1. **What are the names and NetIds of all your team members? Who is the captain?**

   Team Captain: Patrick Asztabski (paszta2)
   Sean Koo (seanhk2)
   Kumarswamy Valegerepura (kv20)
   Kavin Krishnasami (kavinsk2)

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

   We chose the ExpertSearch system. Our subtopics are:
   - Automatically crawling faculty webpages
   - Extracting relevant information from faculty bios
   - Given enough time, implementing Generative Technologies (LLM, GPT)

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

   Datasets:
   - Faculty webpage URLs under MP2.3 on Coursera
   - Directory page URLs from sign-up sheet
   - random URLs from news sites, product sites, etc. for negative examples

   Algorithms/Techniques:
   - Machine learning classifiers for URL classification
   - Advanced NER techniques combined with topic mining for structure extraction

4. **If you are adding a function, how will you demonstrate that it works as expected?**

   We will conduct tests on both seen and unseen data to verify implementation. If given enough time, visual demonstrations showcasing improved results will be presented

5. **How will your code communicate with or utilize the system?**

   Our plan is to build upon the existing system by integrating new advanced techniques and possibly classifiers, thus improving the system through accuracy and robustness.
‘

6. **Which programming language do you plan to use?**

   We plan to use Python given not only the extensive libraries but the former system was built within Python. We will use it for web scraping, machine learning, and data processing.

7. **Please justify the workload of your topic is at least 20*N hours:**

   4 members * 20 = 80 hours of work

Tasks and Estimated:

- System review & planning - 10 hours
- Discovery and processing data - 10 hours
- Implementing and Improving Algorithms - 20 hours
- Extracting structured information - 15 hours
- Integration and testing ExpertSearch - 15 hours
- Preparing demo/report & documentation - 10 hours

Total = 80 hours, but additional refinements, debugging, and review of each other's work will account for possibly hours.

8. **At the final stage of our project, we will deliver:**

    - A comprehensive codebase with appropriate documentation detailing any improvements and implementations
    - A live demonstration comparing our enhancements to the existing system