International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 04 April-2026 | Impact Factor: 3.5

Query - Based AI Notes Assistant

Mrs. B. Sreelatha[1], J. Emmanuel Abishai[2], G. Varun Tej[3], P. Moulika Reddy[4], A. Karthik[5]

[1] Assistant Professor, Department of CSE (Data Science), ACE Engineering College, Hyderabad, Telangana, India

[2345] III B.Tech. Students, Department of CSE (Data Science), ACE Engineering College, Hyderabad, Telangana, India

How to Cite this Article:
Abishai, J. E., Tej, G. V., Reddy, P. M. & Karthik, A. (2026). Query - Based AI Notes Assistant. International Journal of Creative and Open Research in Engineering and Management, 02(04). https://doi.org/10.55041/ijcope.v2i4.241

https://doi.org/ 10.55041/ijcope.v2i4.241

ABSTRACT

The rapid growth of information in the modern era has made it challenging for students and researchers to effectively manage and understand large volumes of data. Traditional note-taking methods are often time-consuming, unstructured, and prone to errors, leading to reduced productivity and difficulty in grasping complex concepts. To address these challenges, the proposed system, *Query-Based AI Notes Assistant*, introduces an intelligent and structured approach to note management.

This system is designed as an interactive web application that leverages advanced Natural Language Processing (NLP) and Retrieval-Augmented Generation (RAG) techniques to generate concise, accurate, and contextually relevant summaries. Beyond simple summarization, it organizes notes by identifying and linking related concepts, enabling users to develop a deeper understanding of the subject matter. The system ensures factual accuracy while maintaining domain relevance, thereby improving the quality of generated content.

By streamlining the processes of summarization and organization, the application reduces cognitive load and enhances learning efficiency. Overall, it serves as a powerful academic support tool that promotes structured learning, better knowledge retention, and more effective interaction with information.

I. INTRODUCTION

The ability to efficiently manage and interpret large volumes of information has become essential for students and researchers. Since, conventional note-taking methods are often time-consuming, The *Query-Based AI Notes Assistant's* aim is to prove an intelligent, user-friendly platform for generating and managing notes. This system allows users to input queries and receive well-structured, concise summaries enriched with relevant contextual information. By integrating modern technologies such as Natural Language Processing and Retrieval-Augmented Generation, the application not only simplifies note-taking but also improves conceptual understanding. The project aims to create a smarter, more efficient learning environment that supports structured knowledge acquisition and enhances overall productivity.

ways like improving system reliability and making sure everything runs smoothly. Overall this tool can simplify the study flow of Students , Researchers.

II. RELATED WORK

Previous research and existing systems in the field of note-taking and summarization primarily include traditional digital tools and modern AI-powered applications. Conventional systems such as basic note-taking apps focus mainly on storing and organizing text but lack intelligent processing and conceptual understanding. With the advancement of artificial intelligence, features such as automatic transcription, summarization, and smart organization of content are introduced. These systems use Natural Language Processing (NLP) to extract key information, generate concise summaries, and categorize notes based on topics or keywords. However, many existing solutions still have limitations, such as reduced contextual understanding, lack of strong conceptual linking, and the need for user intervention to refine outputs. Research systems like AI-assisted note-taking frameworks further attempt to improve organization and efficiency, but challenges remain in ensuring accuracy and meaningful connections between concepts. Therefore, these limitations highlight the need for more advanced systems that can provide deeper contextual insights and structured knowledge representation.

Existing System and its Limitations:

Title	Technology Used	Limitations	Authors
A DL Framework for Extracting and Summarizing Text from Images	Edge Detection Algorithm (Sobel Operator)	Small noise variations can lead to incorrect summaries	A. ElDor, O. Abdul
Verification-Guided Reinforcement Learning for LLMs	Reinforcement Learning from Human Feedback (RLHF)	Expensive and time-consuming due to the need for large-scale human feedback	X. Fang, J. Kang, R. Rocha, S. Ainsworth
Layout-aware Dynamic RAG for Visually-Rich Documents	LayoutLM	Labeled training data has higher model complexity, leading to slower inference	Z. Sourati, Z. Wang, M. Liu, Y. Hu
Fine-Tuning LLaMA2 for Summarizing Discharge Notes	U-LLaMA, other LLMs	May generate incorrect or fabricated information.	M. Dekhordi , Y. Perl, F. Deek, H. Liu
The impact of Concept Mapping on summarizin	TF-IDF, TextRank Algorithm	Cannot capture sentence structure, Poor similarity metric affects ranking	I. Akbarian

III. METHODOLOGY

To get around the problems with other AI assistants, this project combines NLP methods with RAG techniques. It uses a kind of architecture that is very flexible. This approach is divided into four parts to make sure it works fast is easy to understand and keeps data safe.

3.1 Data Input and Query Processing

The system begins with user interaction, where the user provides input in the form of a query or topic through the web interface. This input is preprocessed using Natural Language Processing (NLP) techniques such as tokenization, stop-word removal, and text normalization to ensure that the query is clean and meaningful. The processed query is then analyzed to understand user intent and extract key terms. This step forms the foundation for generating accurate and relevant responses, as it ensures that the system interprets the user's requirements correctly before proceeding further.

3.2 Information Retrieval using RAG

Once the query is processed, the system employs Retrieval-Augmented Generation (RAG) techniques to fetch relevant information from a pre-built knowledge base or external data sources. A vector database is used to store embeddings of documents, allowing efficient similarity search based on the query. The most relevant content is retrieved and passed to the language model. This approach improves factual accuracy and ensures that the generated output is grounded in reliable information rather than purely generative responses.

3.3 Summarization and Presentation

After retrieving relevant data, the system uses advanced NLP models to generate concise summaries. The summarization process focuses on extracting key ideas while maintaining contextual accuracy. Finally, the generated notes are presented to the user through an interactive and user-friendly interface.

IV. MODEL EVALUATION

The Query-Based AI Notes Assistant is tested to ensure that the generated summaries are accurate, relevant, and contextually meaningful. The evaluation process includes both qualitative and quantitative measures. Quantitatively, metrics such as precision, recall, and F1-score are used to assess how well the system retrieves and presents relevant information. For summarization quality the preservation of the key ideas are ensured. Additionally, the performance of the Retrieval-Augmented Generation (RAG) component is evaluated based on retrieval accuracy and response consistency. These metrics help in identifying gaps in information retrieval and generation, allowing continuous improvement of the model.

Alongside automated evaluation, qualitative assessment is conducted through user feedback and manual review. Users evaluate the clarity, coherence, and usefulness of the generated notes, which provides insights into real-world performance. Factors such as readability, logical flow, and the effectiveness of concept linking are carefully analyzed. The system is also tested across diverse topics to ensure robustness and adaptability. By combining both automated metrics and human evaluation, the project ensures a balanced and comprehensive assessment of the model's performance, ultimately enhancing reliability and user satisfaction.

Performance Aspect	Description	Performance
Accuracy	Measures correctness of generated summaries and retrieved information	High – Relevant and factually correct outputs
Precision	Degree to which retrieved information is relevant to the query	Minimal irrelevant content
Recall	Ability to retrieve all important information related to the query	Moderate
Summarization Quality	Preservation of key ideas and reduction of unnecessary details	High – Concise and meaningful summaries
Contextual Understanding	Ability to connect related topics and ideas	Moderate – Accurate interpretation of queries

Table: System Performance Metrics for varying aspects

V. RESULT

5.1 System Interface Screens

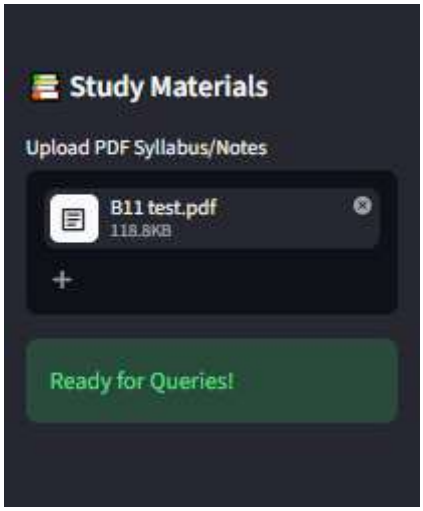

Figure 1: The image shows the sidebar of the UI where the documents are uploaded

Figure 2: The image shows the main dashboard screen, which shows the multiple choices of the way you want to take notes (ie. 5 marks or more), along with RAG-Chat and Flashcards etc. (WIP)

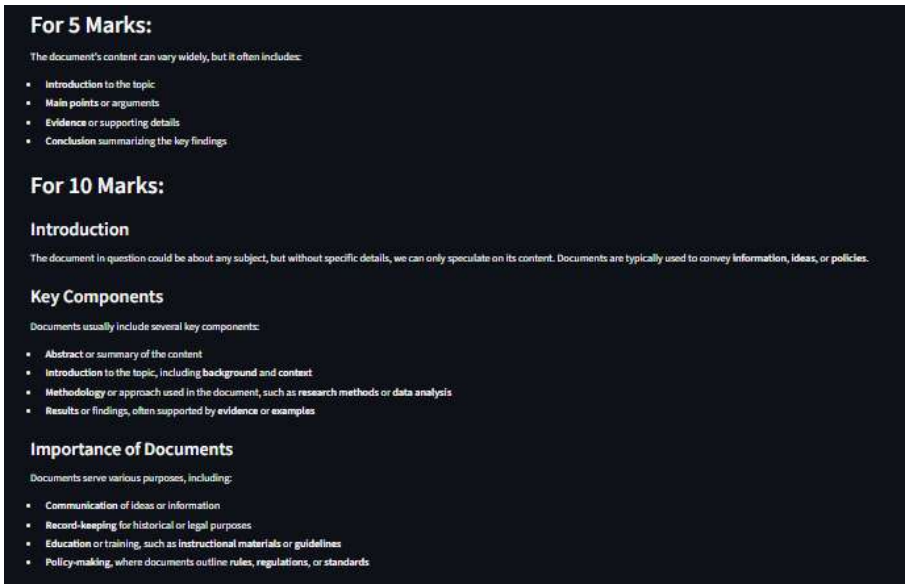

Figure 3: This screen demonstrates how detailed the Assistant can generate the notes based on the user's preference

Figure 4: The "Flashcard" mode where the Assistant generates multiple short QA to quickly revise the topic's keywords to the user

VI. CONCLUSION AND FUTURE SCOPE

The Query-Based AI Notes Assistant successfully demonstrates how artificial intelligence can enhance the process of note-taking, summarization, and knowledge organization. By integrating NLP and RAG techniques, the system provides accurate, relevant information while also establishing meaningful connections between concepts. This not only reduces the time and effort required for managing large volumes of data but also improves learning efficiency and knowledge retention. The structured workflow and user-friendly interface further contribute to a more organized and productive academic experience, making the system a valuable tool for students and researchers.

In terms of future scope, the project can be expanded by incorporating advanced features such as RAG-Chat, Diagram usagw and voice-based query input to improve accessibility and usability. Additionally, implementing personalized learning recommendations and adaptive summarization based on user preferences can further improve user experience. With continuous advancements in AI, the system has the potential to evolve into a comprehensive intelligent learning assistant capable of supporting a wide range of educational and research activities.

ACKNOWLEDGEMENTS

A sincere thanks to our guide and respected faculty for their invaluable support and guidance. We appreciate our institution for providing the resources and the environment to show our potential and grow. Special thanks to our team members for their dedication and teamwork.

VII. REFERENCES

[1] Sourati, Z., Wang, Z., Liu, M. M., Hu, Y., Guo, M., Bharadwaj, S., Han, K., Sheng, T., Ravi, S., Dehghani, M., & Roth, D. (2025). LAD-RAG: Layout-Aware Dynamic RAG for Visually-Rich Document Understanding. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2510.07233

[2] Jiang, W. (2025). RAG-Stack: Co-Optimizing RAG quality and performance from the vector database perspective. *ArXiv.org*. https://doi.org/10.48550/arxiv.2510.20296

[3] Albaroudi, E., Hatamleh, M., Hejazi, S. M., Alshalabi, A. Y., Mansouri, T., & Alameer, A. (2026). COLLAB-LLM: a Communication-Centric Role-Based framework for scalable Multi-Agent LLM collaboration. *Asian Journal of Research in Computer Science*, *19*(1), 152–185. https://doi.org/10.9734/ajrcos/2026/v19i1811

[4] Dehkordi, M. K. H., Perl, Y., Deek, F. P., & Liu, H. (2025). Fine-Tuning LLAMA2 for summarizing discharge notes: Evaluating the role of highlighted information. *Big Data and Cognitive Computing*, *10*(1), 4. https://doi.org/10.3390/bdcc10010004