

Progress Report: ExpertSearch System Enhancement Project

Team Captain: Patrick Asztabski (paszta2)

Sean Koo (seanhk2)

Kumarswamy Valegerezpura (kv20)

Kavin Krishnasami (kavinsk2)

Progress Summary:

System review & planning (10 hours): Completed

- Reviewed existing ExpertSearch system.
- Identified areas for enhancement

Discovery and processing data (10 hours): In Progress

- Collected faculty webpage URLs from MP2.3 on Coursera
- Compiled directory page URLs from the sign-up sheet and negative example URLs from various sources

Implementing and Improving Algorithm (20 hours): Scheduled

- Plan to implement search engine enhancements for URL classification
- Develop NER techniques combined with topic mining to extract faculty structured data

Extracting structured information (15 hours): Not Started

- Awaiting completion of the algorithm implementation phase.

Integration and testing ExpertSearch (15 hours): Not Started

- Integrating new algorithms and techniques into system
- Comprehensive testing using new and known data sets of URLs

Preparing demo/report & documentation (10 hours): Not Started

- Prepare detailed report and documentation
- Plan for live demonstration comparing current enhancements to existing system.

Current Challenges:

- Automating identification of faculty directory pages specific to CS is more complex than anticipated
- Balancing efficiency in web scraping and data extraction.

Next Steps:

- Finalize data processing
- Begin development of our algorithms and new techniques to improve existing system and begin integration/testing