Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 04, April-2025

DOCEASE: A Next-Gen RAG-Based System for Inclusive and Accessible Document Interaction

Akanksha Palve
Dept. of AI & DS
(of Affiliation)
PES' Modern college of Engineering
SPPU
Pune, India

Kanchan Tayade
Dept. of AI & DS
(of Affiliation)
PES' Modern college of Engineering
SPPU
Pune, India

Akshata Kamerkar
Dept. of AI & DS
(of Affiliation)
PES' Modern college of Engineering
SPPU
Pune, India

Shubham Murtadak
Dept. of AI & DS
(of Affiliation)
PES' Modern college of Engineering
SPPU
Pune, India

Prof. Mrs. Priti Malkhede
Dept. of AI & DS
(of Affiliation)
PES' Modern college of Engineering
SPPU
Pune, India

***Abstract*—DocEase indicates a transformative change in the digital document interaction space through Retrieval-Augmented Generation (RAG). It is designed to promote better document user interaction. DOCEASE boasts integrated sophisticated AI techniques: intelligent question-answering, summarization, TTS, language translation, and image analysis. Emerging generative AI and natural language processing trends have reshaped how users interact with different document formats. DocEase personalizes the accessing of information and exploration of content and simplifies the difficult jobs while making them fun. Its dynamic, AI-assisted query-generation process makes it even easier to find what you're looking for more accurately, all while enhancing accessibility and engagement. DOCEASE vision specializes in extracting and utilizing data properly for the purpose of data retrieval in different scenarios. The blend of functionality and convenience turns old document handling into an interactive experience that can engage technology to serve the varied needs of the digital end-users of today.**

***Keywords*—Content Summarization, Large Language Models, Deep Learning, Generative AI, Information Access, Multimedia Analysis, Multilingual Translation, Natural Language Processing, Query Response, Retrieval-Augmented Generation, Text-to-Speech.**

I. INTRODUCTION

In today's rapidly evolving digital landscape, where technological advancements and the expansion of knowledge occur at an extraordinary pace, it has become essential for every individual to cultivate capabilities that help them stay aligned with these swift changes. Traditionally, people have relied on reading books, newspapers, research articles, and other textual sources to stay informed. However, in an age characterized by an overwhelming surge of information from various sources, consuming and processing such a vast amount of data within a limited time frame has become increasingly challenging. This information overload not only hampers efficiency but also diminishes user engagement, as individuals often find themselves sifting through countless pages to extract relevant insights.

To effectively address these challenges of information saturation and declining productivity, we proudly present DOCEASE — an advanced, multi-functional retrieval-augmented generation-based system specifically designed to enhance user engagement and overall efficiency. Harnessing the power of cutting-edge technologies rooted in generative artificial intelligence and natural language processing, DOCEASE offers a comprehensive suite of features tailored to meet the diverse needs of users in a fast-paced world. Among its standout capabilities are content summarization, real-time query-response generation, text-to-speech conversion, multilingual translation, support for regional language PDFs, and in-document image extraction and display. These functionalities work seamlessly together to enable users to quickly retrieve, understand, and engage with applicable information, significantly reducing the time and effort typically required for extensive reading.

At the core of DOCEASE lies a resilient transformer-based architecture that employs an advanced attention mechanism. This enables the system to efficiently handle long-range dependencies in user requests, ensuring accurate and contextually relevant responses. DOCEASE dynamically interacts with user queries, offering personalized and responsive solutions tailored to individual needs. Users can conveniently upload a variety of document formats, including PDFs, PowerPoint presentations, Word documents, and images. Notably, DOCEASE supports regional language PDFs, making it an inclusive solution for users seeking to process documents in their native languages. Once uploaded, users can immediately apply any of the system's features, such as summarization, translation, or text-to-speech

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 04, April-2025

conversion, through an intuitive and user-friendly interface designed for both technical and non-technical users.

A key highlight of DOCEASE is its ability to extract and display images embedded within documents alongside generated responses. This feature transforms how users engage with information by providing a more interactive and visually appealing experience. Whether you are reading a report filled with complex charts or a research paper containing informative visuals, DOCEASE ensures that images are seamlessly retrieved and showcased, enriching the information retrieval process. Furthermore, the integrated data storage system ensures that uploaded documents, along with their corresponding summaries and translations, are securely stored and readily accessible whenever needed. Users can revisit their saved files, making DOCEASE an invaluable tool for researchers requiring quick access to comprehensive information or professionals seeking to manage documents efficiently.

What truly sets DOCEASE apart is its commitment to providing an all-in-one platform that caters to a wide spectrum of users. Whether you are a researcher in search of fast, accurate data extraction, a student aiming to grasp complex material in a shorter time, or a non-technical user seeking greater accessibility and ease of use, DOCEASE is designed to meet your needs. By supporting regional language PDFs, offering real-time responses, and providing image extraction and display, DOCEASE focuses on simplifying the user experience while ensuring no valuable information is overlooked. Its blend of advanced technology with a focus on simplicity ensures that anyone, regardless of technical background, can navigate the platform effortlessly and benefit from its comprehensive features.

In essence, DOCEASE is more than just a document processing tool — it is a game-changing solution crafted to revolutionize how individuals interact with information. By providing quick, accurate, and user-friendly methods to digest large volumes of data, DOCEASE empowers users to make informed decisions, improve productivity, and engage with content in ways that were previously time-consuming or inaccessible. As the digital world continues to accelerate, tools like DOCEASE are not just conveniences but necessities, ensuring that everyone has the means to keep pace with the ever-changing landscape of information and technology.

II. LITERATURE SURVEY

The foundational work by Saad-Falcon on "PDFTriage" [1] lays a strong foundation for organizing and prioritizing information in digital documents. Moreover, the research emphasizes efficient handling of unstructured data but lacks the capability to dynamically respond to specific user queries, a feature essential in today's interactive systems. DocEase builds on this by introducing automated summarization and query responses, enhancing the accessibility and usability of document processing for users.

In addition to Saad-Falcon's work, Gavilanes' research on "Use of LLM for Methods of IR" [2] addresses the dynamic needs of query generation and response, overcoming the limitation in PDFTriage, which didn't leverage machine learning to personalize retrieval tasks. By incorporating large language models (LLMs), this study takes traditional IR methods a step forward, making them more accurate and adaptive. However, it primarily focuses on query improvements and does not explore the benefits of combining search with generation models. DocEase bridges this gap by integrating retrieval-augmented models to combine the best of both retrieval and generation.

Furthermore, in "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks" [16], Lewis et al. take a significant leap by combining search and generation mechanisms. This not only advances the dynamic capabilities introduced by Gavilanes' work but also resolves the issue of retrieval-centric systems lacking context-aware content generation. However, the reliance on complex architectures in RAG can increase computational requirements, which DocEase addresses by optimizing RAG models to balance efficiency and accuracy while still delivering improved user query responses and summaries.

Similarly, Muludi's research on "Retrieval-Augmented Generation (RAG)" [3] enhances retrieval-based models further by emphasizing integration with generative mechanisms for context-aware data augmentation. This addresses the drawback of Lewis et al.'s approach, which lacked scalable adaptation for multilingual tasks. However, the study does not delve into enhancing accessibility for users requiring multilingual translation or hybrid document formats. DocEase leverages this hybrid approach while improving multilingual translation efficiency, making document interactions more accessible for diverse users.

In addition, Rajpurkar et al.'s SQuAD dataset [17] builds a benchmark for testing the comprehension abilities of natural language models, providing a tool to address the lack of standardized evaluation criteria in previous works like Muludi's. However, the dataset focuses narrowly on text-based tasks and lacks the versatility for testing varied document types. DocEase incorporates the advantages of such benchmarks while extending its capabilities to ensure robust query handling across different document structures.

The BART model by M. Lewis et al. [19] introduces a sequence-to-sequence architecture for generating and cleaning textual data, providing advantages over the static methods highlighted in previous benchmarks like SQuAD. However, while it excels in denoising and text generation, it falls short in integrating cross-linguistic capabilities. DocEase incorporates BART-inspired transformations while broadening its functionality to handle multilingual inputs and maintain coherence across varied document types.

Additionally, T. Labruna's research on "Teaching LLMs to Utilize Information Retrieval" [4] introduces methods to improve how models selectively utilize retrieved data, resolving challenges in BART related to content relevance. However, it doesn't optimize models for real-time efficiency or processing at scale. DocEase takes this selective retrieval concept further by streamlining the process to provide both relevance and real-time responses, enhancing user experience.

Lozano's work on an open-source RAG system [5] emphasizes the flexibility of RAG models, addressing Labruna's limitation in accessibility by promoting open frameworks for easier adoption. However, the research does not focus on long-term scalability for handling substantial data loads. DocEase adapts these open-source techniques

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 04, April-2025

while ensuring they scale seamlessly with growing document complexities and user demands.

Similarly, Raffel et al.'s study on a unified text-to-text transformer [18] innovatively transforms all NLP tasks into a single framework, surpassing the modular limitations in Lozano's RAG system. However, it does not address specific tasks like integrating visual or multimedia content. DocEase builds upon this framework while positioning itself to integrate text and multimedia processing for a more inclusive interaction model.

Zhang et al.'s work on "Multimodal Transformers for Image Captioning and Visual Question Answering" [20] expands the scope of NLP to multimodal systems, overcoming the text-centric limitations in Raffel et al.'s approach. However, multimodal systems often introduce new computational complexities. While DocEase currently focuses on text, its architecture is inspired by this study to integrate visual elements efficiently in the future, potentially enhancing document interaction.

Moreover, the work by Vaswani et al. in "Attention is All You Need" [8] provides the foundational architecture for transformer models used in modern NLP. This overcomes the limitations of earlier models like RNNs, particularly in long-range dependency handling. However, early transformer models are not inherently scalable for real-time tasks. DocEase incorporates the advantages of this architecture while optimizing for real-time query responses and handling large document sizes.

Shazeer's "Switch Transformers" [10] and Pope's work on scaling transformers [11] address scalability issues in models like Vaswani's transformers. By focusing on computational efficiency, they resolve challenges related to scaling models for larger datasets. However, the studies remain computationally heavy for low-resource environments. DocEase incorporates these advancements to ensure efficiency without compromising accessibility or computational feasibility.

Furthermore, Hendrycks et al.'s "Natural Adversarial Examples" [22] contributes to robustness by helping models handle unexpected inputs, improving upon earlier works like Switch Transformers, which focused primarily on efficiency. However, this approach doesn't fully account for linguistic and contextual nuances. DocEase integrates these robustness principles while emphasizing accuracy and diversity in query handling.

Additionally, Rombach's research on "High-Resolution Image Synthesis with Latent Diffusion Models" [12] extends possibilities for generating content beyond text, overcoming the single-dimensional nature of textual processing discussed in REACT. While this research is resource-intensive, DocEase incorporates it as a future direction to enrich the user experience with visual elements.

"Scaling Language Boundaries: A Comparative Analysis of Multilingual Question-Answering Capabilities in Large Language Models" looks into the performance disparities of large language models (LLMs) in high-resource and low-resource languages. The study proposes an efficient data collection strategy to improve LLM performance in languages that have been underrepresented, tackling the issue of limited training data. This method could contribute to a more equitable development of AI systems for diverse language contexts.

Correspondingly, the research paper by Gaikwad et al. "Adopting Pre-trained Large Language Models for Regional Language Tasks: A Case Study" examines the use of pre-trained large language models to low-resource languages, focused on sentiment analysis in Marathi. The evaluation of Multilingual BERT, indicBERT, and GPT-3 ADA sheds light on the possibilities and challenges for transferring these models into regional language tasks. This work reiterates the need to tailor AI solutions to accommodate linguistic diversity.

While factors of both aspects combine in the works covered, DOCEASE keeps evolving into a contemporary solution with a competitive edge in seamless document summarization, efficient translation, and unique, interactive query responses. In the course of this progression, the platform is beginning to render a different concept to consultation with regard to digital documents- that is increasingly intuitive, user-friendly, and generalizable.

III. PROPOSED SYSTEM

The development of DOCEASE focuses on creating an effective system that combines document retrieval with advanced language generation to provide users with precise and context-aware responses. The system starts by processing documents in various formats to prepare them for efficient use. This involves steps such as breaking down the text into manageable parts through chunking and embedding the data. Once processed, the content is transformed into vector embeddings using a specialized model designed to capture the core meaning of each document. These embeddings are stored in a Vector Database, where each vector is linked to its original text for quick and accurate retrieval.

When a user submits a query, the system processes it in a similar manner by generating a query embedding. The query vector is then compared to the stored document vectors using similarity measures like cosine similarity. This enables the system to identify the most relevant documents. These retrieved documents, along with any associated metadata, are combined with the user's query to create a contextualized prompt. This prompt is sent to a language model (LLM), which uses the enriched context to generate accurate and relevant responses.

DOCEASE goes beyond just delivering answers; it is designed to support regional language documents, enabling users to work seamlessly with content in multiple languages. This feature ensures accessibility and inclusivity, allowing a broader range of users to interact with the system effectively. Additionally, DOCEASE offers features to enhance user experience and accessibility. These include summarization, which condenses lengthy documents into concise summaries; multilingual translation, which helps break language barriers for non-native speakers; text-to-speech conversion, enabling visually impaired users to listen to the content; and image generation, which creates visual representations of text to aid comprehension.

The system's performance is continuously evaluated based on metrics such as accuracy, response time, etc. Even after deployment, DOCEASE is monitored to ensure it adapts to new challenges and meets user expectations. By iterating based on real-world feedback, DOCEASE remains a reliable, efficient, and user-friendly tool for document retrieval and response generation.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 04, April-2025

Apart from text-based retrieval, DOCEASE should be able to extract and retrieve images in user-uploaded PDFs relevant to the user's query. This capability allows better understanding by giving users textual and visual information, personalized to their needs.

Besides, DOCEASE comes with features that add to user experience and layer of accessibility. Summarization is a feature that shortens long documents into terse summaries. Multilingual translation helps address language barriers with non-native speakers. Text-to-speech creates audio versions of the content for the visually impaired.

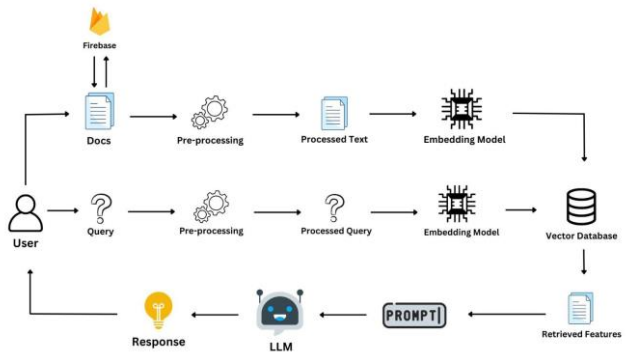

FIG. 1 HIGH LEVEL SYSTEM ARCHITECTURE

IV. CONCLUSION

DOCEASE represents a significant step forward in simplifying and enhancing digital document interaction. By combining Retrieval-Augmented Generation (RAG) techniques with state-of-the-art natural language processing, the system provides accurate, context-aware responses to user queries. Its multimodal features, such as content summarization, multilingual translation, text-to-speech conversion, and in-document image analysis, ensure accessibility and engagement for a wide range of users, including non-native speakers and individuals with visual impairments. It not only streamlines the process of accessing and understanding digital content but also empowers users to navigate complex data efficiently. Through its innovative approach, DOCEASE transforms the way individuals interact with and benefit from digital information systems.

REFERENCES

[1] Saad-Falcon, Jon, et al. "Pdftriage: Question answering over long, structured documents." arXiv preprint arXiv:2309.08872 (2023).
[2] Reddy, V. Madhusudhana, T. Vaishnavi, and K. Pavan Kumar. "Speech-to-Text and Text-to-Speech Recognition Using Deep Learning." 2023 2nd International Conference on Edge Computing and Applications (ICECAA). IEEE, 2023.
[3] Xiao, Guangxuan, et al. "Efficient streaming language models with attention sinks." arXiv preprint arXiv:2309.17453 (2023).
[4] Siriwardhana, Shamane, et al. "Improving the domain adaptation of retrieval augmented generation (RAG) models for open domain question answering." Transactions of the Association for Computational Linguistics 11 (2023): 1-17.
[5] Asai, Akari, et al. "Self-rag: Learning to retrieve, generate, and critique through self-reflection." arXiv preprint arXiv:2310.11511 (2023).
[6] Gao, Yunfan, et al. "Retrieval-augmented generation for large language models: A survey." arXiv preprint arXiv:2312.10997 (2023).
[7] Chen, Xuemin, Martin Gellert, and Wei Yang. "Inner workings of RAG recombinase and its specialization for adaptive immunity." Current opinion in structural biology 71 (2021): 79-86.
[8] Jeong, Cheonsu. "Generative AI service implementation using LLM application architecture: based on RAG model and LangChain framework." Journal of Intelligence and Information Systems 29.4 (2023): 129-164.
[9] Rabowsky, Brent. "Applications of generative ai to media." SMPTE Motion Imaging Journal 132.8 (2023): 53-57.
[10] Phan, Hung, et al. "Rag vs. long context: Examining frontier large language models for environmental review document comprehension." arXiv preprint arXiv:2407.07321 (2023).
[11] Pichai, Kieran. "A retrieval-augmented generation based large language model benchmarked on a novel dataset." Journal of Student Research 12.4 (2023).
[12] Garigliotti, Darío. "Explainable LLM-powered RAG To Tackle Tasks In The Unstructured-structured Data Spectrum." (2023).
[13] Hurtado, Joan Figuerola. "Harnessing Retrieval-Augmented Generation (RAG) for Uncovering Knowledge Gaps." arXiv preprint arXiv:2312.07796 (2023).
[14] Martineau, Kim, A. I. Explainable, and A. I. Generative. "What is retrieval-augmented generation?." IBM Research Blog 22 (2023).
[15] Huang, Wenyu, et al. "Retrieval augmented generation with rich answer encoding." Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers). 2023.
[16] Chen, Wenhu, et al. "Murag: Multimodal retrieval-augmented generator for open question answering over images and text." arXiv preprint arXiv:2210.02928 (2022).
[17] Shuster, Kurt, et al. "Retrieval augmentation reduces hallucination in conversation." arXiv preprint arXiv:2104.07567 (2021).
[18] Ghodratnama, Samira, and Mehrdad Zakershahrak. "Adapting LLMs for Efficient, Personalized Information Retrieval: Methods and Implications." International Conference on Service-Oriented Computing. Singapore: Springer Nature Singapore, 2023.
[19] Deng, Jingcheng, et al. "Regavae: A retrieval-augmented gaussian mixture variational auto-encoder for language modeling." arXiv preprint arXiv:2310.10567 (2023).
[20] Biswas, Debmalya, Dipta Chakraborty, and Bhargav Mitra. "Responsible LLMOps: Integrating Responsible AI practices into LLMOps." (2022).
[21] Tiwari, Apoorva, et al. "Scaling Language Boundaries: A Comparative Analysis of Multilingual Question-Answering Capabilities in Large Language Models." International Conference on Artificial Intelligence and Speech Technology. Cham: Springer Nature Switzerland, 2023.
[22] Gaikwad, Harsha, et al. "Adopting Pre-trained Large Language Models for Regional Language Tasks: A Case Study." International Conference on Intelligent Human Computer Interaction. Cham: Springer Nature Switzerland, 2023.

All authors declare that they have no conflicts of interest.