CSCE 240

Project 6

Anish Ghana

Project 6 consisted of us combining all of the other projects we made this semester seamlessly to create a final chatbot. The project this semester involved reading 10-K files (files that hold the annual financial performance reports of a company) from a company of our choice and then having a chatbot be the interface from which the user extracted concentrated information from that 10-K file. Each student chose two companies for their chatbot to represent. I chose ExxonMobil and Berkshire Hathaway due to my interest in these companies.

This project has given me a huge amount of experience and knowledge on how a chatbot can function and some of the intricacies of Python. I choose Python over C++ or Java due to my unfamiliarity with the language. I wanted to learn a new language and Python is known for being a relatively easy language to learn. I also wanted to take advantage of the numerous libraries in Python. I believe my project was successful as I not only learned a great amount but I also created a chatbot with a multitude of functions. My program can parse through a 10-K text file and retrieve data. It can then take in user input from the console to figure out what information the user wants. It will then display said information in an output.txt file. It also can record chat statistics in a CSV file so that the user can see numerical representations of their conversations with the chatbot.

Some difficulties I had with the project, in the beginning, involved parsing through the 10-K file and finding specific instances of words, as the 10-K files are extremely large and have

multiple instances of multiple instances of the keywords I used to find important sections of the text. I ended up using very particular groups of regular expressions to solve this problem. I also had difficulty making the chatbot more dynamic in its conversation, so I ended up using a library that matched user queries to an extensive list of preset queries I made. All of these issues were difficult to get through, however, they helped me grow as a programmer. If I had more time, I would implement a GUI to make the chatbot more user-friendly. I would also give a greater learning ability by implementing a neural network library from Python so that the chatbot could improve itself. I have had some experience with neural network libraries in Python and it would make my project more similar to popularized versions of chatbots often seen in the news or media

Extra Credit/Reusable Code

PA1

I considered using a peer's web scraper in my project, as it was more useful than my text file parser, however, there was not enough modularity between our code for his web scraper to be effective in my code. I ended up keeping my text file parser, which was crude, but brutally efficient and simple enough to not cause any problems.

PA2

I once again used my code, as I tailor-made PA2 to filter out the parts and items in the 10-K's for my companies and I found that using my peers' PA2s would be difficult and provide little to no benefit as their functionality was very similar to mine. For me, PA2 was the most difficult, as it held the biggest learning curve for Python I have experienced. Also, the logic to

separate the parts and items when there were multiple instances of each in the 10-Ks was difficult for me to grasp at the beginning of the project.

PA3

My PA3 was very similar in function to my peers, so I saw no reason for code reuse. It worked rather well and was reasonably efficient and fulfilled all requirements. I originally had an issue with an infinite loop and not being able to quit the program properly, but I was able to resolve that issue rather quickly. I also used this PA to improve upon the previous functionality from PA2.

PA4

This PA is when I began using code from my peers, in part due to the sheer increase in complexity from PA3 to PA4. PA4 involved numerically and statistically checking the similarity of user queries to preset chatbot queries. This project did highlight the usefulness of the language I chose, as many libraries within Python can take care of many of the aspects of this assignment. My PA4 greatly utilized the SequenceMatcher class from the difflib module to check through user and chatbot queries. I also changed my method to store chatbot preset queries from an array to a dictionary, making my code more readable and making it easier to check through each query. PA4 was approached differently by all of my peers and thus I found some of my peers implemented methods far more efficiently than I did. When browsing through my peers' code I found Dev Patel’s code to be of particular interest to me. His remove words method was significantly more streamlined than mine, and his method made it much easier to add new words to remove.

PA5

PA5 involved taking the information from each chat session and putting it in an easily accessible format. My PA5 presented information about the number of chats, the content of chats, the duration of chats, and individual statistics about chatbots and users. One issue I had was initializing the CSV file. While I was able to resolve this issue, the method I created for this project was inefficient and contained many superfluous lines of code. I ended up switching out my initCSV() method with Dev Patel's, as his method was significantly more efficient than mine.

Conclusion

This project was arduous but hugely rewarding. I learned so much about different programming techniques and some of the intricacies of Python. This project also helped highlight the importance of code reuse, as no programmer can make the most efficient code 100 percent of the time. Overall I gained so much experience from this project and I believe in the future the techniques I learned from this project will help me create efficient, easily readable, and highly functional code.