Career Location Compensation Analysis Tool

Steven Palmer*

CS 2XB3
L04

Abstract

Compensation for a particular career can be highly dependent on location, and choosing to work in areas where compensation is the highest may not be a wise decision due to large variations in the cost of living. This leads to a situation where comparisons cannot easily be made at face value. This project will provide a solution to this problem through the development of a software tool that looks at compensation data for a particular career over a large number of locations and automatically adjusts using cost of living data. The software will then compare the outputs and provide the user with a ranked list of the best locations to work in terms of financial reward.

I. Introduction

In many career fields compensation for the same role can vary substantially based on location [1]. Career opportunities in densely populated urban centers, for example, often come with significantly higher compensation packages than those in smaller cities. Choosing to work in locations that offer the highest salary may seem like an obvious choice, but one must consider differences in the cost of living between locations achieve a realistic comparison. This project aims to shed light on this issue through the development of a software analysis tool that uses publicly available employment and cost of living data to compare and rank the best locations to start a career. The sections that follow will provide details on the objectives and scope of the project, the envisioned project design, and the timeline of deliverables that will be required to complete the project.

II. Objectives and Scope

The objective of this project is to produce a software tool used to assess and rank the most viable locations to start a career in a particular field. The software will produce a list of locations that offer the best income potential for a particular career or field using a mixture of employment and cost of living statistics. In addition, the software will include an option to consider the possibility of commuting in order to reduce the cost of living.

The project will be geographically limited to the United States and will only consider income related factors to arrive at a ranking. These limitations may be expanded upon in the future, but such enhancements are beyond the scope of this project.

III. Architecture & Environment

The software will follow an object-oriented design and employ the model-view-controller architecture to achieve independence of modules/subsystems. The information hiding design principle will be used such that interfaces can be developed for every module prior to beginning any implementation. This will allow for the software to be implemented efficiently

*By virtue of submitting this document I electronically sign and date that the work being submitted is my own individual work.

by a team in parallel.

The software will be written in Java, which offers the following benefits:

- Easy to implement object-oriented design
- Built-in cross platform support
- Access to the JUnit testing framework for efficient test case development

No external libraries are expected to be required for the implementation of this project.

IV. Implementation

The software tool will consist of a simple Java applet GUI frontend linked to the algorithm based analysis backend. The applet will allow a user to input their desired career field, specific role (optional), a location and radius to perform the search (optional), as well as a maximum acceptable commute distance. Based on these inputs, a ranking of the best locations will be computed and displayed. A mockup of the envisioned applet is given in Figure 1.

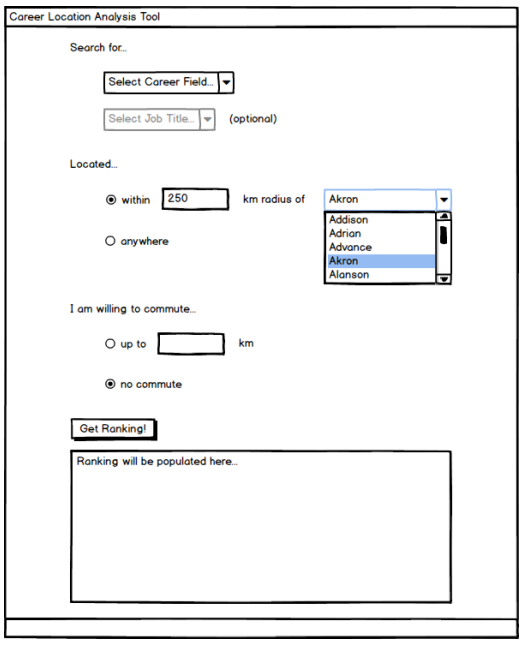

Figure 1: *Applet Mockup*

Open data sets available online will be used in the implementation of the analysis algorithms:

- Occupational employment statistics published by the United States Department of Labor [2] will be used to find location-based average compensation for specific careers
- Location affordability statistics published by the United States Office of the Secretary of Transportation [3] will be used in the assessment of location-based costs of living
- Geographic location data from the GeoName database [4] will be used to create a graph for computing location/commute distances

These data sets are all given as comma separated values.

V. Algorithmic Challenges

Both the employment and the affordability data inputs used by the software contain several different fields. In order to manipulate the data efficiently, the order in which fields are sorted and searched must be carefully considered.

In order to efficiently search for U.S. cities within certain distances of other U.S. cities a graph of the U.S. road network is required. For the purposes of this application, a subset of this network that includes only highways is sufficient and will yield much faster search times. Unfortunately, a public dataset of the U.S. highway system was not found online; instead this will be approximated by creating a graph using the geographic locations of cities. The exact method that will be used to produce this graph is not yet known, but it is expected to present an algorithmic challenge in deciding which edges to include.

In addition to the challenges mentioned above, the size of the input data presents a challenge in itself. The time complexity of algorithm becomes extremely important when

Table 1: *Project Timetable*

Milestone	Deliverable	Completion
1. Software planned and decomposed into modules	Module interfaces	Feb. 21, 2016
2. Input modules fully implemented	Input modules complete and functioning	Mar. 1, 2016
3. Data analysis modules fully implemented	Data analysis modules complete and functioning	Mar. 21, 2016
4. Output modules implemented	All additional modules complete and functioning	Mar 30, 2016
5. Finished implementation	Finished and tested software	Apr. 12, 2016

working with very large data sets. Every algorithm must be carefully thought out such that time complexity does not become a runtime issue.

VI. Deliverables

The software will be broken down into several modules. These modules will include parsers for the input data, data structures with algorithmic methods to analyze the data, as well as modules for the GUI. The implementation of the project will be divided into the following five phases:

1. The first phase of the project will involve planning out the project in detail. The software will be broken down into modules, and interfaces will be designed for each module. The deliverable for this phase will be a set of completed interfaces for all of the software modules.

2. Implementation of the data input modules will begin in the second phase. At the end of this phase, completed modules will allow data to be loaded into memory and ready for manipulation/analysis.

3. In the third phase the modules used in the manipulation and analysis of the data will be implemented. These modules will make up the core of the software, and as such this phase is expected to be the most time-consuming. The end of this stage will result in the completion of the data analysis backend.

4. The fourth phase will involve the implementation of modules associated with the GUI frontend as well as connecting the GUI with the backend that finished in the previous phase. At the end of this phase, a full implementation of the software will be complete.

5. The fifth and final phase will center around quality assurance. Testing will be carried out in this phase to determine whether the software works as intended. Bugs that are discovered during testing will be addressed and corrected. The completion of this phase will yield the release version of the software.

VII. Project Timetable

The deliverables outlined in the previous section are summarized as milestones in Table 1. Preliminary completion dates have been added based on estimations of the expected amount of work involved for phase.

References

[1] E. Torpey. (2015, May). *Same occupation, different pay: How wages vary* [Online]. Available: http://www.bls.gov/careeroutlook/2015/article/wage-differences.htm

[2] U.S. Department of Labor, Bureau of Labor Statistics. (2015, Nov 4). *Occupational Employment Statistics - Employment and Wages* [Online]. Available: http://catalog.data.gov/dataset/occupational-employment-statistics-employment-and-wages

[3] Office of the Secretary of Transportation. (2015, Aug 28). *Location Affordability Index: All Census Places* [Online]. Available: http://catalog.data.gov/dataset/location-affordability-index-all-census-places

[4] GeoNames. (2016, Feb 7). *GeoNames* [Online]. Available: http://download.geonames.org/export/dump/