

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.
 - Team Captain: Sai Ranganathan (sr50)
 - Mriganka Sarma (ms76)
 - Zacharia Rupp (zrupp2)
2. What system have you chosen? Which subtopic(s) under the system?
 - We have chosen to improve ExpertSearch.
3. Briefly describe the datasets, algorithms or techniques you plan to use
 - Datasets:
 - Faculty dataset scraped from MP2.1 for positive examples.
 - Scrape of Alexa Top 500 Domains for negative examples.
 - Techniques:
 - Topic Mining
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?
 - Additional functionality:
 - Topic Mining
 - If it works, users will be able to refine queries by selecting topics from a topic cloud
 - Impact on ndcg@k
 - Improved functionality:
 - Email extraction
 - If it works, more faculty members will have email addresses associated with them.
 - Improved UI:
 - More granular query refinement
 - Top-k associated topics listed under individual faculty members
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
 - Our code will build on the ExpertSearch code by:
 - adding a topic mining function
 - improving email extraction
 - automating scraping process
 - improving UI
6. Which programming language do you plan to use?
 - Python
 - JavaScript
7. 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 the main tasks to be completed, and the estimated time cost for each task.
 - Main tasks:
 - Automatic crawler to identify faculty directory pages (10+ hrs)
 - Automatic crawler to identify faculty webpage URLs (10+ hrs)
 - Improving functionality:
 - Email extraction (10+ hrs)
 - Adding functionality:
 - Topic mining (10+ hrs)
 - UI Improvements:
 - Query refinement options (10+ hrs)

- Topic cloud from mined topics associated with retrieved faculty members.
- Top-5 topics associated with faculty member (5+ hrs)
 - Displayed at the top of the bio excerpt
- Prepopulated email content when a user clicks on a faculty member's email address (5+ hrs)
 - E.g.
"Dear <Faculty Name>,"

It's a pleasure to have gone through some of your research articles. I'd like to connect with you for discussing some ideas in the <Research Area>.

I hope to hear from you soon."