TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

A

PROJECT REPORT

ON

FACULTY PUBLICATION MANAGEMENT SYSTEM

SUBMITTED BY:

RUJA AWAL (PUL075BCT069)

SANTOSH MAKA (PUL075BCT081)

SURAJ POKHREL(PUL075BCT091)

SUYOG DHAKAL (PUL075BCT092)

SUBMITTED TO:

DEPARTMENT OF ELECTRONICS & COMPUTER ENGINEERING

June, 2023

Page of Approval

TRIBHUVAN UNIVERSIY
INSTITUTE OF ENGINEERING
PULCHOWK CAMPUS
DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

The undersigned certifies that they have read and recommended to the Institute of Engineering for acceptance of a project report entitled **"Faculty Publication Management System"** submitted by **Ruja Awal**, **Santosh Maka**, **Suraj Pokhrel**, **Suyog Dhakal** in partial fulfillment of the requirements for the Bachelor's degree in Electronics & Computer Engineering.

.....	.....
Supervisor	External examiner
Arun Kumar Timalsina, Ph.D	**Bibek Ropakheti**
Associate Professor	Director
Department of Electronics and Computer Engineering,	Aadhar group Pvt Limited, Associate Professor
Pulchowk Campus, IOE, TU.	Pokhara University.

Date of approval: June, 2023

Copyright

The author has agreed that the Library, Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering may make this report freely available for inspection. Moreover, the author has agreed that permission for extensive copying of this project report for scholarly purposes may be granted by the supervisors who supervised the project work recorded herein or, in their absence, by the Head of the Department wherein the project report was done. It is understood that the recognition will be given to the author of this report and to the Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering in any use of the material of this project report. Copying or publication or the other use of this report for financial gain without approval of to the Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering and author's written permission is prohibited.

Request for permission to copy or to make any other use of the material in this report in whole or in part should be addressed to:

Head
Department of Electronics and Computer Engineering
Pulchowk Campus, Institute of Engineering, TU
Lalitpur,Nepal.

Acknowledgement

We would like to express our humble and sincere gratitude to the Department of Electronics and Computer Engineering, Pulchowk Campus for providing us this opportunity of major project, where we can apply the knowledge we have gained till date to gap between the academia and real competitive world while using and applying the concept of project management.

We are particularly indebted to our revered supervisor of this project, Dr. Arun Kumar Timalsina, under Pulchowk Campus for the continuous guidance and critique in study and preparation for this project.

Furthermore, We would like to extend our sincere thanks to all our teachers and friends who have directly and indirectly helped in the progress of our project work.

Gratitude,
Ruja Awal (075BCT069)
Santosh Maka (075BCT081)
Suraj Pokhrel (075BCT091)
Suyog Dhakal (075BCT092)

Abstract

Faculty publication management system is an integrated web application software that allows administrators to effectively manage all the faculty publications from record keeping to publication data analytics and visualization. Faculty evaluation system is also a part of the system that allows administrators to evaluate faculty members, lecturers and professors. This report presents design and development of the systems. Due to absence of a unified faculty publication management system, research publications of faculty members were spread across various research publication sites like GoogleScholar, ReserachGate, IEEE, etc and it became a tedious task for administrators to keep track of publications and generate associated reports for administrative purposes. Our system scrapes all the relevant data from such websites and presents user with a unified system to add, edit, approve, search, sort and generate reports of such publications as per faculty members or departments. The system additionally allows a user to perform data analytics and visualizations with bar chart and pie chart of research publications across faculty members and even specific departments. Later, the system uses the publication data of faculty members along with other evaluation metrics like years of service and objective feedback form data provided by students for evaluation of the user. Also, title based clustering of faculty members help to create clusters of members with similar research areas.

Keywords: *research publication, data analytics, visualization, unified system, evaluation*

Contents

Page of Approval **ii**

Copyright **iii**

Acknowledgement **iv**

Abstract **v**

Contents **viii**

List of Figures **x**

List of Abbreviations **xii**

1 Introduction **1**

1.1 Background 1

1.2 Problem Statement 1

1.3 Objectives 2

1.4 Scope 2

2 Literature Review **3**

3 Related Theory **5**

3.1 Web Scraping 5

3.2 IEEE Explore 5

3.3 Google Scholar 5

3.4 ResearchGate 6

3.5 BibTex 6

3.6 Application Programming Interface (API) 6

3.6.1 REST API 6

3.7 JSON Web Tokens 6

3.8 Model View Controller (MVC) Architecture 7

3.9 Data Analytics and Visualization 8

3.10 Clustering 8

3.11 k-means clustering . . . 8
3.12 Elbow Method . . . 9
3.13 Distilbert-base-nli-mean-tokens model of BERT . . . 9
3.14 LDA Model . . . 9

4 Methodology **11**
4.1 Requirements Gathering . . . 11
4.2 System Design . . . 11
4.2.1 Scraping System . . . 12
4.2.2 Faculty Evaluation System . . . 13
4.2.3 Analytics and Visualization . . . 16
4.2.4 Title Based Author Clustering . . . 16

5 Tools and Technologies **20**
5.1 Frontend . . . 20
5.1.1 JavaScript (JS) . . . 20
5.1.2 React . . . 20
5.1.3 Bootstrap and Material UI . . . 20
5.2 Backend . . . 20
5.2.1 Python . . . 20
5.2.2 Django . . . 20
5.3 Database . . . 20
5.3.1 Structured Query Language (SQL) . . . 20
5.3.2 SQLite . . . 21
5.4 Web Scraping . . . 21
5.4.1 Selenium . . . 21

6 Results **22**

7 Epilogue **30**
7.1 Conclusion . . . 30
7.2 Limitations and Future Enhancements . . . 30

References **30**

8 Appendix **33**
8.1 Screenshots . . . 33
8.1.1 Frontend UI . . . 33

8.1.2 Statistics and Reporting 40
8.1.3 Web Scraping 42
8.1.4 Searching and Sorting Feature 43
8.1.5 Approval Hierarchy Feature 44
8.1.6 Personal Website Integration 45
8.1.7 Automatic Profile Summarization 46
8.1.8 Report Generation 47
8.1.9 Faculty Evaluation 50
8.2 Analysis 52
8.2.1 Publication per Year 52
8.2.2 Researchers per Department 53
8.2.3 Publications per Department 54
8.2.4 Publication History 55

List of Figures

3.1 k-means Clustering 9

4.1 Main components of Faculty Publication Management System 12
4.2 Block Diagram of Scraping System 12
4.3 Evaluation based on feedback form 16
4.4 Flowchart of getting Research Topics from each cluster 17
4.5 Mapping of cluster name and author 19

6.1 Department Researchers and Publications count 22
6.2 Comparison of paper published between Department of Electronics and Computer Engineering and Department of Architecture 23
6.3 Comparison of paper published between Department of Electronics and Computer Engineering and Department of Civil Engineering 24
6.4 Comparison of paper published between Department of Electronics and Computer Engineering and Department of Mechanical Engineering 24
6.5 Comparison of paper published between Department of Electronics and Computer Engineering and Department of Applied Science and Chemical Engineering 25
6.6 Comparison of paper published between Department of Electronics and Computer Engineering and Department of Electrical Engineering 26
6.7 Total Publications contribution among all departments 26
6.8 Faculty Evaluation Sample 27
6.9 Title samples of Cluster 3 27
6.10 Research Topics from Cluster 3 28
6.11 Research Topics of every cluster 28
6.12 Title based Author Clustering 29

8.1 Home Page 33
8.2 Account Registration 34
8.3 Login Feature 35
8.4 Add Paper Feature 36
8.5 Paper Detail 37
8.6 Publication Listing 37

8.7 Export Feature for Downloading Publications 38
8.8 Import from BibTex Feature for Adding Papers 39
8.9 Statistics and Reporting of Publications 40
8.10 Bar Graph and Pie Chart Visualizations of Publications 41
8.11 Scraped Data 42
8.12 Search Feature for Searching Authors and Publications 43
8.13 Approval Feature for Admin 44
8.14 Personal Website Integration 45
8.15 Automatic Profile Summarization 46
8.16 MLA Format 47
8.17 APA Format 48
8.18 UGC/TU Format 48
8.19 IEEE Format 49
8.20 Faculty Evaluation Sample 50
8.21 Faculty Evaluation Form 51
8.22 Total number of Publication per Year 52
8.23 Researchers per department 53
8.24 Publications per department 54
8.25 Publications in 2023 55
8.26 Publications in 2022 55
8.27 Publications in 2021 56

List of Tables

4.1 Criteria of Faculty Evaluation System 14
4.2 Evaluation based on Publisher 14
4.3 Evaluation based on citation 15

List of Abbreviations

AJER	American Journal of Engineering Research
API	Application Programming Interface
CSS	Cascading Style Sheet
ECDSA	Elliptic Curve Digital Signature Algorithm
FPMS	Faculty Publication Management System
IEEE	Institute of Electrical and Electronics Engineers
IOE	Institute of Engineering
IADIS	International Journal on Computer Science and Information Systems
JSON	JavaScript Object Notation
JWT	JSON Web Token
LDA	Latent Dirichlet Allocation
MVC	Model View Controller
NLP	Natural Language Processing
RFC	Request For Comments
RSA	Rivest Shamir Adleman
REST	Representational State Transfer
SAGE	Scientific Advisory Group for Emergencies
SQL	Structured Query Language
UI	User Interface

1. Introduction

1.1 Background

Faculty Publication Management System is a part of a university record management system responsible for keeping track of all the publications published by faculty members of the university till the date for easier record keeping. This is necessary for associated faculty individuals and the university itself. Most universities rely on such systems for record management tasks. There has been development of such system but the record keeping and tracking of publications is manual. So, this system automates the collection of publications of faculty members uploaded on the web on platforms like GoogleScholar, Acedemia, ResearchGate, etc. by using web scraping and stores them in a centralized database. This system can be used to view, search and print the publications of the faculty members. This system also provides detailed analysis and visualization to provide insights of the research publications in the institution.

An evaluation system for evaluating the faculty members is also a growing need in the university. Unlike the case of foreign universities, there is no any evaluation system in our university. Evaluation system is used for evaluation of faculty members with the help of students and the evaluation reports are used by the higher committee for promoting or demoting the associated lecturers/professors. But due to lack of such system in our university, evaluation has been a difficult task. Although evaluation is done by administrative processes, there is a lack of feedback from students who are eventually to be benefited by the evaluation. Some lecturers/professors provide printed evaluation forms with a set of objective and/or subjective questions. But this physical task would be a lot easier through a unified web application system. Analysis of their work in research publications from the centralized database in the system also contributes to their evaluation.

1.2 Problem Statement

The major problems are highlighted as:

- There is a need of a unified system for automated gathering and tracking research publications of faculty members.
- In the available system, the administrators in the university have to manually record and update the information about the research publications of faculty members.

- There is a lack of a faculty evaluation system that takes account of feedback from student evaluation form and also analysis of their research works for the evaluation of faculty members.

1.3 Objectives

Our primary objective is to develop a faculty publication management system that uses web scraping to automate the manual record keeping tasks and integrate the faculty evaluation system.
The objectives of this project are:

- To develop a unified system for automated record keeping and tracking of all the research publications present in Internet of faculty members by using web scraping.
- To incorporate the feature of viewing, searching, report generation and analysis and visualization of research publications of faculty members.
- To integrate a comprehensive faculty evaluation system to provide a fair evaluation of faculty members through various metrics.

1.4 Scope

The universities at Nepal currently lack a unified application for gathering, managing and tracking research publications of their faculty members. This system provides a platform for automated record keeping of all the publication works of the faculty members, a detailed analysis and visualizations for tracking the researches and report generation of either individual faculty member or department. Also, the evaluation system integrated in this system can be used for evaluating and assessing the faculty members through their publication data along with other evaluation metrics like years of service and objective feedback form data provided by students for evaluation of the courses taught by faculty members. The clustering of faculty members based on their research publications included in this project helps to group the faculty members with similar areas of expertise and research publications. It is an useful approach to gain insights on the area of researches within a department and allows better collaboration of faculty members and identification of potential interdisciplinary research opportunities.

2. Literature Review

In 2014, a web scraping model based on Transclusion [1] was used for reusing web contents in e-Learning environment. This study focused mainly on the copyright issues of the web contents and learning materials created by teaching instructors and professionals on the web. The model classified scrap-notes into three different types based on copyright restrictions: Trans-quotation, Trans-reference and Trans-annotation to solve the problem of illegal content copying. The University of California publication management system was developed as an online service that covers more than 8,000 faculty and 40,000 articles a year[2] by faculty at all ten University of California campuses which help faculty to view and manage all their publication citations in one place. The project at Touro College and University System library in 2010[3] was focused on building an in-house faculty publications database by collecting the research publications of their faculty and staffs. The project led to publication of an annual book with the bibliography of its faculty publications. In 2018, an overview on web scraping techniques and tools [4] was done to highlight all the techniques and tools existing at the time. The study mainly focused on extracting unstructured data on single or multiple websites into a structured data automatically. The topic of web mining was also promoted on the study. In 2019, a study on personalized content extraction and text classification using effective web scraping techniques [5] was done which mentioned a technique to convert unstructured format into a structured representation. The study mainly focused on text classification and recursive feature elimination techniques for classifying documents. Illinois State University librarians in 2008, developed a project which is an online bibliography to collect the scholarly works by campus faculty to serve the needs of researchers interested in the history of institution[6]. This project was believed to be a great foundation for an institutional repository.

Faculty evaluation is an important factor that contributes to faculty performance, promotion, incentives, accountability and professional development process. By providing detailed and helpful feedback, faculty members can identify areas of opportunity to improve[7]. Student evaluation is mostly used as a single parameter for the faculty evaluation. But, along with student evaluation, faculty evaluation must also consider other aspects like the faculty member's research, years of teaching and service contributions[8]. The Department of Family Practice and Community Health (DFPCH) at the University of Minnesota School of Medicine designed an assessed approach for evaluating faculty performance[9] that helped

faculty members to continuously improve by providing feedback and also used this approach to provide a basis for salary and merit-pay decisions. In practice, students receive a page containing a set of evaluation questions for evaluating the faculty members. This process is prone to human errors and also possibly compromises the records confidentiality, integrity and availability. Such evaluation system is a complex and paper heavy system[10] which can be replaced by online evaluation system. The main challenges in faculty evaluation are carelessness in filling out evaluation forms and evaluation by absent students in evaluation based on student perspective and also the poor mentoring of faculty activities, example, whether or not the faculty member was present at the class at 95% of the time[11].

3. Related Theory

The Faculty Publication Management System is a solution designed to address the challenges faced by academic institutions in managing the research publications of their faculty members. The research publications are spread across a number of research publication platforms like Google Scholar, Springer, Taylor and Francis, Elsevier, Wiley, Emerald Publishing, IEEE Xplore, ResearchGate and so on. Many of the research publication platforms including Taylor and Francis, Elsevier and Wiley operate on a subscription model, where access to their content requires a subscription or institutional affiliation. So, in this project, three research websites, IEEE Xplore, Google Scholar and ResearchGate are used because most of the publications are included in these websites. The research publications of the faculty members used to be collected manually which is cumbersome to the administrators. So, this system automates the collection of publications of faculty members uploaded on these research publication platforms by using web scraping.

3.1 Web Scraping

Web scraping is the process of extraction of data from various websites. The collected information is then exported into various formats useful for the user. Such data can be used for different purposes like data analysis and researches. Web Scraping is used in this project to extract the information of the research papers like title, publication date, publisher, authors, total number of citations and the link to the paper and stores it in a centralized database.

3.2 IEEE Explore

IEEE (Institute of Electrical and Electronics Engineers) is a professional organization for engineers and technologists that publishes journals, conference proceedings, and standards related to the fields of electrical and electronics engineering. IEEE Xplore is the digital library of IEEE and offers a vast collection of scientific and technical publications. It includes journals, conference proceedings, standards, and other related content.

3.3 Google Scholar

Google Scholar is a search engine that provides a simple way to broadly search for scholarly literature. It allows users to search for academic resources and scholarly literature such as abstracts, full-text articles, theses, books, and more from across many disciplines. Also, it

provides various citation metrics and links to full-text versions of articles.

3.4 ResearchGate

ResearchGate is a social networking site for researchers and scientists that allows them to to share papers, discover research, ask and answer questions, and find collaborators. It is one of the largest academic social network in terms of active users. It also provides tools for tracking research impact and metrics. Reading articles in ResearchGate does not require registration.

3.5 BibTex

BibTeX is a reference management tool used in LaTeX documents that allows the users to store and manage their bibliographic data in a structured format which makes it easy to cite and reference sources in a document. It mostly contains fields like author, title, journal, and year. BibTex can also be uploaded to the system to add the research publication in the centralized database.

3.6 Application Programming Interface (API)

An application programming interface (API) is a computing interface that defines interactions between multiple software intermediaries. In our system, it defines the kinds of calls or requests that can be made, how to make them, the data formats that should be used, the conventions to follow, etc. It also provide extension mechanisms so that users can extend existing functionality in various ways and to varying degrees. An API can be entirely custom, specific to a component, or designed based on an industry-standard to ensure interoperability. Through information hiding, APIs enable modular programming, allowing users to use the interface independently of the implementation. The two most common types of web APIs today are REST APIs and GraphQL APIs. REST API is implemented in our system.

3.6.1 REST API

Representational state transfer (REST) is a software architectural style that defines a set of constraints to be used for creating Web services. REST API is used for building our applications because it is simple and straightforward that suits our application as our database does not have to include diverse datasets otherwise GraphQL would be suited.

3.7 JSON Web Tokens

JSON Web Token (JWT) is an open standard (RFC 7519) that defines a compact and self contained way for securely transmitting information between parties as a JSON object. This

information can be verified and trusted because it is digitally signed. JWTs can be signed using a secret (with the HMAC algorithm) or a public/private key pair using RSA or ECDSA. Although JWTs can be encrypted to also provide secrecy between parties, signed tokens are the most common use case. In our system signed tokens verify the integrity of the claims contained within it. When tokens are signed using public/private key pairs, the signature also certifies that only the party holding the private key is the one that signed it. JWT Tokens have been used for following purposes.

1. Authentication:

JSON Web Tokens are a good way of secure authentication. The authentication means confirming that a user is who they say they are. This ensures only those with authorized credentials gain access to secure systems.

2. Authorization:

This is the most common scenario for using JWT. In our system, once the user is logged in, each subsequent request will include the JWT, allowing the user to access routes, services, and resources that are permitted with that token.

In its compact form, JSON Web Tokens consist of three parts separated by dots (.), which are:

1. Header
2. Payload
3. Signature

3.8 Model View Controller (MVC) Architecture

Model, View, and Controller (MVC) architecture is an architectural pattern that divides an application into these three logical parts. The application's data and business logic are represented by the model. In a similar way, the view engages with the user interface and shows the user the data. Input from the user is handled by the controller, which also manages data flow between the model and the view.

In our system, the user interacts with the React components, triggerring actions. The controllers in React handle these actions, update the state, and make API requests to the Django backend. The Django views receive the requests, interact with the models to perform necessary operations on the SQLite database, and return appropriate responses(eg. JSON). React components receive the responses, update the UI, and display the data to the user. MVC pattern ensures a clear separation of concerns and makes easier to reusability and maintainability aspect in the project.

3.9 Data Analytics and Visualization

Data analytics and visualization is used to represent the data in graphical formats or visual formats for making it more human-readable and understandable. Various charts, graphs, maps, and other visual representations are used to understand complex data and identify patterns and trends.

3.10 Clustering

Clustering is a Machine Learning technique whose aim is to group the data points having similar properties and/or features, while data points in different groups should have highly offbeat properties and/or features. In this project, title based author clustering is performed to group the faculty members with similar areas of expertise and research publications. For clustering, k-means algorithm is used.

3.11 k-means clustering

K-means is an unsupervised machine learning algorithm that groups the unlabeled dataset into k number of clusters. It is a centroid-based clustering algorithm, where we calculate the distance between each data point and a centroid to assign it to a cluster. It is an iterative process of assigning each data point to the groups and slowly data points get clustered based on similar features. This algorithm is bound to converge to a solution after some iterations. Goal: Partition data among some "K" number of clusters.

- Choose the number of clusters k.
- Select k random points from the data as centroids.
- Assign all the points to the closest cluster centroid.
- Recompute the centroids of newly formed clusters.
- Repeat the above steps until the algorithm converges.

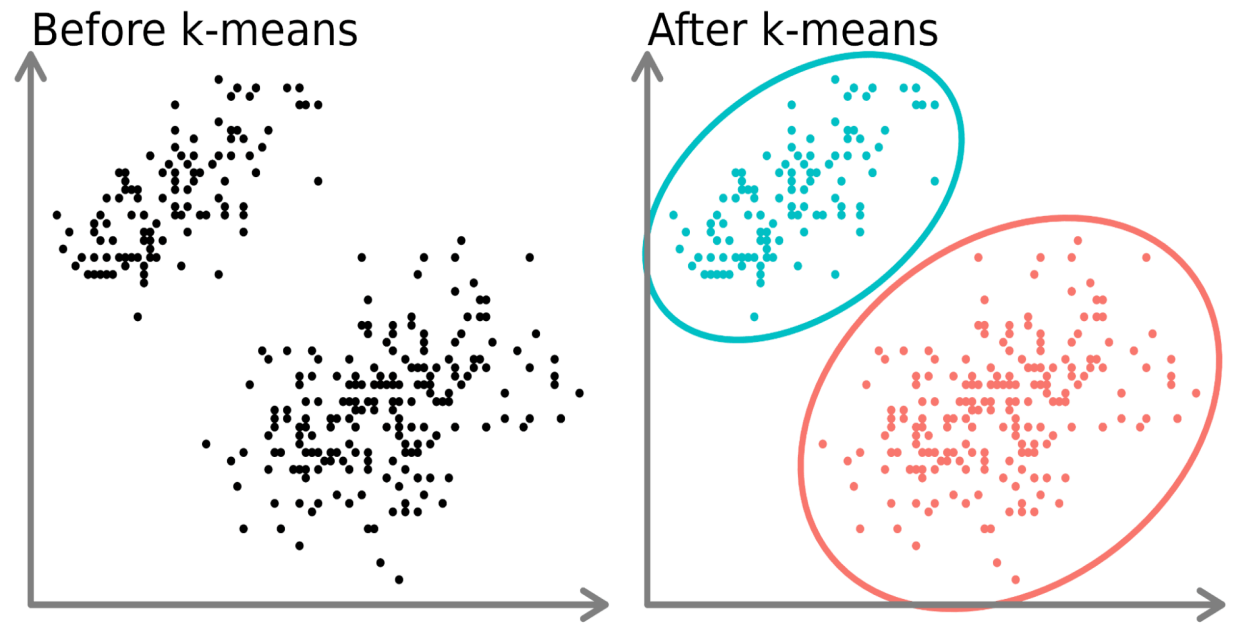


Figure 3.1: k-means Clustering
[12]

3.12 Elbow Method

In the k-means clustering algorithm, k clusters is randomly initialized. Before initializing these clusters, it is important to determine how many clusters are to be used. The Elbow Method is used to determine the optimal number of clusters in a dataset. In this method to determine the k-value, values of k is continuously iterated for k=1 to k=n (n is chosen as per the requirement). For every value of k, the within-cluster sum of squares (WCSS) value is calculated. Now, for determining the best number of clusters(k) a graph of k versus their WCSS value is plotted. When k=1, the WCSS has the highest value but with increasing k value WCSS value starts to decrease. The value of k from where the graph starts to look like a straight line is chosen.

3.13 Distilbert-base-nli-mean-tokens model of BERT

The "distilbert-base-nli-mean-tokens" model, which is a variant of the DistilBERT model, is a model architecture based on BERT framework with a smaller size and fast inference time. It has been used as a vectorizer to encode the text into numeric values.

3.14 LDA Model

The LDA (Latent Dirichlet Allocation) topic modeling algorithm follows the highest probability method to assign topics to documents and words. In LDA, each document is assumed

to be a mixture of topics, and each word in the document is assumed to be generated from one of the topics. This model takes input the pre-processed paper titles and it gives output the research topics extracted from paper titles.

4. Methodology

4.1 Requirements Gathering

The first stage of the methodology involved gathering requirements from the stakeholders, including teachers and staff members to understand the features and functionality required in the app. This involved discussions with he faculty members and administrators to gather feedback and suggestions.

The major requirements in this system were:

1. Record Keeping
2. Adding and Editing Publications
2. Searching and Sorting
3. User Profiling
4. Administrative Approval Management
5. Login and Security
6. Data Analysis and Visualization
7. Administrative Report Generation
8. Course Evaluation
9. Title based Author Clustering

4.2 System Design

Faculty Publication Management System consists of four different components namely, Scraping System, Faculty Evaluation System, Analytics and Visualization and Title Based Clustering as shown in figure.

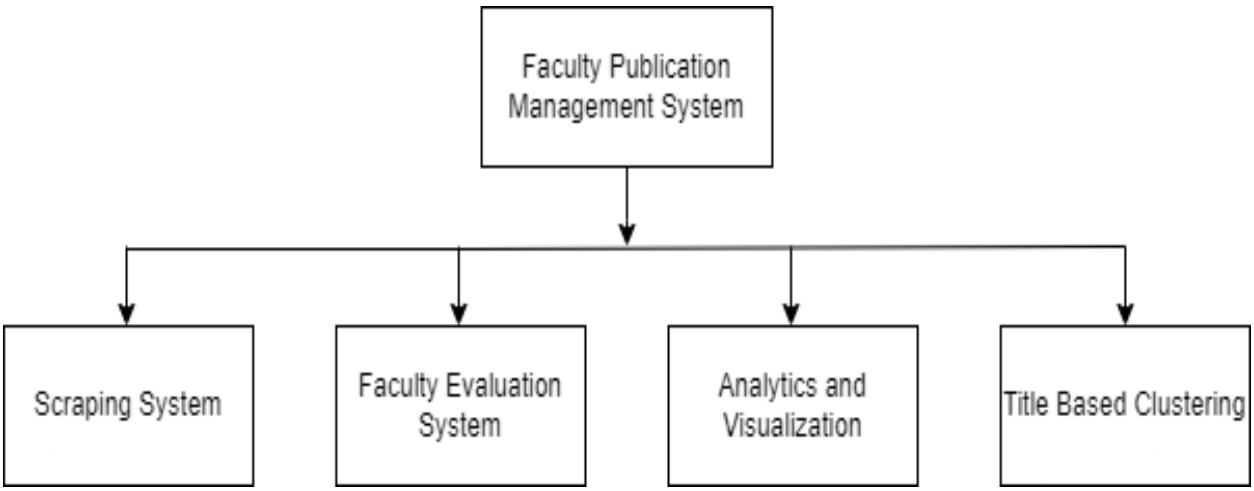


Figure 4.1: Main components of Faculty Publication Management System

4.2.1 Scraping System

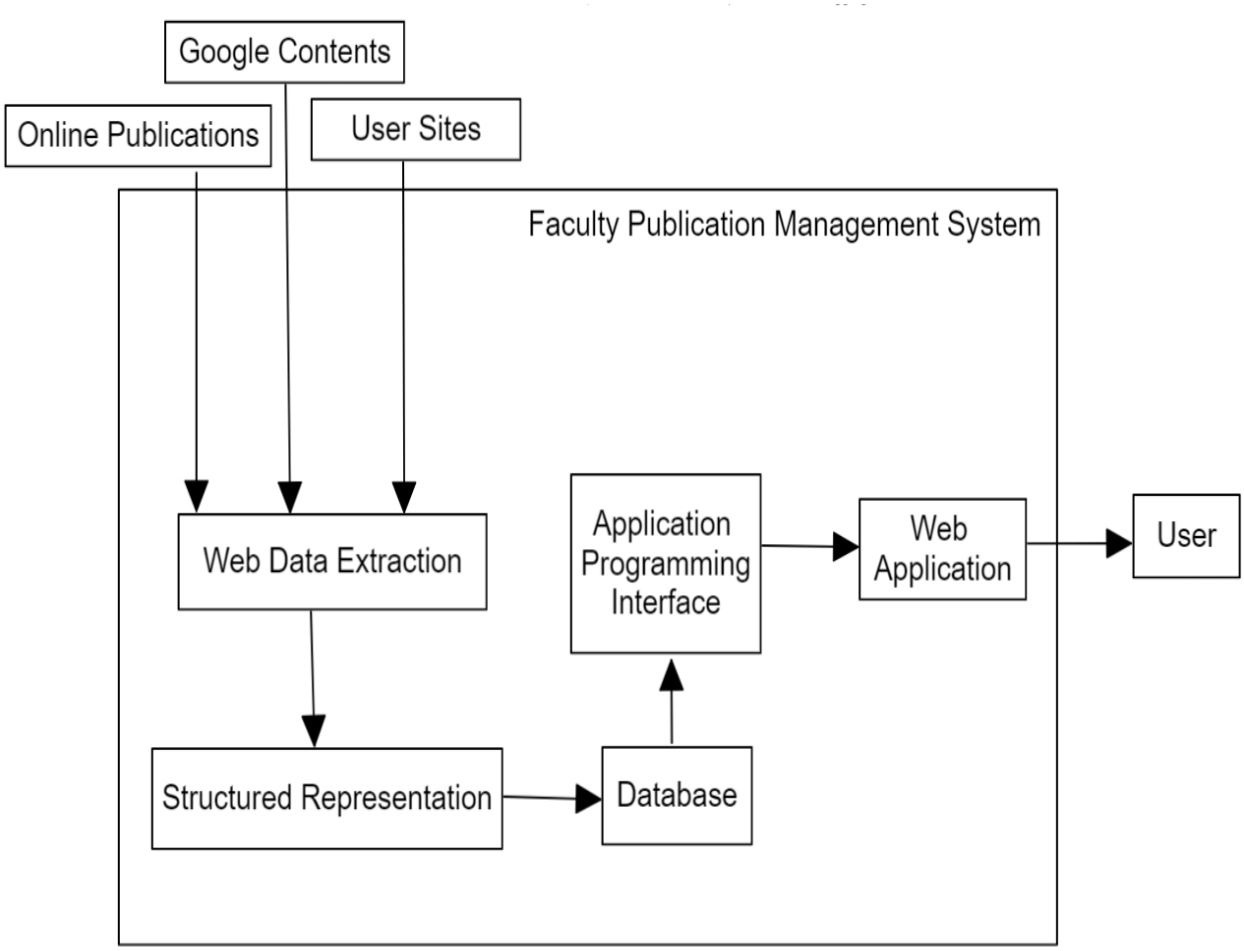


Figure 4.2: Block Diagram of Scraping System

Web Data Extraction

Previously existing publications are stored on publication sites like Google Scholar, Academia, Research-Gate, IEEE etc. Those research publications present in the Internet are scraped. Selenium is used for the web scraping of the research publications. These publication information are extracted from such platforms and fetched into our system.

Structured Representation

In this step, organizing and representing data in a consistent manner is done so that it can be easily understood and manipulated by both human and machines. The contents and information up to this stage, i.e. data obtained after web scraping are unstructured web contents. For these contents to be processed further, they are to be converted into structured representation. The structured representation also allows these contents to be stored in a database. For ease, the scraped data contents are converted and stored in JSON format.

Database

The database is for storing the web extracted components as a form of record management and storage. The database contents will be updated based on the information fetched from the web extraction. The data from web collected are stored in a JSON file which is then stored in the centralized database in a structured and organized way. Not only the scraped data, but the user information are also stored in the database. In this application, Sqlite database is used to store texts, images and scraped data of the users.

Application Programming Interface

An Application Programming Interface (API) is a set of defined rules that enable different applications to communicate with each other and share data. In this application, API transfers the data between front-end and back-end Django server.

Web Application

The web application in this context is mainly concerned with the back-end of the web application. The API provide an interface from the database to the back-end (server) application. The back-end application will provide all the services for any integrated front-end application. The services includes providing stored information, providing user information, updating information, deleting information, etc. These services are executed by the user.

4.2.2 Faculty Evaluation System

For evaluation of faculty members, the faculty evaluation system is integrated where the publications details are obtained from the system itself, course evaluation is performed and

other metrics like number of teaching year experience are analyzed. The course evaluation is done from the data collected from the students through course evaluation form integrated in the system.

The faculty evaluation system is based on 3 criteria. Different target scores for different areas of activity were assumed for experimental purposes without any particular reference.

Table 4.1: Criteria of Faculty Evaluation System

Areas of Activity	**Targeted Score**
Research	100
Teaching	35
Year of Service	25
Total	160

Research

Under Research category , the evaluation of the faculty is based on the citation of the published papers and the platform where paper is published. From entire paper collection in our system, the top 10 publishing platform were picked and each paper was given specific point based on ranking of the platform. The point assigned for individual publishing platform and citation number were assumed by ourselves for random experimental purpose with no any particular reference.

Table 4.2: Evaluation based on Publisher

Publisher	**Point assigned to each paper**
IEEE, Elsevier, Springer	1
SAGE, IADIS	0.9
Oxford University Press, Cambridge University Press	0.8
AJER,World Scientific Publishing , Taylor and Francis	0.7
Others	0.5

Table 4.3: Evaluation based on citation

Citation Number	**Point assigned to each paper**
150 and above	1
100 to 149	0.9
80 to 99	0.8
60 to 79	0.7
40 to 59	0.6
20 to 39	0.5
below 20	0.4

For evaluation of faculty, the points to be assigned to each paper was chosen according to the number of citations for experimental purposes without any particular reference. The research score was calculated as the weighted average of total citation of each paper and rank of paper publisher of each paper.

Teaching

Under Teaching Criteria, the evaluation of the faculty is based on the result of the student feedback form. The overall response from the feedback form is aggregated. The form consist of 7 questions and each question carries 5 marks. Each time when students submits the feedback form, the value is averaged and stored.

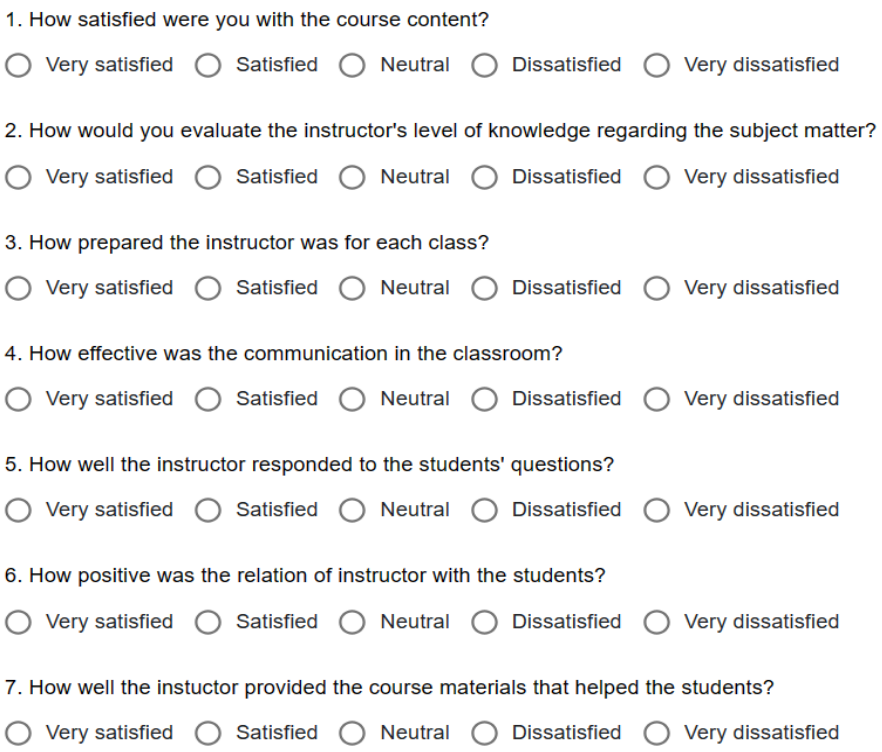


Figure 4.3: Evaluation based on feedback form

Year of Service

Year of Service carries 25 marks. The faculty working tenure in the university is considered in Year of Service criteria. The point is based on the fact that the faculty with higher working year in the university is given higher points.

4.2.3 Analytics and Visualization

Scraped data is visualized using Recharts library which is one of a popular open-source charting library for React. This library provides a wide range of chart types. We have used bar charts, pie charts and line charts to visualize the scraped publication data to get insights in easily understandable form.

4.2.4 Title Based Author Clustering

Title-based author clustering is done to group together academic papers and authors based on similarity index of their titles.

The basic idea is to extract the titles from a set of papers, preprocess the text and then use clustering algorithm to group papers with similar titles into clusters.This approach can be useful for various purposes, such as identifying research areas or communities of scholars, and also detecting emerging trends.

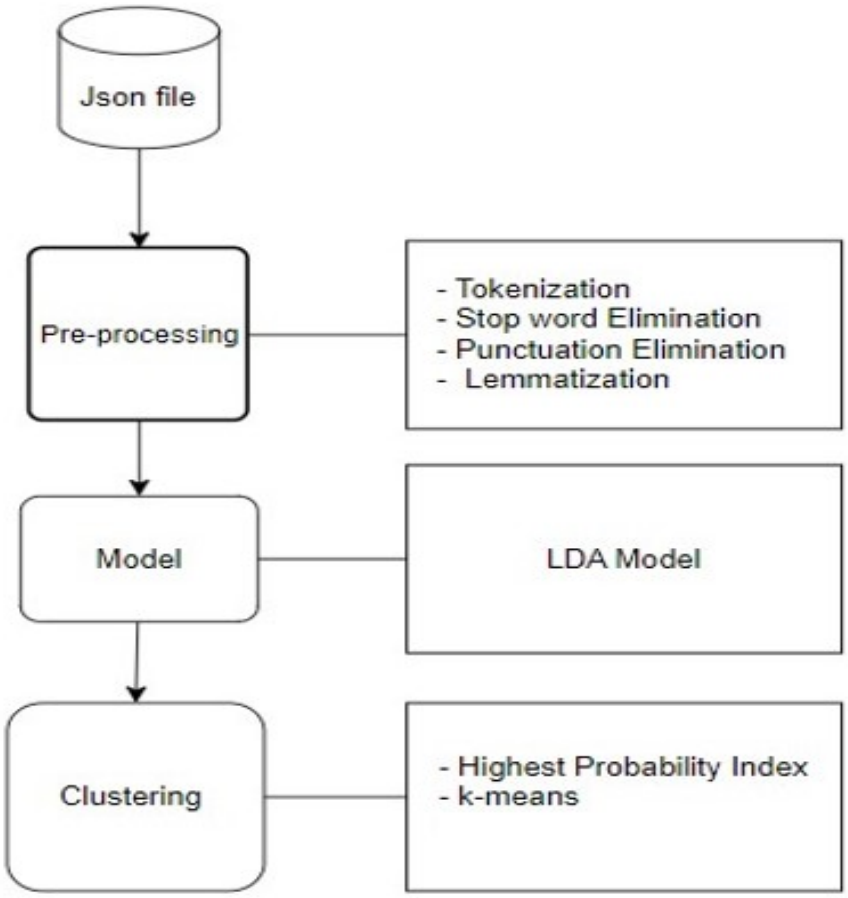


Figure 4.4: Flowchart of getting Research Topics from each cluster

A. Preprocessing phase

Preprocessing is an essential step in Natural Language Processing (NLP) that involves cleaning and transforming raw text data into a format that can be easily analyzed by algorithms. This step plays a crucial role in improving the quality and usefulness of NLP applications. Various steps involved in this phase are:

1. Tokenization
2. Stopword Elimination
3. Punctuation Elimination
4. Lemmatization

1. Tokenization

Tokenization is a preprocessing step, which breaks unstructured data into chunks of information in lower case that can be considered as tokens which is performed using nltk.tokenize

and word-tokenize.

2. Stop words Elimination

Stop words elimination is one of the common preprocessing technique. Stop words usually refers to the most common words in a language, which usually does not bring additional meaning. It involved removing frequently occurring words that carry little semantic meaning, such as "and," "the," and "a." This helped in reducing noise in the data and improving the accuracy of downstream tasks.

3. Punctuation Elimination

Punctuation elimination is another important preprocessing step when punctuation does not brings additional value for text vectorization. It involved removing punctuation marks like commas, periods, and question marks. This helped to standardize the input and simplify the processing of text data. Punctuation elimination was performed using nltk.stopwords and nltk.wordnet.

4. Lemmatization

Lemmatization reduces the inflected words properly ensuring that the root word belongs to the language. In Lemmatization root word is called Lemma. Lemmatization is performed by using nltk.stem and WordNetLemmatizer.

B. k-means Clustering

The main reason of implementing clustering is to group researcher having similar interest based on their previous publications. So, the clustering of the research publications titles is performed first by using k-means clustering algorithm. For now clustering is implemented on Department of Civil Engineering.

Firstly, the SentenceTransformer library is used to encode the pre-processed titles of research publications using the pre-trained 'distilbert-base-nli-mean-tokens' model and a matrix is obtained where each row represents the embedding of a sentence from the corpus which is the title of research publications. These embeddings are then used for k-means clustering. The value of k is selected as 5 from the elbow method, from the total of 845 papers. After the implementation of k-means clustering, 5 different clusters of titles are formed and are extracted in the separate json files.

C. LDA Model

We have used Latent Dirichlet Allocation (LDA) model for topic modeling. It is a probabilistic model used for topic modeling in natural language processing. It takes input the pre-processed paper titles and it gives output the research topics extracted from paper titles. The number of topics is tuned to 5 and alpha is tuned to 'auto' which is the concentration parameter. For every clusters of titles, the research topics are extracted by using this model.

D. Mapping of faculty members to the cluster

To map the faculty members to the cluster, firstly, every researcher's paper is mapped with the cluster name. The corresponding paper count based on similarity of cluster name and the paper title is evaluated and the cluster name with the highest count is mapped to respective researcher.

	full_name	topic_number	paper_frequency
105	Suraj Lamichhane	1	6
106	Suraj Lamichhane	2	3
107	Suraj Lamichhane	3	5
108	Suraj Lamichhane	4	5
109	Vishnu Prasad Pandey	0	7
110	Vishnu Prasad Pandey	1	35
111	Vishnu Prasad Pandey	2	58
112	Vishnu Prasad Pandey	3	22
113	Vishnu Prasad Pandey	4	15

Figure 4.5: Mapping of cluster name and author

Here, faculty member Suraj Lamichhane has total 19 paper published. Out of 19 papers, 6 papers falls under the cluster number 1, 3 papers falls under the cluster number 2, and 5 papers each falls under cluster number 3 and 4. Since out of 19 papers of Suraj Lamichhane, cluster 1 seems to be more dominant. So, Suraj Lamichhane falls under cluster 1.
Finally, in the similar manner all the faculty members are mapped to the respective cluster.

5. Tools and Technologies

5.1 Frontend

5.1.1 JavaScript (JS)

Javascript(JS) is a very popular programming language for web development. All of the frontend is programmed using JS written framework called React. Some of the web scraping part is also done using Selenium library which is also written in JS.

5.1.2 React

React is a very popular frontend framework written in JS for developing web applications. React is used for the entire state management in the frontend.

5.1.3 Bootstrap and Material UI

Bootstrap is a CSS framework used in web programming for designing the visual aspects of the web app. Bootstrap is used along with Material UI, which is a library that provides all the graphical aspects of a web app including icons, svg, images, etc.

5.2 Backend

5.2.1 Python

Python is a high-level programming language for general purpose. Most of the scraping part is done using Selenium library, primarily written in python. All of the backend is programmed using python written framework called Django.

5.2.2 Django

Django is a scalable web programming backend framework. Django strictly follows an MVC architectural pattern which is a very well tested pattern for web development.

5.3 Database

5.3.1 Structured Query Language (SQL)

Structured Query Language (SQL) is a standarized querying language used to manage relational databases and perform various operations like querying, inserting, updating, etc on those databases.

5.3.2 SQLite

SQLite is a small, fast and highly reliable SQL database engine. SQLite is used for storing and retrieving all the research publications data in the system. Relational schemas are designed and implemented for storing various types of structured data. These schemas are normalized to 1NF (1st Normal Form).

5.4 Web Scraping

5.4.1 Selenium

Selenium is an open source scraping tool aimed at supporting browser automation. It provides a playback tool for authoring functional tests across most modern web browsers, without the need to learn a test scripting language. Although it is a testing tool, it can be used for web scraping and web crawling purposes. Selenium uses Chrome webdrivers for its automation tasks.

6. Results

The result includes the comparison of research and publications contribution of various departments upto 2022. In addition, publications data upto May, 2023 is already collected but this analysis only includes the data upto 2022. The paper were extracted from limited sites only which includes Google Scholar, ResearchGate, IEEE Xplore. All the faculty names from each department were collected and researched about their publications from the sites mentioned above. It was found that some faculty member have not published on those sites and some faculty did not have any account on those sites. Civil department had total of 33 faculty members where 3 members did not have any paper published. Architecture department had total of 19 faculty members where 9 members did not have any paper published. Electronics and Computer department had total of 22 faculty members where 6 members did not have any paper published. Mechanical department had total of 19 faculty members where 3 members did not have any paper published. Electrical department had total of 9 faculty members where 2 members did not have any paper published. Applied Science and Chemical department had total of 18 faculty members where 5 members did not have any paper published.

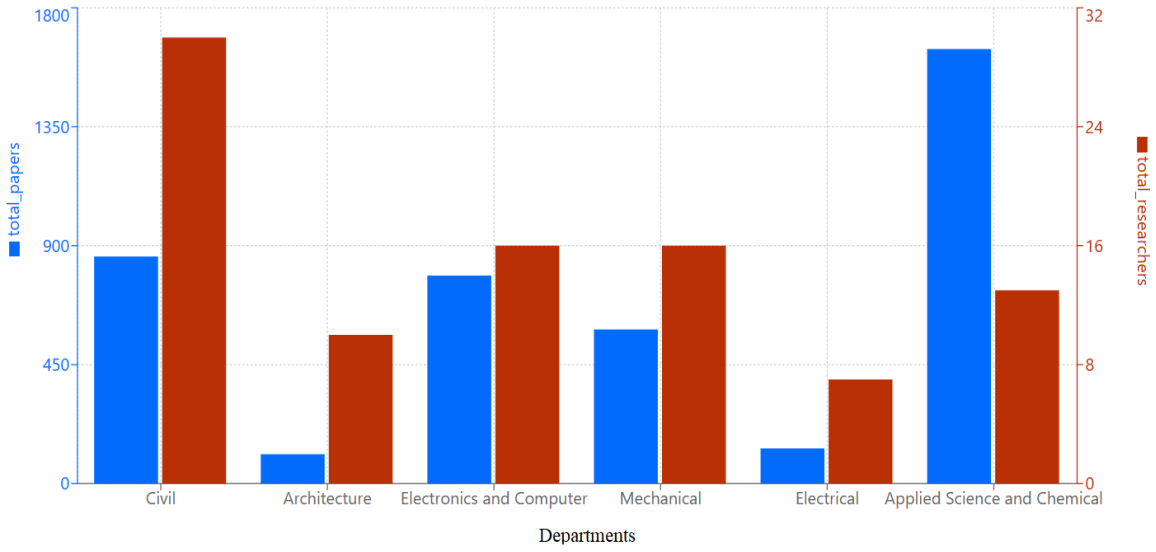


Figure 6.1: Department Researchers and Publications count

As of 2022 A.D., it can be seen from the figure 6.1 that the highest number of researchers is from Department of Civil Engineering with 30 and highest number of paper published is from Department of Applied Science and Chemical Engineering with around 1644 papers.

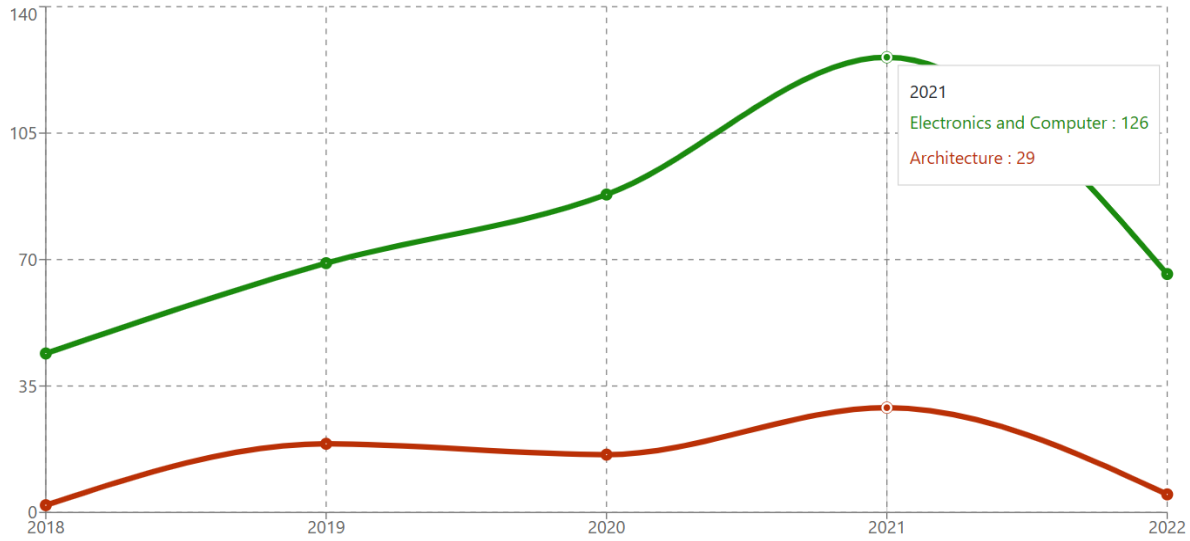


Figure 6.2: Comparison of paper published between Department of Electronics and Computer Engineering and Department of Architecture

The figure 6.2. shows the comparison of paper published between Department of Electronics and Computer Engineering and Department of Architecture. Department of Architecture has around 10 members involved in research and publications while Department of Electronics and Computer Engineering has around 16 members involved in research and publications. As of 2022 A.D. Department of Electronics and Computer Engineering has published around 787 papers while Department of Architecture has published around 111 papers.

The figure 6.3 shows the comparison of paper published between Department of Electronics and Computer Engineering and Department of Civil Engineering. Department of Civil Engineering has around 30 members involved in research and publications while Department of Electronics and Computer Engineering has around 16 members involved in research and publications. As of 2022 A.D. Department of Electronics and Computer Engineering has published around 787 papers while Department of Civil Engineering has published around 859 papers.

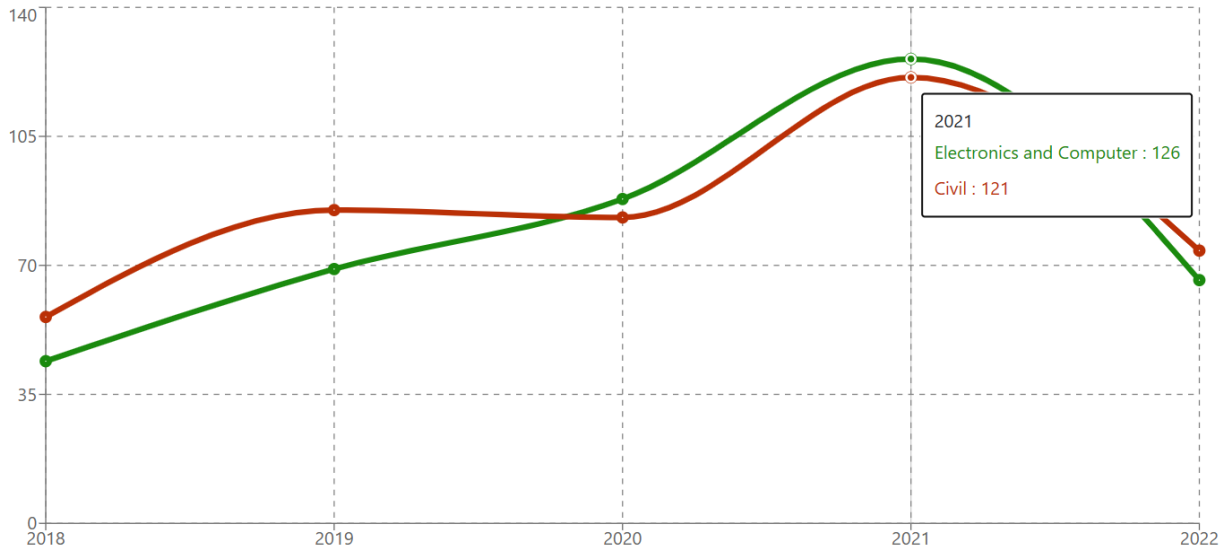


Figure 6.3: Comparison of paper published between Department of Electronics and Computer Engineering and Department of Civil Engineering

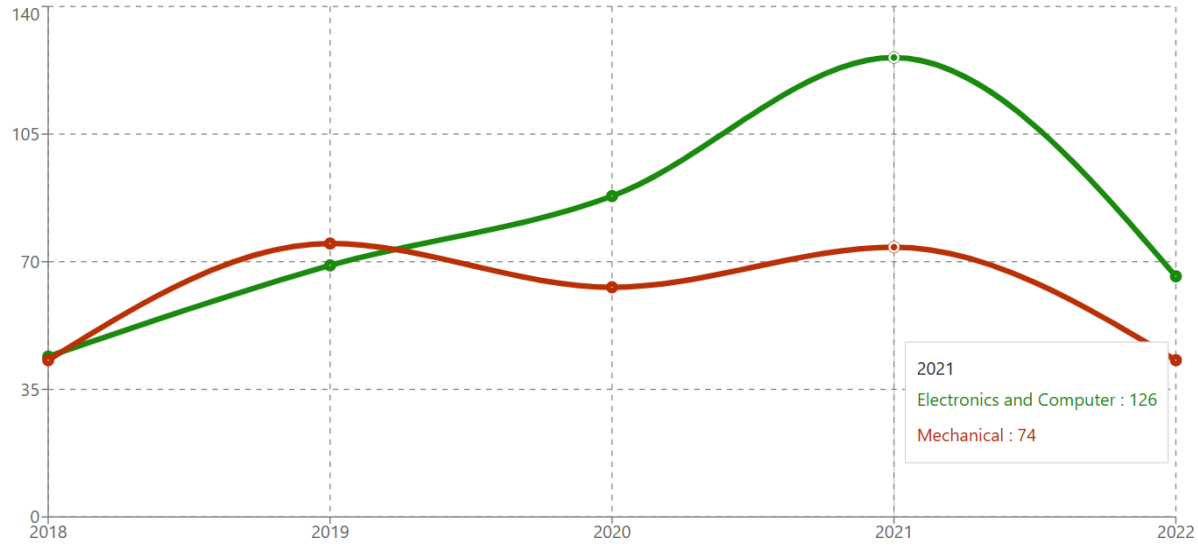


Figure 6.4: Comparison of paper published between Department of Electronics and Computer Engineering and Department of Mechanical Engineering

The figure 6.4 shows the comparison of paper published between Department of Electronics and Computer Engineering and Department of Mechanical Engineering. Department of Mechanical Engineering and Department of Electronics and Computer Engineering both has around 16 members involved in research and publications. As of 2022 A.D. Department of Electronics and Computer Engineering has published around 787 papers while Department

of Mechanical Engineering has published around 583 papers.

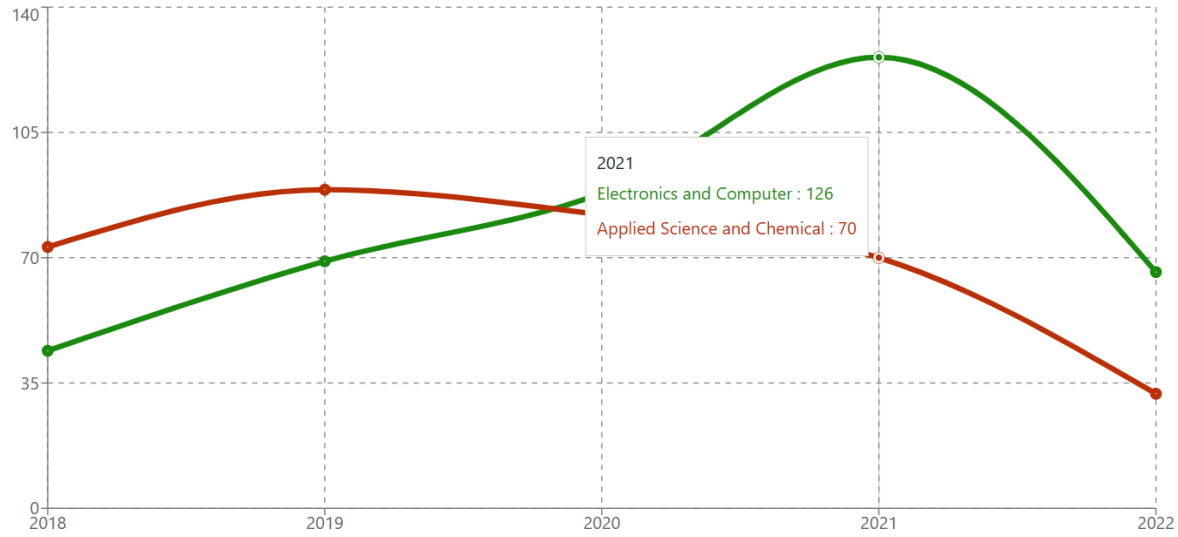


Figure 6.5: Comparison of paper published between Department of Electronics and Computer Engineering and Department of Applied Science and Chemical Engineering

The figure 6.5 shows the comparison of paper published between Department of Electronics and Computer Engineering and Department of Applied Science and Chemical Engineering. Department of Applied Science and Chemical Engineering has around 13 members involved in research and publications while Department of Electronics and Computer Engineering has around 16 members involved in research and publications. As of 2022 A.D. Department of Electronics and Computer Engineering has published around 787 papers while Department of Applied Science and Chemical Engineering has published around 1644 papers.

The figure 6.6 shows the comparison of paper published between Department of Electronics and Computer Engineering and Department of Electrical Engineering. Department of Electrical Engineering has around 7 members involved in research and publications while Department of Electronics and Computer Engineering has around 16 members involved in research and publications. As of 2022 A.D. Department of Electronics and Computer Engineering has published around 787 papers while Department of Electrical Engineering has published around 133 papers.

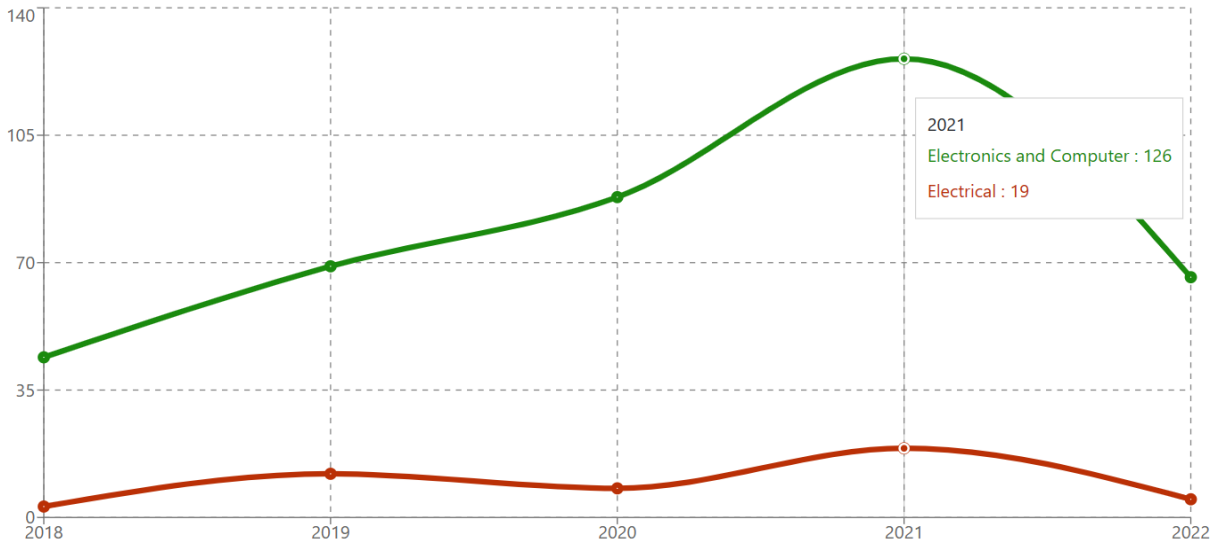


Figure 6.6: Comparison of paper published between Department of Electronics and Computer Engineering and Department of Electrical Engineering

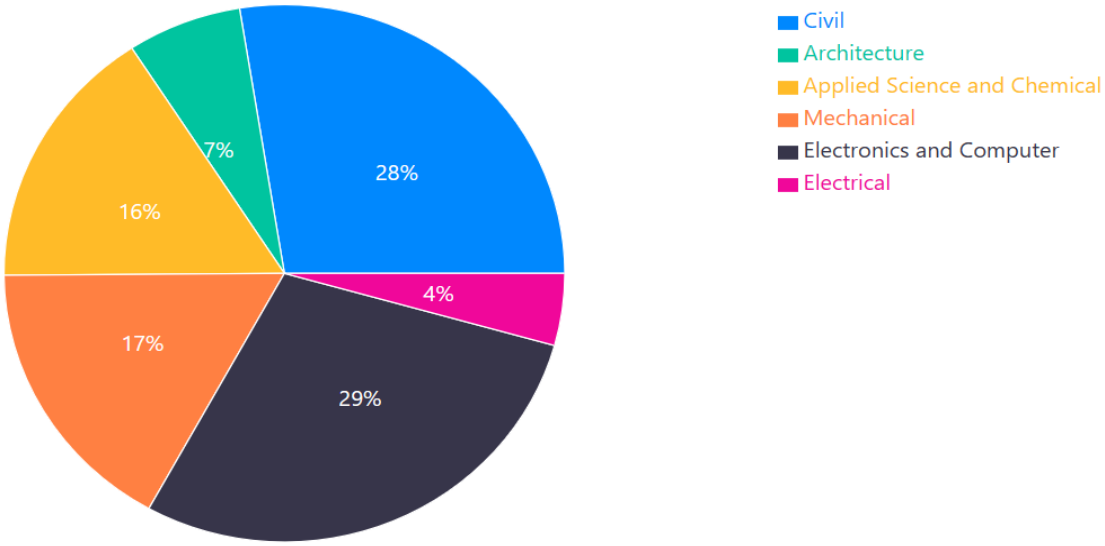
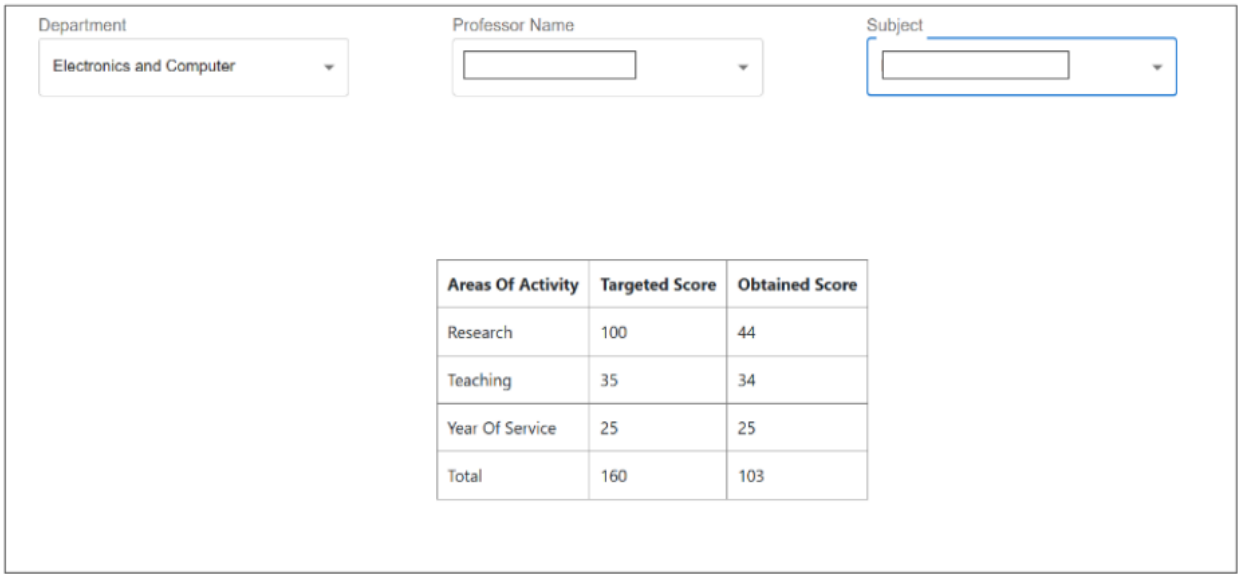


Figure 6.7: Total Publications contribution among all departments

In the year 2022 A.D., all-together 296 papers were published. From the Pie chart in figure 6.7, it can be seen that Department of Electronics and Computer Engineering along with Department of Civil Engineering has higher contribution in research and publications with 29 and 28 percent respectively while Department of Electrical Engineering has least

contribution with 4 percent.



Areas Of Activity	Targeted Score	Obtained Score
Research	100	44
Teaching	35	34
Year Of Service	25	25
Total	160	103

Figure 6.8: Faculty Evaluation Sample

One of the sample of the evaluation of the faculty members is as shown in figure 6.8. The faculty obtained Research score of 44 based on the paper published and its citation number. The faculty obtained Teaching score of 34 based on the response of the feedback form filled by the students. The Faculty obtained Year of Service score of 25 based on the number of years the faculty has been involved in the university.

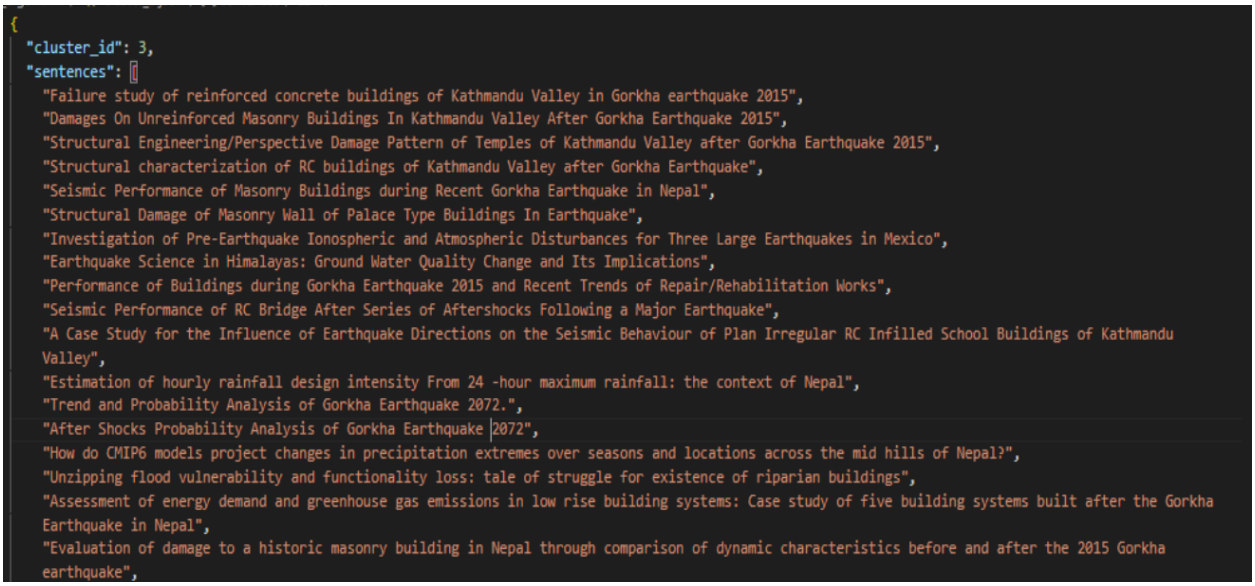


Figure 6.9: Title samples of Cluster 3

Using K-means clustering, the titles were divided into 5 clusters. The sample of cluster 3 is as shown in figure 6.9.
From the cluster 3 sample, the research topics were extracted using the LDA method as shown in the figure 6.10.

```
Research Topic of cluster 3: earthquake, landslide, building, analysis, study
```

Figure 6.10: Research Topics from Cluster 3

Similarly, from every cluster the research topics were extracted. The figure 6.11 shows the research topics obtained from each cluster.

```
{ 0: ['analysis', 'hazard', 'study', 'road' ,'case'],
  1:['masonry', 'concrete', 'effect', 'reinforce','building'],
  2: ['water', 'climate', 'groundwater','change', 'use'],
  3 : ['earthquake','landslide','building','study', 'scale'],
  4: ['soil', 'slope', 'use case', 'study', 'effect']
   }
```

Figure 6.11: Research Topics of every cluster

Below is the list of Faculty members with their top research interest topic.

full_name	topic	paper_frequency
Akarsha Khawas	['analysis', 'hazard', 'study', 'road', 'case']	1
Anil Marsani	['masonry', 'concrete', 'effect', 'reinforce', 'building']	8
Basanta Raj Adhikari	['masonry', 'concrete', 'effect', 'reinforce', 'building']	52
Bharat Mandal	['analysis', 'hazard', 'study', 'road', 'case']	5
Bharat Raj Pahari	['analysis', 'hazard', 'study', 'road', 'case']	7
Bhim Kumar Dahal	['water', 'climate', 'groundwater', 'change', 'use']	5
Bhola Nath Sharma Ghimire	['earthquake', 'landslide', 'building', 'study', 'scale']	16
Devi Prasad Bhattarai	['earthquake', 'landslide', 'building', 'study', 'scale']	3
Gokarna Bahadur Motra	['analysis', 'hazard', 'study', 'road', 'case']	14
Hari Darsan Shrestha	['water', 'climate', 'groundwater', 'change', 'use']	3
Hari Ram Parajuli	['analysis', 'hazard', 'study', 'road', 'case']	15
Indra Prasad Acharya	['masonry', 'concrete', 'effect', 'reinforce', 'building']	20
Jagat Kumar Shrestha	['soil', 'slope', 'use case', 'study', 'effect']	17
Kshitij Shrestha	['soil', 'slope', 'use case', 'study', 'effect']	28
Mahendra Raj Dhital	['earthquake', 'landslide', 'building', 'study', 'scale']	1
Mukesh Raj Kafle	['water', 'climate', 'groundwater', 'change', 'use']	6
Narayan Prasad Gautam	['soil', 'slope', 'use case', 'study', 'effect']	2
Pawan Kumar Bhattarai	['water', 'climate', 'groundwater', 'change', 'use']	7
Pradeep Kumar Shrestha	['masonry', 'concrete', 'effect', 'reinforce', 'building']	2
Prem Chandra Jha	['analysis', 'hazard', 'study', 'road', 'case']	1
Rajan Suwal	['analysis', 'hazard', 'study', 'road', 'case']	14
Ram Chandra Tiwari	['analysis', 'hazard', 'study', 'road', 'case']	11
Santosh Kumar Shrestha	['water', 'climate', 'groundwater', 'change', 'use']	7
Santosh Kumar Yadav	['earthquake', 'landslide', 'building', 'study', 'scale']	5
Saraswati Thapa	['water', 'climate', 'groundwater', 'change', 'use']	7
Shubh Narayan Pathak	['soil', 'slope', 'use case', 'study', 'effect']	1
Shukra Raj Paudel	['water', 'climate', 'groundwater', 'change', 'use']	14
Sunita Ghimire	['analysis', 'hazard', 'study', 'road', 'case']	2
Suraj Lamichhane	['masonry', 'concrete', 'effect', 'reinforce', 'building']	6
Vishnu Prasad Pandey	['water', 'climate', 'groundwater', 'change', 'use']	58

Figure 6.12: Title based Author Clustering

7. Epilogue

7.1 Conclusion

In conclusion, the Faculty Publication Management System has been developed and implemented to keep track of faculty members' research publications for record keeping. It provides a unified platform for automated gathering, managing and tracking the research publications of the faculty members published in major publishing platforms i.e. IEEE, Google Scholar and ResearchGate which reduces the administrative burden by replacing manual work. Using this system, one can view and search the research publications and also generate the report of all the publications. Additionally, the system provides valuable insights about the faculty publication from analysis and visualization of data of the research publications from centralized database. By incorporating various metrics such as course evaluation, research output from the system itself, and year of services, the Faculty evaluation system provides a comprehensive assessment of faculty performance and help identify areas for improvement. With all the three objectives have been achieved, the Faculty Publication Management System project is a valuable contribution to the universities in Nepal, and it has the potential to ease the faculty members and students by providing them with easy access to information about their research publications and a way to improve themselves from evaluation system.

7.2 Limitations and Future Enhancements

Some of the limitations of our project and how this project will be enhanced in the future to deal with them:

1. Right now, the application is limited to web browsers. This application can be extended to mobile applications as well.

2. Only objective questions are present in the evaluation form filled by students for the evaluation of faculty performance. Text fields can be added in the evaluation form which can be used for text analysis to mine valuable insights.

3. Clustering of authors is performed only of the Department of Civil Engineering Department, Pulchowk Campus. Furthermore, author clustering of other departments can be performed.

References

[1] Ja-Ryoung Choi and Soon-Bum Lim. Web scraping model based on transclusion in e-learning environment. pages 62–65, 08 2014.

[2] University of California, Irvine. UC Publication Management System. https://www.lib.uci.edu/dss/uc-publication-management-system. Accessed on 19th June 2023.

[3] Sara Tabaei, Yitzchak Schaffer, Gregory McMurray, and Bashe Simon. Building a faculty publications database: A case study. *Public Services Quarterly*, 9(3):196–209, 2013.

[4] Anand V Saurkar, Kedar G Pathare, and Shweta A Gode. An overview on web scraping techniques and tools. *International Journal on Future Revolution in Computer Science & Communication Engineering*, 4(4):363–367, 2018.

[5] T Karthikeyan, Karthik Sekaran, D Ranjith, JM Balajee, et al. Personalized content extraction and text classification using effective web scraping techniques. *International Journal of Web Portals (IJWP)*, 11(2):41–52, 2019.

[6] MA Vanette Schwartz MLS and MUP Bruce Stoffel MLS. Building an online faculty publications database. *College & Undergraduate Libraries*, 14(3):1–25, 2007.

[7] Michelle Dennis, James Halbert, Donna DiMatteo-Gibson, Chuks Agada, and Sarah Fornero. Implementation of a faculty evaluation model. *Journal of Leadership, Accountability and Ethics*, 17:30–35, 11 2020.

[8] Samita Bai, Sharaf Hussain, Quratulain Rajput, and Shakeel Khoja. Faculty performance evaluation system: An ontological approach. volume 2015, 11 2014.

[9] C. J. Bland, L. Wersal, W. VanLoy, and W. Jacott. Evaluating faculty performance: A systematically designed and assessed approach. *Academic Medicine: Journal of the Association of American Medical Colleges*, 77(1):15–30, 2002.

[10] Rodelio Fuente and Romel Fuente. Faculty performance evaluation system with application of data analytics. *South Asian Journal of Engineering and Technology*, 12:24–31, 03 2022.

[11] Hossein Didehban, Azim Mirzazadeh, and Hamid Reza Khankeh. The perspective of medical faculty members about the challenges of faculties' evaluation system in iranian context: A qualitative study. 11:47–53, 03 2020.

[12] Kevin Babitz. Introduction to k-means clustering with scikit-learn in python, Jul 2018.

8. Appendix

8.1 Screenshots

8.1.1 Frontend UI

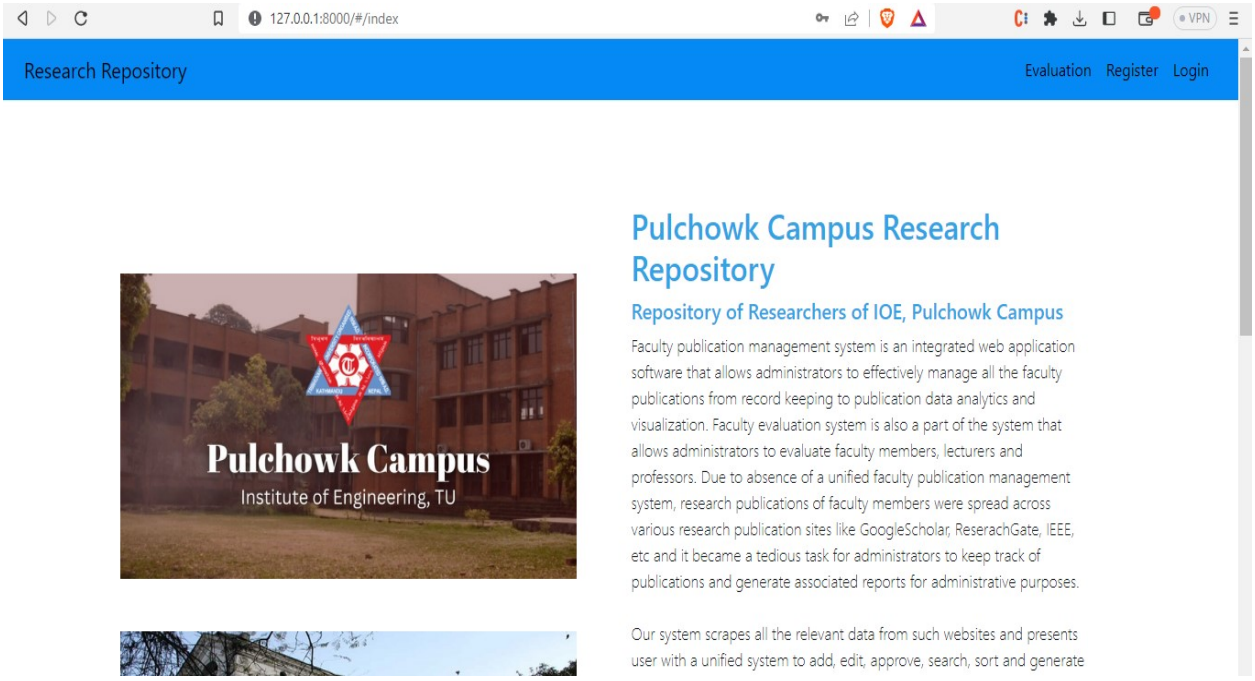


Figure 8.1: Home Page

Creating Account

In order to upload any resources, the faculty members have to first create the account. This account creating feature helps the user to create an account. After creating account, the user can upload the resources and the information are then shown in the respective accounts.

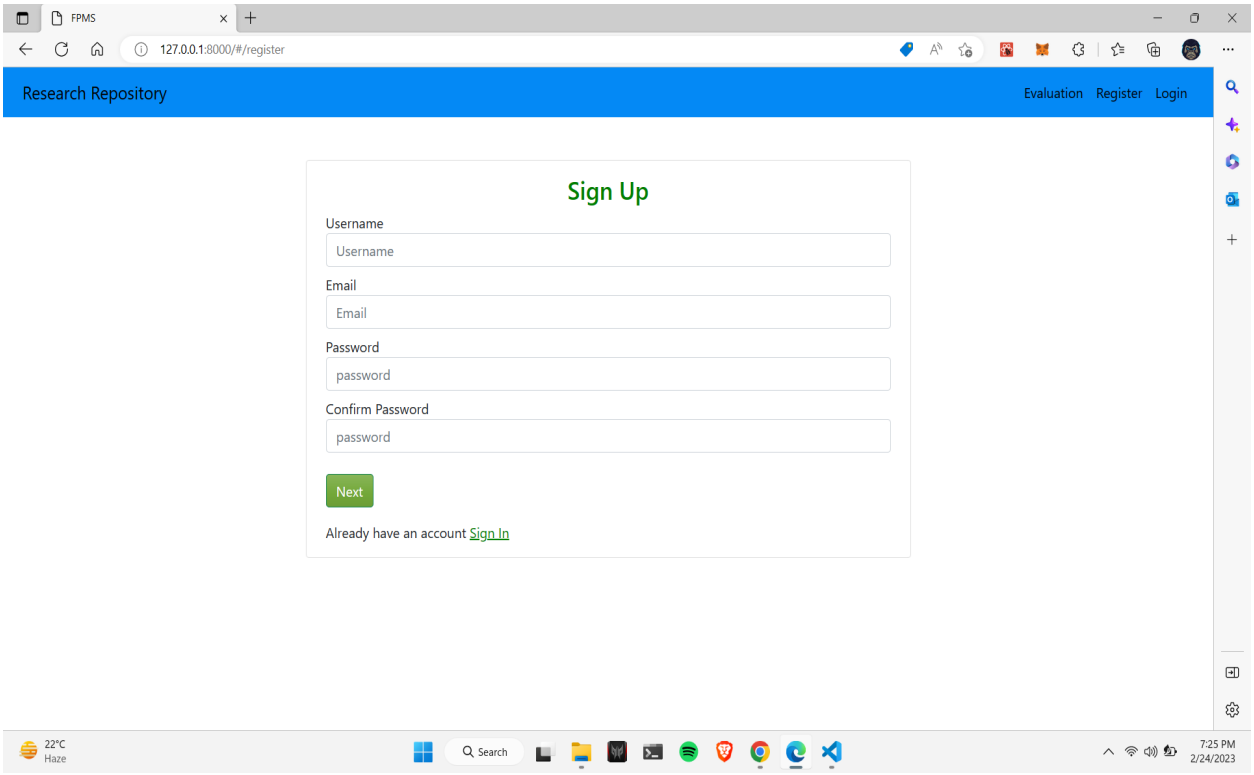


Figure 8.2: Account Registration

Login Feature

This is the login feature of the system. To upload the resources by the respective faculty members, one must login to the system.

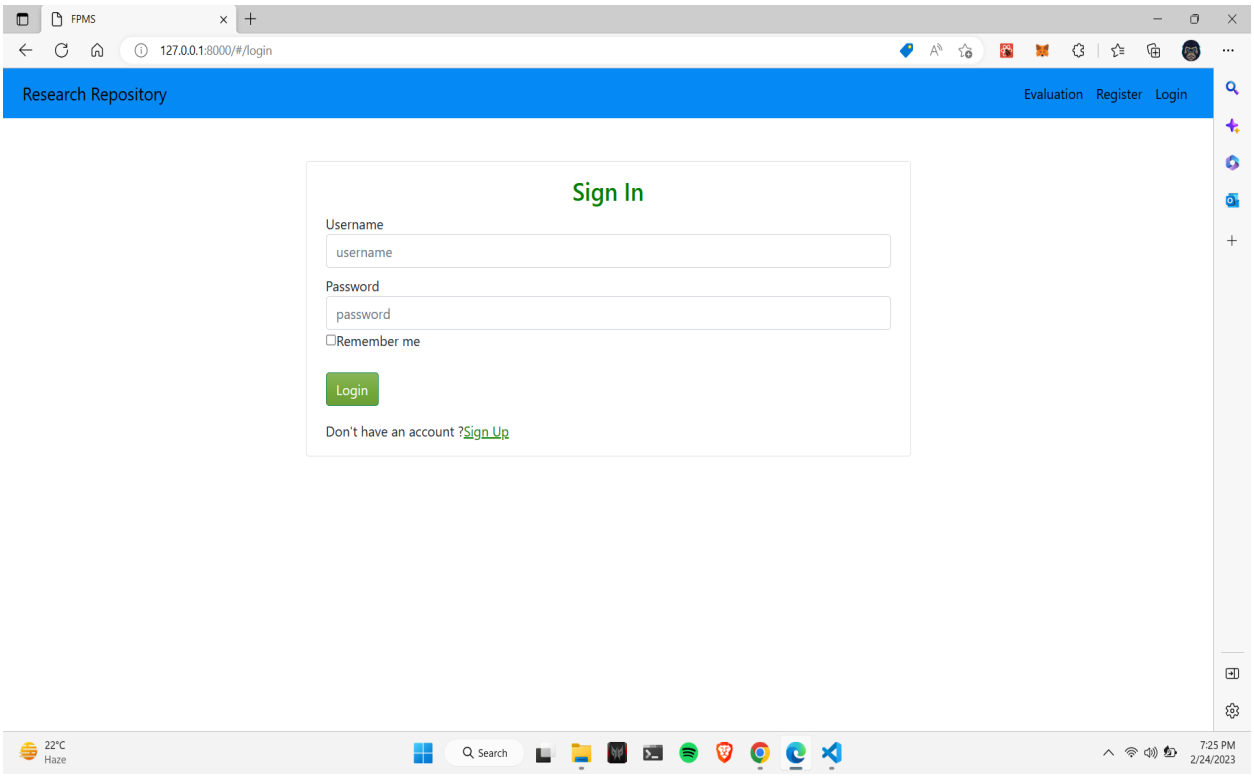


Figure 8.3: Login Feature

Record Management UI

This system provides a standard way for the record management of the publications of the faculty members. The publications which are uploaded by the faculty members after creating their respective account in this system can be viewed at a single place. This features helps to manage the records in standard order.

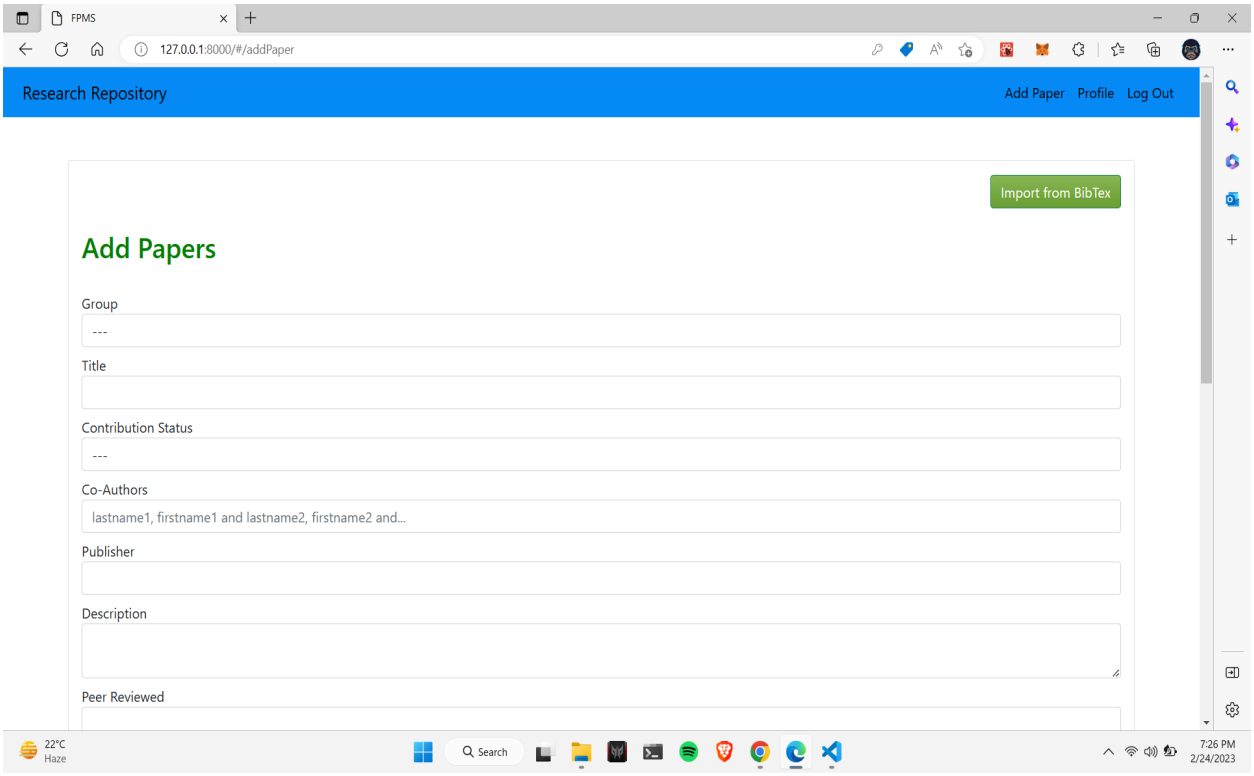


Figure 8.4: Add Paper Feature

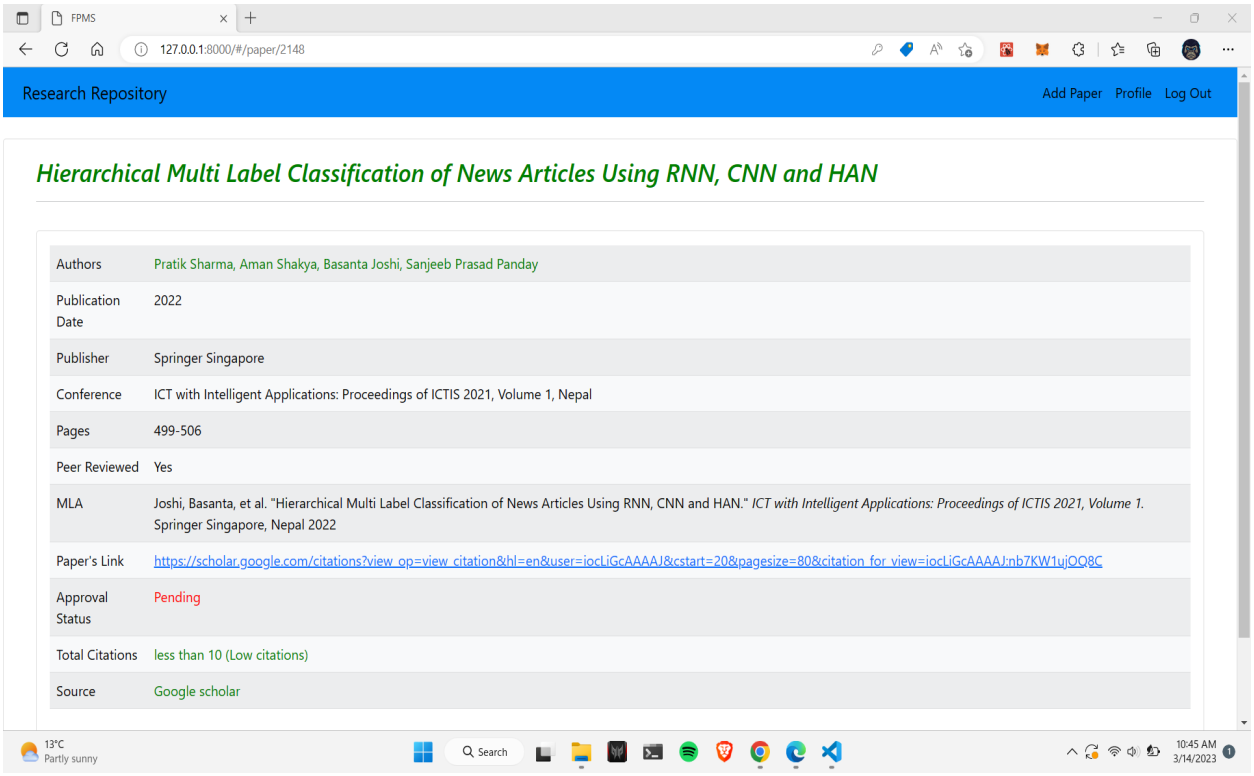


Figure 8.5: Paper Detail

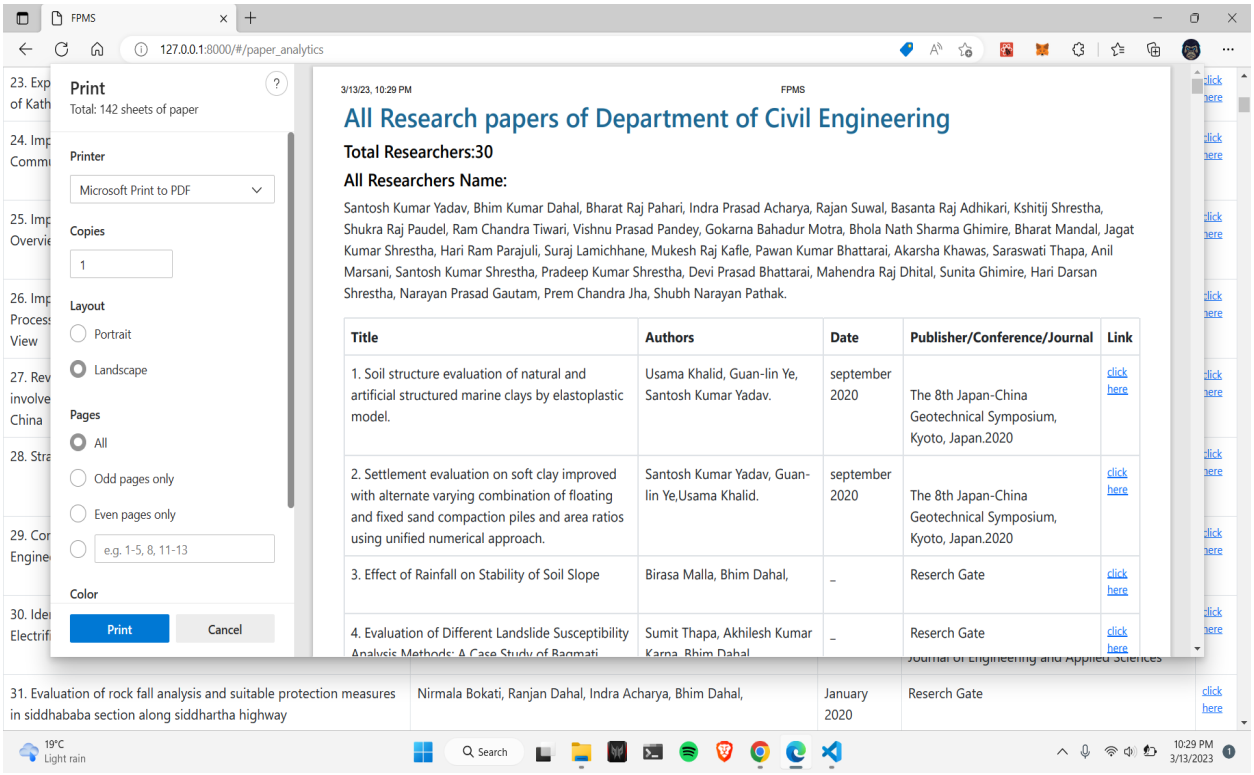


Figure 8.6: Publication Listing

Uploading and Downloading Feature

The user is able to upload the resource only after signing in and downloading feature is available for every user who uses the application.

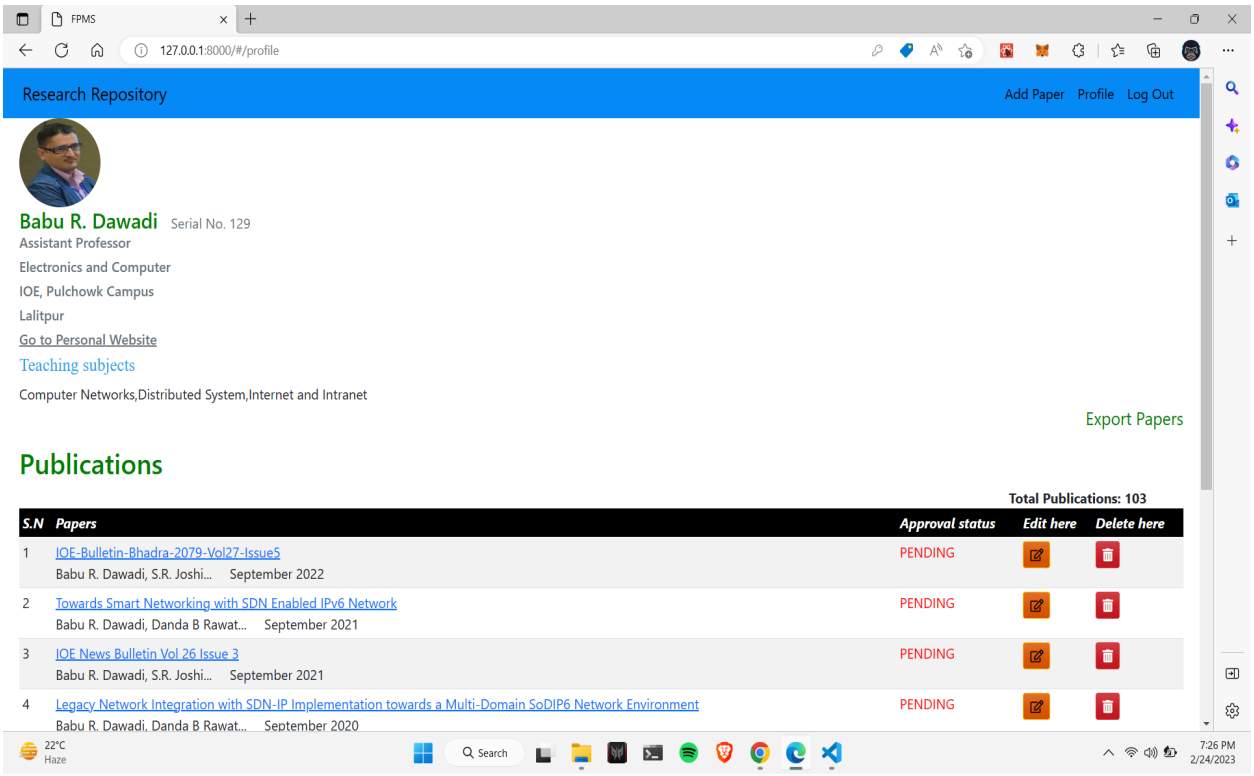


Figure 8.7: Export Feature for Downloading Publications

BibTex Importing Feature

This feature lets the user to interact with the BibTex for the reference. The user can directly import the BibTex file for the purpose of reference from the system.

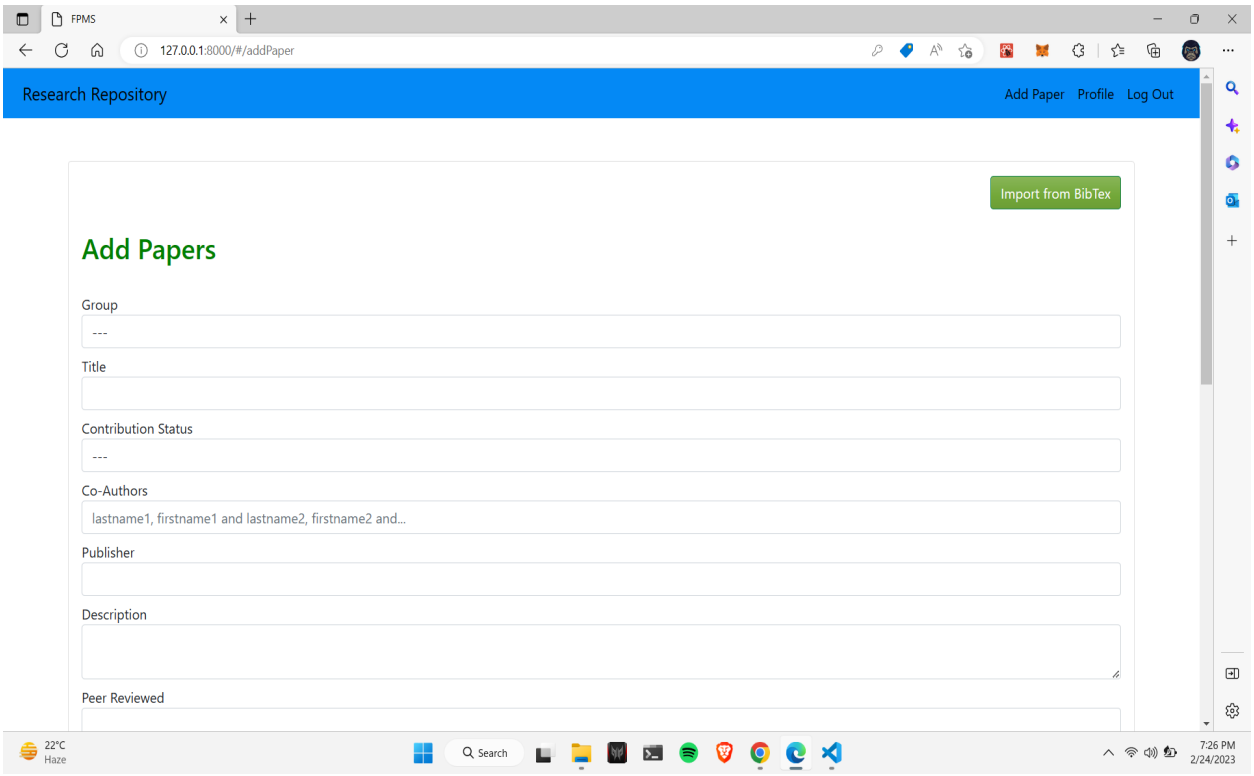


Figure 8.8: Import from BibTex Feature for Adding Papers

8.1.2 Statistics and Reporting

According to various stakeholders interested in the Faculty Publication Management System, it was suggested that there should be a feature to view statistics and reports of all the available works published by a researcher, who is the main user of the system. It makes the process of reviewing a researcher easy for any interested organization.

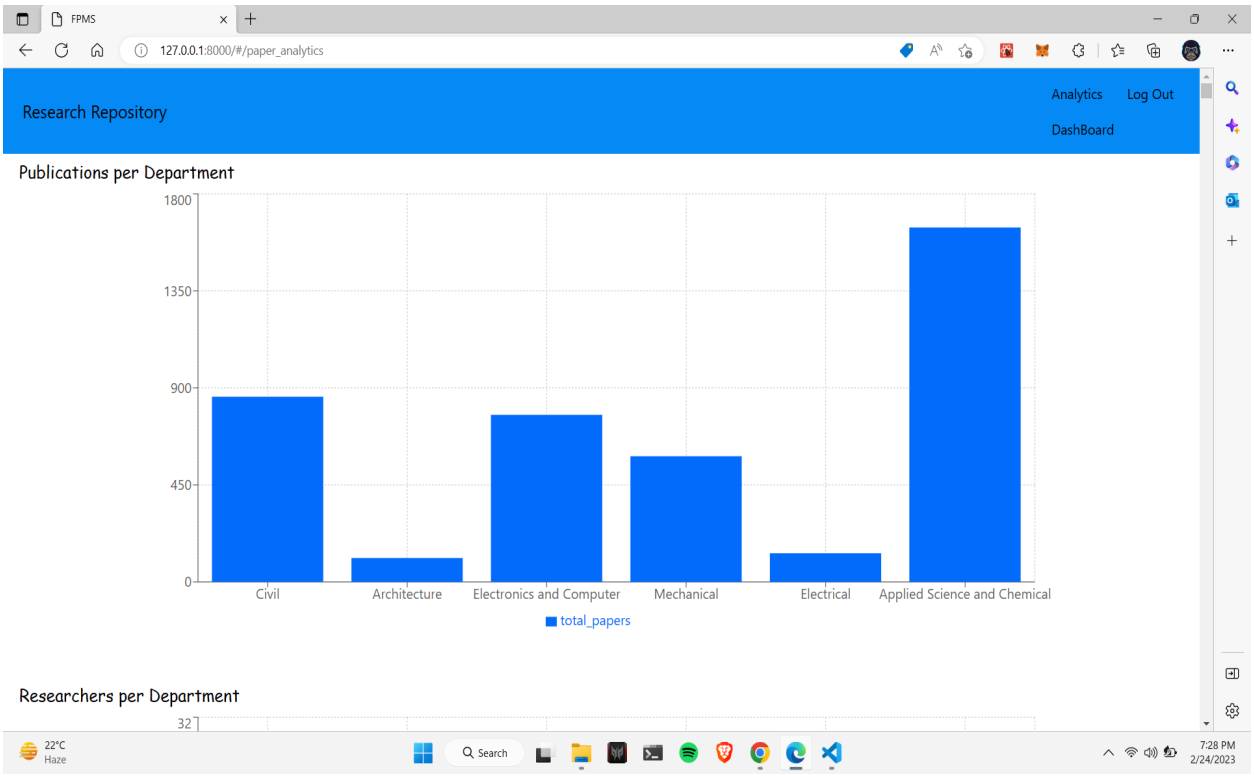


Figure 8.9: Statistics and Reporting of Publications

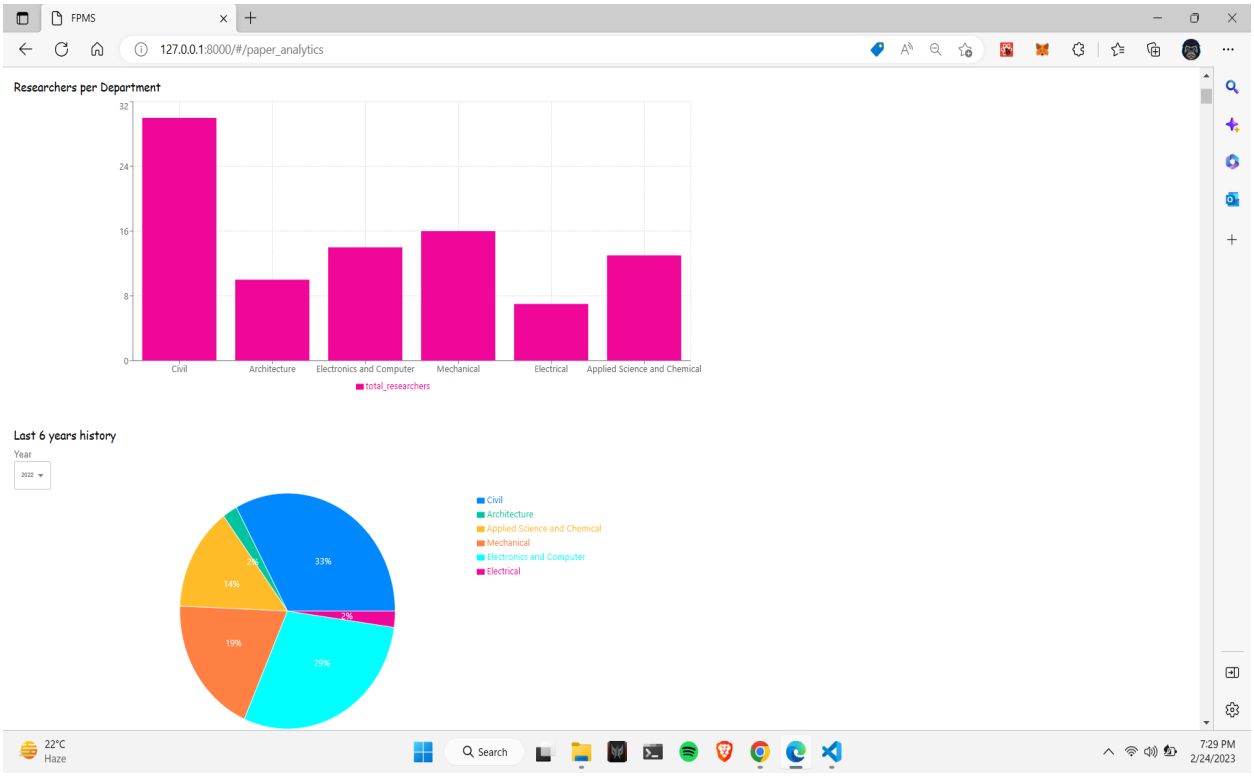


Figure 8.10: Bar Graph and Pie Chart Visualizations of Publications

8.1.3 Web Scraping

The data is scraped from the user account available on the web. After successful scraping, the data is stored on the system's database and later shown in the frontend as shown in the figure.

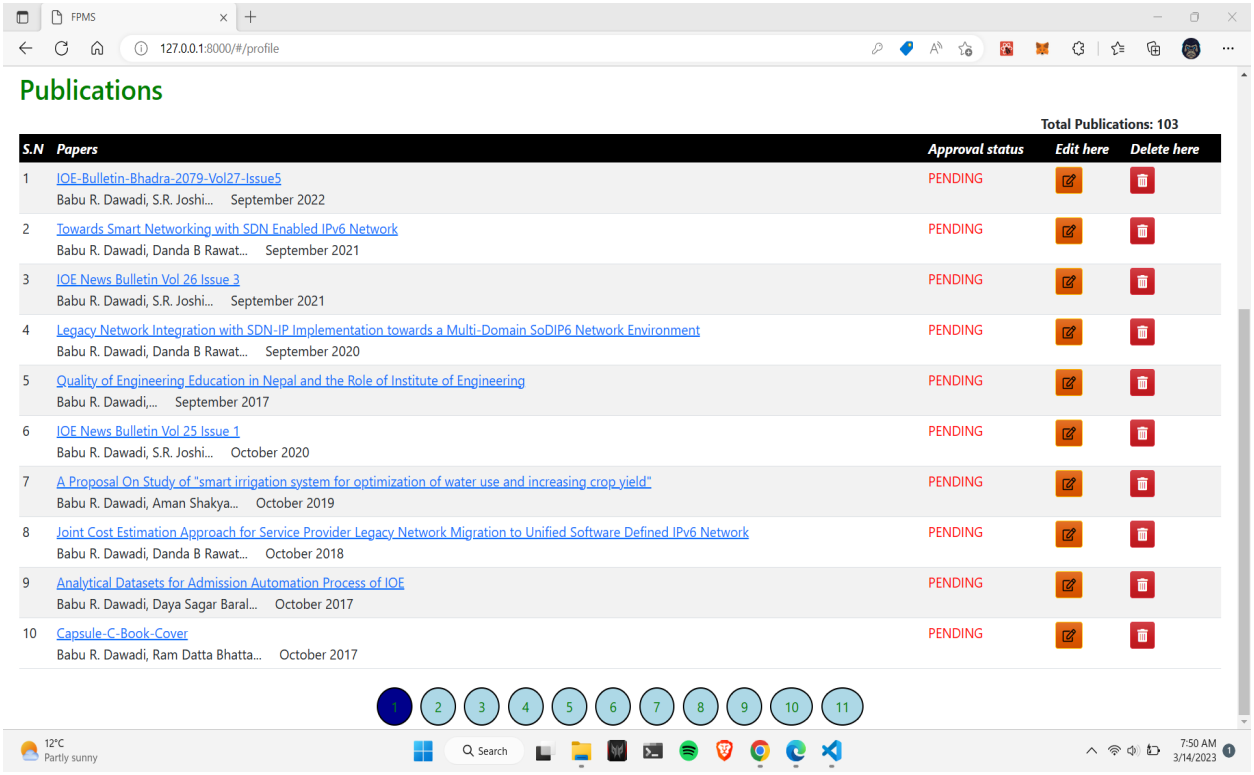


Figure 8.11: Scraped Data

8.1.4 Searching and Sorting Feature

The system searching and sorting feature in the frontend of the application allows users and organizations to search specific contents or publications and order them in a specific manner. This makes it easy to use the application when the researcher is involved in large number of publications.

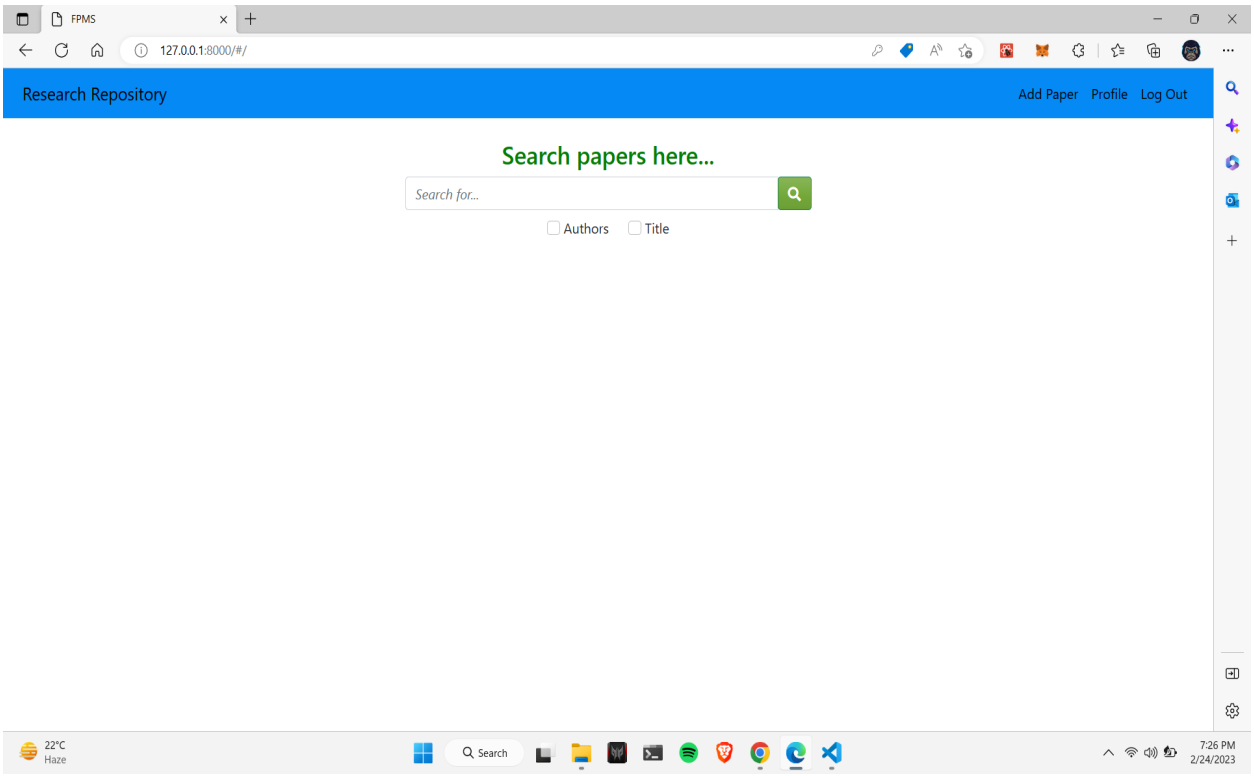


Figure 8.12: Search Feature for Searching Authors and Publications

8.1.5 Approval Hierarchy Feature

The Admin approval feature allows the resources to be approved or be rejected based on the admin. This would allow users and organizations to check the validity of the research publication and take appropriate actions accordingly.

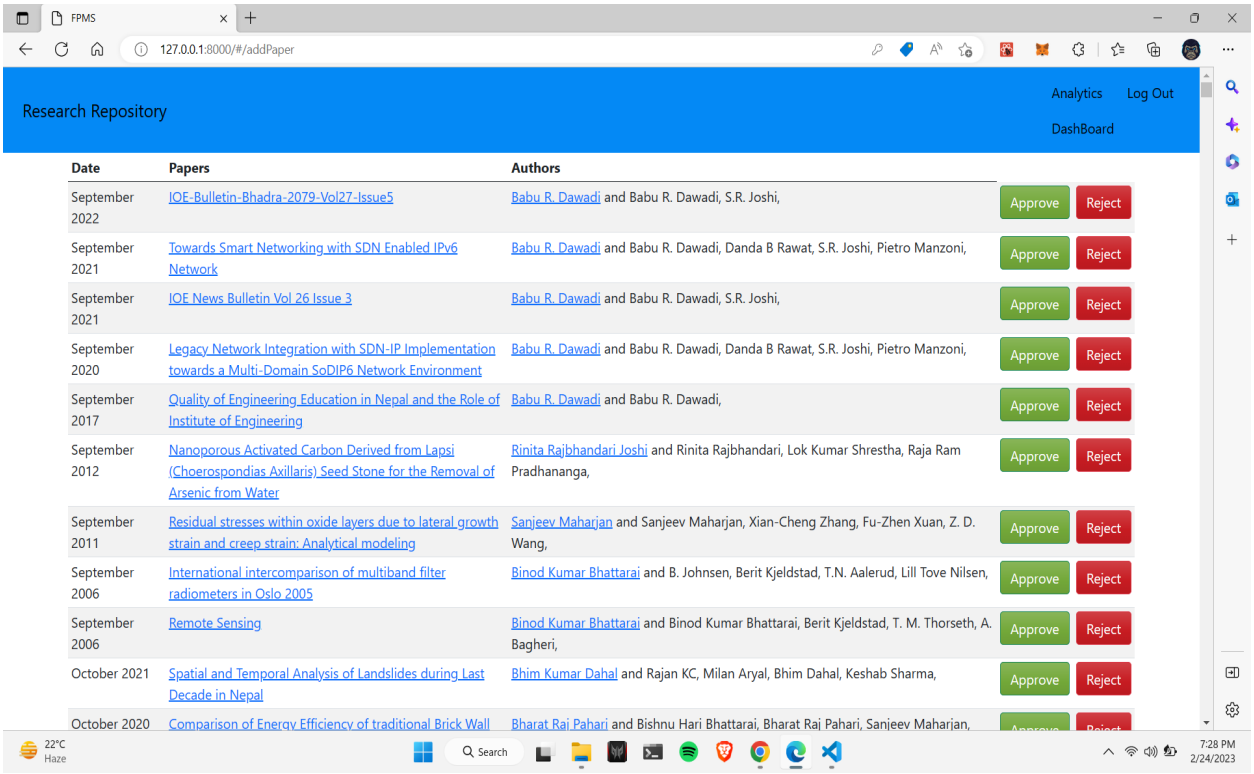


Figure 8.13: Approval Feature for Admin

8.1.6 Personal Website Integration

The 'Go to Personal Website' component in the system directs the user to his/her personal online repository. This makes it possible for the researcher to make an impression to an interested organization and also eliminate the task of redundant downloads of various records.

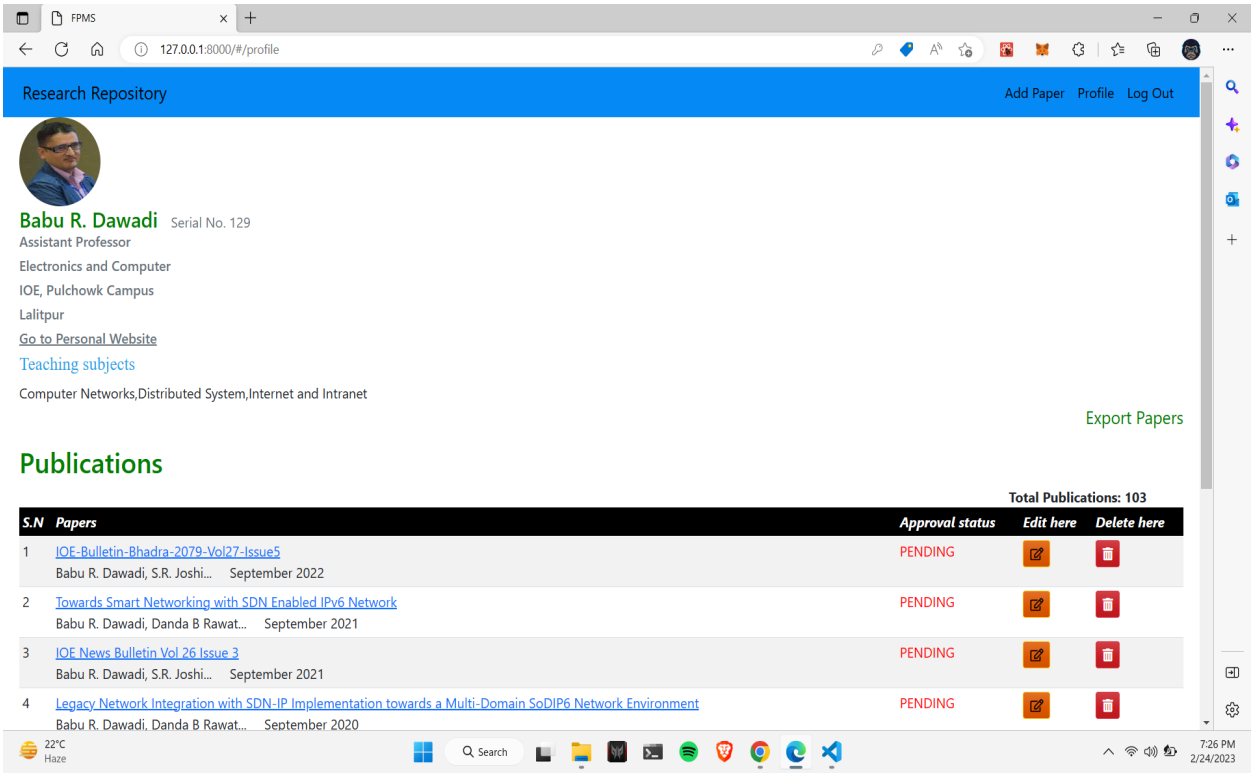


Figure 8.14: Personal Website Integration

8.1.7 Automatic Profile Summarization

The automatic profile summarization is given when the user exports his/her publications report.

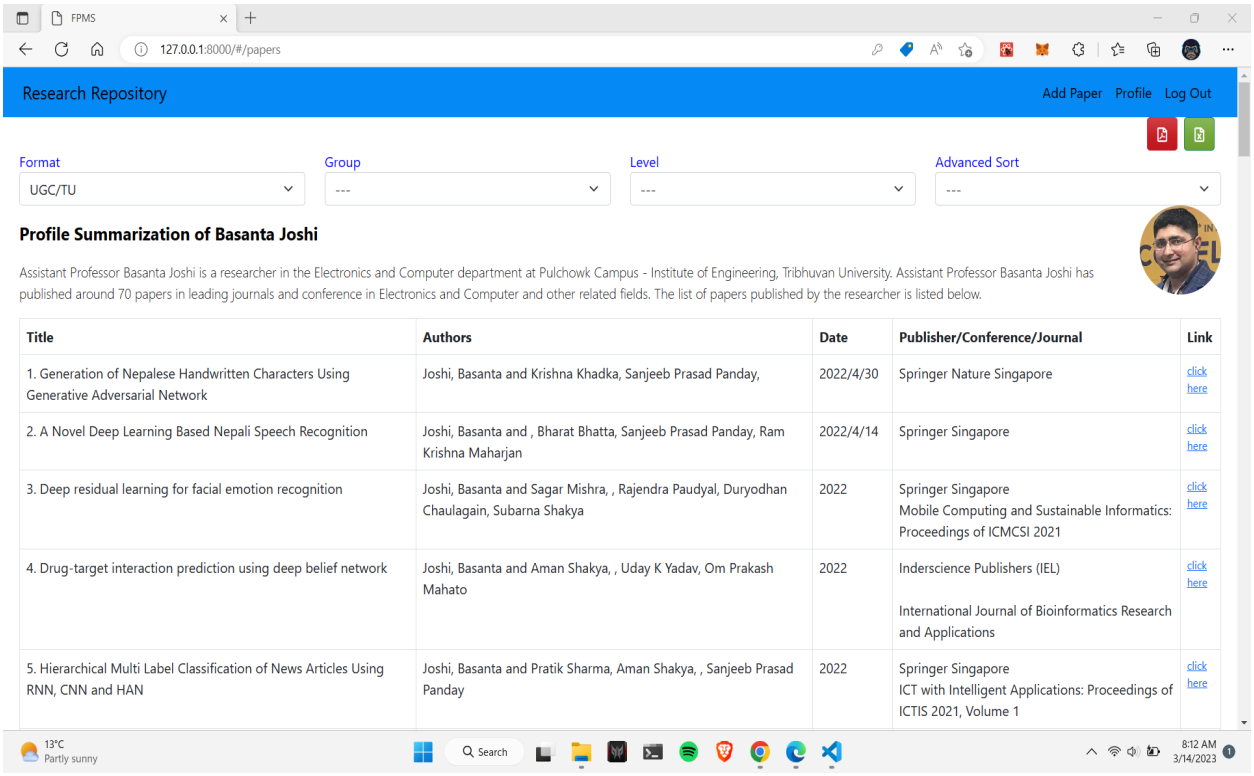


Figure 8.15: Automatic Profile Summarization

8.1.8 Report Generation

The report generation feature helps to download/print ones report on different format like MLA, APA, UGC/TU and IEEE.

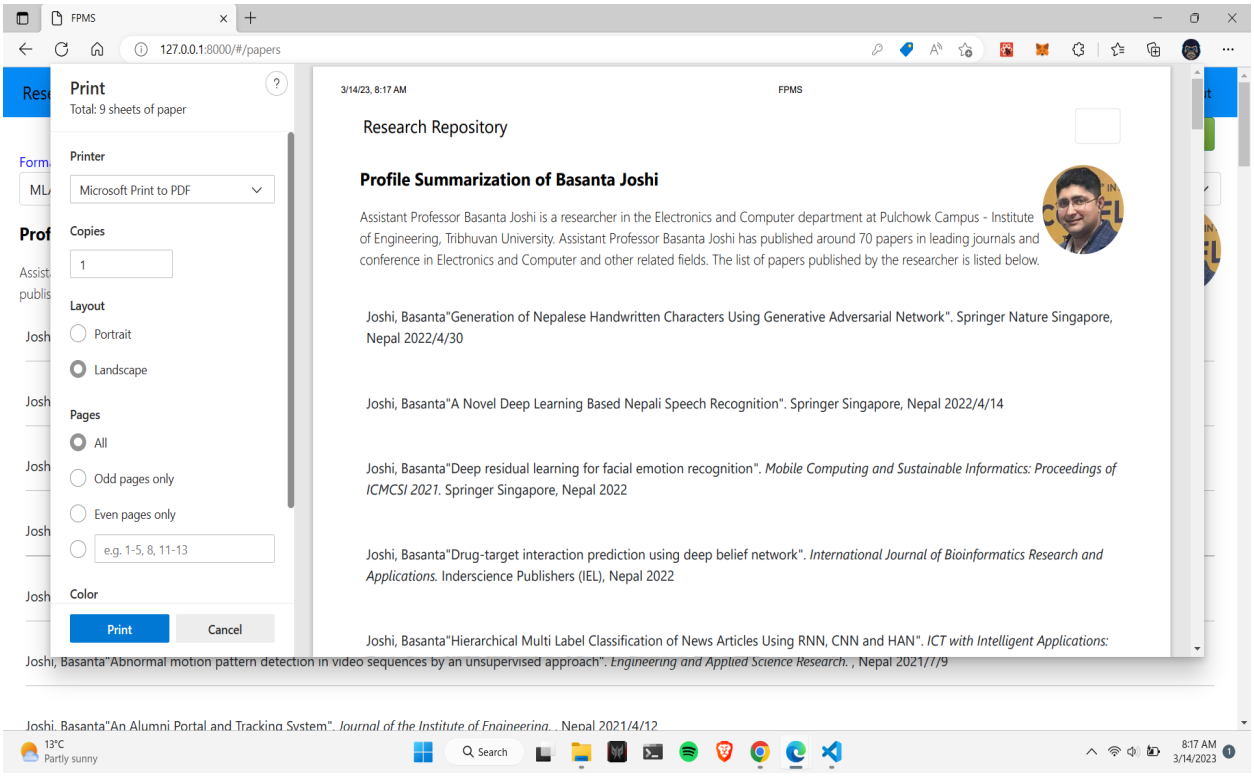


Figure 8.16: MLA Format

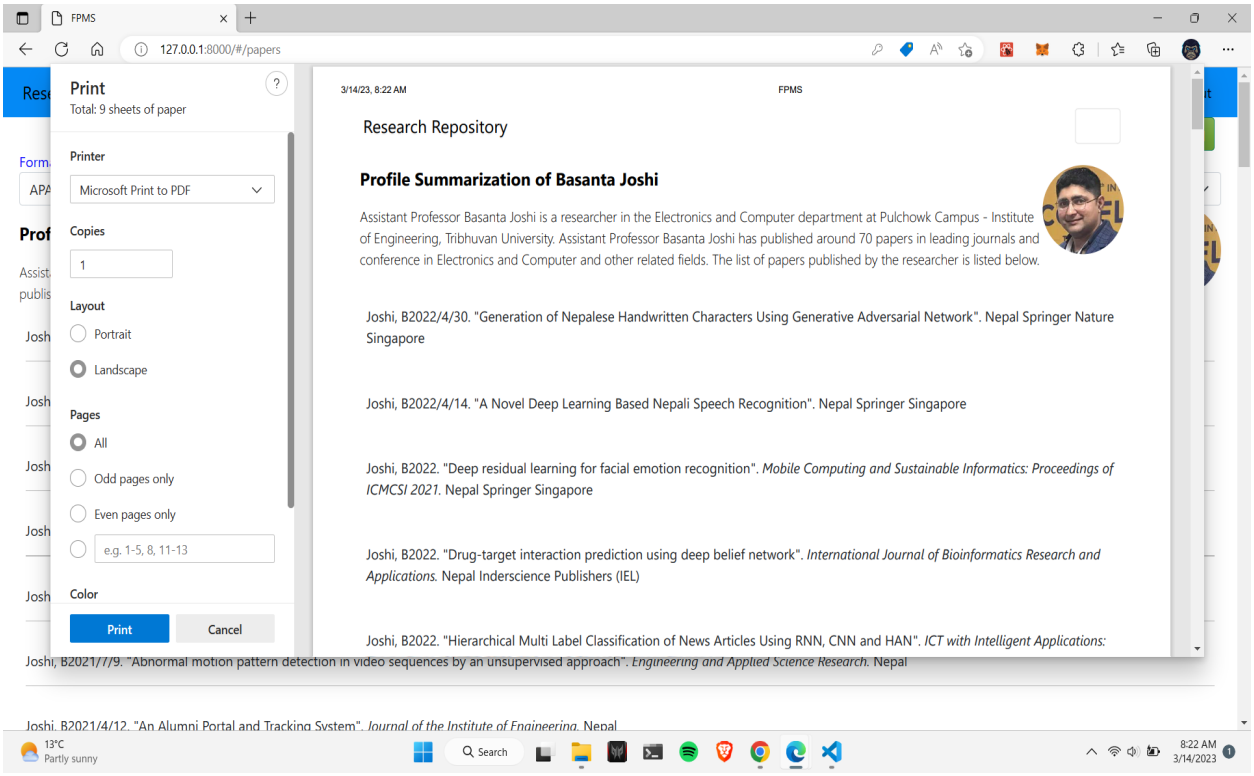


Figure 8.17: APA Format

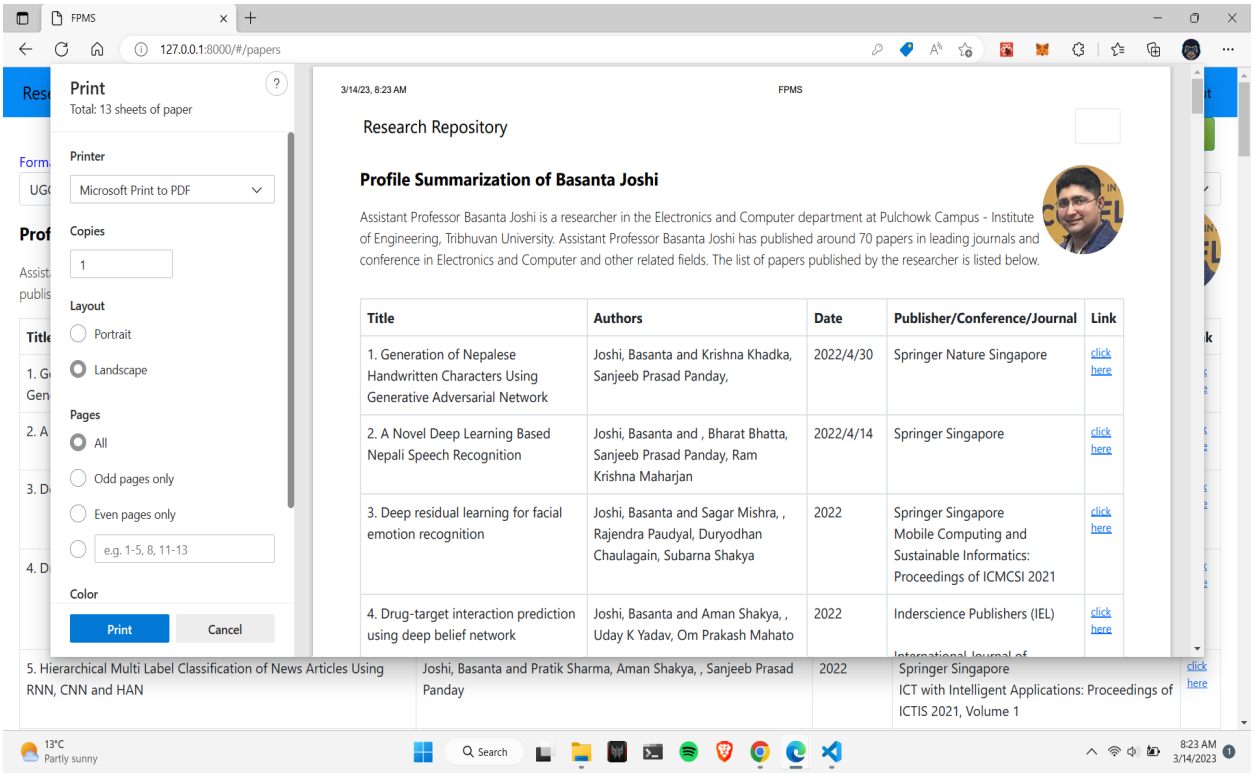


Figure 8.18: UGC/TU Format

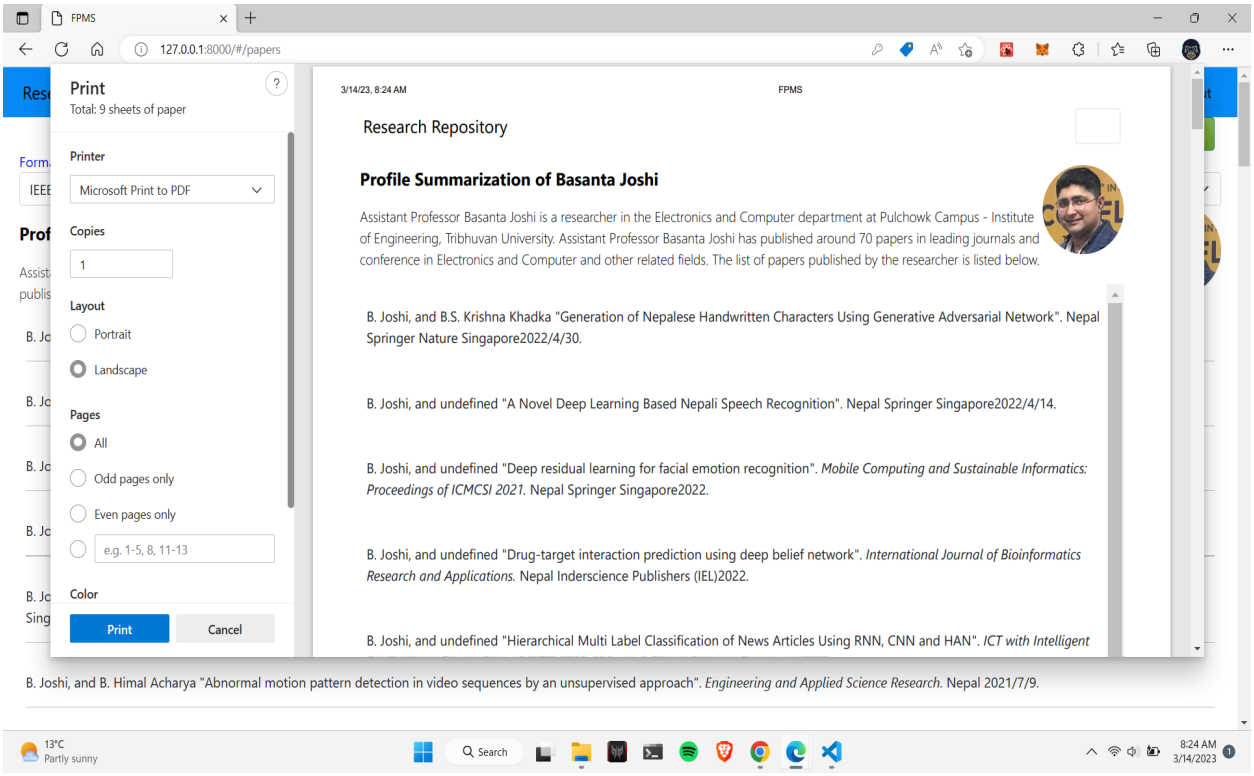


Figure 8.19: IEEE Format

8.1.9 Faculty Evaluation

For evaluation of faculty members, a prototype of the evaluation form has been prepared. The data collected from the students with these forms are stored at the database which is later used by evaluation engine.

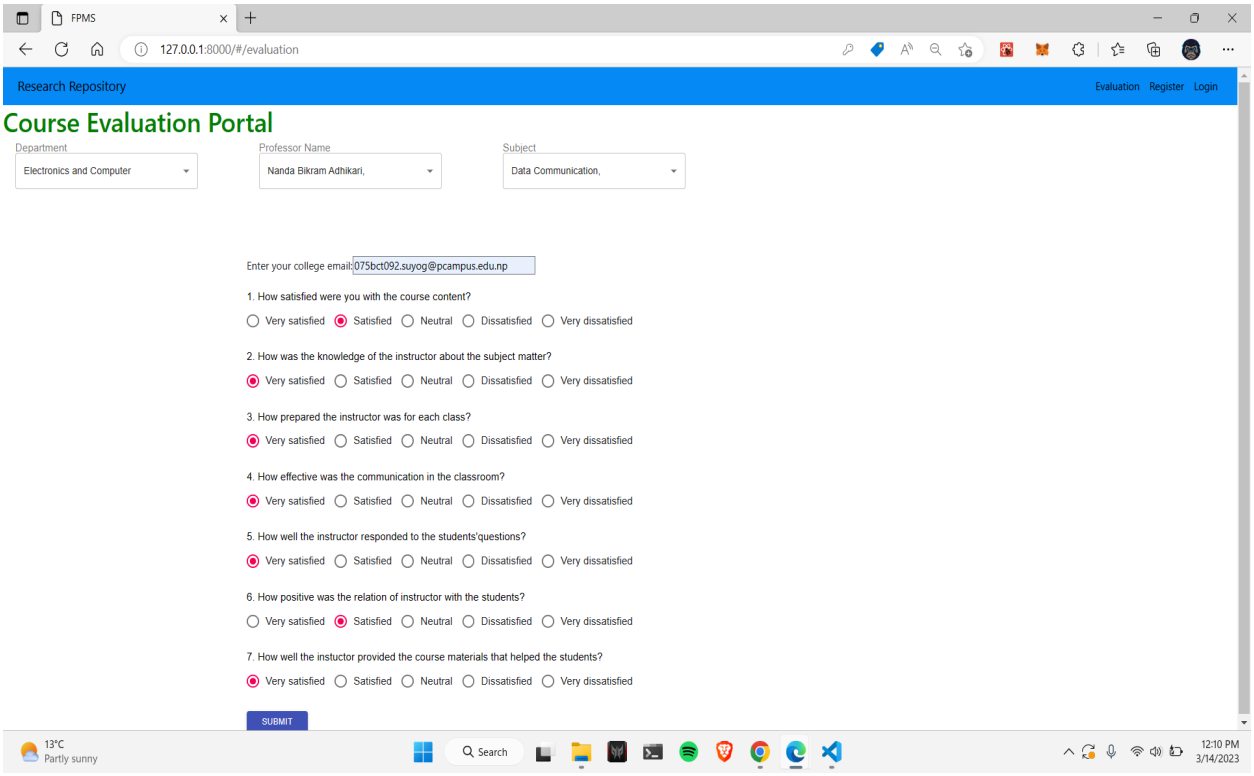
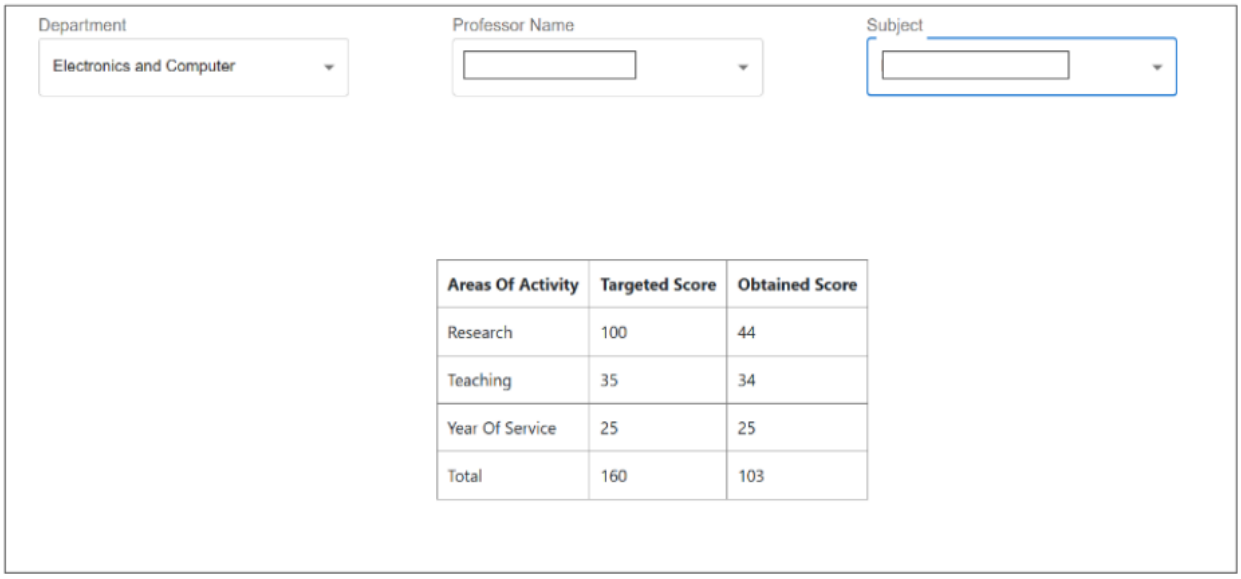


Figure 8.20: Faculty Evaluation Sample

The overall evaluation of the professor based on various criteria like number of publications, teaching feedback from students and number of teaching year experience can be analysed only by the admin of the system.



Areas Of Activity	Targeted Score	Obtained Score
Research	100	44
Teaching	35	34
Year Of Service	25	25
Total	160	103

Figure 8.21: Faculty Evaluation Form

8.2 Analysis

8.2.1 Publication per Year

From our system data , it can be seen that in the year 2021, the highest number of papers were published while analyzing past six years data.

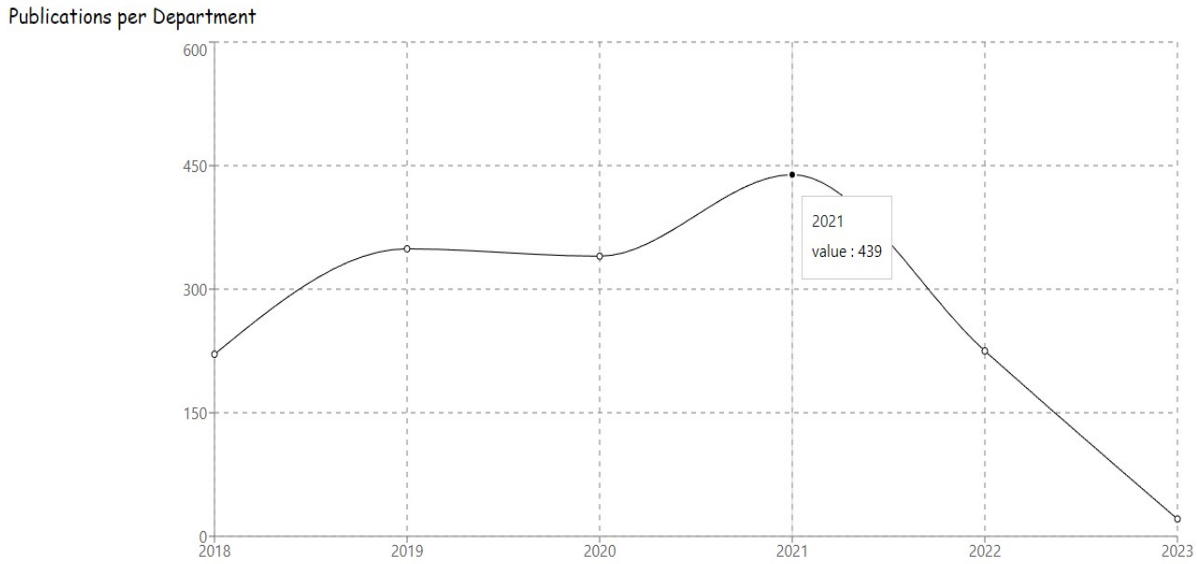


Figure 8.22: Total number of Publication per Year

8.2.2 Researchers per Department

From our system data , it can be seen that Civil department has the highest number of researchers and electrical department has the lowest number of researchers.

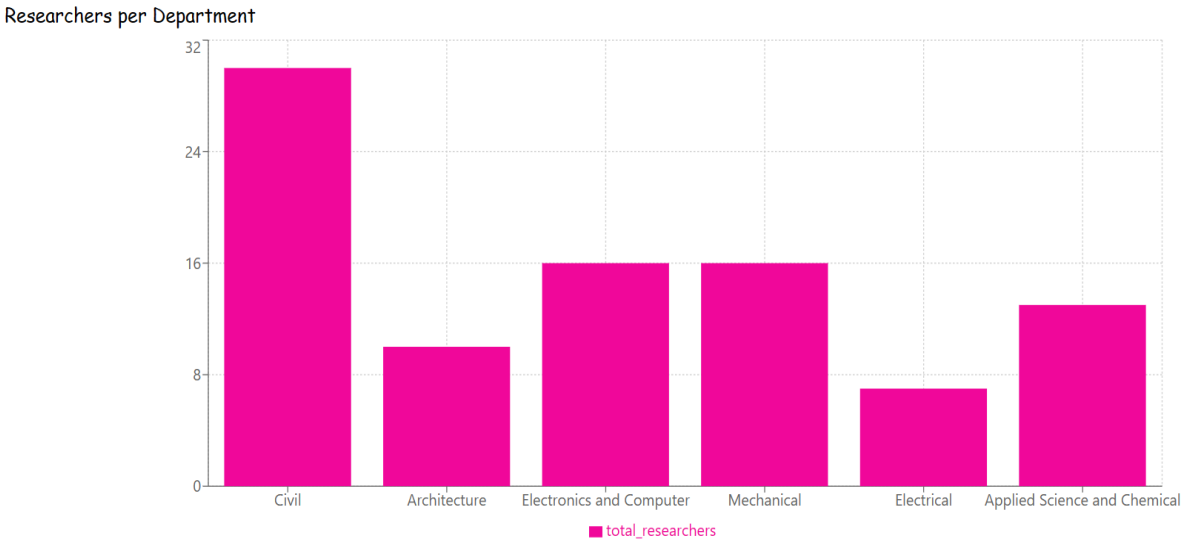


Figure 8.23: Researchers per department

8.2.3 Publications per Department

From our system data , it can be seen that Applied science and chemical department has the highest number of publications and Architecture department has the lowest number of publications.

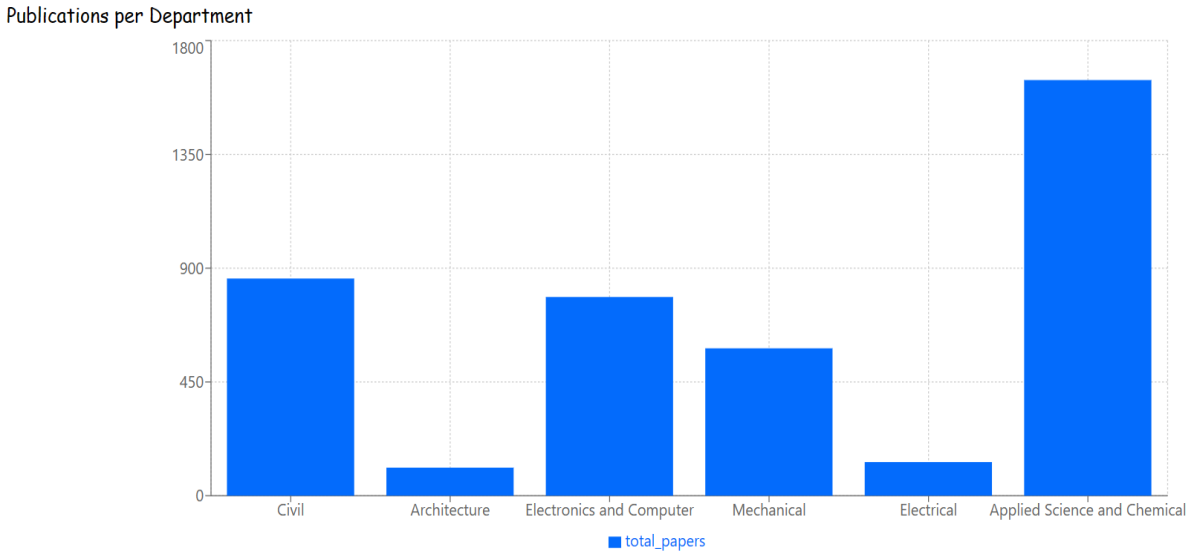


Figure 8.24: Publications per department

8.2.4 Publication History

From our system data , it can be seen that in 2023, Civil department researchers have published more paper than other department researchers which covers about more than one third of the campus data.

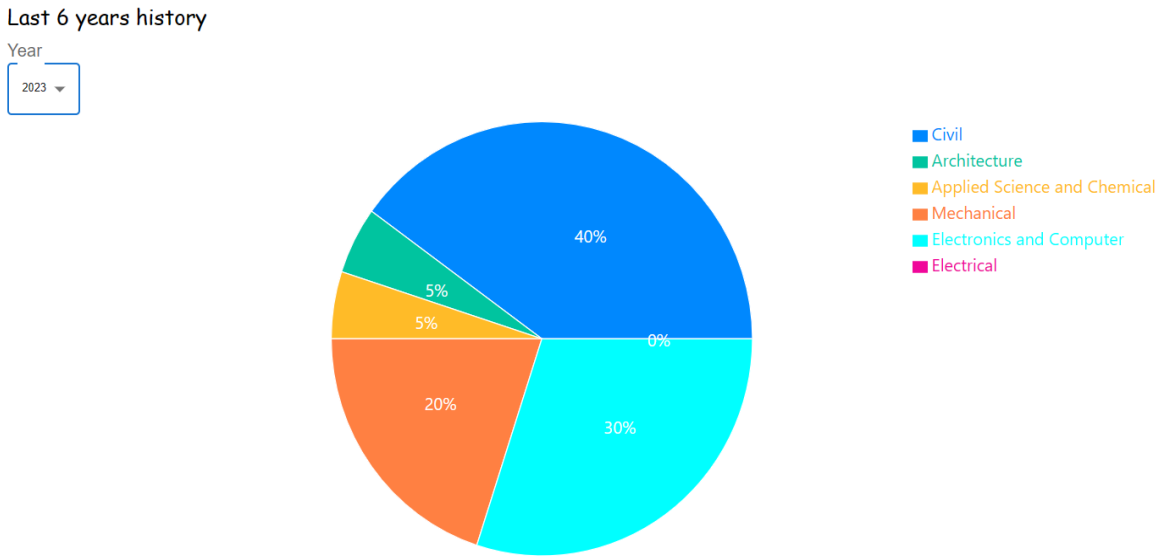


Figure 8.25: Publications in 2023

From our system data , it can be seen that in 2022 also, Civil department researchers have published more paper than other department researchers which is followed by Electronics and Computer department in second.

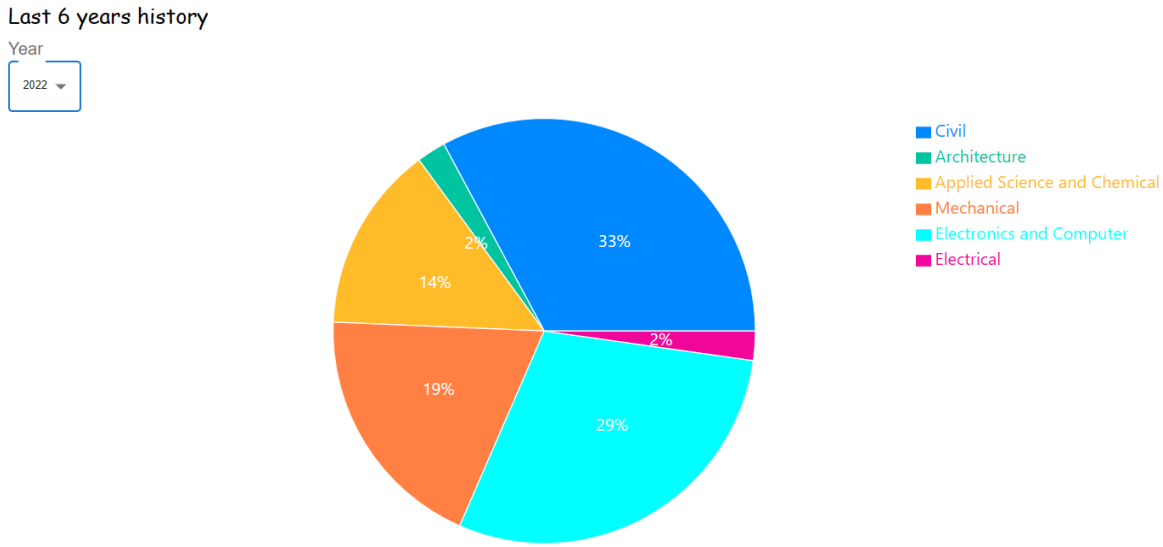


Figure 8.26: Publications in 2022

From our system data , it can be seen that in 2021, Electronics and Computer Engineering department researchers have published more paper than other departments researchers followed by Civil department in second.

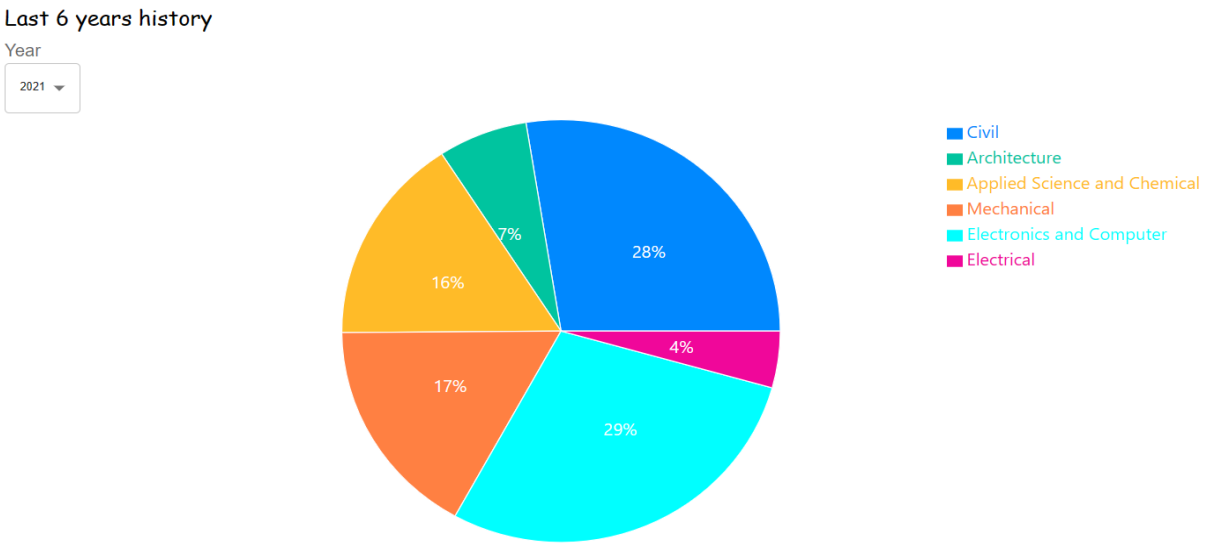


Figure 8.27: Publications in 2021