1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.
2. What system have you chosen? Which subtopic(s) under the system?
   1. If it is not listed above, how is it related to the class?
3. Briefly describe any datasets, algorithms or techniques you plan to use
4. 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?
5. 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
6. Which programming language do you plan to use?
7. Please justify that the workload of your topic is at least 20\*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

**Proposal**

1. Joseph Babbo, jbabbo2
2. ExpertSearch System; Automatically crawling faculty webpages
3. Use the MP2 signup sheet as a dataset for directory page URLs alongside crawled miscellaneous university website pages, to get non-directory page URLs. This data would be used to create a university directory page classifier. Additionally, as time allows, a similar thing could be done with faculty webpage URL. These classifiers will then be wrapped in a frontend, where when the top page of a university is pointed to, the faculty and directory pages are correctly returned.
4. To show that this new classifier works as expected our dataset will be split into training and test sets. Additionally, once the wrapper is completed the function can be demonstrated by simply pointing it to a university website.
5. The code will not communicate with or utilize the ExpertSearch system. However, it will be written in a way where that it could hopefully in the future be easily integrated into the system.
6. Python
7. Scrapping Relevant Data – 5 hrs.

Creating Functioning Classifiers to Identify Pages – 15 hrs.

Wrapper to point at University URL and Return Correct Pages – 5 hrs.