Alchemist

NSERC Grant Search and Forecasting Tools
Requirements Specification Document

Version 3.0.0
March 30, 2016
Group Number: 29

Danish Khan: 1217176
Haris Khan: 1413511
Michael Bitzos: 1405050
Junhao Wang: 1215428

SFWR ENG 2XB3
Course: Software Engineering Binding Theory to Practice
Department of Computing & Software
McMaster University

Revision History

By virtue of submitting this document we electronically sign and date that the work being submitted by all the individuals in the group is their exclusive work as a group and we consent to make available the application developed through [CS] or [SE]-2XB3 project, the reports, presentations, and assignments (not including my name and student number) for future teaching purposes.

Version 1.0.0

- Completed initial introduction, functional requirements, non-functional requirements, domain definitions and requirements on the development and maintenance process.

Version 2.0.0

- Edited function requirements to better suite the purpose of the application.
- Edited the domain to narrow the field of customers to professors, faculty, staff and students.
- Restricted search queries to be as defined by the NSERC grant datasets when it comes to forecasting for a field of research.

Version 3.0.0

- Added application icon, edited title page to meet requirements.
- Added precise list of functional requirements
- Added precise list of non-functional requirements
- Edited introduction to include domain, stakeholders, knowledge of domain that informs the requirements.

Table of Contents

1 Introduction

The Alchemist toolset is an iOS, Android, and web browser enabled application. It's primary use is to determine approximately the financial size of field of research in Canada through the use of NSERC datasets. NSERC is the largest financier of research in Canada and more than half of their grants are co-funded by private entities. Aside from providing forecasting services, the Alchemist toolset also provides searching and data visualization services that allow users to interact and view data more critically and easily.

1.1 Domain Overview

The Alchemist toolset is outfitted to be used specifically by students, professors, faculty, staff, businesses and the public. The person of interest this toolset is meant to serve is anyone looking for a general overview of and metrics on the state of the research industry. Thus, it is critical for the application to be accessible through mobile phones and modern web browsers without the need for any additional add-ons.

1.2 Goals

The goal of this program is to help students and interested parties in industry visualize the trends in research, help identify stagnation and growth periods and estimate the growth of research. The endeavour is to provide a toolset that is accessible to a wide range of users; ranging from confused high school students entering University with the intent to find realtime information on a few subject interest to a journalist looking for the history of grant awards under a particular keyword. The secondary motivation behind this project is to provide specifically highschool students seeking careers in research a platform from which they can access first hand an overview of statistics that are otherwise sought for in popular magazines such as MacLean's.

2. Domain

This application is designed to describe the relationship between research grant awards and the specific field of research to which the grants were awarded. The application analyzes the relationship between the number of grants awarded in a field of study over time. By comparing trends of the fields of research and grants awarded to these fields of research over time, the application produces a projected investment amount specific to that field of research that the user is interested in. The overall goal of this application is to reduce the financial waste incurred through poor investment decisions and inform investors of the potential strengths of the field of which they are interested in investing in as well as provide the students, faculty, and staff an overview of the metrics present within the NSERC grant datasets.

The primary stakeholder for this application would be the Natural Sciences and Engineering Research Council of Canada (NSERC). The NSERC is a major investor in research in a wide variety of fields. This institution makes decisions on a regular basis regarding which fields of

research are suitable for available funds. The NSERC also partners with various companies interested in investing in specific fields. Thus, the NSERC stands to have the greatest benefit from this application which both reduces cost from the project cancellations and cost from investigating specific fields in search of an optimal field to invest in.

The secondary stakeholders that this application can appeal to are large bodies of students that are investigating various fields of research to participate or major in. These students can benefit from the prediction model presented by the application. Students would be interested in having an indication of research performance of a specific field in order to decide their specializations. However, not every student that uses this application would be a stakeholder. In general, the secondary stakeholder should be an organization or group that represents the interest of the larger student body.

3. Functional Requirements

The goal of this project will be to find a model that describes the relationship between research grant awards and the specific field of research to which that grant was allocated, to find a relationship between the number of grants awarded in a particular field of study over time, and how a particular field of research total funding overtime changes as a function of grants and time. The motivation for this project is to reduce the amount of financial waste that occurs as a result of the misallocation of resources in the field scientific research and development.

The NSERC grant dataset will serve as the input for our algorithm. The planned output for this program will consists of a graph connecting each participating University in the database to all the available fields of research in the database over annual time intervals. This would rank each University by the type of research most associated and conducted within it, allowing policy makers to see each University's development in research over time. The second output will be contain each universities grant amount changing over time in the various fields of research adjusted for inflation. The final output would be a total grant amount in each field of research listed over time and adjusted for inflation. This output will also show the program's expected projection into the next five years.

Table 1: Functional Requirements

Functional Requirements
0. The toolset must contain two visually distinct spaces of operation; a query page from which the user will be able to conduct a query and a data presentation/visualization page where the user will be able to see the result of a query. The user must be able to go back to the query page from the presentation/visualization page to conduct another query.
1. The program must forecast the financial size of a valid field of research as categorized by the NSERC grant datasets and produce a forecasted financial size of a field for every year from 2015 to 2020.

2. The program must forecast the number of grants awarded for a valid field of research as categorized by the NSERC grant datasets and produce a forecasted number of grants awarded to that field for every year ranging from 2015 to 2020.
3. The program must present a visual depicting the result of requirements 1 and 2 in the form of a bar graph showing the change in the size of the field of research and number of graphs awarded from 2015 to 2020.
4. The program must allow users to plot the total amount of grant dollars awarded to a field of research on a provincial distribution map of Canada for a given field of research as categorized by the NSERC grant datasets.
5. The program must allow users to plot the total amount of grant dollars awarded to a field of research on a distribution bar graph listing of all of the most funded institutions for a given field of research as categorized by the NSERC grant datasets.
6. The program must allow users to plot the total number of grants awarded to a field of research on a distribution bar graph listing of institutions for a given field of research as categorized by the NSERC grant datasets.
7. The program must allow users to plot the total amount of grant dollars awarded to a field of research on a distribution bar graph listing of all the most funded subjects within the field of research as categorized by the NSERC datasets.
8. The program must allow users to plot the total number of grants awarded to a field of research on a distribution bar graph listing of all the most awarded subjects within the field of research as categorized by the NSERC datasets.
9. The program must allow users to plot the total number of grants awarded to a field of research on a distribution bar graph listing of all of the Canadian provinces and territories for a valid field of research as categorized by the NSERC grant datasets.
10. The program must allow users to plot the total number of grants awarded to a field of research on a distribution bar graph listing of all of the Canadian provinces and territories for a valid field of research as categorized by the NSERC grant datasets.
11. The program must provide a visual distribution of all the entire financial history of a given field of research as categorized by the NSERC grant datasets.
12. The program must provide a visual distribution of all the entire history of the number of awarded grants to a given field of research as categorized by the NSERC grant datasets.
13. The program must list professors, alumni, or researchers who've received a grant whose subject of interest is within the valid field of research as categorized by the NSERC grant datasets.
14. All queries made within the toolset should return a result within 1.5 seconds.
15. The toolset server must be able to handle a client load maximum of 100 queries per second.
16. The toolset must be accessible from a modern browser running HTML5 and with Javascript enabled regardless of the operating system.

17. The user must be able to search the NSERC datasets for other than forecasting purposes. For example, the user should be able to look for the number of grants awarded to a specific professor, institution, subject or province, organization, or search for any professor to whom was granted an x amount of grant dollars.
18. The user must be able to select from a list of the following grant column options when performing a query: 1) Field of Research 2) Organization 3) Province 4) Professor 5) Subject, 6) Year awarded 7) Amount awarded
19. The query page must contain a “search for” button that returns the result of the query and displays it in the visualization page.
20. The query page must show the number of results that were retrieved for a specific query.
21. The user must be able to download any bar graphs or maps presented in the visualization page as a .jpeg, .png, .pdf, .svg format directly within the toolset.
22. The user must be able to print any visual directly within the toolset.
23. The client side interface must be able to function within and be accessible through any mobile phone running Android 3.0 Honeycomb or later and iOS 7.0 or later.
24. The application must be accessible through the following web browsers; FireFox, Opera, Chrome, Safari, and Internet Explorer.
25. The application must not require the installation or any additional software downloads to function on the client side.

4. Non-functional Requirements

Reliability is the most important non-functional requirement. There will be a margin of error in the functionality of the final product; the prediction of the future grants should be within a margin of error of 5%. This should be mentioned in the software’s disclaimer. Alongside reliability is security; if the data that the software relies on is insecure or originates from somewhere insecure, then the predictions themselves will not be accurate as the baseline of the information is false and would overall be useless if that was the case.

The target audience is investors who are not in a hurry, therefore, performance is not the most important. Even though performance is not the most important factor, having software that takes minutes to load would not be optimal. The software is considered a viable product as long as the software functions within a reasonable several seconds.

It would be best to allow the software to be accessible on as many platforms as possible so the target audience is not upheld to platform requirements. This will also alleviate portability issues and some physical constraints in terms of hardware. Whether it's Windows, Mac, Linux, iPhone, Android, Alchemist should be available to all. Moreover, parallel to this multiplatform requirement, it should be easily operable; the software should respond with feedback to notify the user of progress and have a quick and easy UI.

Table 1: Non-Functional Requirements

1. The toolset must be easy to use and pleasing to the eyes. The colour scheme of the application must be universal.
2. The toolset must present all graphs with multiple colours to allow users to be able to easily discern between the entries.
3. A search query must be easily conducted without the use of a user manual or a visual instruction.
4. A search query must be able to predict what information the user is looking for and suggest visualizations for the search query automatically.
5. The app must show all results within a single web page or scrollable frame.
6. The frame on which data is presented must be scrollable via a mouse connected to a desktop or through a touch screen device.
7. The user must be able to see the online or offline status of server.
8. The user must be able to hover through the bars in the bar graph and see a numerical representation of its magnitude in percent and respective unit.
9. The transition from the query to the data visualization page must be smooth and visually appealing.
10. The application must be advertisement free.

5. Requirements on the Development and Maintenance Process

The quality control procedures of this application will mainly be based on the idea of updating the information the program gets. This means that the database the application uses should be updated as much as possible to lessen the margin of errors of the predictions of the trends of grants. Making sure the predictions are as accurate as possible is a crucial aspect of our program, having valid up to date information is the most important part of our program.

Another quality control procedure that our company will ensure is version control and updating of the application’s algorithm with the newest most powerful ones. Making sure the program is as fast as it can is additionally very important factor in the development as the database is hundreds of thousands of rows of data. Additionally updating prediction algorithms that develop

over time will also be another quality control procedure as lowering the margin of error with more accurate prediction.

As for system testing procedures, we plan to test every new implementation in our program after it is completed to make sure that testing can be organized and pipelined along with the development. Additional to module testing, there will be rigorous system wide testing that makes sure each of the implemented modules work as intended according to the design specification. Making sure the system works functionally, error free will be accomplished with standard software testing procedures.

Finally with development requirements, we will be placing each functionality in a priority of from most to least importance:

1. Sorting of the database:
   - The searching assumes the information is originally sorted to work
   - One of the programs functionality is to give an ordered representation (Top ten) which corresponds to the model.
2. Searching the database:
   -The model will offer the function to search with a parameterized search
3. Model
   -The model is main logic behind the project, making it a priority
4. Graphing the database
   -Graphing of the database is dependent on the rest of prioritized functions but only the user interface is dependent on the graphing part.
5. User Interface
   -Even though the user interface is crucial in the success of the program, it is the lowest priority since it depends so heavily on the logic of the model and the database.

6. Appendix

6.1 List of Tables