

ExpertSearch - Extracting relevant information

As stated by the project topics document, the goal of this project is to enhance the ExpertSearch app (<https://github.com/CS410Fall2020/ExpertSearch/>) to convert the unstructured text in faculty webpages into more structured text, so that we can more accurately extract email and faculty names from the faculty member bios crawled data.

Progress

- Setup a vm running a Linux distro and set it up with Python 2.7
- Get code from git repo
- Run up with gunicorn and necessary packages on localhost
- Started modifying the code around the ranker for the structured text update

Remaining tasks

- Determine best text structure
- Update results structure in server.py's search method to use this new structure
- Push code to this repo
- Host website on the cloud

Challenges

- This App is made to run using Python 2.7 on Linux. Had some issues setting up the vm I'm using to work on and getting sure python, pip, gunicorn and other packages were installed and running using the correct version
- I can see potential issues related to python 2.7 when deploying to the cloud, but Python 2.7 still has support until next year, enough to deliver the project.