

Project Proposal

What are the names and NetIDs of all your team members? Who is the captain?

Individual Project:

Kamal Mahmood (mahmood3) - Captain

What system have you chosen? Which subtopic(s) under the system?

System: System Extension

Sub-Topic: Automatically Crawling Faculty Webpages

Briefly describe any datasets, algorithms or techniques you plan to use.

I plan to use the dataset of the faculty directory pages and faculty webpage URLs from MP2. I plan to use the text classification technique.

If you are adding a function, how will you demonstrate that it works as expected? If you are improving a function, how will you show your implementation actually works better?

I plan to add a new function. I will demonstrate it by giving a tutorial presentation of the added functionality.

How will your code communicate with or utilize the system? It is also fine to build your own systems, just please state your plan clearly.

I plan to make my crawler tool be its own system.

Which programming language do you plan to use?

Python

Please justify that the workload of your topic is at least $20 \cdot N$ hours, N being the total number of students in your team. You may list main tasks to be completed and estimated time for each task.

Estimated Main Tasks	Estimated Time
Faculty Directory Pages Text Classification	6 hours
Faculty Webpage URLs Text Classification	5 hours
Build Crawling Function	7 hours
Progress Report	2 hours
Testing Code	4 hours
Project Documentation	4 hours
Prepare and Edit Tutorial Presentation	4 hours
Total:	32 hours