



INDIAN INSTITUTE OF TECHNOLOGY,
BOMBAY

COMPUTER SCIENCE AND ENGINEERING
SOFTWARE LAB
CS699

RESEARCH PAPERS EASY ACCESS

TEAM NAME: LEARNERS

DEEPTI MITTAL 193050025

MONIKA SROHA 193050087

SWARIL SINGHAL 193050039

Contents

1	ABSTRACT	2
2	INTRODUCTION	2
3	MOTIVATION	2
4	PRIOR WORK	3
5	FEATURES	3
6	TECHNOLOGIES USED	4
7	WORKING PROCESS	6
8	IMPLEMENTATION	7
9	FUTURE SCOPE	12
10	CONCLUSION	12
11	REFERENCES	13

1 ABSTRACT

We are generating "word cloud" for Research Papers of professors .The Research papers are extracted from the homepage of professor. The cloud gives greater prominence to words that appear more frequently in the Research Papers. The word-cloud will be clickable, by clicking on any word in the word cloud it will display the list of Research papers on which it appeared.

2 INTRODUCTION

We are representing Research papers published by a professor in the form of word cloud. Word Cloud is basically an image composed of words used in particular text. Our word cloud will be based on the terms appearing in the research papers published by a professor. These terms will be extracted from the links of research papers provided on the homepage of professor. Words in word cloud are clickable link displaying the list of research papers containing that word.The word cloud will be formed using frequently occurring words in the research papers. The word-cloud assigns weight to the words as per their frequencies in the research paper.

3 MOTIVATION

Word Clouds are effective tool to represent what is emphasized in your text. This can be used for displaying the field of interest of different professors. For example, if you construct a word cloud for an Algorithm professor, you can instantly see what words are used more frequently in his research papers. This can help you spot words that perhaps he is focusing on the most or other key areas. Just like an info-graphic and other compelling pictorial representations, they:

- Make an impact
- Are easy to understand
- Can easily be shared
- ease in handling papers published by the professor

These Word-Clouds are clickable. On clicking any word in the Word-cloud, it displays the list of Research Papers that contain that word so that the user can find

the topics of his/her interest that intersect with professor research interest and from there he/she can redirect to related Research paper directly.

4 PRIOR WORK

There is no prior work done in representing the research paper in the form of Word-Cloud and doing word-wise classification of research papers of a professor. Also ease in handling the research paper by a professor.

5 FEATURES

- Clickable Word-Cloud classifying Research papers word-wise.
- Each word will display the list of Research papers containing that word.
- Each Research Paper in the list will be a link pointing directly to the Research Paper.
- Ease in handling papers published by a professor
- good GUI to handle research papers.

6 TECHNOLOGIES USED

Django

Django is a open source Web framework that encourages rapid development and clean, pragmatic design. Django takes security seriously and helps developers avoid many common security mistakes. Some of the busiest sites on the Web leverage Django's ability to quickly and flexibly scale. It takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel.

Python Tools and Libraries

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages. It supports functional and structured programming methods as well as OOPs. It can be used as a scripting language or can be compiled to byte-code for building large applications. It provides very high-level dynamic data types and supports dynamic type checking. It can be easily integrated with C, C++, COM, ActiveX, CORBA, and JavaScript. Libraries used in the project are:

Beautiful Soup

Beautiful Soup is used for extracting Research Papers from homepage of professors given its URL. Beautiful soup is basically used for web scrapping. Web scraping is the process of downloading data from websites and extracting valuable information from that data.

PDF-Miner

PDF-Miner is a tool for extracting information from PDF documents. Unlike other PDF-related tools, it focuses entirely on getting and analyzing text data. PDF-Miner allows one to obtain the exact location of text in a page, as well as other information such as fonts or lines. It includes a PDF converter that can transform PDF files into other text formats.

Pattern.en

The pattern.en module contains a fast part-of-speech tagger for English (identifies nouns, adjectives, verbs, etc. in a sentence), sentiment analysis, tools for English verb conjugation and noun singularization pluralization, and a WordNet interface. This library helps in selecting valid words from the papers and merging frequency of similar words such as merging plural form of word to its singular form.

JavaScript, HTML and CSS

JavaScript is used in the frontend. JavaScript is a high level, dynamic, untyped, and interpreted programming language. JavaScript is used for displaying the words, making it clickable and finally showing the links to Research papers.

7 WORKING PROCESS

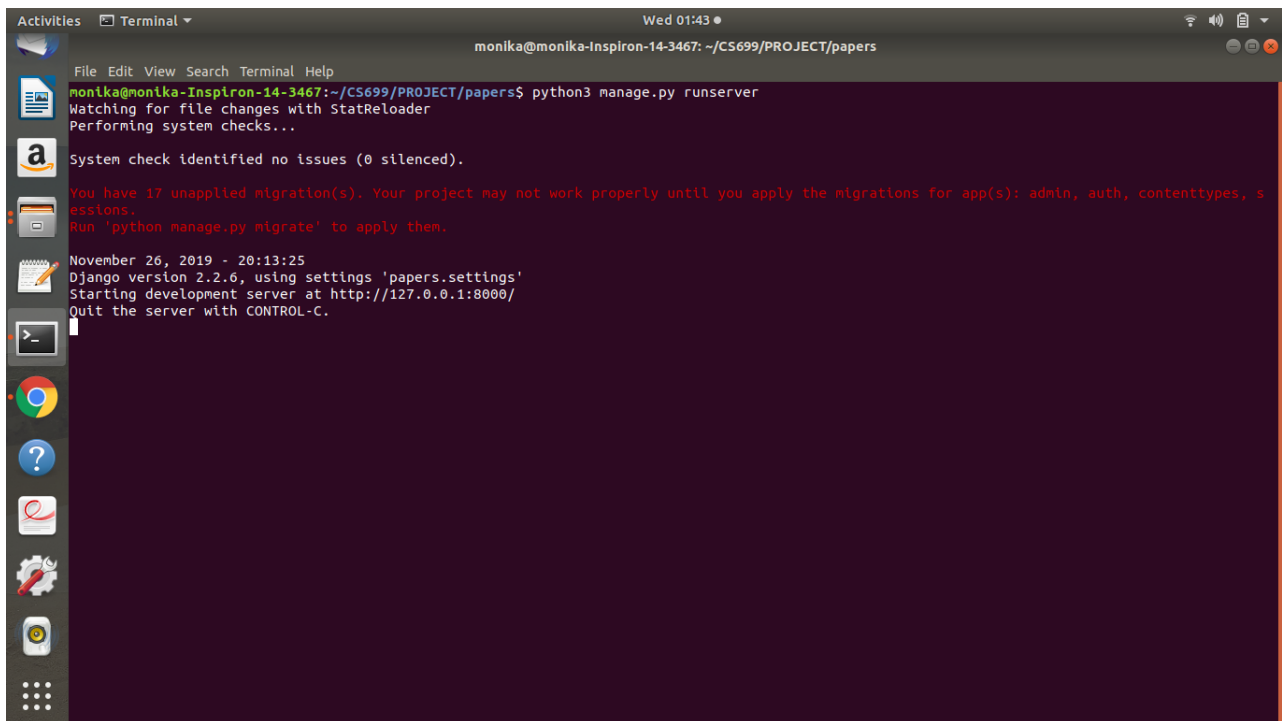
This is a web application with easy to use interface. Application will ask for the URL of the homepage of professor. On providing the URL, application will extract the link for actual page of the publications. Then, extract all the links for downloadable PDFs for the given professor by reading the structure of the professor's homepage and thus finally downloading the PDFs in the machine using the extracted links. Frequency counter is applied on each research paper to count the word with maximum frequency and thus finding 50 top most occurring words in the research papers while removing all the useless words such as adjectives, conjunctions, articles, etc termed as stop-words. We are using above 1000 stopwords to extract meaningful words. It will then create a final CSV of the top most 50 words on the basis of frequency. This CSV is used to generate a word-cloud by assigning weights to words as per their frequencies and display it in the web app. This word-cloud is clickable. Each word in the word-cloud is a clickable link. On clicking any word in the word-cloud, application will display the list of Research papers of the professor which contain that word in it.

8 IMPLEMENTATION

To run Research Papers Easy Access follow following steps:

1. To host server open terminal in the given directory and give command

`python3 manage.py run server`



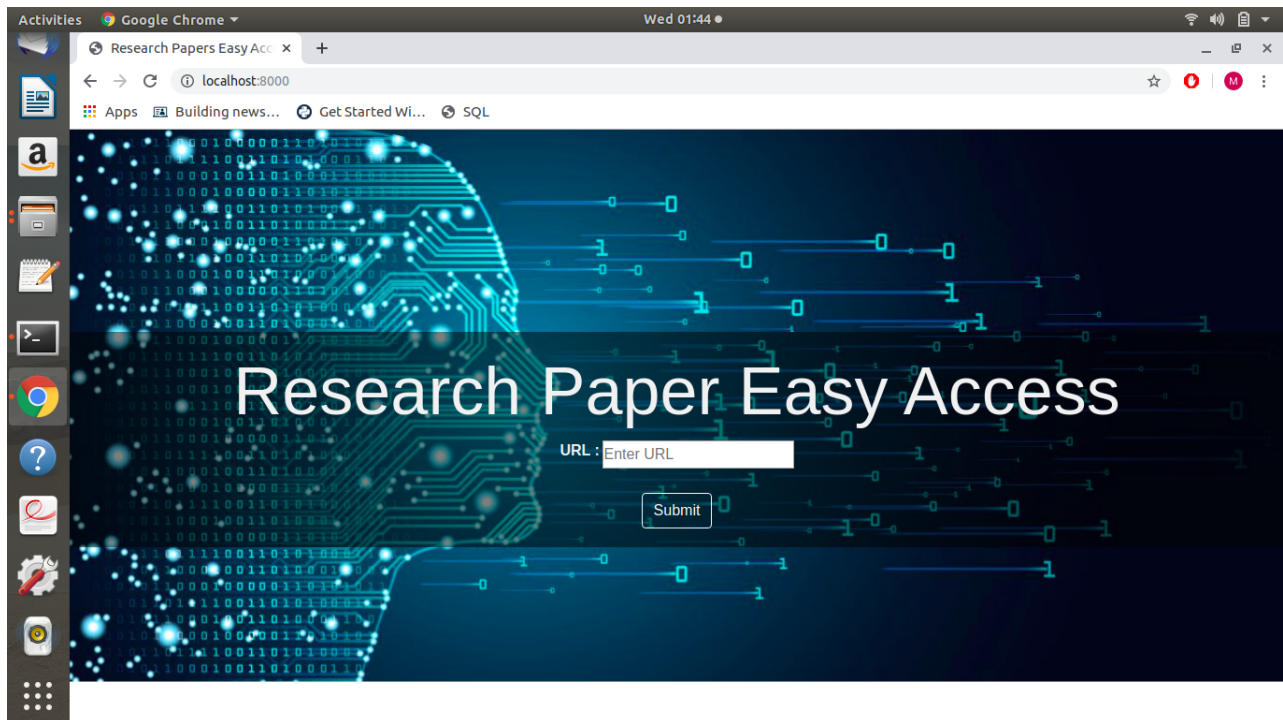
```
monika@monika-Inspiron-14-3467: ~/CS699/PROJECT/papers$ python3 manage.py runserver
Watching for file changes with StatReloader
Performing system checks...

System check identified no issues (0 silenced).

You have 17 unapplied migration(s). Your project may not work properly until you apply the migrations for app(s): admin, auth, contenttypes, sessions.
Run 'python manage.py migrate' to apply them.

November 26, 2019 - 20:13:25
Django version 2.2.6, using settings 'papers.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CONTROL-C.
```


-
2. Now open any Browser and go to *localhost:8000*. This will display Homepage of Research Paper Easy Access.



-
3. Enter URL of homepage of a professor. Currently we have implemented it on seven professors. They are:

Pushpak Bhattacharya : <https://www.cse.iitb.ac.in/pb/>

Rohit Gurjar : <https://www.cse.iitk.ac.in/users/rgurjar/>

Varsha Apte : <https://www.cse.iitb.ac.in/varsha/>

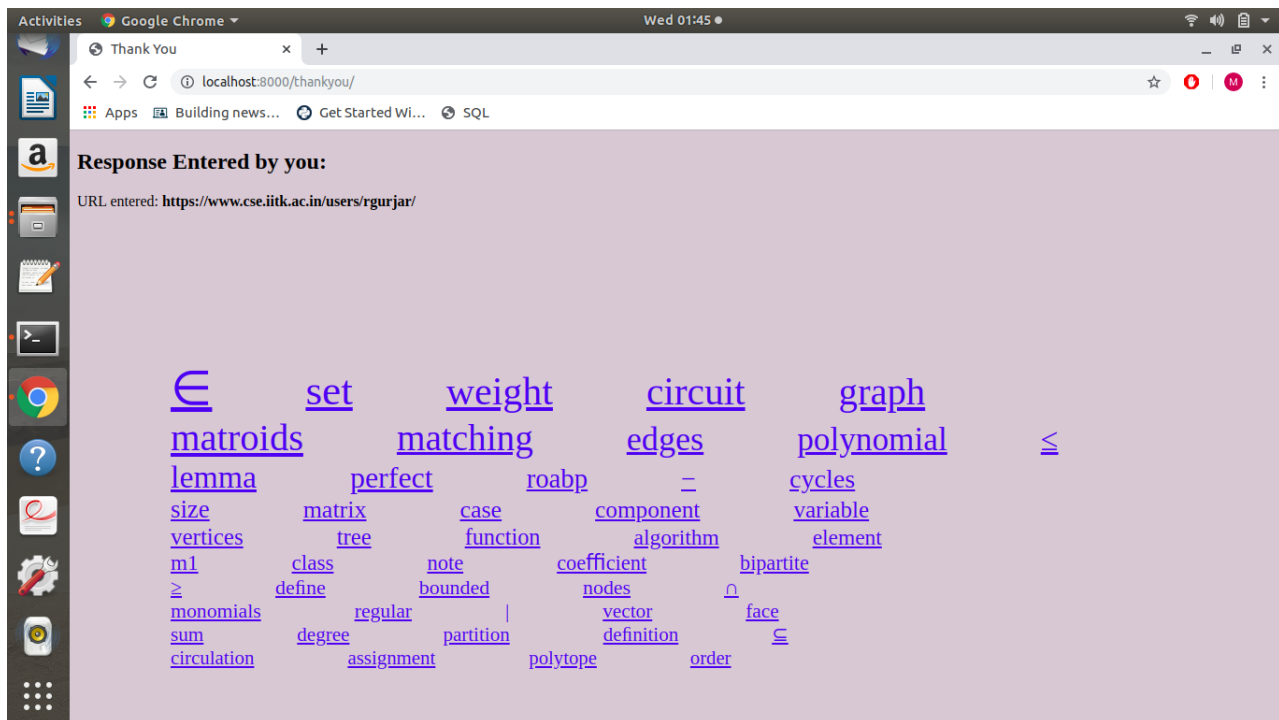
Om Damani : <https://www.cse.iitb.ac.in/varsha/>

Mythili Vutukuru : <https://www.cse.iitb.ac.in/mythili/>

Ajit Diwan : <https://www.cse.iitb.ac.in/aad/>

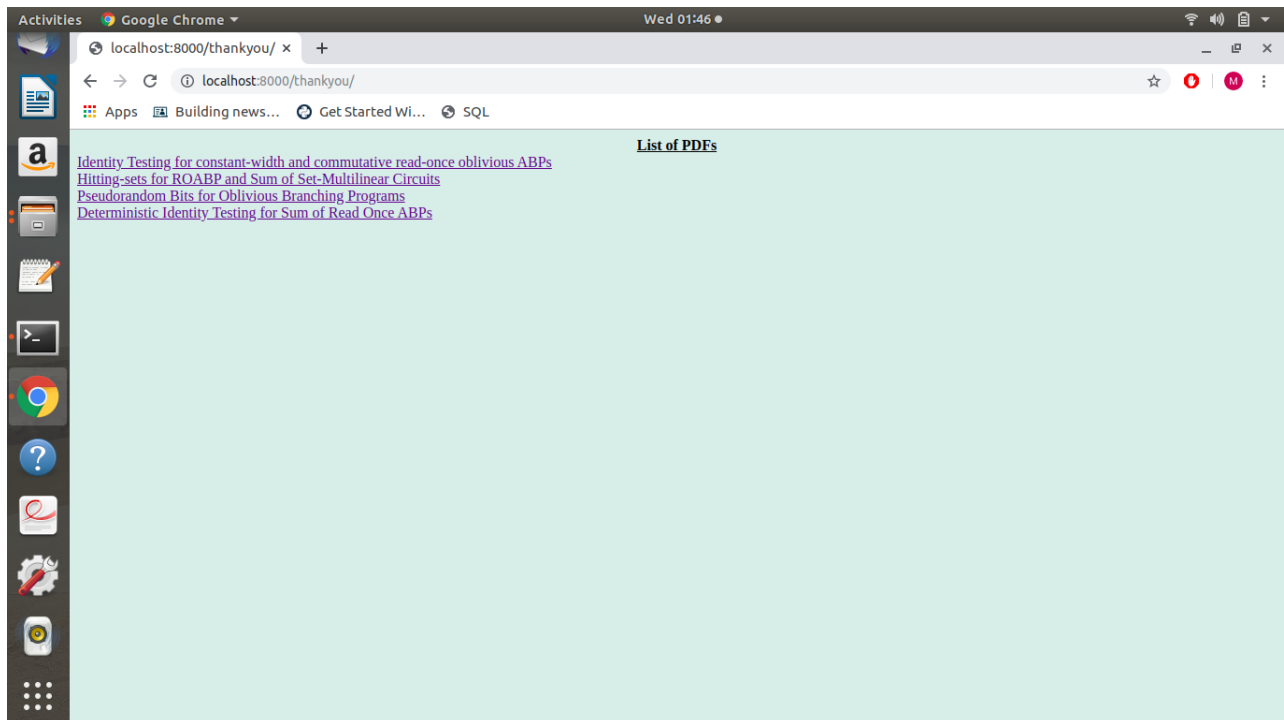
Ganesh ramkrishnan : <https://www.cse.iitb.ac.in/ganesh/>

On entering URL you will get a word-cloud of research paper of that professor.



Word-Cloud of Prof. Rohit Gurjar

-
4. Now on clicking any word, It will display the links to Research papers containing that word.



9 FUTURE SCOPE

We have implemented Research Paper Easy Access on seven professors till now, so implementing it on all the professors will be the next step but some professors are using DBLP for displaying their research paper so for that next we will extract research papers from DBLP links.

10 CONCLUSION

The Word-Cloud for research papers is going to be really beneficial from the student's point of view. Earlier, it was very time consuming task to find the professor and his/her Research papers that intersect with ones research interest or for some academic work. Thus, it provide easier GUI based web application that it is adaptive and reliable. This application along with reducing the hard work in finding some research paper provide many more advantages such as speedy analysis of professor's research interest, ease in handling the research paper by professor, word-wise classification of research paper.

11 REFERENCES

- Prof. Kavi Arya
- Senior Teacher Assistant Diptesh Kanojia
- www.w3schools.com
- www.cse.iitb.ac.in
- www.codingforum.com
- geeksforgeeks.org
- stackoverflow.com