

Project Proposal

Project Name: *2.2 ExpertSearch System*

Team member:

Ri Xu (Captain), NetID: rixu2

Jinou Yang, NetID: jinouy2

The system we are choosing is Option2: Improving A System, specifically the ExpertSearch system.

The main dataset we will be using is the faculty bio from **MP2.3**. Our project will make improvements to the ExpertSerach system in the following ways:

1. Add extractions on experts' phone numbers using regex pattern matching. (*5 hours*)
2. Add extractions on experts' research interests in the format of keywords using topic mining. (*10 hours*)
3. Improve search results preview by adding phone numbers and keywords. (*5 hours*)
4. Build experts recommendation system based on users' search queries and browsing history. Users' interaction data can be stored and retrieved from browser cookies, BM25 can be used to find the top relevant experts to recommend. (*20 hours*)

We will implement web scraping scripts to extract phone numbers in python and store them in a local file, which will be shown in the preview on frontend. Another python script is to perform topic mining on faculty bio and extract top research interest keywords and store them in a local file. Search related function and ranker will also be implemented in python. For web frontend improvements, HTML, CSS and Javascript will be used to achieve the objectives.