International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 10 | Oct - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

SaaS Platform for Context-Aware PDF Summarization using Generative NLP

Prof. Dhanashri Londhe[1], Jeevan Dhatbale[2], Mohan Dinkar[3], Anuj Dhole[4], Sujal Dhavale[5]

[1]*Prof. Dhanashri Londhe, Computer Engineering, Zeal college of Engineering and Research*
[2]*Jeevan Dhatbale, Computer Engineering, Zeal college of Engineering and Research*
[3]*Mohan Dinkar, Computer Engineering, Zeal college of Engineering and Research*
[4]*Anuj Dhole, Computer Engineering, Zeal college of Engineering and Research*
[5]*Sujal Dhavale, Computer Engineering, Zeal college of Engineering and Research*

Abstract – The modern digital landscape is overwhelmed with lengthy, complex documentation, especially in professional, academic, and legal sectors, primarily existing in the Portable Document Format (PDF). The exponential growth of documents creates a significant information retrieval bottleneck, where professionals spend excessive time manually sifting through files to extract key context and critical data points. Existing summarization tools are often limited to simple text compression, failing to understand document structure, tables, figures, or the overall context of complex, multi-page PDFs, which leads to generic, non-actionable summaries. This deficiency necessitates a robust, intelligent, and scalable solution that moves beyond simple summarization to genuine document understanding. Our solution is a SaaS Platform for Context-Aware PDF Summarization with Document Intelligence using Generative NLP, designed to address this problem by offering an accessible, web-based service. The platform is engineered to first apply Document Intelligence techniques to analyze the PDF's layout, extract structured data from tables and figures, and understand the hierarchical flow of information. This comprehensive context is then fed to a Generative NLP model, specifically a fine-tuned open-source Transformer model from the Hugging Face ecosystem, ensuring specialized and cost-effective performance, instead of relying on generic commercial APIs. The core innovation lies in the 'context-aware' nature of the summarization, where the generated output is tailored to the document's structure and the user's implicit intent, producing highly relevant and accurate summaries and key insights.

Key Words: Natural Language Processing (NLP), Software as a Service, Document Intelligence, Fast API (Python Web Framework).

1.INTRODUCTION

In the contemporary information economy, the volume of digital documentation, particularly in the Portable Document Format (PDF), has reached an unprecedented scale. Across industries—from legal and finance to engineering and research—professionals are increasingly burdened by the necessity of rapidly synthesizing critical information from extensive, complex, and highly structured documents. The inherent limitations of traditional text processing tools, which treat PDFs as simple streams of text, have exacerbated this challenge. They fail to interpret visual and logical cues, such as the relationship between figures, tables, and the surrounding text, or the hierarchical structure of a long report. This deficiency results in generic, low-quality, or entirely context-less summaries. The industry's reliance on rudimentary algorithms for summarization, rather than a deep, context-aware understanding, creates a significant gap between raw data and actionable intelligence. The future of document analysis requires a move beyond simple summarization to a specialized, scalable, and intelligent service capable of mimicking human-level document comprehension. The ambitious task at hand is the creation of a Software as a Service (SaaS) platform that seamlessly integrates Document Intelligence (DI) with Generative Natural Language Processing (Generative NLP). This endeavor is not merely an incremental improvement on existing tools; it is a foundational effort to redefine document interaction. By building a unified service architecture using a high-performance technology stack (React frontend, FastAPI backend) and a cost-efficient, fine-tuned Hugging Face model, we aim to provide an enterprise-ready solution.

3. FUNCTIONAL REQUIREMENTS

3.1 Authentication and Data Segregation

This section defines the mandatory requirements for secure user access and the logical separation of user data on the multi-tenant SaaS platform.

FR1: User Authentication - The system shall provide user registration and login functionality using email/password and third-party institutional credentials (e.g., Google login) thus ensuring authenticity and management.

FR2: Authorization and Role Management - The system shall enforce role-based access control (RBAC), distinguishing between User roles (Students, Professionals, Enterprise Clients) and the Admin role. The Admin shall be able to manage users, plans, and usage limits (e.g., documents/month).

FR3: Data Segregation (Multi-Tenancy) - The platform shall ensure strict segregation of user data such that a User shall not be able to access or view documents, summaries, or history belonging to any other user. This applies to both the file storage and the database records.

3.2 Document Intelligence Pre-Processor

This section details the functions required for handling input documents, extracting structure and text, and preparing the context for the generative model.

FR4: PDF Ingestion and Validation - The system shall accept user uploads of PDF files. The system shall perform file validation checks on upload, including format confirmation (must be PDF), size constraints, and integrity checks.

FR5: Text and Layout Extraction - The system shall utilize Document Intelligence (IDP/OCR) techniques to accurately extract all text content from the PDF, preserving the overall structural and hierarchical flow of the document (e.g., headers, sections, bullet points).

3.3 Generative Summarization Core

This section covers the central functions of the platform, enabling context-aware summary generation and customization.

FR6: Core Summarization Logic - The system shall feed the context-rich input (segmented text + structural data) to a specialized Generative NLP Model to produce a coherent and factually grounded output summary.

FR7: Customization Parameters - The system shall allow users to define key parameters for the generated output: Summary Length: (Short, Medium, Detailed)

Summary Type: (Bullet Points, Paragraph, Hybrid)

Context Preference/Style: (Academic, Business, General) to tailor the model's output tone and focus.

2.1 SYSTEM ARCHITECTURE

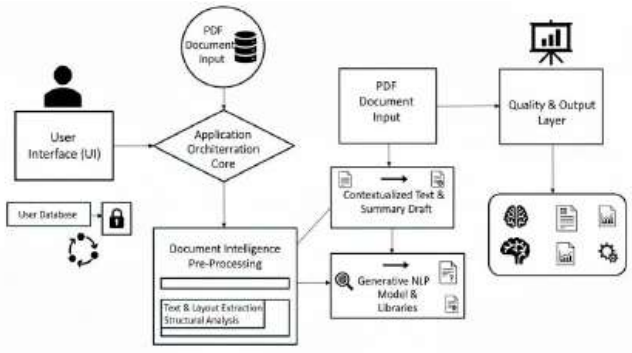

Fig. 4.1 Proposed System Block Diagram

Description

1. **Start:** Initiates the summarization process.
2. **Upload PDF:** The user uploads the target PDF document to the system.
3. **Valid PDF?:** Checks the uploaded file's integrity and format, and verifies user access limits.
    a. **Yes:** If all validation checks pass, the process moves to Extract Text.
    b. **No:** If the file is invalid or the user's limits are exceeded, the process shifts to Show Error.
4. **Extract Text:** The Document Intelligence pipeline runs to extract raw text and structural metadata (tables, figures) from the validated PDF.
5. **Summary Type:** This decision represents determining the core synthesis approach: selecting between Extractive drafting or Abstractive drafting based on user-defined parameters or internal logic.
6. **Extractive/Abstractive:** Based on the selected approach, the process moves to the core synthesis step.
7. **Generate Summary:** The Generative NLP Core takes the extracted context and defined method (Extractive/Abstractive) to synthesize the summary draft.
8. **Show Summary:** The final, checked summary is retrieved from temporary storage and presented to the user on the UI for review.
9. **Show Error:** Displays an appropriate error message (e.g., Invalid File, Processing Failure). The flow includes a loop back to Upload PDF to allow the user to try again.
10. **End:** The entire summarization session and process concludes.

3. CONCLUSIONS

In conclusion, this project successfully addresses the significant market gap left by generic summarization tools, providing a specialized solution that overcomes the limitations of processing complex, unstructured data like PDF documents. By developing the Brevity.AI SaaS platform, we deliver an accessible and easy-to-use tool for professionals, researchers, and students. The project's contributions lie in the sophisticated design and integration of a Document Intelligence pipeline with a Generative NLP core, utilizing a scalable React/FastAPI architecture. This innovative approach ensures that user feedback and performance metrics are continuously integrated, enhancing the platform's utility.

ACKNOWLEDGMENT

It gives us great pleasure to present our work on the **"SaaS Platform for Context-Aware PDF Summarization using Generative NLP."** We express our sincere gratitude to our guide, **Prof. Dhanashri Londhe**, for her invaluable technical expertise and dedicated support throughout this endeavor. We also thank the Head of the Department, **Prof. A. V. Mote**, and the Principal, **Dr. Ajit M. Kate** (ZCOER), for their motivation and for furnishing the necessary resources. This work would not have been possible without their continuous encouragement.

REFERENCES

[1] I. Awasthi, S. S. Anand, K. Gupta, P. S. Bhogal, and P. K. Soni, "Natural Language Processing (NLP) based Text Summarization - A Survey," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, IEEE, 2021, pp. 1310-1317.
[2] M. Ulker and A. B. Ozer, "Abstractive Summarization Model for Summarizing Scientific Article," *IEEE Access*, vol. 12, pp. 91252–91262, Jun. 2024.
[3] A. Ramprasad and P. Sivakumar, "Context-Aware Summarization for PDF Documents using Large Language Models," in *2024 International Conference on Expert Clouds and Applications (ICOECA)*, Coimbatore, India, IEEE, 2024, pp. 186–191.
[4] C. Ma, Z. Wu, J. Wang, S. Xu, Y. Wei, F. Zeng, Z. Liu, X. Jiang, L. Guo, X. Cai, S. Zhang, T. Zhang, D. Zhu, D. Shen, T. Liu, and X. Li, "An Iterative Optimizing Framework for Radiology Report Summarization with ChatGPT," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 5, pp. 4163-4175, May 2024.
[5] N. Moratanch and S. Chitrakala, "A Survey on Extractive Text Summarization," in *2017 International Conference on Computer, Communication and Signal Processing (ICCCSP)*
[6] M. F. Mridha, A. A. Lima, K. Nur, S. C. Das, M. Hasan, and M. M. Kabir, "A Survey of Automatic Text Summarization: Progress, Process and Challenges," *IEEE Access*, vol. 9, pp. 156043-156070, Dec. 2021.
[7] Y. Kumar, K. Kaur, and S. Kaur, "Study of automatic text summarization approaches in different languages," *Artificial Intelligence Review*, vol. 54, no. 8, pp. 5897–5929, Dec. 2021.
[8] S. Munnangi, "Revolutionizing Document Workflows with AI-Powered IDP in Pega," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 11s, pp. 570-580, Dec. 2023.