

PORTFOLIO MANAGEMENT SYSTEM

Harshit Pratap Singh¹, Purnima², Er. Ankita Agarwal³

¹UG Student of Department of Computer Science and Engineering, Shri Ramswaroop Memorial College of Engineering and Management Lucknow, Uttar Pradesh, India

²UG Student of Department of Computer Science and Engineering, Shri Ramswaroop Memorial College of Engineering and Management Lucknow, Uttar Pradesh, India

³Professor, Department of Computer Science and Engineering, Shri Ramswaroop Memorial College of Engineering and Management Lucknow, Uttar Pradesh, India

Abstract— Research paper management is a crucial aspect of academic and professional work, as it involves organizing, storing, and retrieving research papers in a convenient and efficient manner. This research paper proposes a portfolio management software that aims to streamline the research paper management process for students. The software provides a range of features, including OCR functionality for digitizing research papers, summarization algorithms for creating short and concise summaries of papers, meta tag extraction for easy search and retrieval of papers, title extraction for efficient categorization of papers, and post generation for better readability. The proposed software is designed to be user-friendly and intuitive, allowing students to easily upload, manage, and share their research papers. The software can be a valuable tool for students in managing their research papers and improving their productivity.

Keywords— Portfolio Management System, Research papers, Summarization, OCR, Meta tags, AI based Digitization, Instagram-like interface, React, Material UI.

I. INTRODUCTION

Managing research papers can be a daunting task for students and researchers, as it involves organizing, storing, and retrieving papers in a convenient and efficient manner. To address this challenge, this research paper proposes a portfolio management software that streamlines the research paper management process. The software is designed to provide a range of features such as OCR functionality, summarization algorithms, meta tag and title extraction, and post generation, to help users manage their research papers more effectively.

The proposed software is user-friendly and intuitive, allowing students and researchers to easily upload, manage, and share their research papers. In this paper, we describe the features and functionality of the proposed software and demonstrate its effectiveness in managing research papers. The proposed software is expected to improve productivity and efficiency in research paper management and can be a valuable tool for students and researchers alike.

II. LITERATURE SURVEY

A. Mehdi Allahyari, Seyedamin Pouriyeh, Mehdi Assefi, Saeid Safaei^[6]

This paper suggests increasing amount of text data available from various sources has made it necessary to effectively summarize the information for practical usage. Automatic text summarization has thus become an important area of research, which aims to produce a concise and coherent summary while retaining the key information and meaning. This survey of the literature gives a general overview of the many methods and strategies employed in automatic text summarization. The different processes involved in summarization are discussed, and the strengths and weaknesses of various methods are described.

The review also highlights the challenges of automatic text summarization, which include the lack of human knowledge and language capability by computers. The article presents the historical development of text summarization, beginning with the pioneering research of Luhn et al. in the 1950s. The review concludes by emphasizing the importance of automatic text summarization in various domains, including search engines, news websites, and knowledge extraction approaches. The study provides valuable insights into the state-of-the-art in automatic text summarization, offering a useful reference for researchers and practitioners in the field.

B. A. Titipat Achakulvisut, Daniel E. Acuna, Tulakan Ruangrong, Konrad Kording^[7]

This paper describes the development of a content-based recommendation system for scientific publications called Science Concierge. The authors highlight that finding relevant scientific literature is a daunting task due to the exponentially increasing number of scholarly publications. The recommendation system implemented here is designed to provide near-real time suggestions, adapt to new content, and be open source. The authors tested the system on a dataset of 15,000 posters from the Society of Neuroscience

Conference in 2015 and found that the algorithm outperformed keyword-based suggestions.

The authors compare their approach with other recommendation systems and emphasize that scientific literature requires specific considerations. For example, scientists are highly specialized, and understanding the differences in fine-grained topics is challenging. Additionally, scientific discovery systems need to quickly adapt to new fields of research, which can appear overnight. The authors introduce their Python library, which combines a large-scale Rocchio Algorithm with online Latent Semantic Analysis to create scalable vectorization of documents. The library has the ability to speed up and improve the accuracy of scholarly research.

C. Chirag Patel, Atul Patel and Dharmendra Patel^[11]

This paper suggests that Optical Character Recognition (OCR) method is widely used for converting printed or handwritten text into editable text. However, the accuracy of OCR can be influenced by the text preprocessing and segmentation algorithms, as well as other factors such as image size, style, orientation, and complex background. This paper introduces the OCR method, the history of the open-source OCR tool Tesseract, and its architecture. The experiment results of OCR performed by Tesseract on different kinds of images are also discussed. The study concludes with a comparative analysis of Tesseract with other commercial OCR tools by considering the vehicle number plate as input.

The paper also provides a detailed overview of OCR applications, including license plate recognition, image text extraction from natural scene images, and extracting text from scanned documents. Many OCR tools are available in the market, but only a few of them are open-source and free. Tesseract is one such tool that is written in C++, is platform-independent, and can be used in other applications in the form of Dynamic Link Library (DLL). The history, functionality, and architecture of Tesseract are discussed in detail, along with a step-by-step description of how Tesseract OCR works. This paper provides valuable insights into the OCR method and the open-source OCR tool Tesseract, which can be useful for researchers and practitioners in this field.

III. EXISTING SYSTEM

As the project focuses on the development of a portfolio management system for research papers, it is important to review existing systems to gain insight into the state of the art.

One existing system that is worth mentioning is Mendeley, a reference manager and academic social network that allows users to organize, share, and discover research papers. Mendeley offers features such as reference management, collaboration tools, and an online community where users can connect and discuss research. However, it does not provide

automated summarization, meta-tagging, OCR, or image digitization features.

Another system worth considering is Zotero, a free, open-source reference manager that allows users to store, organize, and share research papers. Zotero also offers browser extensions, collaboration tools, and the ability to generate citations and bibliographies. However, it does not provide automated summarization, meta-tagging, OCR, or image digitization features.

Other existing systems such as EndNote, RefWorks, and Papers may offer similar features to Mendeley and Zotero, but they also do not provide automated summarization, meta-tagging, OCR, or image digitization features.

Given the limitations of existing systems, our proposed portfolio management system aims to provide a more comprehensive solution for managing research papers by integrating automated summarization, meta-tagging, OCR, and image digitization features. This will help users save time and effort in organizing and summarizing their research papers while also improving the accessibility and readability of their research portfolio.

IV. PROPOSED METHODOLOGY

The proposed project is a portfolio management software that is designed to assist students in organizing and managing their research papers. The software will provide a comprehensive set of features such as OCR, metadata extraction, paper summarization, and presentation of research papers in a more readable format. Additionally, the software will offer collaboration and discovery tools, enabling students to share their research with others and discover new research in their field.

The software will include a login page, registration page, and forgot password page, which will be built using React with Material UI. The registration page will require users to enter their name, email, password, and confirm password to create an account. The login page will allow users to access their account using their email and password. The forgot password page will enable users to reset their password by sending a password reset link to their email address.

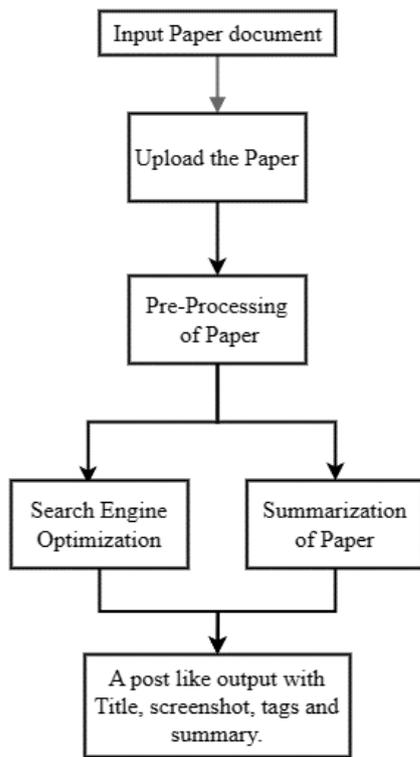


Fig 1. Flowchart

Agile and Waterfall approaches will be combined to construct the project. Several sprints will be employed to break up the development process, with each sprint concentrating on creating a particular system feature. The Waterfall methodology will be utilized to keep the project on track and on schedule, while the Agile methodology will be used to ensure flexibility and adaptation.

A combination of technologies, including the following, will be used to create the system:

- ReactJS
- Golang
- Python
- PostgreSQL
- Tesseract
- Spacy
- MaterialUI

Once logged in, users will be directed to a homepage similar to Instagram, with a top navigation bar offering options such as profile, messages, and home. The profile page will display the user's information, such as their name and profile picture, and provide a section to display their research papers in a more readable format. Users will be able to interact with their research papers by adding metadata, summarizing papers, and performing OCR on images of papers.

The software will be designed to make it easier for students to manage and organize their research papers, enabling them to focus more on their studies and research. With the

comprehensive set of features offered by the software, students will be able to streamline their research paper management process and collaborate with others more effectively.

V. EXPERIMENT AND RESULT

As a result, the user uploaded research papers will be stored in server that s/he can access it from anywhere, and they will also be processed for better digitization, the auto summarization of the content, meta tag generation, and title extraction will be performed. The paper screenshot will also be presented as a carousel that the user can easily navigate through using arrow buttons.



Fig.2 Portfolio Management System.

All the processed papers will be stored on the server and the information related to them will be stored on the database. Each processed paper will be presented as a post in which screenshot will work as carousel, title will be heading, meta tags will be the hashtags like twitter, and a small summary as a description. This whole processing will be automated, so no manual work is needed, users just need to register, login and upload papers.

VII. CONCLUSION

In conclusion, the portfolio management software for research papers presented in this project offers a range of functionalities that can greatly enhance the management and accessibility of research papers. With the OCR feature, users can digitize their papers, making them easily searchable and allowing for summaries and meta tags to be added for efficient organization. Additionally, the title extraction feature and presentation of papers as Instagram-like posts further improves readability. Overall, this software has the potential to significantly streamline research paper

management for students and professionals alike, saving time and increasing productivity.

VIII. FUTURE PLANS

There are several future plans that could be considered for the portfolio management software project. Some of these future plans include:

1. Integration with academic databases: The software could be integrated with academic databases such as Google Scholar and PubMed to allow users to search and access research papers directly from the platform.
2. Collaboration tools: The software could be enhanced with collaboration tools such as discussion forums, peer review, and citation management to enable users to collaborate and communicate effectively.
3. Mobile application: The software could be developed into a mobile application, which would enable users to access the platform on-the-go and manage their research papers from anywhere.
4. Integration with citation management tools: The software could be integrated with citation management tools such as Zotero and Mendeley to enable users to manage their references and citations more efficiently.
5. Machine learning-based recommendation system: The software could be enhanced with a machine learning-based recommendation system that provides personalized recommendations based on the user's research interests and reading history.
6. Integration with social media platforms: The software could be integrated with social media platforms such as Twitter and LinkedIn to enable users to share their research papers and connect with others in their field.
7. Expansion to other domains: The software could be expanded to other domains such as business, medicine, and engineering to provide a comprehensive solution for researchers in various fields.

These are just a few potential features for the project that could be explored in further research. Each of these plans could help to enhance the capabilities and usefulness of the portfolio management software, making it an even more valuable tool for students managing their research papers.

IX. REFERENCE

- [1] Shahbaz Syed, Dominik Schwabe, Martin Potthast (2022) "Summary Workbench: Unifying Application and Evaluation of Text Summarization Models".
- [2] Chuhan Wu, Fangzhao Wu, Tao Qi, Yongfeng Huang, Xing Xie (2021) "Fastformer: Additive Attention Can Be All You Need".
- [3] Ming Zhong, PengFei Liu, Yiran Chen, Danqing Wang, Xipeng Qiu, Xuanjing Huang (2020) "Extractive Summarization as Text Matching".
- [4] K. Karpagam, A. Saradha (2019). "a framework for intelligent question answering system using semantic context-specific document clustering and Wordnet".
- [5] A. Kumar, A. Sharma (2019) "Systematic literature review of fuzzy logic based text summarization. Iran. J. Fuzzy Syst.?".
- [6] Mehdi Allahyari, Seyedamin Pouriyeh, Mehdi Assefi, Saeid Safaei (2017) "Text Summarization Techniques: A Brief Survey"
- [7] A. Titipat Achakulvisut, Daniel E. Acuna, Tulakan Ruangrong, Konrad Kording (2016) "Science Concierge: A Fast Content-Based Recommendation System for Scientific Publications"
- [8] O. Alsharif and J. Pineau (2014). "End-to-end text recognition with hybrid HMM maxout models", ICLR.
- [9] A. Bissacco, M. Cummins, Y. Netzer, and H. Neven. (2013) "Photo OCR: Reading text in uncontrolled conditions", in ICCV.
- [10] Horacio Saggion and Thierry Poibeau. (2013) "Automatic text summarization: Past, present, and future. In Multi-source, Multilingual Information Extraction and Summarization", Springer, 3–21.
- [11] Chirag Patel, Atul Patel and Dharmendra Patel (2012) "Optical Character Recognition by Open-Source OCR Tool Tesseract"
- [12] Beaux Sharifi, Mark-Anthony Hutton, and Jugal Kalita (2010) "Summarizing microblogs automatically".
- [13] Dragomir R Radev, Hongyan Jing, Małgorzata Styś, and Daniel Tam. (2004) "Centroid-based summarization of multiple documents. Information Processing & Management" 40, 6, 919–938.
- [14] Simone Teufel and Marc Moens. (2002) "Summarizing scientific articles: experiments with relevance and rhetorical status", Computational linguistics 28, 4, 409–445.