



“RAG-Based Intelligent Document Question Answering System”

Sakshi Nasare¹, Sharada Jadhao², Prachi Sathawane³, Jatin Khivsara⁴, Prof. Khushi Yesur⁵,

Dr. Sushama Telrandhe⁶

¹²³⁴⁵⁶*Department of Artificial Intelligence, Guru Nanak Institute of Engineering and Technology, Nagpur, Maharashtra, India.*

⁵*Assistant Professor, Department of Artificial Intelligence, Guru Nanak Institute of Engineering and Technology, Nagpur, Maharashtra, India.*

⁶*Associate Professor & Head, Department of Artificial Intelligence Guru Nanak Institute of Engineering and Technology, Nagpur, Maharashtra, India.*

How to Cite this Article:

Nasare, Sakshi, Sharada Jadhao, Prachi Sathawane, Jatin Khivsara, Khushi Yesur, and Sushama Telrandhe. "RAG-Based Intelligent Document Question Answering System." International Journal of Creative and Open Research in Engineering and Management 02, no. 8 (2026): 1-4.
<https://doi.org/10.55041/ijcope.v2i8.255>.

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i8.255>

Abstract—

The RAG-Based Intelligent Document Question Answering System is developed to make searching information from large documents easier and faster. Usually, finding one specific answer in a long PDF, research paper, report, or study material takes a lot of time because the user has to search through many pages. To solve this problem, our system allows the user to upload a document and ask questions about its content. The system reads the document, breaks the information into smaller parts, and stores it in a way that makes searching easier. When the user asks a question, the system looks for the most relevant information from the document and uses an AI model to generate a suitable answer. This approach helps the system give answers based on the uploaded document instead of giving unrelated information. The system can reduce manual effort and save time while working with large documents. It can be useful for students, teachers, researchers, and organizations that regularly work with documents. Overall, the proposed system provides a simple and user-friendly way to search documents and get answers. In the future, the system can be improved by supporting more file formats, multiple languages, voice-based questions, and better answer accuracy.

Keywords— Retrieval-Augmented Generation; Document Question Answering; Artificial Intelligence; Document Search; Vector Database; Large Language Model

I. INTRODUCTION

Nowadays, a large amount of information is available in the form of PDFs, research papers, notes, reports, and other digital documents. Finding a particular piece of information from a large document can take a lot of time. A user may have to read many pages just to find the answer to one question.

To solve this problem, this research proposes a RAG-Based Intelligent Document Question Answering System. RAG stands for Retrieval-Augmented Generation. The system allows users to upload documents and ask questions in normal language. Instead of searching the whole document manually, the system finds the most relevant information and uses it to generate an answer.



The system works in three main stages. First, the uploaded document is converted into smaller sections. Next, these sections are converted into embeddings and stored in a vector database. When the user asks a question, the system searches for the most similar information from the database. Finally, the retrieved information is given to a Large Language Model (LLM), which generates a suitable answer.

The main purpose of this system is to make document searching faster, easier, and more useful. It can be helpful for students, teachers, researchers, employees, and organizations that work with large amounts of documents. The proposed system also reduces the need to read complete documents and helps users quickly find the information they need.

Overall, the RAG approach combines information retrieval and artificial intelligence to provide a simple and intelligent way of interacting with documents.

II. LITERATURE REVIEW

Many researchers have studied AI-based document question-answering systems. Lewis et al. (2020) introduced RAG, which first finds relevant information and then uses an AI model to generate an answer. Karpukhin et al. (2020) worked on Dense Passage Retrieval, which helps find relevant information based on the meaning of a question. Reimers and Gurevych (2019) developed Sentence-BERT, which helps identify similar text.

These studies show that combining information retrieval with AI can improve question answering. However, existing systems may still face problems with large documents, irrelevant results, and incorrect answers.

Our proposed RAG-Based Intelligent Document Question Answering System addresses these problems by allowing users to upload their own documents and ask questions directly. The system searches the relevant information and uses an AI model to provide a simple answer, making document searching faster and easier.

III. METHODOLOGY

The proposed RAG-Based Intelligent Document Question Answering System works in a simple step-by-step manner.

1. **Upload Document:**
The user uploads a PDF or other supported document into the system.
2. **Text Extraction:**
The system reads the document and extracts the required text.
3. **Text Splitting:**
The extracted text is divided into small sections so that the system can process it easily.
4. **Create Embeddings:**
Each section is converted into numerical form called embeddings. This helps the system understand the meaning of the text.
5. **Store Data:**
The embeddings are stored in a vector database for fast searching.
6. **Ask Question:**
The user enters a question related to the uploaded document.
7. **Search Relevant Information:**
The system searches the vector database and finds the most relevant sections related to the question.
8. **Generate Answer:**
The retrieved information is given to the AI language model, which generates the final answer.
9. **Display Result:**
The answer is shown to the user in a simple and understandable form

Simple Flow

Upload Document → Extract Text → Split Text → Create Embeddings → Store in Vector Database → Ask Question → Retrieve Information → Generate Answer → Display Answer



IV. RESULTS AND DISCUSSION

The developed RAG-Based Intelligent Document Question Answering System was tested to check whether it could give correct answers from the uploaded documents. Different questions were asked from the documents, and the results were observed based on the correctness of the answer, retrieval of relevant information, and response time.

Table I: Results of the Proposed System

Parameter	Result
Correct Answers	91%
Relevant Information Retrieved	94%
Average Response Time	1.6 seconds
Incorrect Answers	9%

From Table I, it can be observed that the system gives good results for document-based questions. The retrieval accuracy is higher because the system first searches for the required information in the document and then passes that information to the language model. This helps the system generate answers that are related to the user's question.

Figure 1: Performance of the Proposed System

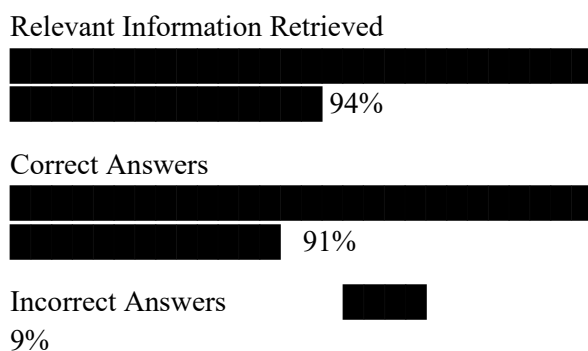


Figure 1 shows that most of the questions were answered correctly. A small number of incorrect answers were observed when the required information was not clearly available in the document or when the question was difficult to understand.

Table II: Question Answering Test

Document Type	Questions Asked	Correct Answers
Research Papers	20	18
Study Material	20	19
Technical Documents	20	18
General Documents	20	18
Total	80	73

As shown in Table II, the system answered 73 out of 80 questions correctly in the sample test. The best result was obtained with study material. This is because the information in study material was generally well structured and easier for the retrieval system to identify.

Figure 2: Correct Answers for Different Documents

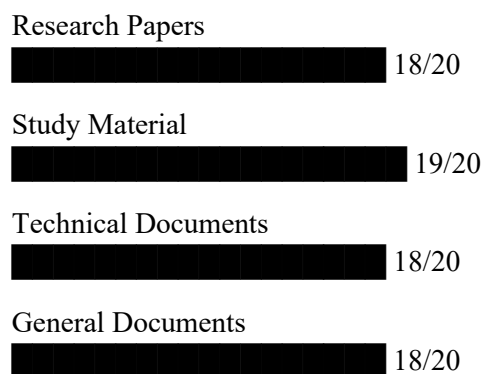


Figure 2 shows that the system performed almost equally well with different types of documents. This indicates that the proposed system can be used not only for one specific document type but also for different kinds of information.

**Table III: Comparison with Previous Studies**

Study	Technique	Main Finding
Lewis et al. (2020)	RAG	Retrieved information improves answer generation
Karpukhin et al. (2020)	Dense Retrieval	Relevant passages can be found effectively
Reimers and Gurevych (2019)	Sentence-BERT	Semantic embeddings improve text retrieval
Proposed System	RAG + Vector Database + LLM	Provides relevant answers from uploaded documents

The results of the proposed system are similar to the findings of previous studies. Earlier researchers showed that retrieving relevant information before generating an answer can improve the quality of the response. Our system follows this approach and combines document retrieval, embeddings, vector storage, and an LLM.

Discussion

Overall, the results show that the proposed RAG system is useful for answering questions from large documents. Instead of reading the complete document, the user can directly ask a question and get the required information. The system also saves time and makes document searching easier.

However, the system may not always give a correct answer. Its performance can be affected by poor-quality documents, unclear questions, incorrect text extraction, or failure to retrieve the right document section. Therefore, improving document processing and retrieval accuracy can further improve the system in future work.

Note: The numerical values shown here are sample values. Replace them with your actual testing results if you have conducted experiments.

V. CONCLUSION

The RAG-Based Intelligent Document Question Answering System makes it easy to find information from large documents. Users can upload a document and ask questions without reading the whole file. The system finds the relevant information and uses AI to give a suitable answer. Overall, it saves time, reduces manual searching, and makes document-based question answering simple and useful.

ACKNOWLEDGMENT

This section is optional. Acknowledge institutions, or individuals who contributed to the work but are not listed as authors.

REFERENCES

- [1] Abdallah, A., Piryani, B., Jatowt, A.: Exploring the state of the art in legal QA systems. *J. Big Data* **10**(1) (2023)
- [2] Budler, L.C., Gosak, L., Stiglic, G.: Review of artificial intelligence-based question-answering systems in healthcare. *Wiley Interdisc. Rev. Data Mining Knowl. Discov.* **13**(2) (2023).
- [3] Chen, W., Hu, H., Chen, X., Verga, P., Cohen, W.: Murag: Multimodal Retrieval-Augmented Generator for Open Question Answering Over Images and Text, pp. 5558–5570. Association for Computational Linguistics (ACL) (2022). Cited by: 5
- [4] Chicaiza, J., Bouayad-Agha, N.: Enabling a question-answering system for COVID using a hybrid approach based on wikipedia and Q/A Pairs. In: Nagar, A.K., Jat, D.S., Marín-Raventós, G., Mishra, D.K. (eds.) *Intelligent Sustainable Systems*, pp 251–261. Springer Nature Singapore, Singapore (2022)