**Final Project Report**

Charlie Grant

CSCE 240

**Introduction**

**Problem**

The goal of this project was to create a chatbot that uses the 10-K filed by two companies to parse financial and business information. The chatbot needs to be able to handle any kind of input by the user, and answer most basic questions about each company. It also is suggested to be able to handle casual conversation like “how is the weather”. Finally it needs to be able to report its usage statistics, show statistics of specific sessions, and display every interaction that occurred in a session.

**Introduction**

**Scope**

Using a googled definition, the scope of a project is the work required to complete the project including the tasks, time, and resources. The task will be explained below, and time will also be discussed as each specific project is discussed. Finally the resources will be covered in the language section, other than the IDE. I used Eclipse for C++ programmers for the entire project.

**Language**

I chose to use C++ because I wanted to try a new language, and I wanted the homeworks we were working on to be easily translatable to the project. Because the homeworks were required to be in C++ and somewhat because I thought more people would have chosen C++, I selected this language. The other language I considered was Java. This is because I am most comfortable with Java because of CSCE 145 and CSCE 146. However, I actually think that had I chosen this language I would have had an even harder time because I would have been less malleable and willing to change my approach to tackle each project.

**Project 1**

For the first project I chose to use an input file of the two stored text files. It had a global variable for each word, line, character, and part. It used the getline function to count lines, then I found a way online to use substrings and delimiters to count words and characters. Finally for Parts I kept track of pages, and restricted when a part could be marked. This was to use only the table of contents to then track how many times the word part appeared using the to lower case function and the find function of strings. Originally I really struggled to find a way for the system to count parts, but on the last day of the project I talked to other classmates about what their approach had been and I created a tactic that made sense to me. And one I thought I could easily accomplish. That last aspect will be a very common occurrence. I spent maybe around an hour or two, most of which was learning C++ and its differences from the languages I knew.

**Project 2**

For the second project I used the technique from PA1 to restrict counting items to the table of contents. I organized the items into a linked list. This is because this structure was something I used a lot when coding in Java and I was confident I could translate this previous knowledge into C++. I then used regex to match the structure of how each company started an Item’s section. This meant I printed all information of the selected item using an ofstream. I also added the print everything function. Originally I had a very specific regex for each item, which had a better result if I was hyper specific, but I knew moving forward I needed a better system so I made the regex more wide spread.

This project took the most time because it is what the entire premise of the chatbot is based off of. This project is essentially building the chat bot, especially because it was asked to handle not matching with a topic for the users question. I was up till 3 AM for the three nights before this project was due.

**Project 3**

For the third project I used a function I called qloop. Essentially, unless the user chose to quit, a while loop would always be called. I added the function to switch companies using a global boolean for true being one company and false for the other. I had already done this function when working on PA2 so the time devoted to this project was much shorter.

**Project 4**

For the fourth project I added a function I called QMap. This used a for loop to check every word for appearance in each item title. I weighted words like risk and executive to have a higher match percentage. I also added a chat function as was asked but this section was a separate method from the qloop so the chats can not be done from the main interaction hub.

**Project 5**

For the fifth project I used an ofstream to output to a file called ChatStats that stored the information with each session being an ifstream to read the needed information and manipulate it. I used the given statistics commands given in the lecture description of project 5.

**Reuse and Integration**

I reused Marcus’s PA5 to display each chat session. I reused this section because it was just more succinct and well organized.

I believe my best component is the linked list and regex to find the information. I believe where my chatbot struggles is the PA4 aspect of matching intent, but once intent is found it always prints the correct information with no struggle. When it comes to final integration I found it quite easy because i would just copy all code from the previous project into a new project and continue to work from there.

**Evaluation**

I was able to get an F1 score of 82%. I do however believe it is likely slightly lower than this because it is possible I somehow phrased some questions in a way more beneficial to the code and less casual.

**Discussion**

**Significance**

The significance to this project for me was first learning a new programming language. Secondly, I was able to learn how to implement regex and how to structure a chatbot. I refined my use of for loops and of linked lists. I also became very well versed in creating handles for exceptions.

The significance to the general user is that implementing the code used for the project from any student will allow them to parse information from a similarly structured 10-K to easily learn information that is normally not easily understandable.

**Challenges**

I mostly struggled with PA2 because as I stated, this was honestly 80% of the code for me. I struggled with a language I was not used to and I struggled with a decent lack of direction on how to tackle the assignment. My experience after the 3rd project when I was still refining PA2, was one of honestly very little stress and very little difficulty. In the future I think delaying PA2 to give us extra time or giving example output would alleviate stress when working on this project.

**Future Work**

If I were to still work on this project I would hope to improve how the information is printed and displayed to the user. I would also hope to narrow what information would be allowed. For the questions about executives I would aim to target specific members of the company instead of printing all information in the item.