iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** X **Month of publication:** October 2026

DOI: https://doi.org/10.22214/ijraset.2026.85328

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 14 Issue X Oct 2026- Available at www.ijraset.com

Smart Q Gen: An AI-Powered Question Paper Generation System Using Hybrid Information Retrieval

Harees Jayadev T J[1], Aravinth M G[2], Venkata Lakshmi S[3]
Department of Computer Science and Engineering, K.L.N. College of Engineering, Tamil Nadu, India

Abstract: ***Preparing examination question papers is a repetitive and time-consuming task for faculty members. The conventional process requires faculty to review educational materials, identify important concepts, formulate questions, determine suitable difficulty levels, and arrange questions according to examination requirements. Smart Q Gen is an AI-assisted web-based question paper generation system designed to reduce this manual effort by generating questions from faculty-provided educational documents. The system accepts PDF, DOCX, and TXT files and performs text extraction, cleaning, structure detection, and text chunking. A hybrid information retrieval approach combines BM25 keyword retrieval and FAISS semantic retrieval using Sentence Transformer embeddings. Reciprocal Rank Fusion (RRF) is applied to combine the rankings produced by the two retrieval methods. The retrieved context is supplied to the FLAN-T5-Small model together with parameters such as question type, number of questions, difficulty level, and Bloom's Taxonomy level. Generated questions undergo validation and duplicate checking before being organized into a structured question paper. The system also provides faculty authentication, question preview, PDF export using ReportLab, and MongoDB Atlas-based storage. The proposed system integrates document processing, hybrid retrieval, AI-based question generation, validation, and paper generation into a single faculty-assisted workflow.***

Index Terms: Artificial Intelligence, Automated Question Generation, Retrieval-Augmented Generation, BM25, FAISS, Reciprocal Rank Fusion, FLAN-T5, Bloom's Taxonomy.

I. INTRODUCTION

Assessment is an important component of the educational process because examinations provide a structured method for evaluating student understanding. Question papers must therefore cover relevant concepts from the syllabus while maintaining suitable variation in question type, difficulty, and cognitive level.

In conventional academic environments, faculty members manually review lecture notes, textbooks, presentations, and other educational resources before preparing examination questions. They must identify important topics, formulate suitable questions, prepare alternatives or options when required, check for repetition, and finally organize the questions into the required examination format. Repeating this process for different subjects and examinations requires considerable time and effort.

The development of Natural Language Processing and transformer-based language models has created opportunities for automating parts of question generation. However, direct question generation without appropriate source context may produce questions that are weakly related to the actual learning material. Therefore, a question-generation system should retrieve relevant information from the source document before generating questions.

Retrieval-Augmented Generation [9] provides a suitable approach for this requirement. Instead of relying only on information stored within the language model, relevant passages can first be retrieved from the faculty-provided document and supplied as context to the generation model.

However, a single retrieval method may not perform equally well for every type of query. BM25 [5] is effective when important terminology appears directly in the document, whereas semantic retrieval can identify conceptually similar passages even when different words are used. Smart Q Gen therefore combines BM25 and FAISS [11] retrieval and uses Reciprocal Rank Fusion [12] to obtain a combined ranking.

The retrieved context is then passed to FLAN-T5-Small [4], along with the requested question type, number of questions, difficulty level, and Bloom's Taxonomy level. The generated questions are validated and checked for duplication before being included in the final question paper.

©IJRASET: All Rights are Reserved

The main contribution of Smart Q Gen is the integration of established document processing, hybrid retrieval, question generation, validation, and PDF-generation techniques into a single document-grounded workflow. The system is intended to assist faculty members rather than completely replace human academic review.

II. LITERATURE REVIEW / RELATED WORK

A. Automatic Question Generation

Automatic Question Generation (AQG) has been explored as a method of reducing the effort required to prepare educational questions [3]. Early approaches commonly relied on predefined templates and linguistic rules. These methods can provide predictable outputs but are limited by the rules and templates available. Mitkov and Ha [17] investigated computer-aided generation of multiple-choice tests, while Heilman and Smith [7] studied statistical ranking methods for question generation. These studies demonstrated the importance of generating and ranking candidate questions, although rule-based and statistical approaches may have limited flexibility for varied educational material.

B. Neural and Transformer-Based Question Generation

Neural question-generation methods introduced sequence-to-sequence learning for generating questions from source text. Du et al. [8] demonstrated neural question generation for reading comprehension, showing that neural models could generate questions with greater linguistic flexibility. Transformer architectures [1] subsequently improved natural-language generation. The T5 framework [2] formulated NLP tasks as text-to-text transformation, while instruction-tuned models [4] improved the ability of language models to follow task-specific instructions. Smart Q Gen uses FLAN-T5-Small as a lightweight generation model suitable for the project's resource constraints.

C. Retrieval-Augmented Generation

Retrieval-Augmented Generation [9] combines information retrieval with language generation. Relevant passages are retrieved from an external knowledge source and provided to the language model as context. This approach is particularly suitable for Smart Q Gen because the educational document uploaded by the faculty member should act as the primary source for question generation. Retrieving relevant portions before generation helps maintain a connection between the generated questions and the supplied material.

D. Keyword and Semantic Retrieval

BM25 [5] is a lexical information-retrieval method that considers term frequency, inverse document frequency, and document length. It is useful when important terminology in the query appears directly in the source document. Semantic retrieval represents text as vectors and searches for passages with similar meanings [6]. Smart Q Gen uses the all-MiniLM-L6-v2 Sentence Transformer model [10] to create embeddings and FAISS [11] for vector similarity search. The two approaches have complementary characteristics: BM25 is effective for exact terminology, while semantic retrieval is useful when the query and source text use different wording.

E. Difficulty-Controlled Question Generation

Tomikawa, Suzuki, and Uto [14] proposed an adaptive question-answer generation method using Item Response Theory and pretrained Transformer models. Their work focuses on generating questions with controlled difficulty and relating question difficulty to learner ability. Huang et al. [15] proposed EQGG: Automatic Question Group Generation, which focuses on question-group generation while considering aspects such as coverage, similarity, and diversity. Tomikawa and Uto [16] later investigated difficulty-controllable multiple-choice question generation using large language models and Direct Preference Optimization.

These studies provide important research foundations for automated educational question generation. However, Smart Q Gen does not implement IRT or DPO. Its current difficulty control is instruction-based and selected by the faculty member.

F. Research Gap

Existing research has investigated question generation, difficulty control, question-group generation, and language-model-based educational assessment. Smart Q Gen focuses on integrating these practical requirements into a document-based workflow: Document Processing → Hybrid Retrieval → RRF → FLAN-T5-Small → Validation → Duplicate Checking → Question Paper Generation. The system therefore focuses on practical faculty-assisted question-paper preparation from uploaded educational documents.

TABLE I
COMPARISON OF RELATED APPROACHES

Approach	Main Technique	Advantage	Limitation
Traditional QG [7], [17]	Rules / Templates	Simple and predictable	Limited flexibility
Neural QG [8]	Seq2Seq	Flexible generation	Context dependent
Transformer QG [1], [2], [4]	Pretrained models	Improved language generation	May require suitable context
RAG [9]	Retrieval + Generation	Better source grounding	Depends on retrieval
Smart Q Gen	BM25 + FAISS + RRF + FLAN-T5	Integrated document-grounded workflow	Requires faculty review

III. PROPOSED METHODOLOGY

The proposed Smart Q Gen methodology consists of multiple stages that transform an educational document into a structured question paper.

A. Document Upload

The system accepts educational materials in PDF, DOCX, and TXT formats. Faculty members upload the source material through the web interface.

B. Text Extraction

PDF documents are processed using PyMuPDF, while DOCX files are processed using python-docx. TXT files are processed using standard Python text-processing functionality. The extracted content is converted into a common textual representation for further processing.

C. Text Cleaning

Extracted documents may contain unnecessary spaces, formatting artifacts, broken line structures, and other unwanted content. The cleaning stage normalizes the extracted text before retrieval.

D. Structure Detection

The system identifies meaningful document structures such as headings, sections, and other content regions. This helps preserve the organization of the educational material.

E. Text Chunking

The processed document is divided into smaller chunks. Chunking allows the retrieval system to search focused portions of the document instead of treating the entire document as one large unit.

F. BM25 Retrieval

BM25 performs keyword-based retrieval over the generated document chunks. It identifies chunks containing important terms related to the selected topic or query.

G. FAISS Semantic Retrieval

The all-MiniLM-L6-v2 Sentence Transformer converts text into numerical vector representations. FAISS then searches these vectors to identify semantically similar document chunks. This allows the system to retrieve relevant content even when the exact query terms are not present in the document.

H. Reciprocal Rank Fusion

BM25 and FAISS produce separate ranked lists. Smart Q Gen combines these rankings using Reciprocal Rank Fusion [12]. The RRF score is represented as:

(1)

where *d* represents a document chunk, *R* represents the retrieval systems, *k* is a ranking constant, and $rank^r(d)$ represents the position of the document in retrieval ranking *r*. The resulting ranking is used to select the most relevant context for question generation.

I. Question Generation

The retrieved context is supplied to FLAN-T5-Small together with the question type, the number of questions, the difficulty level, and the Bloom's Taxonomy level [13]. The model generates questions based on the retrieved educational context and the selected parameters.

J. Question Validation

The generated questions are checked for completeness, relevance, and appropriate structure before being included in the final question paper.

K. Duplicate Checking

The system checks generated questions for repetition or similarity. This reduces unnecessary duplication in the final question paper.

L. Question Paper Assembly

Validated questions are arranged according to the required examination configuration.

M. PDF Generation

The final question paper is converted into a formatted PDF using ReportLab. Faculty members can preview the generated paper before exporting it.

IV. SYSTEM DESIGN AND ARCHITECTURE

Smart Q Gen follows a layered architecture consisting of the frontend, document-processing layer, retrieval layer, generation layer, validation layer, database, and output module. The overall workflow, from the faculty member to the previewed and exported question paper, is shown in Fig. 1.

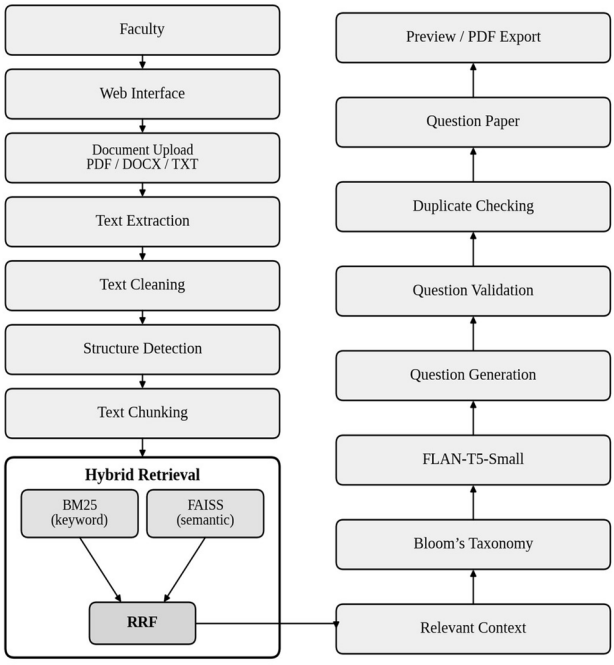


Fig. 1. Overall architecture and workflow of the proposed Smart Q Gen system.

A. Frontend Layer

The frontend is implemented using HTML, CSS, and JavaScript. It provides interfaces for faculty registration, login, document upload, question configuration, preview, and PDF export.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue X Oct 2026- Available at www.ijraset.com

B. Document Processing Layer

The document-processing layer extracts and cleans content from uploaded files. Structure detection and chunking prepare the content for retrieval.

C. Hybrid Retrieval Layer

The retrieval layer consists of two parallel retrieval mechanisms: BM25 for keyword ranking and FAISS for semantic ranking. The rankings are combined using RRF to obtain the final relevant chunks.

D. Generation Layer

The retrieved context is passed to FLAN-T5-Small with the selected generation parameters.

E. Validation Layer

Generated questions are checked and duplicate questions are removed or avoided before final assembly.

F. Database Layer

MongoDB Atlas stores faculty information, uploaded-document metadata, and generated question-paper records.

G. Output Layer

The generated question paper can be previewed and exported as a PDF using ReportLab.

V. IMPLEMENTATION AND EXPERIMENTAL SETUP

Smart Q Gen is implemented as a Python-based web application using Flask. The technologies used are summarized in Table II.

TABLE II
TECHNOLOGIES USED

Component	**Technology**	**Purpose**
Programming Language	Python	Backend and processing
Web Framework	Flask	Web application
Frontend	HTML, CSS, JavaScript	User interface
PDF Processing	PyMuPDF	PDF text extraction
DOCX Processing	python-docx	DOCX text extraction
Embeddings	Sentence Transformers	Text embeddings
Embedding Model	all-MiniLM-L6-v2	Semantic representation
Keyword Retrieval	BM25	Lexical retrieval
Vector Retrieval	FAISS	Semantic retrieval
Rank Fusion	RRF	Combine retrieval rankings
Generation	FLAN-T5-Small	Question generation
Database	MongoDB Atlas	Data storage
PDF Export	ReportLab	Question-paper generation

A. Experimental Evaluation

The system can be evaluated using three major aspects. First, retrieval effectiveness: BM25, FAISS, and Hybrid RRF can be compared using Precision@5 and Mean Reciprocal Rank (MRR). Second, question quality: generated questions can be evaluated for topic relevance, contextual grounding, question validity, option relevance, and Bloom's Taxonomy alignment, using a five-point rating scale for evaluator assessment. Third, processing efficiency: the processing time can be measured for document extraction, text cleaning, retrieval, question generation, validation, and PDF generation. The evaluation measures are defined in the project documentation; the graphs presented in Section VI use illustrative values and are to be replaced by measured values.

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

VI. RESULTS AND DISCUSSION

A. Retrieval Performance

The retrieval experiment compares BM25, FAISS, and Hybrid RRF. BM25 is expected to perform effectively for exact terminology, while FAISS can identify semantically related content. Hybrid RRF combines the rankings from both methods. The purpose of this experiment is to determine whether combining lexical and semantic retrieval provides more useful context than either method individually. Fig. 2 shows the layout for the Precision@5 and MRR comparison; the plotted values are illustrative and do not represent measured results.

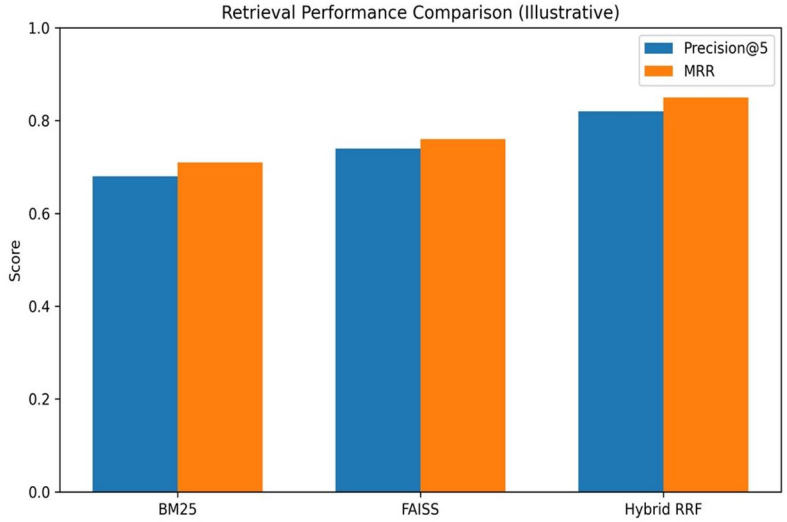


Fig. 2. Comparison of BM25, FAISS, and hybrid RRF retrieval performance (illustrative values).

B. Generated Question Quality

The generated questions should be evaluated for their educational usefulness rather than only their grammatical correctness. The evaluation considers topic relevance, contextual grounding, question validity, option relevance, and Bloom's Taxonomy alignment, and a reviewer can assign scores from 1 to 5 for each criterion (Fig. 3, illustrative values).

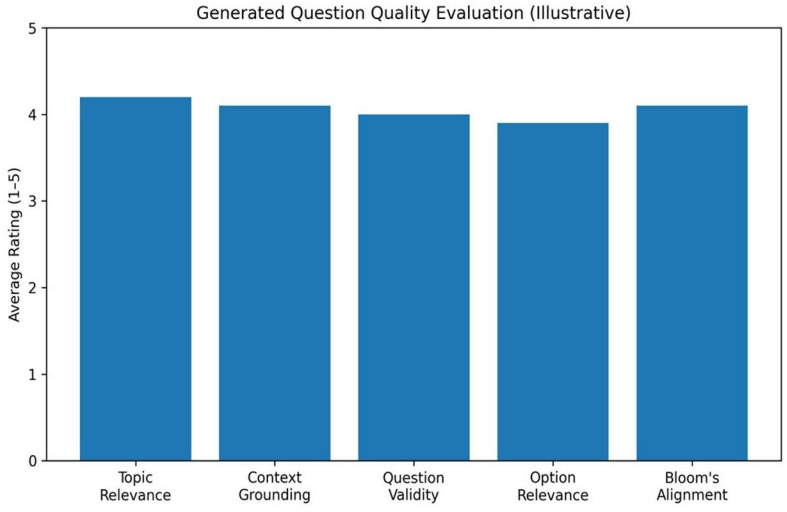


Fig. 3. Evaluation of generated question quality (illustrative values).

C. Processing Time

Processing-time evaluation identifies the computational requirements of each stage. The measurements are to be recorded from the application for extraction, cleaning, retrieval, generation, validation, and PDF generation (Fig. 4, illustrative values).

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

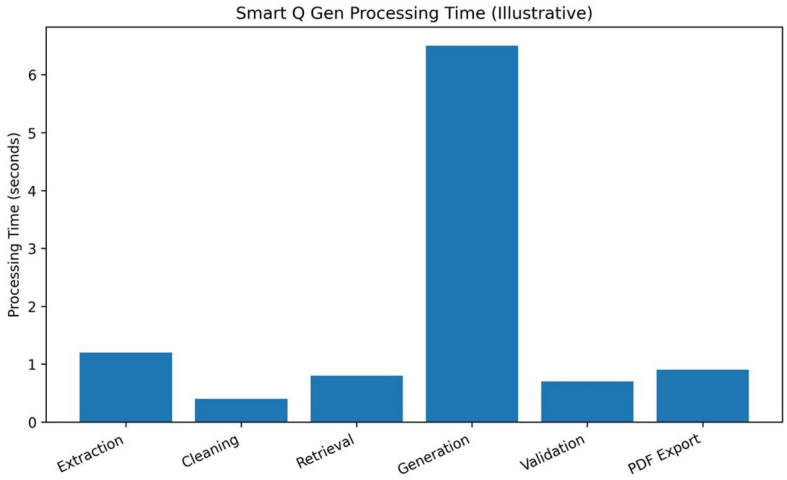


Fig. 4. Processing time of Smart Q Gen components (illustrative values).

D. Discussion

The major strength of Smart Q Gen is the integration of multiple processing stages into a single workflow. Document processing prepares the source material, hybrid retrieval identifies relevant context, FLAN-T5-Small generates questions, and validation and duplicate checking improve the final question set.

BM25 and FAISS address different retrieval requirements. BM25 is useful when important technical terminology occurs directly in the source material, while FAISS can identify semantically related passages with different wording. RRF provides a ranking-based method for combining these complementary retrieval results.

The retrieved context is particularly important because FLAN-T5-Small is a lightweight generation model. Providing relevant passages helps keep generated questions connected to the supplied educational material. Bloom's Taxonomy and the selected difficulty setting provide additional control over the intended characteristics of generated questions. The system is therefore designed as a faculty-assistance tool, with faculty members retaining responsibility for reviewing the final questions before examination use.

VII. LIMITATIONS

The current system has several limitations. First, the quality of generated questions depends on the clarity, completeness, and organization of the uploaded educational material. Poor-quality documents or extraction errors can affect subsequent retrieval and generation. Second, text chunking can influence retrieval quality because related information may be separated across different chunks. Third, FLAN-T5-Small is a lightweight model and may have limitations in complex reasoning, question diversity, and multiple-choice distractor quality compared with larger language models.

Fourth, difficulty is currently selected as an instruction-based generation parameter. It is not statistically calibrated using Item Response Theory. The current system also does not implement Direct Preference Optimization (DPO) and does not use project-specific fine-tuning. The generated questions therefore require faculty review before being used in an actual examination. Finally, the quantitative evaluation of retrieval performance, question quality, and processing time still needs to be completed using actual experimental measurements.

VIII. FUTURE WORK

Several improvements can be considered for future versions of Smart Q Gen.

1) *Multilingual question generation.* The system can be extended to support educational content and question generation in multiple languages.
2) *Larger language models.* More capable language models can be investigated to improve reasoning, fluency, question diversity, and distractor generation.
3) *Automated difficulty estimation.* Future versions can estimate question difficulty automatically instead of relying only on a faculty-selected instruction.

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

4) *Improved question diversity.* Generation strategies can be improved to reduce repetitive question structures and increase variation.
5) *More question formats.* Additional formats such as short-answer, descriptive, fill-in-the-blank, true/false, and other examination patterns can be incorporated.
6) *Question bank.* A dedicated question-bank module can store previously reviewed questions according to subject, unit, topic, difficulty, and Bloom's level.
7) *Faculty templates.* Faculty-specific examination templates can be introduced to automatically maintain institutional formatting requirements.
8) *Analytics.* Analytics can provide information about generated questions, topic coverage, difficulty distribution, and question usage.
9) *Improved retrieval.* Future versions can investigate improved chunking, reranking, and semantic retrieval techniques.
10) *Large-scale deployment.* The system can be optimized for larger document collections and multiple concurrent faculty users.

IX. CONCLUSION

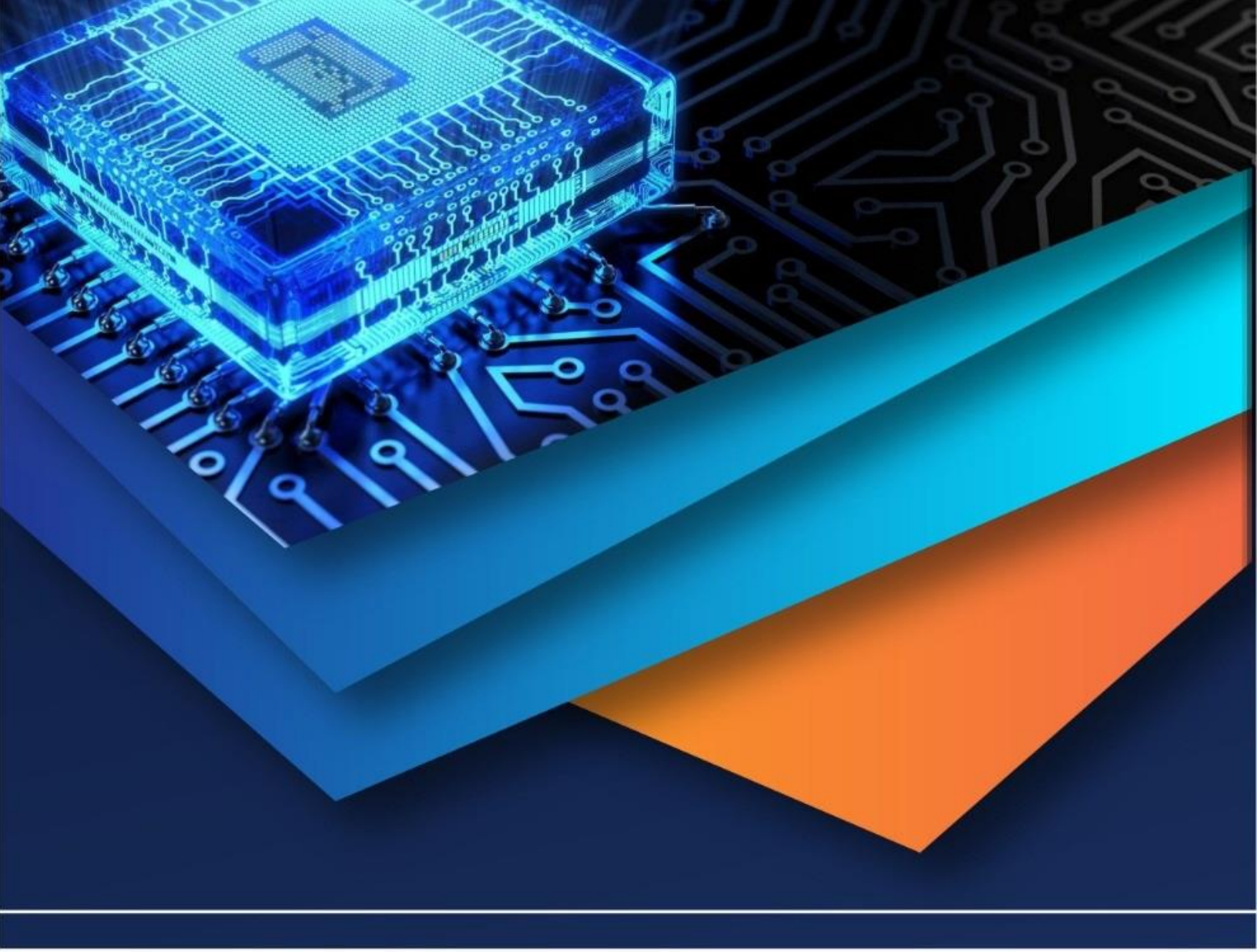
Smart Q Gen provides an AI-assisted approach for reducing the manual effort involved in examination question-paper preparation. The system accepts PDF, DOCX, and TXT educational documents and processes them through text extraction, cleaning, structure detection, and chunking. Relevant information is retrieved using a combination of BM25 keyword retrieval and FAISS semantic retrieval, while Reciprocal Rank Fusion combines the resulting rankings. The selected context is supplied to FLAN-T5-Small together with the required question type, number of questions, difficulty level, and Bloom's Taxonomy level.

Generated questions are validated and checked for duplication before being assembled into a structured question paper. The system provides question preview and PDF export using ReportLab, while MongoDB Atlas stores faculty information, document metadata, and generated-paper records. The primary contribution of Smart Q Gen is the practical integration of established document-processing, hybrid-retrieval, and language-generation techniques into a document-grounded faculty-assisted workflow. The system is designed to assist faculty members rather than completely replace human academic judgment. Completing the quantitative evaluation with measured retrieval, question-quality, and timing results is the immediate next step of this work.

REFERENCES

[1] A. Vaswani et al., "Attention Is All You Need," Advances in Neural Information Processing Systems, 2017.
[2] C. Raffel et al., "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," Journal of Machine Learning Research, vol. 21, no. 140, pp. 1–67, 2020.
[3] G. Kurdi, J. Leo, B. Parsia, U. Sattler, and S. Al-Emari, "A systematic review of automatic question generation for educational purposes," International Journal of Artificial Intelligence in Education, vol. 30, pp. 121–204, 2020.
[4] H. W. Chung et al., "Scaling Instruction-Finetuned Language Models," arXiv preprint arXiv:2210.11416, 2022.
[5] S. Robertson and H. Zaragoza, "The Probabilistic Relevance Framework: BM25 and Beyond," Foundations and Trends in Information Retrieval, vol. 3, no. 4, pp. 333–389, 2009.
[6] V. Karpukhin et al., "Dense Passage Retrieval for Open-Domain Question Answering," in Proc. EMNLP, 2020.
[7] M. Heilman and N. A. Smith, "Good Question! Statistical Ranking for Question Generation," in Proc. NAACL-HLT, 2010.
[8] X. Du, J. Shao, and C. Cardie, "Learning to Ask: Neural Question Generation for Reading Comprehension," in Proc. ACL, 2017.
[9] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in Advances in Neural Information Processing Systems, vol. 33, 2020.
[10] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks," in Proc. EMNLP-IJCNLP, 2019.
[11] J. Johnson, M. Douze, and H. Jégou, "Billion-Scale Similarity Search with GPUs," IEEE Transactions on Big Data, vol. 7, no. 3, 2021.
[12] G. V. Cormack, C. L. A. Clarke, and S. Buettcher, "Reciprocal Rank Fusion Outperforms Condorcet and Individual Rank Learning Methods," in Proc. SIGIR, 2009.
[13] B. S. Bloom et al., Taxonomy of Educational Objectives: The Classification of Educational Goals, 1956.
[14] Y. Tomikawa, A. Suzuki, and M. Uto, "Adaptive Question–Answer Generation With Difficulty Control Using Item Response Theory and Pretrained Transformer Models," IEEE Transactions on Learning Technologies, vol. 17, pp. 2186–2198, 2024.
[15] P.-C. Huang, Y.-H. Chan, C.-Y. Yang, H.-Y. Chen, and Y.-C. Fan, "EQGG: Automatic Question Group Generation," IEEE Transactions on Learning Technologies, vol. 17, pp. 1994–2007, 2024.
[16] Y. Tomikawa and M. Uto, "Difficulty-Controllable Multiple-Choice Question Generation Using Large Language Models and Direct Preference Optimization," IEEE Access, vol. 14, pp. 44698–44706, 2026.
[17] R. Mitkov and L. A. Ha, "Computer-Aided Generation of Multiple-Choice Tests," in Proc. HLT-NAACL Workshop on Building Educational Applications Using NLP, 2003, pp. 17–22.

©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)