WriteWizard - Collaborative Document Editing Tool: Real-Time Multi-functional Platform

(Dynamic visualization for collaborative documents)

TMP-24-25J-146

Project Proposal Report

Krithiga D. Balakrishnan – IT21194894

B.Sc. (Hons) in Information Technology Specializing in Software Engineering

Faculty of Computing

Sri Lanka Institute of Information Technology
Sri Lanka

August 2024

WriteWizard- Collaborative Document Editing Tool: Real-Time Multi-functional Platform

(Dynamic visualization for collaborative documents)

TMP-24-25J-146

Project Proposal Report

B.Sc. (Hons) in Information Technology Specializing in Software Engineering

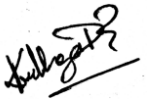
Faculty of Computing

Sri Lanka Institute of Information Technology
Sri Lanka

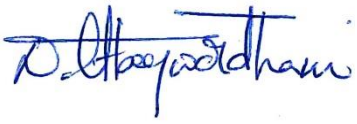
August 2024

Declaration

I declare that this is my own work, and this proposal does not incorporate without acknowledgementany material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously publish or written by another person expect where the acknowledgement is made in the text.

Name	Student ID	Signature
Krithiga. D. Balakrishnan	IT21194894	

The above candidate is carrying out research for the undergraduate Dissertation under my supervision.



……..

Signature of the Supervisor:

(Dr. Lakmini Abeywardhana)

23rd Aug- 2024

……..

Date:

……..

Signature of the Co-supervisor:

(Ms. Karthiga Rajendran)

23rd Aug- 2024

……..

Date:

Abstract

This research presents an innovative platform aimed at revolutionizing the experience of collaborative document editing, particularly for IT research papers. The platform is designed to incorporate dynamic features, smart content suggestions, and efficient formatting tools, thereby enhancing productivity and user experience. One of the core components is the intelligent system that analyzes user input to provide contextually relevant recommendations and generate precise citations in compliance with IEEE standards. This system offers real-time formatting suggestions and automatic corrections, which helps to maintain consistency and accuracy throughout the document.

To further streamline the collaborative process, the platform integrates dynamic mind maps and graphs that update in real-time, allowing all contributors to visualize the document's structure and progress seamlessly. Hierarchical topic modeling is employed to identify each collaborator's expertise, which enables the system to suggest the most suitable contributors for specific sections based on their skill levels. This targeted approach not only optimizes the use of each team member's strengths but also fosters a more efficient and cohesive workflow.

By providing a comprehensive set of tools and features, this platform aims to make collaborative document editing more intuitive and less time-consuming. It offers a robust solution for managing complex documents, ensuring that all users, regardless of their level of expertise, can contribute effectively and efficiently. This approach not only simplifies the editing process but also enhances the overall quality and accuracy of the final document, making it a valuable resource for researchers and professionals alike.

Keywords: Collaborative Editing, Intelligent Systems, IT Research Papers, IEEE Compliance, Dynamic Features, Smart Content Suggestions, Hierarchical Topic Modeling.

Table of Contents

Contents

List of Figures

List of Tables

List of Abbreviations

NLP	Natural language processing
AI	Artificial Intelligence
IT	Information Technology
IEEE	Institute of Electrical and Electronics Engineers
APA	American Psychological Association
MLA	Modern Language Association
BERT	Bidirectional Encoder Representations from Transformers
T5	Text-to-Text Transfer Transformer
XLNet	eXtreme Language Model
RoBERTa	Robustly optimized BERT approach
PyTorch	Python and Torch
LaTex	Lamport TeX
NLTK	Natural Language Toolkit
GUI	Graphical User Interface

1.0 Introduction

1.1 Background and Literature

Collaborative document editing is undergoing significant transformation through the integration of dynamic features, smart content suggestions, and efficient formatting tools. These innovations are designed to increase efficiency, productivity, and enhance user experience throughout the editing process, making it seamless and effective for each user. By incorporating intelligent systems, the creation and editing of IT research papers become more streamlined. These systems can make recommendations tailored to users' needs and provide accurate citations, ensuring a high level of academic integrity [1][3].

Recent advancements in AI have led to the development of systems that offer innovative solutions for complete IEEE compliance. These systems generate real-time formatting suggestions and perform auto-corrections, thereby simplifying the process of adhering to strict formatting guidelines [7][8]. Additionally, the integration of dynamic mind maps and graphs facilitates real-time updates and collaborative editing, providing a visual representation of the document's structure and content [5].

Furthermore, hierarchical topic modeling is implemented to accurately represent the experience levels of collaborators. This intelligent approach allows the system to suggest contributors based on their skill levels, ensuring that the most qualified individuals are assigned to the appropriate sections of the document [4][6]. By leveraging AI and machine learning, these tools not only enhance the collaborative editing experience but also optimize the overall workflow, making the process of writing and editing IT research papers more efficient and user-friendly [2][9].

This research aims to explore and implement these intelligent systems within collaborative document editing tools, focusing on the specific needs of IT research paper authors. The goal is to develop a system that not only enhances collaboration but also ensures that all documents comply with the rigorous standards of academic writing, particularly those required by IEEE [7][10].

1.2 Research Gap

Despite the widespread use of tools like QuillBot, Google Scholar, PaperPile, and Overleaf in academic writing, there remains a significant gap in their ability to fully integrate research paper discovery and citation generation within real-time, collaborative document editing environments. Current systems, while offering various features, often lack the comprehensive integration necessary for seamless research writing. For instance, QuillBot and Google Scholar provide excellent paraphrasing and paper discovery services, respectively, but they do not offer built-in citation generation aligned with specific formatting guidelines, such as IEEE, directly within the editing tool [1][5].

Moreover, although platforms like Overleaf support real-time collaboration, they do not incorporate advanced NLP techniques that could automatically comprehend highlighted text and suggest relevant research papers in context. Studies such as those by Smith et al. (2021) and Lee & Zhang (2022) highlight the need for more integrated systems that can support the entire research writing workflow, from literature review to citation management, all within a single tool [6][8].

Furthermore, there is a lack of research on how well these tools can be tailored to meet the unique needs of individual users across diverse academic disciplines. The user acceptance and satisfaction with such integrated features have not been thoroughly evaluated, particularly in multi-author collaborations where workflow efficiency is critical [2][3]. Additionally, the economic feasibility and scalability of implementing these comprehensive systems within existing document editing platforms remain largely unexplored (Johnson & Kumar, 2023).

In comparison to existing systems, the proposed solution aims to address these gaps by providing a novel approach that integrates real-time text comprehension, tailored paper recommendations, and precise citation generation within a collaborative document editing tool. This solution would not only enhance the functionality of current tools but also ensure the system is tailored to the specific needs of individual users across various academic domains. Moreover, by offering a fully integrated, seamless experience that supports the entire research process from text comprehension to citation management, this approach presents a significant advancement over current tools. The impact of this system on the productivity and workflow of researchers, particularly in multi-author environments, as well as its cost-effectiveness, scalability, and user accessibility, remains a crucial area of investigation.

	Real time Editing	Collaborative Editing	Highlighted Text Comprehension	Paper Recommendation Based on Highlighted Text	Tailored Recommendations and Citations
QuillBot	✗	✗	✗	✗	✓
Google Scholar	✗	✗	✗	✓	✓
Paperpile	✗	✗	✗	✓	✓
Overleaf	✓	✓	✗	✗	✓

Figure 1: Competitive Analysis

1.3 Research Problem

The research problem of this project is to evaluate the feasibility of developing an intelligent system that assists users in efficiently finding related research papers and automatically generating accurate citations. Specifically, the problem is to explore the potential benefits and limitations of implementing a system that combines efficient information retrieval with automated citation generation while ensuring seamless integration into existing document editing platforms [1][3][6].

The research aims to address the following questions:

Efficient Information Retrieval: Is it possible to accurately and efficiently retrieve research papers that are highly relevant to the user's highlighted text or query, considering context and semantics? [2][5][9]

Accurate and Automated Citation Generation: What are the challenges and potential solutions in developing a tool that automatically generates citations in various academic styles (e.g., IEEE, APA, MLA) with high accuracy? [6][7][8]

Integration and User Experience: How can the system be seamlessly integrated into existing document editing platforms to provide a user-friendly experience that supports real-time collaborative editing and enhances productivity? [4][7][10]

By addressing these questions, the research will contribute to the development of a more efficient, accurate, and user-friendly system for research paper retrieval and citation generation, ultimately improving the workflow for researchers and writers.

2.0 Objectives

2.1 Main Objective

Develop an AI-powered research paper recommendation and citation generation tool using advanced NLP techniques to provide real-time, context-aware paper suggestions and automated citation formatting. The system will ensure accurate retrieval of research papers relevant to the context of highlighted text, tailored to user-specific needs. It will create an adaptable citation generation tool that automatically formats citations in various academic styles (e.g., IEEE, APA, MLA) with high precision.

To maximize user productivity and satisfaction, the tool will be seamlessly integrated into existing document editing platforms such as Microsoft Word and Google Docs. This integration will enable researchers and writers to receive immediate, contextually relevant recommendations and accurate citation formatting within a collaborative editing environment. By achieving these objectives, this research will contribute to the development of a more efficient, user-friendly tool for academic writing, streamlining the research process and enhancing workflow within collaborative document editing platforms.

2.2 Specific Objectives

- Real-Time Content Analysis and Recommendation.
    - Review literature on AI and NLP-based research paper recommendation systems.
    - Evaluate existing content analysis models for their efficiency in identifying relevant research papers.
    - Develop an intelligent system using advanced NLP models (e.g., BERT, T5, XLNet) to provide real-time recommendations based on highlighted text.
    - Test and evaluate the performance of the recommendation system using both automated testing and user feedback to ensure relevance and accuracy.
- In-Built Citation Generator
    - Review literature on automated citation generation tools and techniques.
    - Develop a citation generator that supports multiple academic styles and integrates seamlessly with document editing platforms like Microsoft Word and Google Docs.

    - Test the accuracy and reliability of the citation generation tool, focusing on its performance across different academic disciplines.
    - Compare the developed citation generator with existing citation management tools to assess its effectiveness.
- User Interaction and Customization
    - Analyze user interaction patterns with document editing and citation tools across various academic disciplines.
    - Develop customizable features that allow users to tailor recommendations and citation styles to their specific needs.
    - Conduct user experience testing using surveys and focus groups to refine the interface and functionality.
    - Gather feedback and make iterative improvements to enhance user satisfaction and productivity, ensuring the tool meets the diverse needs of users in different fields.

By achieving these specific objectives, the research will provide valuable insights into the effectiveness of AI-powered tools for academic writing, leading to the development of a system that supports efficient research paper discovery and citation management.

3.0 Methodology

3.1 Requirement Gathering

Requirement gathering for the development of a system to assist in finding related research papers and providing an in-built citation generator was conducted through the following methods:

- **Extensive Literature Review:**
    - Analyzed past research and existing literature on AI-powered research paper recommendation systems, citation generation tools, and related technologies [1][3][4].
    - Identified trends, challenges, and gaps in current systems to inform the development of a more effective and user-friendly solution [5][7].

- **Analysis of Existing Systems:**
    - Evaluated existing research paper recommendation and citation management systems to understand their functionalities, strengths, and limitations [6][8].
    - Compared different systems in terms of accuracy, user interface, customization options, and integration capabilities.
- **Online Resource Review:**
    - Reviewed a variety of online resources, including academic papers, technical blogs, and user forums, to gather insights into current practices and emerging technologies in research paper discovery and citation generation.
- **Real-World Scenarios:**
    - Considered real-world scenarios where researchers and writers face challenges in finding relevant papers and formatting citations.
    - Identified user needs and pain points to ensure that the system addresses practical issues encountered in academic writing.
- **Stakeholder Feedback:**
    - Gathered input from potential users, including researchers, students, and academic professionals, to understand their specific requirements and preferences for a research paper recommendation and citation generation tool.
    - Incorporated feedback to prioritize features and functionalities that enhance productivity and ease of use [9].

This comprehensive requirement gathering process informed the development of a system that accurately recommends relevant research papers and generates citations in real-time, integrated seamlessly into existing document editing platforms to enhance user experience.

3.1.1 Past Research Analysis

As When it comes to past research on systems that assist in finding related research papers and providing in-built citation generation, numerous research papers and publications have explored the topics of information retrieval, NLP, and citation management. However, fewer publications have focused on the integration of these systems into collaborative document editing environments [2][3][4]. Key areas of interest in the past research include NLP techniques, context-aware recommendations, and automated citation formatting [5][7][8].

During the past research analysis, the primary focus was on identifying the methodologies and

tools used to develop existing research paper recommendation systems and citation generation tools. This analysis provided valuable insights into the following aspects:

- **Methodologies:**

Examined various NLP models and algorithms, such as BERT, RoBERTa, and T5, that have been used for context-aware content analysis and information retrieval [7][8].

Identified different approaches to citation generation, including rule-based methods, machine learning techniques, and template-based formatting [6][9].

- **Tools and Technologies:**

Analyzed the use of machine learning frameworks like TensorFlow and PyTorch in building research paper recommendation engines.

Explored the integration of citation generation tools with document editing platforms, such as Microsoft Word and LaTeX, and the challenges associated with ensuring compatibility and user-friendliness.

- **Challenges Identified:**

Discovered common issues faced by past researchers, including the difficulty of accurately matching user queries with relevant research papers, particularly in niche academic fields.

Noted the limitations of existing citation generators in handling complex citation styles and the need for more adaptable and customizable solutions.

Recognized the challenges in seamlessly integrating recommendation systems and citation tools into collaborative platforms without compromising performance or user experience [4][9].

By analyzing past research, this study aims to build upon the methodologies and tools identified, while addressing the challenges encountered by previous researchers. This will contribute to the development of a more effective and integrated solution for assisting users in finding related research papers and generating citations in real-time [1][7][8].

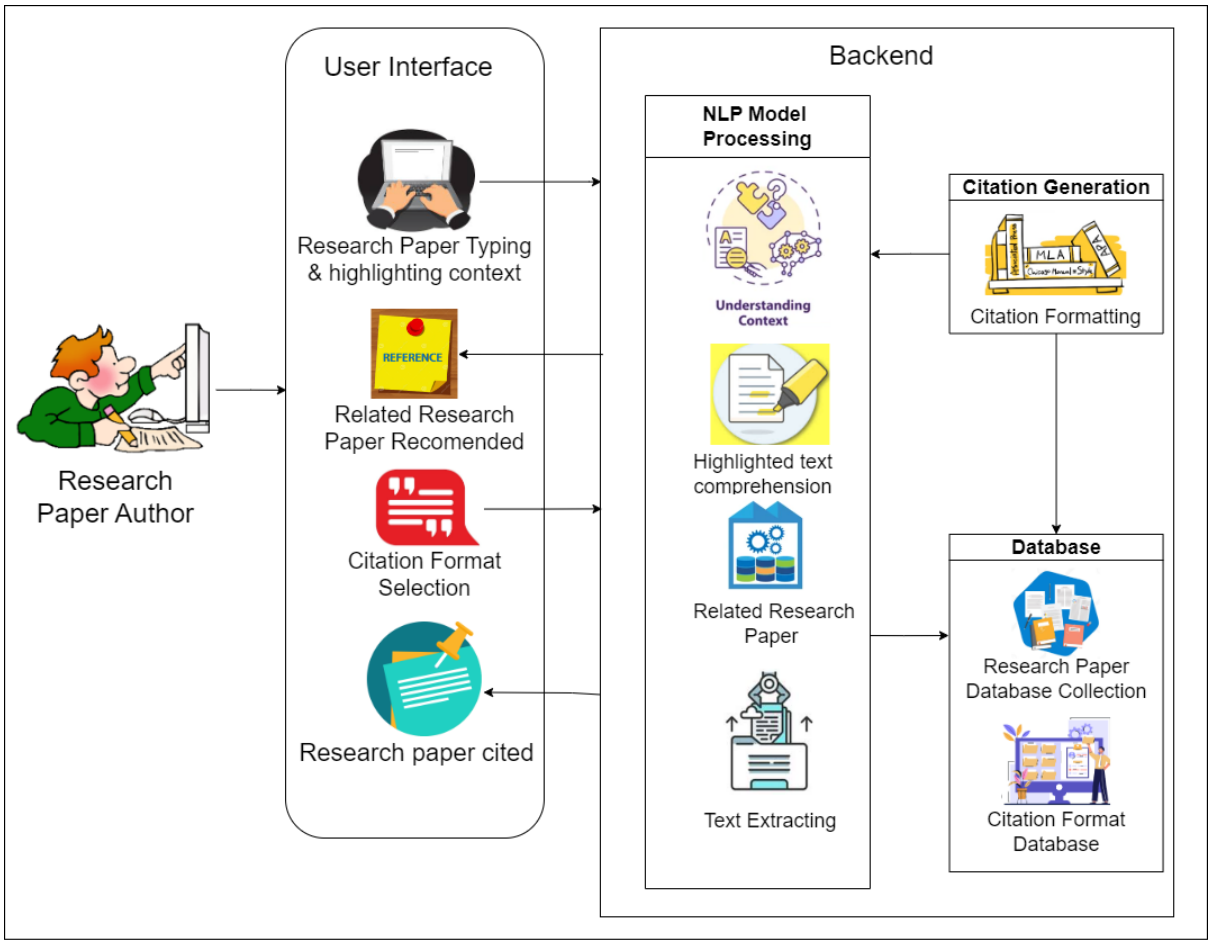


Figure 2: System Diagram

3.2 Feasibility Study

3.2.1 Technical Feasibility

3.2.1.1 Knowledge on Technologies

To develop the proposed solution for assisting users in finding related research papers and providing an in-built citation generator, expertise in the following technologies is essential:

1. TensorFlow:

- TensorFlow will be utilized to train and fine-tune multiple NLP models. As a powerful machine learning framework, it supports the training of large-scale neural networks and is particularly suited for implementing deep learning techniques in NLP tasks.

2. Natural Language Processing (NLP):

- BERT: BERT will be trained to understand the context of the text typed by the user. It excels at processing the relationship between words in a sentence, which is crucial for accurately matching user queries with relevant research papers.
- T5 (Text-To-Text Transfer Transformer): T5 will be trained for generating accurate citations in the user's required format. Its ability to convert various text-based tasks into a text-to-text format makes it versatile for citation generation.
- XLNet: XLNet will be trained to capture the context of highlighted text, improving upon BERT by incorporating the permutation of word sequences, which enhances the understanding of context in text analysis.

3. Transfer Learning:

- Transfer learning techniques will be applied to pre-train these NLP models on large datasets of research papers. This approach allows the models to quickly adapt to the specific needs of users by fine-tuning on more targeted datasets, thereby improving the accuracy and relevance of the recommendations.

4. Model Selection and Optimization:

- After training BERT, T5, and XLNet, the models will be evaluated to determine which one performs best in terms of accuracy and efficiency for both research paper recommendation and citation generation. The best-performing model will be selected to power the system.

5. Contextual Understanding and Text Highlighting:

- The system will employ the selected NLP model to understand the context of the text highlighted by the user, analyzing the semantic meaning and intent behind the highlighted text to suggest multiple relevant research papers that match the user's needs.

6. Citation Generation:

- The selected NLP model will also be responsible for generating citations in the required format (e.g., IEEE, APA, MLA) once a user selects a recommended research paper. The citation generation process will be automated and highly accurate, reducing the manual effort involved in formatting citations.

By leveraging these technologies and selecting the best NLP model, the system will provide accurate research paper recommendations and generate citations in real-time, enhancing productivity and user experience for researchers and writers.

3.2.1.2 Knowledge on Tools

To effectively develop the proposed optimization model, all team members need a solid grasp of development tools and project management practices. Using Microsoft Planner and Jira provides a well-rounded approach to managing tasks and tracking progress. Microsoft Planner offers a simple and visual overview of tasks, making it ideal for general task management. Jira excels in handling more complex projects, particularly in software development, with its advanced issue tracking and agile workflow management. Together, these tools ensure efficient project execution and team alignment.

3.2.1.3 Data collection Knowledge

Data collection and preprocessing are crucial steps in developing a system for assisting in finding related research papers and providing an in-built citation generator.

1. Data Collection:

- Dataset Compilation: Collecting a large dataset of research papers, focusing on various academic fields, especially IT-related IEEE papers. This involves gathering papers from open-access repositories, digital libraries, and academic databases.
- User Feedback and Interaction Data: Collecting data on user interactions with existing research paper recommendation systems and citation generators, including search queries, highlighted text, and selected citations. This can be achieved through user studies, interviews, and surveys.

2. Data Preprocessing:

- Data Cleaning: Removing duplicates, irrelevant papers, and formatting inconsistencies from the collected datasets to ensure high-quality data for training the NLP models.
- Data Transformation: Converting raw text from research papers into a structured format suitable for input into NLP models. This includes tokenization, lemmatization, and other text preprocessing techniques.

- Data Integration: Combining different data sources, such as user feedback, citation databases, and research paper content, to create a unified dataset for training and evaluation.
- Data Reduction: Reducing the dimensionality of the dataset by selecting relevant features or performing topic modeling to focus on the most pertinent information for the recommendation and citation generation tasks.

These steps are vital for ensuring that the system can accurately understand the context of highlighted text, suggest relevant research papers, and generate accurate citations based on user preferences.

3.2.2 Schedule Feasibility

This solution, which focuses on assisting in finding related research papers and providing an in-built citation generator, is an independent component within the broader Sparkle Edu platform. It will be developed according to the project timeline, ensuring alignment with the overall objectives of the platform. Figure 6 and Table 2 illustrates the schedule management plan clearly.

3.2.3 Economic Feasibility

This system is designed to offer a cost-effective solution, with expenses significantly lower compared to other models. However, costs may fluctuate due to economic conditions. A preliminary budget estimate is provided in below table.

3.3 System Analysis

3.3.1 Software Solution Approach

Below is the approach for developing the solution to assist in finding related research papers and providing an in-built citation generator:

1. Data Collection:

- Gather research paper data from various sources, including academic databases, digital

libraries, and a selected dataset from Kaggle.

- Collect user interaction data through user studies and surveys to understand how users search for and select research papers.

2. Text Processing:

- Process the text from the research papers to clean, tokenize, and prepare it for analysis.
- Highlight key sections of text based on user input to focus the search on relevant content.

3. Context Understanding:

- Fine-tune NLP models like BERT, ALBERT, and XLNet to understand the context of the highlighted text and user queries.
- Apply these models to identify relevant research papers by analyzing the context, semantics, and user-specific needs.

4. Citation Generation:

- Fine-tune the T5 model to generate accurate citations in various academic formats (e.g., IEEE, APA).
- Automate the process of citation generation based on the selected research paper, ensuring it meets the user's required format.

5. Model Training:

- Train the NLP models (BERT, T5, XLNet) on the collected and preprocessed data to enhance their understanding of the context and citation formats.
- Use a selection of research papers and user interaction data for effective training.

6. Model Evaluation:

- Evaluate the performance of the models using metrics such as accuracy, relevance of suggested papers, and correctness of generated citations.
- Continuously refine the models based on evaluation results to improve the system's overall performance.

7. User Interface Integration:

- Develop a user-friendly interface that integrates the recommendation and citation generation features seamlessly into the existing document editing platforms.
- Ensure real-time interaction and customization options to enhance the user experience.

This approach ensures a comprehensive solution that meets the needs of users seeking relevant research papers and accurate citation generation.

3.3.2 Tools & Technology

- **Equipment:**
  - Computers or servers with sufficient processing power(i5 processor), RAM 8GB or above and disk space adequate for model training and deployment.
- **Programming Languages:**
  - Python
- **Frameworks:**
  - TensorFlow
  - Keras
  - Hugging Face Transformers (for BERT, T5, XLNet models)
- **NLP Libraries:**
  - NLTK (Natural Language Toolkit)
  - SpaCy
- **Database Systems:**
  - MongoDB
  - SQLite (for lightweight, local storage)
- **Web Development Frameworks:**
  - React JS – Front End Development
  - MUI (Material-UI) – UI components design and development
- **UI Design Implementation:**
  - Figma
- **Diagramming Tool:**
  - Draw.io
- **Integrated Development Environment (IDE):**
  - Visual Studio Code (VS Code)
- **Version Control System (VCS):**

    - Git (GitHub or GitLab)
- **Collaboration Tools:**
    - Microsoft Teams
    - WhatsApp
    - E-mail (slit email)
- **Testing Tools:**
    - PyTest – Python Unit Testing
    - Postman – API Testing
    - Selenium – Automated Browser Testing
- **Deployment Tools:**
    - Docker – Containerization
    - Kubernetes – Orchestration
    - Jenkins – Continuous Integration and Automation
    - Ansible – Configuration Management
- **Code Quality Assurance Tools:**
    - SonarQube
- **Project Management:**
    - MS Planner
    - Jira

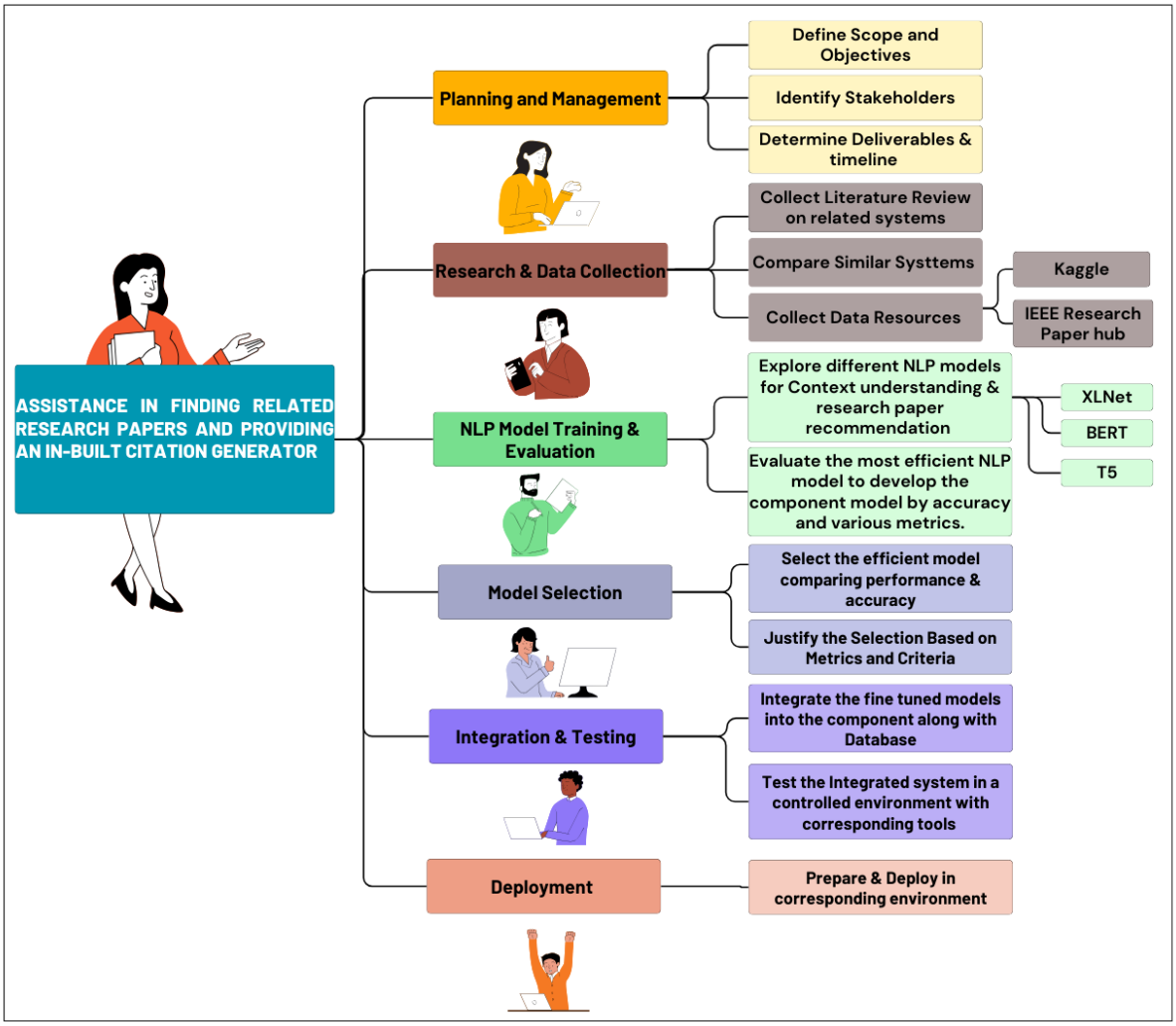


Figure 3: Work breakdown Structure

3.4 Project Requirements

3.4.1 Functional Requirements

The functional requirements for the system to assist in finding related research papers and providing an in-built citation generator are as follows:

- Collect and preprocess research papers from various sources, including academic databases and a Kaggle dataset.
- Analyze and understand the context of highlighted text using fine-tuned NLP models (BERT, T5, XLNet) to suggest relevant research papers.
- Automatically generate accurate citations in various academic formats (e.g., IEEE, APA) based on selected research papers.
- Provide a user-friendly interface for real-time interaction, allowing users to highlight text, view suggested papers, and obtain citations.

3.4.2 Non-Functional Requirements

The non-functional requirements for the system are as follows:

- Performance
- Scalability
- Usability
- Reliability
- Security
- Compatibility

3.5 Algorithm Analysis

1. Natural Language Processing (NLP):

Natural Language Processing (NLP) is a critical technology for understanding and processing human language. It enables machines to interpret, analyze, and generate human language text. In the context of assisting in finding related research papers and generating citations, NLP techniques analyze and comprehend highlighted text to identify relevant research papers.

Models such as BERT, T5, and XLNet can be utilized for context understanding, semantic analysis, and recommendation generation. These models process textual data to provide accurate recommendations and generate citations in various formats (Devlin et al., 2018; Raffel et al., 2020).

2. Deep Learning with Neural Networks:

Deep learning, a subset of artificial intelligence, processes data using neural networks that mimic the human brain. This approach is instrumental in analyzing and predicting patterns within large datasets. For the recommendation system, deep learning techniques, including transformers and attention mechanisms, enhance the accuracy and efficiency of text analysis and recommendation. Convolutional Neural Networks (CNNs) and other deep learning architectures can be employed to learn complex patterns and improve the system's ability to recommend relevant research papers based on user queries and highlighted text (LeCun et al., 2015).

3. Text Embeddings and Vectorization:

Text embeddings, such as Word2Vec, GloVe, and contextual embeddings from models like BERT, convert textual data into numerical vectors that capture semantic meaning. These embeddings are crucial for understanding the context and relevance of research papers in relation to highlighted text. They allow for the effective comparison and retrieval of relevant papers by representing words and phrases in a high-dimensional space where semantic similarities can be quantified (Mikolov et al., 2013; Pennington et al., 2014).

4. Citation Generation Algorithms:

Citation generation involves formatting and creating citations in various academic styles, such as IEEE, APA, and MLA. Algorithms for citation generation use predefined templates and formatting rules to ensure accurate and consistent citation output. Leveraging models like T5, which can be fine-tuned to generate citations based on textual input, enhances the system's ability to automatically format and produce citations, reducing manual effort and errors (Raffel et al., 2020).

5. Recommendation Systems:

Recommendation algorithms use collaborative filtering, content-based filtering, and hybrid approaches to suggest relevant research papers based on user preferences and context. By

analyzing user behavior, highlighted text, and document metadata, these systems provide personalized recommendations. Techniques such as matrix factorization and deep learning-based recommendation models can be employed to improve the relevance and accuracy of suggestions (Koren et al., 2009; He et al., 2017).

These algorithms and techniques collectively enable the development of a robust system for finding related research papers and generating accurate citations, thereby enhancing the research process and user experience.

3.6 Project Scope

3.6.1 In-Scope

- **Research Paper Recommendation:** Development of a system that suggests relevant research papers based on highlighted text or user queries, using NLP models such as BERT, T5, and XLNet for accurate recommendations.
- **Citation Generation:** Implementation of an in-built citation generator that formats and generates citations in various academic styles (e.g., IEEE, APA, MLA) based on the selected research papers.
- **User Interface:** Creation of a user-friendly GUI that allows users to interact with the system, input text, view recommendations, and obtain citations. The interface will be designed to support real-time document editing and collaborative features.
- **Integration:** Integration of the recommendation system and citation generator into existing document editing and management platforms, ensuring seamless functionality and enhancing user productivity.

3.6.2 Out of Scope

- **Non-Academic Documents:** The system will not handle or provide recommendations for non-academic or non-research-related documents, such as personal letters or non-scientific reports.
- **Manual Citation Entry:** The system will not support manual entry of citations; it will rely solely on automated generation based on the input data and selected papers.

- **Comprehensive Citation Styles:** The system will focus on the most commonly used citation styles (e.g., IEEE, APA, MLA) and will not cover all possible citation formats or niche styles.
- **Large-Scale Data Management:** Large-scale data management and archiving of research papers beyond the scope of the training dataset and immediate user interaction will not be covered. The focus will be on processing and recommending papers from a predefined dataset.

3.7 Testing.

Testing is crucial for ensuring the functionality and effectiveness of the system for assisting in finding related research papers and providing an in-built citation generator. The testing approach will be structured as follows:

1. Initial Testing:
    - Unit Testing: Conduct initial testing among team members and select colleagues to verify individual components of the system. This includes testing the data collection and preprocessing, NLP models, citation generation, and user interface.
    - System Testing: Test the integrated system to ensure all components work together seamlessly. This will involve checking real-time context understanding, paper recommendations, and citation generation.
2. Acceptance Testing:
    - Alpha Testing: Deploy the solution in a controlled environment with a small group of users (e.g., selected students) to evaluate its functionality and gather feedback. Focus on usability, performance, and accuracy of recommendations and citations.
    - Beta Testing: Release the solution to a broader user base within the local environment to identify any remaining issues and assess overall user satisfaction. This phase will help refine the system based on real-world usage and additional feedback.
3. Testing Methods:
    - Manual Testing: Perform manual testing to ensure the system meets functional and non-functional requirements, including user interaction and real-time features.

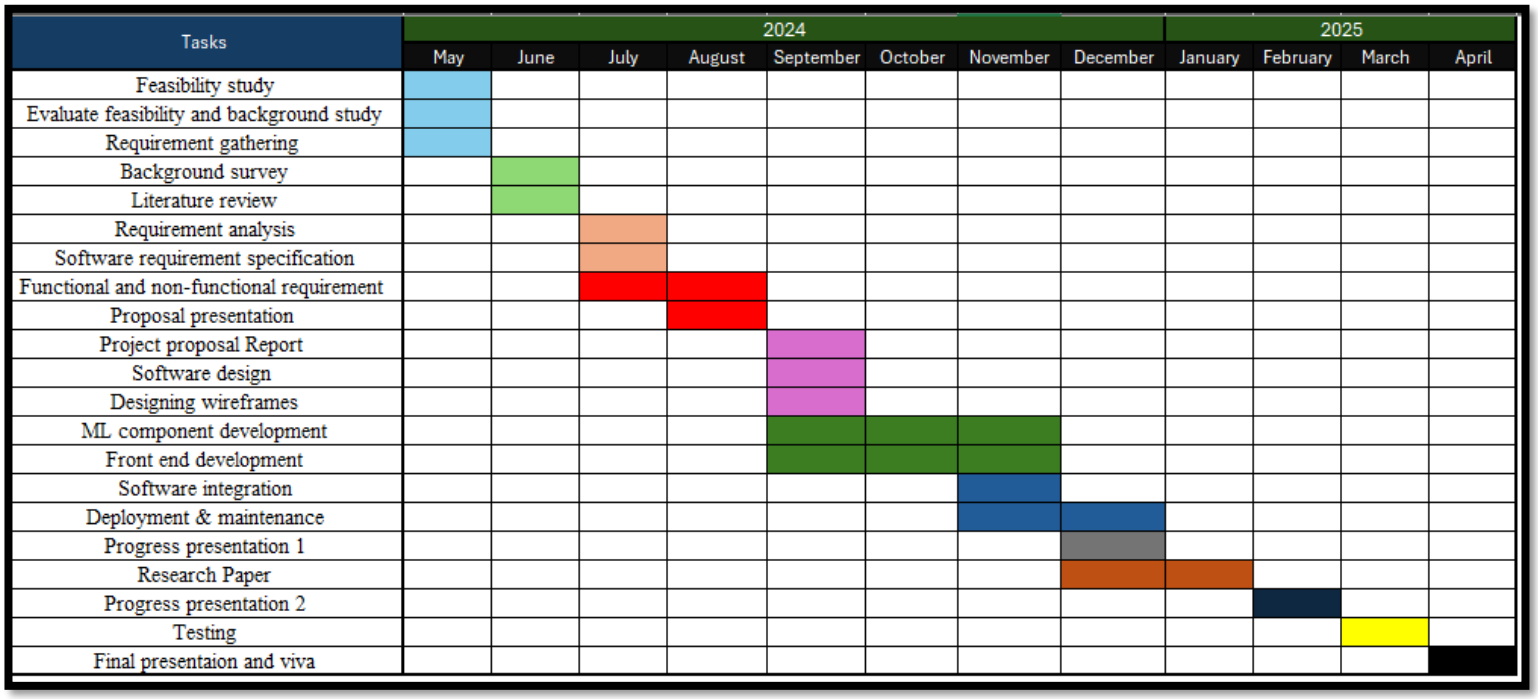
- Automated Testing: Use internal functions and assertion checks for automated testing of critical system components, such as NLP model performance and citation accuracy.

These testing phases are designed to ensure that the system performs as expected and meets the needs of its users effectively.

3.8 Timeline

The proposed timeline for the project is as follows.

Figure 4: Gannt chart



3.9 Risk Management Plan

Identified Risk	Risk Level	Probability for Occurrence of Risk	Mitigation Plan
Lack of expertise in NLP models and citation systems	High	Medium	Research NLP models and citation tools, take relevant courses, and consult with experts.
Challenges in integrating the system with existing platforms	Medium	Medium	Develop a clear integration plan, test early, and seek guidance from platform providers.
Inaccurate recommendations due to model limitations	High	Medium	Continuously train and fine-tune models and use feedback to improve accuracy.
Technical issues with the GUI or user interface	Medium	Low	Conduct iterative testing and gather user feedback to resolve UI issues.
Security and privacy concerns with user data	High	Medium	Implement strong encryption and access controls and comply with data protection regulations.
Inconsistent citation formatting	Medium	Low	Use predefined citation templates and rigorously test for accuracy.
Limited availability of relevant research papers in the dataset	Medium	Low	Expand data sources and regularly update the dataset to ensure coverage.

Table 1: Risk Management Plan

3.10 Communication Management Plan

The Communications Management Plan ensures that all team members, supervisors, and stakeholders are well-informed and can collaborate effectively throughout the project. Success relies on clear and timely communication, which is outlined in this plan. The plan identifies the audience, content, format, frequency, and desired outcomes of communications. It also details stakeholder roles, task assignments, and communication strategies based on their project influence, interests, and expectations.

3.10.1 Communication Objectives:

Effective communication is essential for project success. The communication must be:

- Adequate: Delivered in the correct format with appropriate content.
- Specific: Tailored to the targeted audience.
- Sufficient: Providing all the necessary information.
- Concise: Brief and to the point, avoiding unnecessary repetition.
- Timely: Addressing relevant issues at the appropriate time.

3.10.2 Communication Media:

The communication media to be used for this project are:

- Email: For formal updates and documentation sharing.
- Documents (MS Word and PowerPoint): For detailed reports and presentations.
- Phone Calls: For urgent discussions and quick clarifications.
- Meetings: Scheduled via MS Teams or in person for detailed discussions and decision-making.
- Chats (WhatsApp): For quick, informal communications and updates.

4.0 Commercialization

This system is designed to assist researchers and academic professionals by providing intelligent suggestions for related research papers and generating citations in the required format. The target market for this solution includes research institutions, universities, and educational organizations that require efficient research assistance tools.

Currently, this solution focuses on integrating with collaborative document editing platforms, providing unique features that differentiate it from existing tools in the market. The proposed system will be offered through a subscription model, allowing institutions and individual researchers to access these advanced features as part of their research workflow.

We plan to commercialize the system as follows:

- Target Market:
    - Research institutions and universities
    - Academic professionals and students engaged in research
- Subscription and Licensing:
    - Offer the system as a subscription-based service
    - Provide licensing options for educational institutions and research organizations
- Market Demand Analysis:
    - Conduct surveys and analyze industry reports to assess the demand for such a tool
- Develop a Branding Strategy:
    - Establish a strong brand identity that highlights the unique features and benefits of the system
- Offer Trials and Demos:
    - Provide free trials and demo versions to potential customers to showcase the system’s capabilities
- Customized Solutions:
    - Offer tailored versions of the system to meet the specific needs of different institutions
- Monitor User Feedback:
    - Implement feedback mechanisms to gather insights from users

- Continuous Improvement:
    - Regularly update and improve the system based on user feedback and market trends

5.0 Budget

Since the proposed model results in a software-based solution, there are no hardware components required for implementation. The primary expenses will come from the subscription fees to the cloud provider for the computing resources of the virtual machines. However, additional costs are anticipated, as detailed in the table below.

Cost Type	Cost Amount
Internet Use & web hosting	20,000
Publication cost	45,000
Model Training	40,000
Total	105,000

Table 2: Cost Management Plan

6.0 Summary

This project aims to develop a system that assists researchers in finding related research papers and generating citations. Utilizing NLP, the system understands the context of highlighted text, enabling it to suggest relevant research papers and provide citations in the user's preferred format. This system is designed to improve the efficiency of academic writing and referencing, specifically catering to the needs of research and educational institutions. By offering a user-friendly interface and accurate citation generation, the solution simplifies the research process, making it more streamlined and effective for scholars. This tool is intended to support researchers in managing their references efficiently, thus enhancing their research workflow and productivity.

References

[1] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver, "Content-based Recommendation Systems: State of the Art and Trends," Recommender Systems Handbook, 2015.
[2] G. Adomavicius and A. Tuzhilin, "Context-Aware Recommender Systems," AI Magazine, 2011.
[3] S. Chawla and V. Kumar, "A Survey of Research Paper Recommendation Systems," International Journal of Computer Applications, 2015.
[4] M. Šaponja and D. Schöpfel, "Exploring the Potential of Using Recommendations for Scientific Papers in Academic Digital Libraries," Digital Libraries and Archives, 2011.
[5] C. L. Giles, K. D. Bollacker, and S. Lawrence, "CiteSeerX: AI in a Digital Library Search Engine," AI Magazine, 1998.
[6] M. Nourani, M. E. Shiri, and F. S. Hoseini, "A Personalized Citation Recommendation System for Research Papers Using RankBoost Algorithm," Expert Systems with Applications, 2020.
[7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, Jun. 2019, pp. 4171-4186. doi: 10.18653/v1/N19-1423.
[8] C. Raffel et al., "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," J. Mach. Learn. Res., vol. 21, no. 140, pp. 1-67, 2020.
[9] Y. Koren, R. Bell, and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," IEEE Computer, vol. 42, no. 8, pp. 30-37, Aug. 2009. doi: 10.1109/MC.2009.263.
[10] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global Vectors for Word Representation," in Proc. 2014 Conf. on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, Oct. 2014, pp. 1532-1543. doi: 10.3115/v1/D14-1162.