Different Pos Taggers, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, PP 51-55
[25] Deokate Harshada G., Jogdand Prasad P, Satpute Priyanka S., Shaikh Sameer B., Automatic Question Generation from Given Paragraph, USRD International Journal for Scientific Research & Development Vol. 7, Issue 03, 2019 | ISSN (online): 2321-0613
[26] Automatic Generation of Question Bank Based on Pre-defined Templates by Ahmed Ezz Awad and Mohamed Yehia Dahab in International Journal of Innovations & Advancement in Computer Science IJIACS ISSN 2347 – 8616 Volume 3, Issue 1 April 2014.
[27] Automatic question generation in multimedia-based learning by Yvonne SKALBAN , Le An HA , Lucia SPECIA , Ruslan MITKOV.
[28] G-Asks: An Intelligent Automatic Question Generation System for Academic Writing Support by Ming Liu and Rafael A. Calvo
[29] Mihir Joisher, Swapnil Ghagare, Mittal Patel, and Ritesh Rathi, "Automatic Question Paper Generation System" International Journal of Advanced Research in computer and communication Engineering (IJARCCE), vol.4 Dec 2015.
[30] Mrunal Patangare, Rushikesh Pangare, Shreyas Dorle, Uday Biradar, Kaustubh Kale, "Android Based Exam Paper Generator" Proceeding of the Second International Conference on Inventive Systems and Control (ICISC 2018).
[31] Noor Hasimah Ibrahim Teo, Nordin Abu Bakar and Moamed RezduanAbd Rashid, "Representing Examination Question Knowledge into Genetic Algorithm", IEEE Global Engineering Education Conference (EDUCON), 2014.
[32] Vijay Krishnan Purohit, Abhijeet Kumar, Asma Jabeen, Saurabh Srivastava, RH Goudar, Shiwangowda, "Design of Adaptive Question Bank Development and Management System", 2nd IEEE International Conference on Parallel, Distributed and Grid Computing, 2012.
[33] Suraj Kamya, Madhuri Sachdeva, Navdeep Dhaliwal and Sonit Singh, “Fuzzy Logic Based Intelligent Question Paper Generator” IEEE International Advance Computing Conference (IACC),2014.
[34] AUTOMATED QUESTION GENERATOR SYSTEM USING NLP LIBRARIES Priti Gumaste1, Shreya Joshi2, Srushtee Khadpekar3, Shubhangi Mali41-4BE Computer Engineering, Sandip Institute of Technology and Research Centre, Nasik, Maharashtra, India.
[35] Automatic Question Paper Generation, according to Bloom’s Taxonomy, by generating questions from text using Natural Language Processing Shivali Joshi*, Parin Shah*, Sahil Shah*

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)