

Progress Report For Professor's Finder

Providing UIUC Professor's Profiles Based on Specific Topics

1. Project Overview

On our platform, based on the user's input of research field or topic, we will display a list of professors with the highest relevance to the search query. This is to assist users in finding desired mentors or collaborators. We are committed to providing optimal search results that caters to multiple professional domains and a user-friendly interface.

2. Progress Made

Data Collection

- Through web scraping, we have obtained two sets of results. Firstly, we have created a txt file for each professor, containing their personal information, research interests, achievements, and courses taught.

Alejandro Dominguez-Garcia.txt	2023/11/1 21:57	TXT 文件
Andreas C. Cangellaris.txt	2023/11/1 21:57	TXT 文件
Andrew Miller.txt	2023/11/1 21:58	TXT 文件
Andrey Mironov.txt	2023/11/1 21:59	TXT 文件
Anu Aggarwal.txt	2023/11/1 21:57	TXT 文件
Arijit Banerjee.txt	2023/11/1 21:57	TXT 文件
Arne Fliflet.txt	2023/11/1 21:57	TXT 文件
Bill O'Brien, Jr..txt	2023/11/1 21:59	TXT 文件
Bin Hu.txt	2023/11/1 21:58	TXT 文件
Brian T. Cunningham.txt	2023/11/1 21:57	TXT 文件
Bruce Hajek.txt	2023/11/1 21:58	TXT 文件
Can Bayram.txt	2023/11/1 21:57	TXT 文件
Chandrasekhar Radhakrishnan.txt	2023/11/1 21:59	TXT 文件
Chester S. Gardner.txt	2023/11/1 21:58	TXT 文件
Daniel M. Liberzon.txt	2023/11/1 21:58	TXT 文件
David M. Nicol.txt	2023/11/1 21:59	TXT 文件
Deming Chen.txt	2023/11/1 21:57	TXT 文件
Dimitrios Katselis.txt	2023/11/1 21:58	TXT 文件

Fig1. Professor information documents

- Secondly, we acquired another index json file containing information or links that will be displayed for each professor on the webpage, like names, department, photo URL, etc.

```
{
  "name": "Anu Aggarwal",
  "department": "ece",
  "title": "Teaching Assistant Professor",
  "photo_url": "https://ws.engr.illinois.edu/directory/viewphoto.aspx?id=96320&s=400&type=portrait",
  "home_page_url": "https://ece.illinois.edu/about/directory/faculty/agganu",
  "doc_name": "anu-aggarwal"
},
{
  "name": "Narendra Ahuja",
  "department": "ece",
  "title": "Research Professor, formerly Donald Biggar Willett Professor ",
  "photo_url": "https://ws.engr.illinois.edu/directory/viewphoto.aspx?id=30761&s=400&type=portrait",
  "home_page_url": "https://ece.illinois.edu/about/directory/faculty/n-ahuja",
  "doc_name": "narendra-ahuja"
},
{
  "name": "Jont Allen",
  "department": "ece",
  "title": "Professor",
  "photo_url": "https://ws.engr.illinois.edu/directory/viewphoto.aspx?id=3617&s=400&type=portrait",
  "home_page_url": "https://ece.illinois.edu/about/directory/faculty/jontalle",
  "doc_name": "jont-allen"
},
}
```

Fig2. Professor index

Front-end Development

- Utilizing the FastAPI framework for the front-end development of the website. Completed the design of the user interface and implemented basic functionalities, including search, display of professor listings, and navigation to individual professor detail pages.

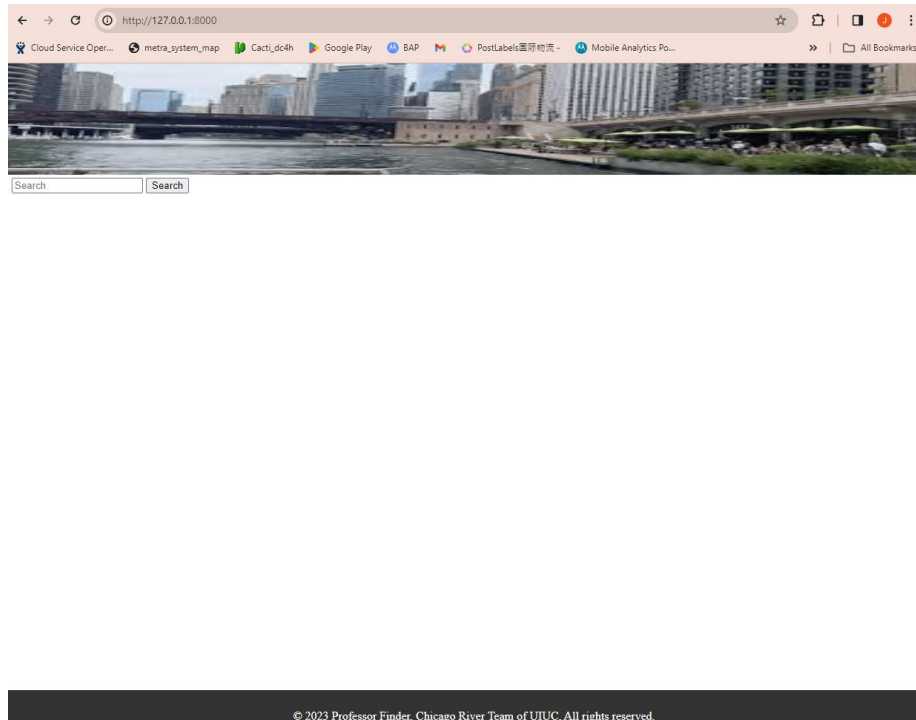


Fig3. Search page

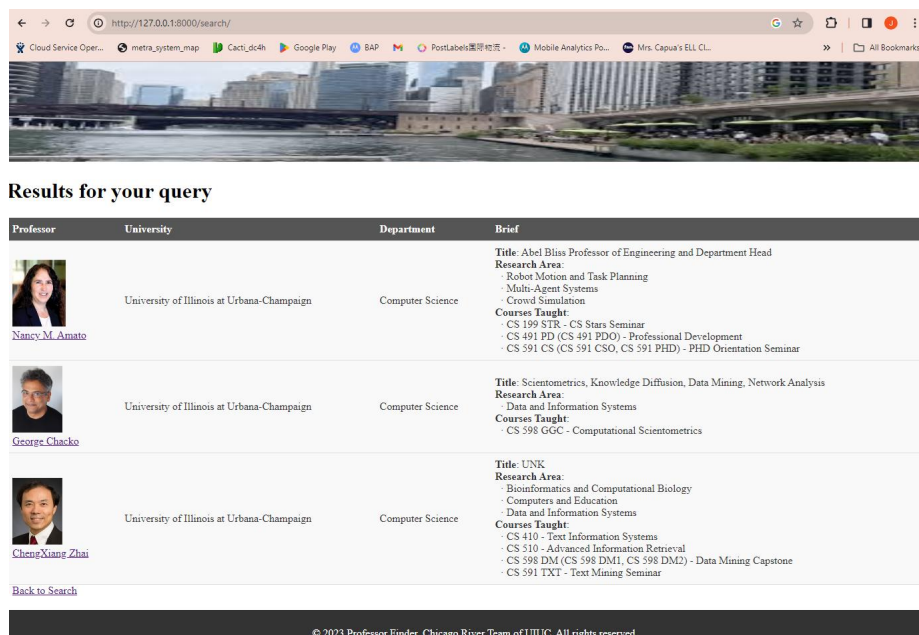


Fig4. Result page

3. Remaining Tasks

Data Processing and Algorithm applying

Implement a TF-IDF model and/or BM25 model for ranking professor's relevance with user-inputted topics.

More User-Friendly Interaction

We will optimize the design of the search interface and the aesthetics of the overall website.

Testing

After the completion of website development, design test cases and conduct comprehensive testing to ensure the usability and quality of the website.

There are also some work pending for exploration such as implementing synonym and near-synonym queries, as well as automatic correction of typo errors in searches.

4. Challenges/Issues

Compatibility Issues

The websites of different colleges have differences in layout and content, and naturally, their source codes vary, making a single web scraping script unable to retrieve information from all college websites. Currently, my web scraping script is only compatible with the websites of the nine colleges within the College of Engineering. In the future, compatibility improvements will be considered.

Optimizing Search

To enhance the optimization of search results, we will attempt to use specific term collections, matching search query terms with synonyms or near-synonyms to provide more comprehensive and reliable results. Additionally, we are exploring the challenge of automatically correcting typo errors, aiming to make user queries more convenient and intelligent.