PUBLICATIONS SUMMARY GENERATOR FOR FACULTY MEMBERS PROFILE BUILDING

Somya Agarwal, Pratik Lohar, Arpit Menaria

Assistant Professor, Student, Student

Department of CSE,
Geetanjali Institute of Technical Studies, Udaipur.

Abstract :
In the academic domain, maintaining an up-to-date, concise summary of faculty research publications is a critical yet often neglected task. This project presents a Publications Summary Generator using Python, aiming to automate the extraction, classification, and summarization of research publications from various sources such as Google Scholar, Scopus, or local datasets. The tool generates structured summaries for individual faculty profiles to facilitate institutional reporting, research assessments, and online visibility. The system uses web scraping, Natural Language Processing (NLP), and data visualization libraries to produce automated outputs. This paper discusses the system architecture, development methodology, implementation details, and the effectiveness of the tool in real academic scenarios**.**

I. INTRODUCTION

Faculty members across educational and research institutions are engaged in publishing research work across a variety of disciplines. Maintaining these records manually for profile building, institutional audits, or national-level accreditation systems like NAAC/NBA is time-consuming. This project introduces an automated Publications Summary Generator that simplifies the generation of summarized profiles. The tool is developed using Python and is capable of collecting publication data, categorizing it by publication type, year, and indexing platform, and generating readable outputs for each faculty member.

II. LITERATURE REVIEW

Prior research has explored the use of web scraping and NLP to automate bibliometric tasks. For example, tools like Publish or Perish extract data from Google Scholar but require manual input. Bibliometric dashboards in academic institutions often rely on structured databases like Scopus or WoS, which are expensive or restrictive. NLP tools like BERT and SpaCy have been applied in summarizing academic texts. However, a lightweight, customizable Python-based solution focused on individual faculty profiles is not widely available, creating a gap this project aims to fill.

III. METHODOLOGY

The development of this project follows the Agile model. The key modules include:
Data Collection: Using Python's `requests`, `BeautifulSoup`, or `Selenium` for scraping publication data.
- Data Preprocessing: Cleaning and converting raw data into structured format using `pandas`.
- Text Summarization: Using NLP libraries such as `NLTK`, `spaCy`, or `TextRank` for generating summaries of abstracts or publication impact.
- Profile Builder: A script that formats all data into a readable summary categorized by publication type, year, and indexing.
- Output Generation: The final output is exported in PDF or HTML format for integration into institutional websites or internal systems.

IV. SYSTEM ARCHITECTURE/DESIGN

The system is designed with modularity in mind:

- Frontend (Optional): A minimal GUI using `Tkinter` or web interface using `Flask`.
- Backend:
  - Scraper Module: Fetches publication data.
  - NLP Processor: Summarizes abstracts or extracts keywords.
  - Database (optional): SQLite or JSON for storing records.
  - Formatter Module: Converts processed data into summary format.

Workflow Diagram:
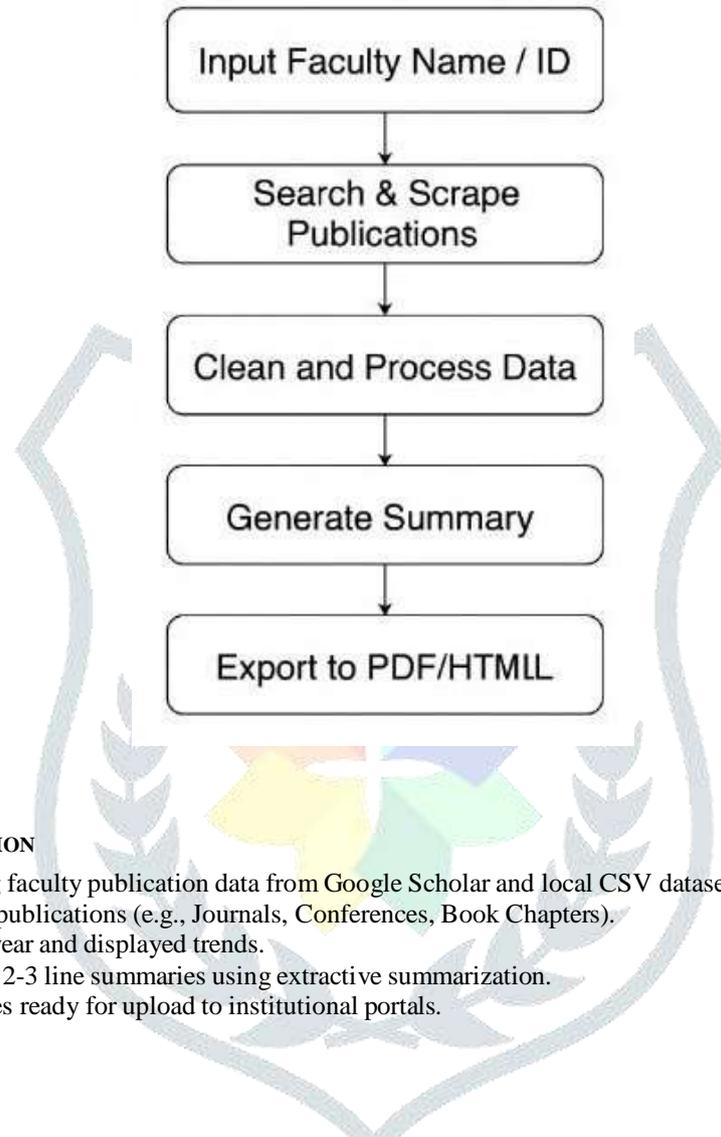
V. RESULTS / IMPLEMENTATION

The system was tested using faculty publication data from Google Scholar and local CSV datasets. The prototype successfully:
- Extracted and categorized publications (e.g., Journals, Conferences, Book Chapters).
- Counted publications per year and displayed trends.
- Summarized abstracts into 2-3 line summaries using extractive summarization.
- Produced profile summaries ready for upload to institutional portals.

Sample Output:


```
Dr. John Doe
Total Publications: 23
Top Keywords: Machine Learning, Image Processing, AI
Recent Publications:
- "Efficient CNN Models for Image Classification" (Springer, 2023)
- "AI Trends in Healthcare" (IEEE, 2022)
```

VI. CONCLUSION

This Python-based Publications Summary Generator automates the manual task of compiling and formatting faculty research outputs. It improves accuracy, saves time, and enhances the visibility of academic contributions. The system is scalable and can be integrated with institutional ERP or LMS platforms. Future enhancements include adding support for Scopus/ORCID integration and deploying the tool as a full-stack web app.

VII. REFERENCES

1. Harzing, A.W. (2007). Publish or Perish. https://harzing.com/resources/publish-or-perish
2. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media.
3. Google Scholar API (Unofficial). https://github.com/OrganicIrradiation/scholar.py
4. Scopus API. Elsevier Developer Portal. https://dev.elsevier.com/
5. SpaCy NLP Library. https://spacy.io/
6. Fister, I. Jr., et al. (2016). "A bibliometric analysis of research on data mining in agriculture." Computers and Electronics in Agriculture.