Project Progress Report

Which Tasks Have Been Completed?

- Built a list of the unique characteristics of faculty directory pages and faculty webpages URLs to help in identifying them
- Implemented a function that classifies a page as directory vs. non-directory
- Created a list of the most common attributes that do not appear in faculty directory pages and faculty webpages URLs to avoid crawling them
- Implemented a function that classifies a URL as a faculty webpage or not

Which Tasks are Pending?

- Implement a crawler function that gets the list of the faculty webpages
- Test the added functionality, which is a crawler that automatically crawls a faculty directory webpage
- Create project documentation

Any challenges faced?

There have been no challenges faced.