A Search Engine To Find CS Professors

Akshat | Aslam | Gaurav | Shamir

Features:

1. Search For Professors.
    a. By Research Interest.
    b. By University Name
    c. By Professor Name
2. Ranking the Professors.
3. Summary Of Professors.
4. Similar Research Interests

Data Collection:

1. Obtained Google Scholar Ids of around 18000 Professors from CS Ranking.

2. Used Python Based Scholarly Library to get the meta data like Name, Profile Picture, H-index, Total Citations, List of Publications, etc.


```
{'affiliation': 'Professor, IIT Gandhinagar',
 'citedby': 7752,
 'citedby5y': 3907,
 'email_domain': '@iitgn.ac.in',
 'filled': False,
 'hindex': 29,
 'hindex5y': 23,
 'i10index': 48,
 'i10index5y': 38,
 'interests': ['algorithms', 'data mining', 'social networks'],
 'name': 'Anirban Dasgupta',
 'scholar_id': 'plJC8R0AAAAJ',
 'source': 'SEARCH_AUTHOR_SNIPPETS',
 'url_picture': 'https://scholar.google.com/citations?view_op=me
photo&user=plJC8R0AAAAJ'}
```


Ranking Professor:

- Each Professor has a score.
- Professors are ranked based on the score.
- Higher the score more the ranking.

$$score = \frac{\alpha * (prof_{TotalCitations})}{a} + \frac{\beta * (prof_{CitationsFiveYear})}{b} + \frac{\gamma * (prof_{h-index})}{c}$$

Normalization Numbers:

- a = 520787 (Max value of total citations)
- b= 353903 (Max value of total citations in last five year)
- c = 345 (Max value of H-index)

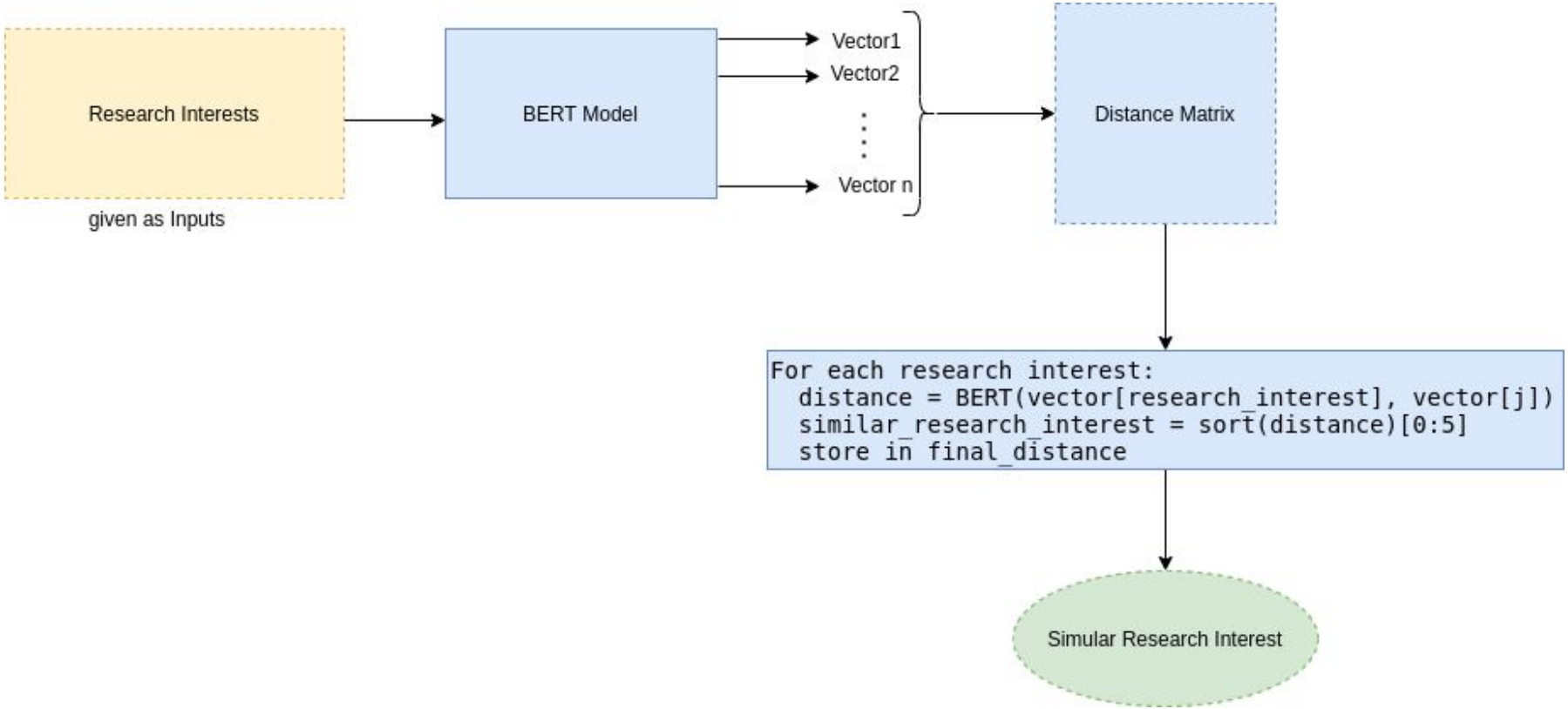
α = 0.3
β = 0.2
ɣ = 0.5

Summary Generation:

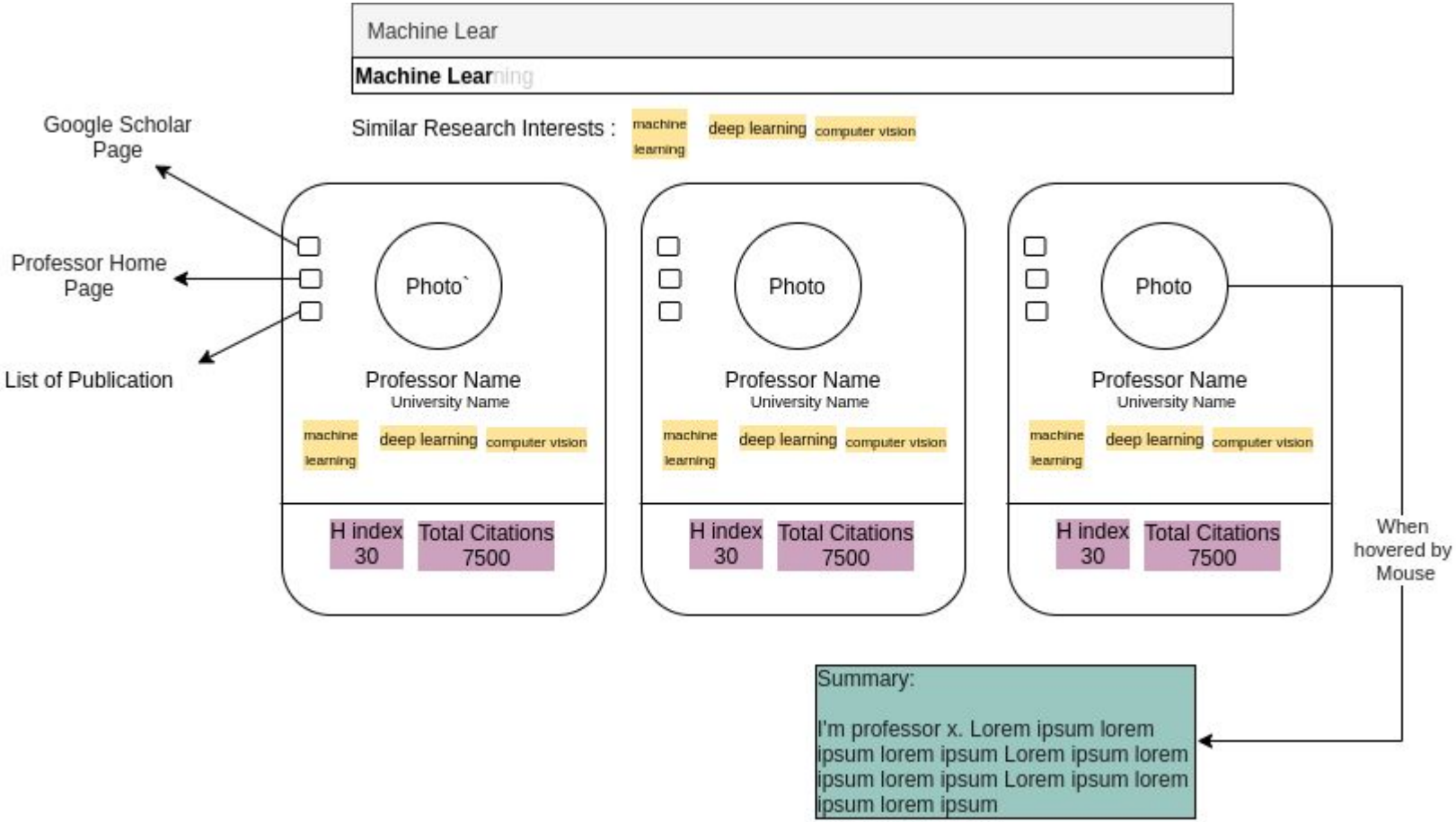
- Python’s pysummarization library was used to generate summaries from Homepage URLs of the professors.
- It uses natural language processing to summarize the webpage.
- This library applies accel-brain-base to implement **Encoder/Decoder based on LSTM** improving the accuracy of summarization by **Sequence-to-Sequence**(**Seq2Seq**) learning.


```
    "U2NUj90AAAAJ": " I am MAYANK SINGH, an Assistant
Professor in the Department of Computer Science and
Engineering, Indian Institute of Technology, Gandhinagar.
I joined IITGN as Assistant Research Professor on 2nd July 2018.
Currently, I am leading the Computational Linguistics and
Complex Social Networks Group (LINGO. My primary research
interests include Natural Language Processing, Text Mining,
Information Retrieval, and Evolving Large-scale Networks.
Specifically, I am highly interested in application areas
such as information extraction (both text and visual) from
scientific articles, curating and analyzing social media
text in resource constraint Indian languages, citation
networks, and Twitter and WhatsApp user interaction networks."
}
```

Similar Interest Using BERT:



Interface:


Machine Lear
Machine Learning
Google Scholar Page
Professor Home Page
List of Publication
Similar Research Interests :
machine learning
deep learning
computer vision
Photo`
Photo
Photo
Professor Name
University Name
machine learning
deep learning
computer vision
H index 30
Total Citations 7500
When hovered by Mouse
Summary:
I'm professor x. Lorem ipsum lorem ipsum lorem ipsum Lorem ipsum lorem ipsum lorem ipsum Lorem ipsum lorem ipsum lorem ipsum

Database Schema:

invertedIndex

{

"_id" : "research_interest"

"profs": [scholar_id's of prof's]

}

Publications

{

"_id" : "google_scholar_id"

"publications": []

}

profData

{

"_id" : "google_scholar_id" ;

"Name" : " " ;

"Affiliation" : " " ;

"Pic_URL" : " " ;

"Homepage_URL" : " " ;

"Research_Interests" : [" ", " ",...] ;

"Google_Scholar_URL" : " " ;

"Cites_per_year" : { " " : __," " : __,...} ;

"Total_Citations" : ____ ;

"Citations_last_5_years" : ____ ;

"hindex" : __ ;

"hindex5y" : __ ;

"i10index" : __ ;

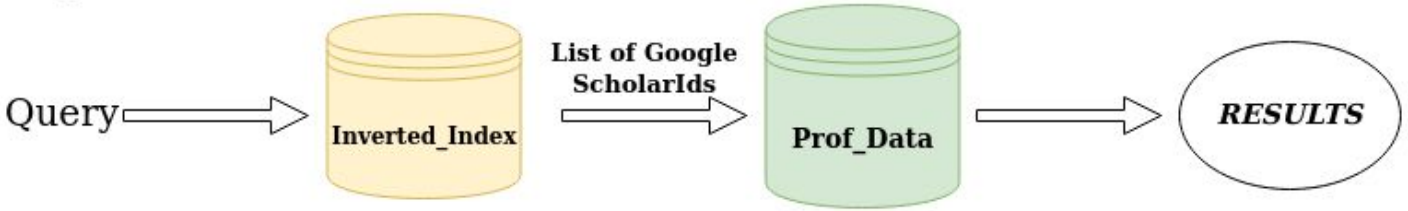
"i10index5y" : __ ;

"Score" : _____ ;

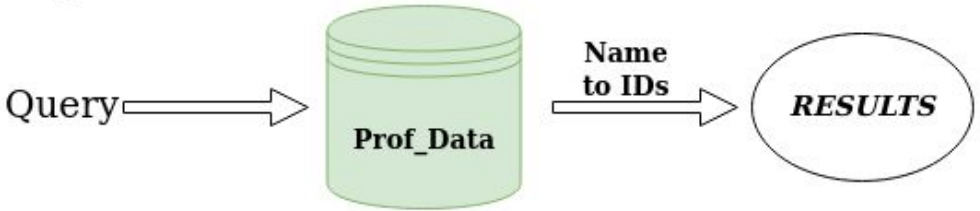
}

Database Queries:

Search By Interest:



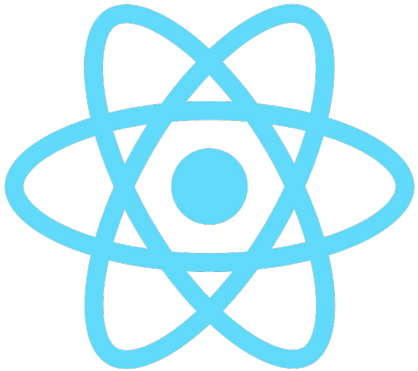
Search By Professor Name:



Search By University:

Tech Stacks Used:

Github: https://github.com/mshamir11/profSearch



React Frontend

MongoDB Database

Flask Backend

Demo

References:

1. Scholarly : https://pypi.org/project/scholarly/
2. Summarization :https://pypi.org/project/pysummarization/
3. BERT Vectors : https://pypi.org/project/sent2vec/