

Requirements

In your proposal, please answer the following questions:

- 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.**

Nikita Volynskiy (nikitav2) Captain

James Rockey (jrockey2) Project Designer

- 2. What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?**

For our project, we want to create a program that will rank different jobs based on a user's resume. We are interested in this project because we are looking for jobs, and we want to figure out which jobs best fit our skillset. Our planned approach is to implement a back end text retrieval system, treating job descriptions as documents and a resume as a query. We plan on utilizing public job descriptions datasets available on Kaggle. We also plan on utilizing one of the available python libraries such as Metapy to rank the different documents based on a user's input resume. We plan on implementing different versions of this system using different ranking algorithms so we can evaluate which system is the most accurate. The expected outcome of this project is creating a tool that will allow us to find the best jobs to apply for so we can minimize the time we spend searching for jobs.

We are going to evaluate our system by measuring the accuracy, efficiency, and usability of the application. To measure the effectiveness of the different algorithms we plan on comparing the precision and recall using different F-measures. Since we care about the relevance of the suggested job descriptions and also that our system has good recall and does not exclude any relevant jobs we believe that using the F1 score will work well, but we are open to exploring different measures as we get deeper into the project. We will create a set of test job documents, and run our system with resumes tailored towards different industries and academic disciplines so we can see how our system works. To measure the efficiency of our system we plan on tracking the runtime of running these tests. Once we are satisfied with the performance, we can work on improving the usability by sharing this project with our friends and asking their input.

3. Which programming language do you plan to use?

We plan on using Python to implement this program and plan on it mainly being a command line application. However if we have extra time we want to implement a front end using Javascript, HTML, and CSS.

4. 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.

Project Tasks Estimated Time:

1. Implementing different ranking algorithms (10hrs)
2. Work on Command Line interface (25hrs)
3. Create test datasets (8hrs)
4. Run tests on different versions of our system to measure effectiveness (8hrs)