Individual Capstone Assessment

Our project aims to create a search engine on a database of medical documents related to COVID-19 and a web interface for users to query the database using the search engine. Through the project, we want to help users get access to information about academic work related to COVID-19 more easily. We can also use the tool to collect user interactions, which can help us improve our search algorithm. Our project is a collaboration between Computer Science and Biomedical disciplines. For our part (Computer Science), we need to implement the search algorithm and the web interface. We use Elasticsearch and several tools for the search algorithm and Flask for the web interface.

What I learned through my CS courses are crucial for the development of this project. While I consider all courses are relevant to some extent, there are four most relevant courses: Python Programming (CS2021), Design and Analysis of Algorithms (CS4071), Database Design and Development (CS4092), and Software Engineering (EECE3093C). Python Programming is essential since the whole project is implemented in Python. Design and Analysis of Algorithms provides me necessary knowledge in algorithm development and complexity analysis to implement the search algorithm and backend logic. Database Design and Development helps me know how to design the database (documents and web database) for the project and use SQL to interact with the database. Software Engineering helps me know how to approach a big project and design the programming paradigms to make it organizable.

My co-op experiences are directly related to this project. In my last three co-ops, I worked for Dr. Danny Wu, who is now my project advisor. Since this project is directly from his lab, Dr. Wu can help me immensely in the process. While I worked with Dr. Wu, I learned many technical skills necessary for this project like Elasticsearch and text-retrieval algorithms. As I worked on many Biomedical-Informatics projects during my co-ops, I have also been familiar with some biomedical background knowledge for this project. I also frequently worked in a 2-man team during my co-ops, so this time I have experience with how to organize tasks and communicate with my teammate.

I feel motivated to participate in this project as it is related to my academic and work experiences. I have always been fond of algorithm development, and for this project I can utilize my algorithm skills to implement a good search engine. I am also interested in pursuing a post-graduate program related to Natural Language Processing, and this project is perfect for me to approach the field by learning some text retrieval and processing techniques. For the project itself, I think it is a very useful application that can help much in COVID-19 research. If the project is made into a final product, I hope we can find users to make good use of the product. Through the project, I can also learn some medical knowledge, which I have been interested in lately after COVID-19.

For the project, our main goal is completing the final product: a web interface that allows users to search for medical documents related to the query terms and saves user interactions into the database. We consider we reach the expected results if we can make a functional website that meets all of our use cases. There are two main tasks of this project: website development (frontend and backend) and algorithm development (pulling data and building a search engine). Since our team has two members, each of us takes care of a main task; my task is algorithm development. We plan to develop the project in an AGILE model - dividing the product into different phases from simple to complete and setting a timeline to complete each phase. If we can complete all phases in the timeline and arrive at the final product, we will consider ourselves having done a good job.