Course Project Proposal

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.

- Shai Yusov (syusov2@illinois.edu).
  It's a team of one and I will be the captain.

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

- System: EducationalWeb System
- Subtopic: Automatically creating teaching material for in-demand skills - Identifying in-demand skills

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

- I plan to crawl and analyze job postings for software engineers in New York, NY, on the job board Indeed. After obtaining these postings, the dataset will be pre-processed and fed into a component that will be responsible for identifying the emerging keywords/topics. The identification process will likely involve keyword extraction and/or 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?

- I will demonstrate that the function identifies in-demand skills by running it on the input dataset, obtaining the top emerging keywords/topics, and validating the results through human judgement and known industry trends.

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

- I plan to build a standalone system/component for this project. Specifically, the system will ingest a dataset and output the top emerging keywords/topics from that dataset. Although outside the scope of this project, this functionality is generic and can be plugged into other tools so long as the input specification to the system is decoupled from the source or actual content and the outputs are emitted in a standard, easy-to-digest format.

6. Which programming language do you plan to use?

- I plan to use Python for this project.

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.

- Writing, testing, and executing a customized web crawler: at least 10 hours
- Investigating and defining the right algorithms/techniques: at least 5 hours
- Writing and testing the component performing the identification: at least 10 hours