Publications Indexer Tool: Using Web Crawler and JavaScript

Pundrik Mishra[1], Sneh Jain[2], Disha Dahake[3], Nilakshi Jain[4]

[1, 2,3,] *B.E. Student, Dept. of IT Engineering, Shah & Anchor Kutchhi Engineering College, Mumbai, India*
[4] *Professor, Dept. of IT Engineering, Shah & Anchor Kutchhi Engineering College, Mumbai, India*

---***---

***Abstract* - There** is no common tool available on the World Wide Web for indexing Publications with an API documentation for external usage. Publications related things like Journals, Conferences, Books, Copyrights, Patents, Technical Paper, and Certifications are needed to be indexed for showcasing them wherever needed. The users should be able to search and traverse through all the publications that are available and filter according to their needs. Using this system, users will be able to search, submit and view their Publication. This system is dynamic and uses AJAX to fetch publication data from the database without having to load the page. This system also integrates Google scholar publications using a web crawler to fetch the user's publication for the feasibility of not having to add every single one of their Journal's details again. This system has the potential to become the centre of all the publications on the Internet.

***Key Words*: Publications, PHP Curl, Database, Indexing, Google Scholar crawler.**

I. INTRODUCTION

The Publication Indexer Tool using web crawling and JavaScript is an approach to make publications ubiquitous and widespread on the World Wide Web. It is highly dynamic, has many capabilities and potential to be the centre of all the publications available worldwide. The profile of the user is not private like Scopus and other private publication indexers, our system is public and provides filtering by year without having to load the page. Basic functionalities include – displaying all publications record – Journal, Conference, Technical Paper, Books, Copyrights, and Patents, providing a filter, giving an option to add your own publication.

II. LITERATURE SURVEY

A survey on how people find papers to read when they cannot do a search. [1]

A survey published by Dynamic Ecology collected responses from over 150 people from year 2018-19. According to the survey, the single most popular way that respondents, is by looking for journals online, used by 82% of respondents.

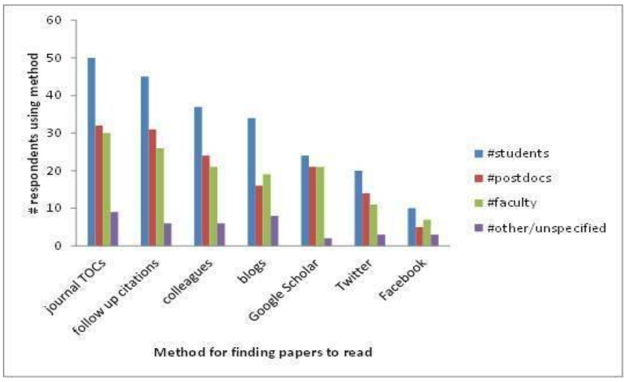


Fig 1: *Survey showing how people find papers online*

Respondents were unable to find their desired journals online because of no central database of publications. Everything is scattered on the internet in terms of publications. It is very hard to find the exact publication that the user is looking for. With our system, all users have a central database to look at all the publications they need. Over 73% of the respondents used follow up citations across other papers to find their required journal. Another source by respondents were blogs. 52% of the respondents chose blog as their answer.

III. SOFTWARE REQUIREMENT SPECIFICATION

A. Introduction:

1. *Purpose:* The primary purpose of this system is to provide a central database for everyone to use and provide public display of publications records synced with their Google Scholar profile.

2. *Scope of the Project:* Publication Indexer Tool can be used by any college to get publication information about their faculty on cross platform. Using this system, it will be easy to filter and search through various publications entered by different users. By using this, one can commemorate their faculty on our system. It can be used to keep a track of all the publications they have published.

3. *Intended Audience:* The intended audience of this system are the college, and their various faculties of various departments.

B. Overall Description

1. *Product Perspective:* The online publishing indexer tool is a web application. This application can be used to index the publications and can be used to provide appreciation to the users about their publications. It offers full automation of publication sorting and indexing without having to load the page using AJAX. It can also be used to sync user's google scholar profile to provide backwards compatibility support to the user to deduct the redundant tasks. There is google OAuth login for every user who wants to add their publication on the central database.

2. *Product Functions:* Some of the features that are included into the application are:
   a) View all the Publications
   b) Sort Publications
   c) Submit your Publications
   d) Sync your profile with Google Scholar
   e) User's Google OAuth login
   f) Edit your submitted Publications

3. *User classes and characteristics:* The users of this application are the college faculty or the college students and the end users are the random visitors of the website who want to find a research work to read or reference someone's work. Faculties and the students need to know the basic functionalities of research work publications. Website will automate the indexing process, no website administrator required.

4. *Operating Environment:* The application is a web application. The application can be run from any device that has access to the internet and can operate a web browser on the system with basic support of JavaScript.

5. *Design and Implementations Characteristics:* Works only with an internet connection and a browser with JavaScript support. Available only in English language.

6. *Assumptions and Dependencies:* It is assumed that all the users are from one GSuite organization and have the same domain name for email. Apart from that, all the software and hardware mentioned above are assumed to be available to all end users.

IV. SYSTEM REQUIREMENTS

A. User Interfaces

Bootstrap and JavaScript is used along with a custom theme to provide a beautiful UI for all the users.

B. Hardware Requirements

1. Any Device with a browser
2. 512 MB Ram

C. Software Requirements

1. Frontend UI: Bootstrap
2. Backend: MySQL
3. Code Igniter Web Framework
4. Working internet connection
5. Google OAuth
6. PHP Spreadsheet
7. Simple HTML Dom Parser
8. JavaScript

D. Communication Interfaces

1. Internet Connection
2. WIFI
3. Browser

V. PRODUCT DESIGN

A. Use Case Diagram

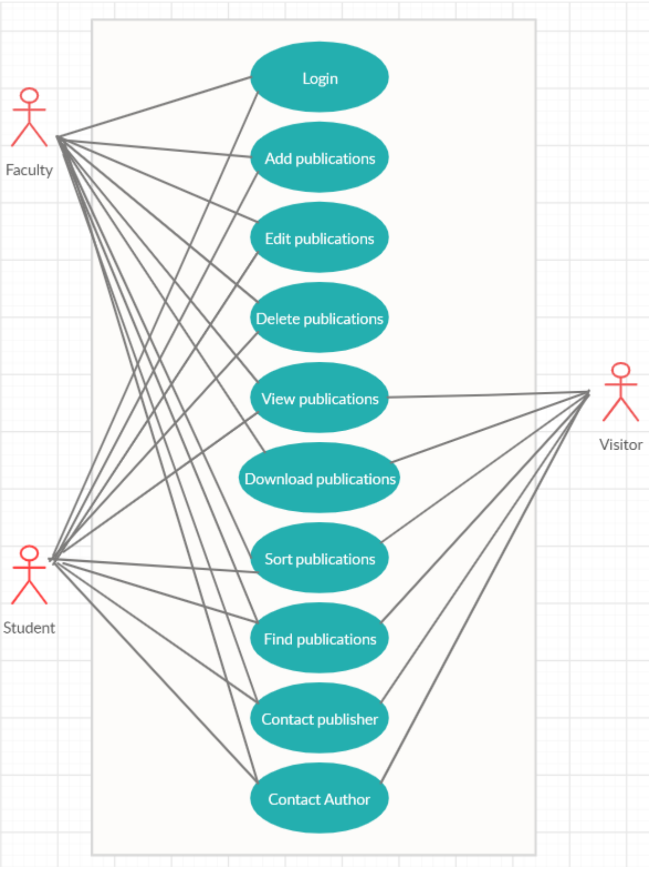


Fig 2: *Use Case Diagram*

B. Module Diagram

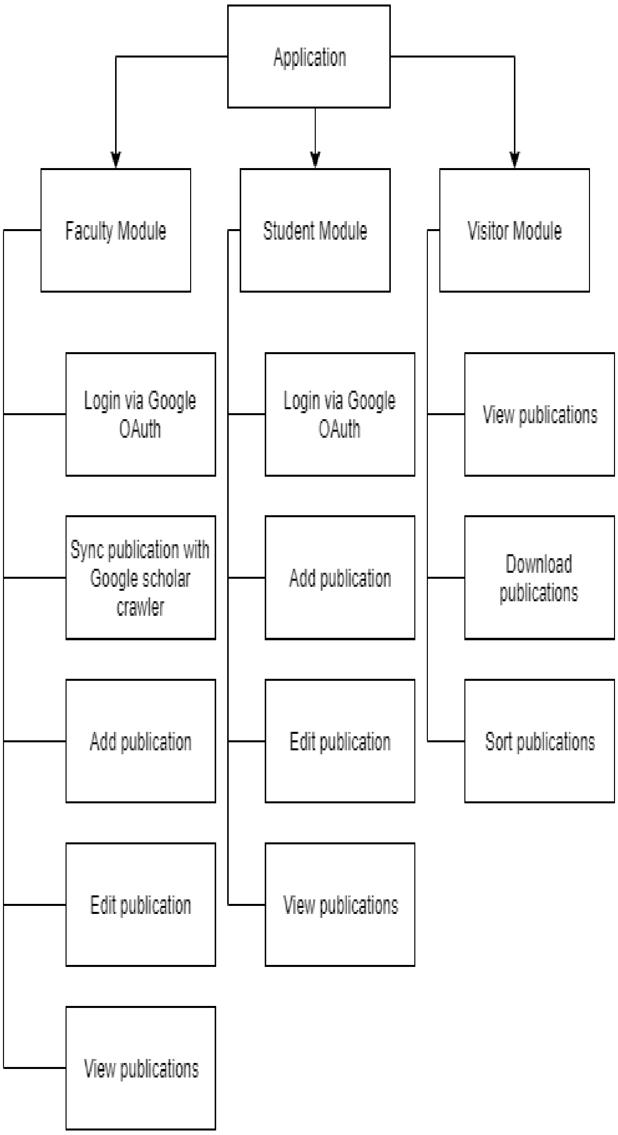
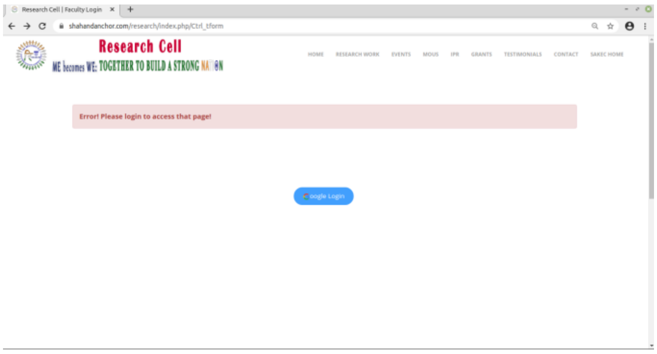


Fig 3: *Module Diagram*

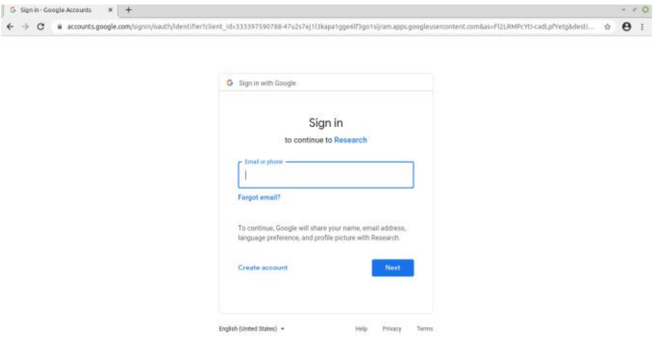
VI. USER INTERFACES


Research Cell | Faculty Login
shahandanchor.com/research/index.php/Ctrl_tform
Research Cell
ME becomes WE: TOGETHER TO BUILD A STRONG NATION
HOME
RESEARCH WORK
EVENTS
MOUS
IPR
GRANTS
TESTIMONIALS
CONTACT
SAKEC HOME
Google Login

Error! Please login to access that page!

Fig 4: *Login screen*

The login screen of our application using google OAuth library.

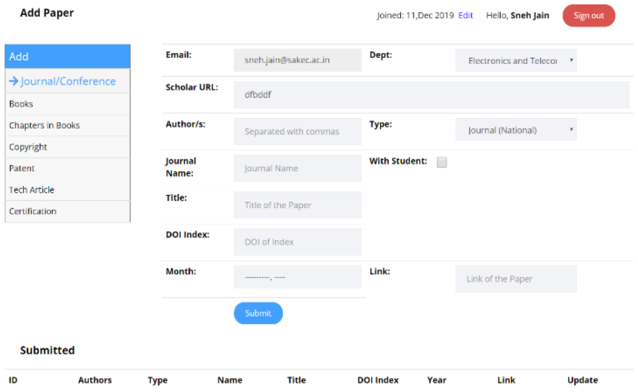

Sign in - Google Accounts
accounts.google.com/signin/oauth/identifier?client_id=333397590788-47u2s7ej1l3kapa1gge6lf3go1sjjram.apps.googleusercontent.com&as=Fl2LRMPcYtl-cadLpfYetg&desti...
Sign in with Google
Email or phone
Forgot email?
Create account
Next
English (United States)
Help
Privacy
Terms

Sign in
to continue to Research

To continue, Google will share your name, email address, language preference, and profile picture with Research.

Fig 5: *Login into Research application*

The college faculties and students are provided with a unique Gmail account using GSuite application provided by Google. Using that faculty and student can login to add their unique publication data. Only the users with this Gmail account of that college's domain can login into the application.



Add Paper

Joined: 11,Dec 2019 Hello, Sneh Jain

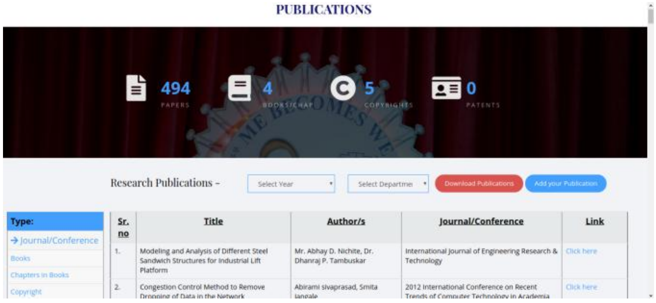
Email:	sneh.jain@sakec.ac.in	Dept:	Electronics and Telecoı
Scholar URL:	dfbddf		
Author/s:	Separated with commas	Type:	Journal (National)
Journal Name:	Journal Name	With Student:	□
Title:	Title of the Paper		
DOI Index:	DOI of Index		
Month:	---------, ----	Link:	Link of the Paper

Submitted

ID	Authors	Type	Name	Title	DOI Index	Year	Link	Update

Fig 6: *Adding publication*

After authentication, user is provided with publication adding forms and they can also edit their previously added publications on the website. The form is dynamic and is updated with AJAX without having to load the page.

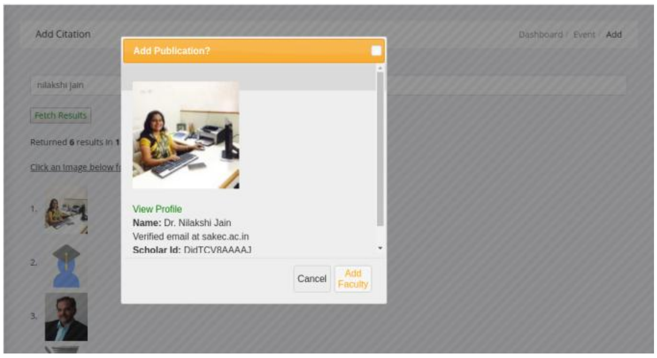


Research Publications -

Sr. no	Title	Author/s	Journal/Conference	Link
1.	Modeling and Analysis of Different Steel Sandwich Structures for Industrial Lift Platform	Mr. Abhay D. Nichite, Dr. Dhanraj P. Tambuskar	International Journal of Engineering Research & Technology	Click here
2.	Congestion Control Method to Remove Dropping of Data in the Network	Abirami sivaprasad, Smita Jangale	2012 International Conference on Recent Trends of Computer Technology in Academia	Click here

Fig 7: *Viewing Publication*

After adding your publication, you can view it on the front end where all the added publications are displayed. You can find your publication by scrolling or via filtering it.



Name: Dr. Nilakshi Jain
Verified email at sakec.ac.in
Scholar Id: DidTCV8AAAAJ

Fig 8: Google Scholar Web Crawler

Users can also fetch all their papers from Google scholar with the crawler to add it to the central database. This significantly reduces the tasks for users about having to add every single one of their papers again.

VII. ANALYSIS

This Web application helps college to commemorate their faculty and students with recognition and display their Publication records to the public, all on one website that can be opened on a browser with no external software required. If every faculty and students use this web application and add their respective publication records, it will increase College's recognition along with its reputation. Visitors of the web application can easily see and sort through the publications to get a rough idea about the College's research work. Faculties should collaborate with their fellow students to write publications and then add it on the website for recognition which can be used by the students in their respective resume. This web application will help out everyone by giving them recognition on a web based platform on the World Wide Web.

VIII. FUTURE WORK

- Develop Android and iOS Application for this web based application to expand the domain.
- Provide a REST API and document it so that other services can use the website data easily.

IX. CONCLUSION

Having a Publications central application on the World Wide Web is open to many possibilities. The deep-rooted Google Scholar application can be modified and updated a lot to live up to the standards of this modern era.

X. ACKNOWLEDGEMENT

We would like to thank our honourable principal Dr. Bhavesh Patel, for giving us such an opportunity to carry on such innovative research work. We would also like to thank our Project Guide, Dr. Nilakshi Jain, for guiding us through all the hardships of this Project. We are also very grateful to Ms. Swati Nadkarni, the Head of the Information Technology Department, for her guidance and support.

REFERENCES

[1] "Survey on how do people find papers to read for research work" – By Dynamic Ecology, https://dynamicecology.wordpress.com/2013/10/29/survey-results-how-do-you-find-papers-to-read-when-you-cant-do-a-search/ , Oct 2018

[2] Dr. Jain Nilakshi, Patel Reeva, Dhuri Siddhesh, Gada Preet, et. al.: "Digital Forensics Capability Analyzer: A tool to check forensic capability", IEEE Technically Co-Sponsored 3rd Biennial International Conference on Nascent Technologies in Engineering, January, 2019

[3] Nilakshi Jain, Dhruvin Mehta, Nerkar Kuldeep, Chetan Barot, Nilesh VK, "Information Management System for Faculty and Students" - https://www.researchgate.net/publication/27650034_INFORMATION_MANAGEMENT_SYSTEM_FOR_FACULTY_AND_STUDENTS ,March, 2015

[4] Dr. Nilakshi Jain, Taha Bohari, Nir Jaharia, Ketan Desai, Rushikesh Parab. "Placemate – Sakec Portal" – IRJET - https://irjet.net/archives/V5/i4/IRJET-V5I4432.pdf ,April 2018

[5] https://www.w3schools.com/bootstrap/
[6] https://www.javatpoint.com/jquery-tutorial
[7] https://www.w3schools.com/w3css/
[8] https://codeigniter.com/docs
[9] https://www.tutorialspoint.com/mysql/index.htm
[10] https://www.tutorialspoint.com/javascript/index.htm

BIOGRAPHIES

Mr. Pundrik Mishra is currently pursuing Bachelor of Engineering (B.E.) in the field of Information Technology from Shah & Anchor Kutchhi Engineering College (SAKEC), affiliated to Mumbai University, Mumbai, India. His areas of interest include Web Development, Machine Learning, Software Development and Cyber Security. He is currently working on the research cell of SAKEC as a Web Master.

Ms. Disha Dahake is currently pursuing Bachelor of Engineering (B.E) in the field of Information Technology from Shah and Anchor Kutchhi Engineering College (SAKEC), affiliated to Mumbai University, Mumbai, India. Her areas of interest include Web Development, Database and Cyber Security. She is currently working on the research cell of SAKEC as a Web Master. She is also a member of Computer Society of India (CSI).

Mr. Sneh Jain is currently pursuing Bachelor of Engineering (B.E.) in the field of Information Technology from Shah & Anchor Kutchhi Engineering College (SAKEC), affiliated to Mumbai University, Mumbai, India. His areas of interest include Web Development, Artificial Intelligence, Software Development. He is currently working on the research cell of SAKEC as a Web Master.

Dr. Nilakshi Jain graduated from the Pacific Academy of Higher Education and Research University's Faculty of Computer Engineering PhD program. She is currently working as an Associate Professor and Research Coordinator at Shah & Anchor Kutchhi Engineering College in the Information Technology Department. She is a Certified Ethical Hacker (EC-Council-USA). She has published more than 25 research papers in reputed international journals and conferences including IEEE, ACM, Springer, etc. and published books on Digital Forensics and Artificial Intelligence under Wiley Publication. Her areas of interest include Digital Forensics, Artificial Intelligence and Usability Engineering.