Milestone 0 Report

Vincent Chen, Eunseo Lee, BK Kang, Andrew Lu

Our application is a database-driven web application tailored towards students (Waterloo co-op students are the users of the application) at the University of Waterloo to allow them to explore co-op salaries across a variety of companies, locations, roles/positions, terms, etc, which helps students make informed decisions on which companies and positions they would apply and compare salaries across programs and over time. The dataset we will use is a self-reported co-op salaries datasheet retrieved from a Reddit thread. This dataset contains companies that offer at least one job posting on WaterlooWorks along with their hourly salary and work term number, and additional information such as benefits, year of job posting, and any companies put on a separate blacklist. However since the hourly salary for a company is likely to be an aggregation of individual job positions/roles within the company and many records within the dataset are empty, we will transform the Reddit dataset into a compatible relational format and populate the empty entries with synthetically generated numbers that reflects the average values observed in the current dataset. This populated and transformed dataset will become our production dataset. We (the project group) will be the administrators of the database system and will maintain the performance and availability of the database, ensure backup and recovery, handle blacklisted companies (via constraints and triggers), etc. The functionalities we would like to support are: listing salaries by program, faculty, or location, list top-paying companies by role/position, display salary across terms, flag companies with very low salary, etc.

The platform we will develop our application is MySQL database on the school server. On the backend we will use Java/PHP, Python, and Flask for cleaning up the data and handling MySQL connections and queries, and on the frontend we will use Javascript, HTML, and CSS as it provides a lightweight web-based user interface that can easily be deployed locally or on a server later on. We will also make use of additional libraries and tools such as pandas, matplotlib, React, etc. Users will interact with our application through a web-based interface via a browser or local host network and the UI will have dropdown menus that filters search results by faculty, company, or role, a search bar that searches the database by keywords, and displays output through tables or charts (distribution or trends).

The four members of our team and the work they did as of now are:

- Vincent Chen: Worked on the schema of the database and found the dataset.
- Eunseo Lee: Designed tests, queries, and worked on the README file.

- BK Kang: Setted up the GitHub repository and the code environment.
- Andrew Lu: Worked on the Milestone 0 Report.

GitHub repository: https://github.com/bkctrl/cs348-project