AI Based Question Paper Generation

Author's Names: Dr. J. Jeya Bharathi ,P. Kirankumar, K. Jeevan Sai ,P. V. Sai Rohit ,K. Deva Chandan.

Department And Affiliation: Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Anand Nagar, Krishnankoil-626126, Tamil Nadu, India.

Author's Email-id : j.jeyabharathi@klu.ac.in , 99220041553@klu.ac.in , 99220041499@klu.ac.in , 99220041555@klu.ac.in , 99220041503@klu.ac.in .

Abstract
The project introduces an AI-based system for question paper generation to automate exam creation processes. The traditional school method requires extensive time and human effort while frequently failing to maintain balanced question difficulty levels. The system combines NLP capabilities with Machine Learning algorithms to produce and organize questions which automatically evaluate according to syllabus input and subject parameters and question types and marks and difficulty criteria. The system will preserve variety by applying semantic similarity approaches and will use MongoDB to store its data. The system provides a complete web interface that allows secure role-based access for faculty members to manage their tasks with ease. The system produces template-based question papers that users can save as PDF and Word files
Keywords: Machine Learning, NLP, MongoDb, NoSQL,PDF&WORD files, SDG goal:4 Quality Education

Introduction

These The conventional process of developing scholarly tests, including question papers, has been a mainstay of the education system for many years. This process is therefore usually known for being time-consuming and strenuous for instructors, with a heavy workload. Developing an excellent exam involves attentively considering several aspects, such as a clear fit with the syllabus, correct levels of difficulty, and an even distribution of marks between different subject matters. Manually meeting these requirements is not only time-consuming but prone to human error, creating inconsistencies in the level and overall quality of the test. The absence of a reliable and objective process for equalizing these factors can undermine the fairness and validity of the assessment process.
Over the past decade, the swift development of artificial intelligence (AI) has offered a revolutionary chance to mechanize and streamline intricate, knowledge-intensive processes. Education is especially well-suited to take advantage of these technologies. The research here suggests the creation of an AI-driven question paper generation system that is geared to overcome the very basic problems of conventional exam development. By incorporating advanced technologies like Natural Language Processing (NLP) and Machine Learning (ML), the system is intended to automate the whole process, from question generation to final paper compilation. The system to be proposed will scan the given syllabus and other important parameters, such as subject, question type, distribution of marks, and required difficulty criteria, to generate and categorize questions automatically. Ensuring the difficulty levels remain balanced across all the assessments. In addition, to avoid repetition and attain rich variety of questions, the system will utilize semantic similarity methods to

generate varied and different assessments. All information, including questions and system parameters, will be stored and handled efficiently using MongoDB, a scalable NoSQL database. To ensure greater usability, the system will have a full-stack web interface with role-based secure access, where faculty members will be able to handle the whole workflow with unprecedented ease and efficiency. This secure, easy-to-use space will allow educators to monitor the process while the AI performs the complex, repetitive work. Lastly, the system will generate question papers in a standardized, template-based output that is easily exportable as PDF or Word documents, making them easily applicable for academic use.

Literature Review

There has been an increase in research on education technology, with AI being noted to play a significant role in the automation of assessments. There have been a number of approaches explored: Question Generation with NLP: Experiments have indicated that it is possible to use NLP methods like Part-of-Speech (POS)tagging, Named Entity Recognition (NER), and syntactic parsing to produce informative questions from course content. The methods guarantee that produced questions are contextually appropriate (Liu et al., 2020). [2]Machine Learning for Classification: Decision Trees, Support Vector Machines (SVM), and Neural Networks are used to classify questions into levels of difficulty (easy, medium, hard). Balanced question distribution is facilitated over examinations (Zhang et al., 2021). [3]Semantic Similarity within Question Banks: For providing diversity while minimizing redundancy, semantic similarity methods like cosine similarity and Word2Vec embeddings have been utilized. These methods ensure similar questions are not repeatedly picked (Mitra & Baral, 2019).Databases for Educational Use: NoSQL databases such as MongoDB have been implemented in new education frameworks because of their scalability, schema adaptability ,and effective retrieval of unstructured information (Patel & Mehta, 2022).

Proposed Methodology

Creation and Storage of Question Bank

The The core of the system is to create an exhaustive and well organized question repository. Faculty members will be uploading questions into the repository, properly classified by subject, topic, question type (for example, MCQ, descriptive, true/false, or short/long answer), marks, and difficulty level. The information will be housed in MongoDB, which supports high scalability, flexibility, and optimal retrieval of big data.This way, faculty can easily update and manage questions while keeping the database organized and easily accessible.

Question Paper Generation Engine

The question paper generation engine is at the center of the system. Faculty members will provide input parameters like subject, syllabus coverage, distribution of marks, the total number of questions, and the desirable difficulty ratio (e.g.,40% easy, 40% medium, and 20% hard). With these inputs, the engine will generate appropriate questions from the database dynamically, with a balanced and well-structured paper. Besides, the engine offers template-based formatting, which offers professional layouts and designs that comply with standards and can be tailored according to faculty needs.

Question Classification and Difficulty Prediction

To enable machine learning to automatically judge question difficulty, models like Random Forest, Support Vector Machines (SVM), or Neural Networks will be trained on pre tagged datasets. The models will be trained to classify new questions into difficulty levels based on features that they extract. Some of these features are sentence length, linguistic complexity, richness of vocabulary, and cognitive categorization based on Bloom's Taxonomy. Automating difficulty prediction eliminates subjectivity and provides consistency in question classification.

Semantic Similarity for Diversity

One of the main issues with question paper generation is prevention from repetition and keeping the coverage broad for syllabus topics. In this regard, the system adapts NLP-based semantic similarity methods like TF-IDF vectorization, BERT embeddings, and cosine similarity measures. These measures assist in the detection and elimination of duplicate or highly similar questions during selection. As a result, the papers generated become more diverse, leading to increased fairness and comprehensiveness.

Web-Based Interface and Security

An easy-to-use, full-stack web application is the central interface for system interaction. The system provides role based access to various stakeholders. Administrators have access to system users, templates, and database settings, while faculty members can upload and manage questions and generate papers. A potential future add-on is a student module to view practice papers. Security is addressed through user authentication, stored data encryption, and secure access controls to ensure sensitive educational content is kept safe.

Export Options

Lastly, the system offers export options with flexibility for academic purposes. The produced question papers can be exported in PDF or Word formats for easy sharing digitally and printing hard copies, respectively. This feature integrates smoothly into academic workflows already in place while minimizing time and effort for educators. Finally the

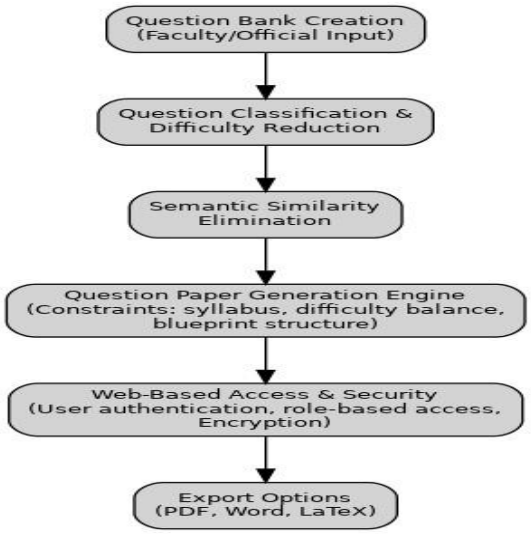

Figure1 Workflow of project

Results and Discussions:

While the project is in its development phase, the designed system is expected to provide a number of valuable advantages that can improve the process of creating examination papers in educational institutions. One of the most striking benefits is efficiency. Since the paper generation is automated, the system can save the time and effort of the faculty members by up to 70%. Historically, instructors take a few hours to choose, organize, and structure questions, but with this system, it is done virtually instantly after entering the parameters needed. This instantaneous aspect saves time for teachers and allows them to give greater attention to teaching and guiding students instead of paperwork.

Although these potential advantages are highly desirable, there are some challenges that need to be met to make the system robust. One such challenge is maintaining high accuracy in difficulty classification, especially for subjective or domain specific questions where linguistic features are not enough to define complexity. Another challenge is dealing with domain specific question types like numerical problems in mathematics or case studies in management, which need to be dealt with by special processing methods. In addition, the integrity of exam data must be ensured, as unauthorized access or data leakage would risk invalidating the integrity of tests. In order to overcome these risks, ongoing training and calibration of ML models, sophisticated NLP pipelines, and the implementation of cybersecurity features including encryption, authentication, and access control will be inevitable.

Figure 2: Interface of the project

Figure 3: Generated papers

Conclusion

This work presents a new AI-driven question paper generation system that harmoniously combines natural language processing (NLP), machine learning algorithms, semantic similarity metrics, and a secure full-stack web interface to streamline and improve the process of exam paper generation.Through the use of these cutting-edge technologies, the system efficiently minimizes the task load of the faculty members, who historically have devoted considerable time and energy to creating equitable and balanced exams, while at the same time maintaining fairness and objectivity by using automateddifficulty rating and diversity screening for question selection.The integration of MongoDB as the database framework allows the system to scale seamlessly, accommodating large question repositories across many subjects, without compromising on quick and reliable retrieval systems. In addition, integrating template-based formatting and export flexibility in common formats like PDF and Word increases usability, making itfeasible for institutions to implement the system within theexisting academic workflows without added complexity.

Acknowledgements: I would like to express my appreciation to Kalasalinagam Academy of research and Education for providing us with a place and resources to conduct our project entitled "AI-Based Question Paper Generation". A very special thank you to Dr. J. Jeya Bharathi who provided us with continuous support, direction and motivation throughout such a complex project. Without the backing of Department of Computer Science And Engineering, and all the instructors of the Department of Computer Science and Engineering, we would not have successfully concluded this project. Finally, I want to express my appreciation to every family member and friend for their many forms of assistance in helping me achieve the final objective of successfully completing this project.

References

[1] S. M. Metev and V. P. Veiko, Laser Assisted Microtechnology, 2nd ed., R.M.Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[2] J. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer,1989, vol. 61.

[3] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," IEEE Electron Device Lett., vol. 20, pp. 569-571, Nov.1999.

[4] M. Wegmuller, J. P. von der Weid, P. Oberson, and N.Gisin, "High resolution fiber distributed measurements with coherent OFDR," in Proc. ECOC'00, 2000, paper 11.3.4, p. 109.

[5] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "Highspeed digitalto-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.

[6] (2002) The IEEE website. [Online]. Available:
http://www.ieee.org/

[7] M. Shell. (2002) IEEEtran homepage on CTAN.[Online]. Available:http://www.ctan.org/texarchive/macros/latex/contrib/supported/IEEEtran/

[8] A. S. Singh, "AI-Powered Question Paper Generation With NLP," Int. J. Creative Research Thoughts, vol. 11, no. 12, pp. 1–6, 2023. [Online].
Available: https://ijcrt.org/papers/IJCRT2312530.pdf

[9] H. Nguyen et al., "Towards Automated Generation and Evaluation of Questions in Educational Domains," Proc. NAACL 2022, pp. 1–11, 2022. [Online].
Available: : https://www.cs.cmu.edu/~hn1/papers/NAACL2022_TowardsAutomated.pdf

[10] J. González, S. González, Y. Rodríguez, and K.Zuluaga, "Classification of mathematical test questions using machine learning," PLoS ONE, vol.18, no. 10, e0286989, 2023. [Online].
Available: https://journals.plos.org/plosone/article?id=
10.1371/journal.pone.0286989.