Final Documentation

Ethan Hammer

The goal of this project was to retrieve financial information from two different companies via their respective SEC filed 10k documents. The companies I chose were originally Wendys and McDonalds. However, I eventually switched McDonalds to Restaurant Brands International due to it being closer in format to Wendy's 10k file. To accomplish this goal, we were required to create a user-accessible chatbot that could map user queries to desired information from the 10k documents. This includes accounting for misinput such as spelling errors, strange wording, etc. I have not done anything similar to this in the past, except for some basic file parsing. This made the project a great learning experience.

To accomplish the goal of creating a fully functional chatbot, we were required to assemble five separate projects into one fully functional project. These projects were assigned roughly every two weeks and were essential building blocks that a chatbot requires such as text parsing and query mapping. For this project, I chose to use Java. This is the language in which I have the most experience in and thus could apply my concept knowledge without having to worry so much about my language knowledge. Java is also much easier to handle when it comes to string manipulation or searching as compared to the course language of C++. Here is a summary of the functionality of each major assignment:

PA1

PA1 simply ensured that the files needed from the SEC website could be recovered, accessed, and read through. To do this, I used a java scanner that took the respective files that were already in .txt files and showed their respective word and character counts. Originally, I attempted to make a web scraper to automatically take the website page and convert it to a .txt

file with just an URL. However, I had trouble doing this and found it often was not formatted properly. So, I chose to manually copy the text and put it into the company's respective .txt files.

PA2

PA2 was used to map user requests to different section headers and data in the 10k files. To do this, I created a key system, where certain keywords would be matched to a section and that section would then be found in the text. The keys were recognized via regex patterns, where each word in the user query was tested to see if it matched a key. I also used regex patterns to try to find a company name in the user input, so the chatbot could decipher which file the user wanted. One cool thing I implemented in this project was an option that asked the user which key they wanted if multiple were to be picked up.

PA 3

The only difference between PA3 and PA2 is I implemented a Java Swing GUI to make the bot more user friendly. This was rather difficult though due to the complete change in when and how responses should be carried out and taken in. This resulted in the removal of the feature that asked users which key they meant if multiple were picked up.

PA4

In PA4, I changed how the keys were detected. Instead of using a regex pattern matcher, I opted to use a string similarity coefficient. This basically compares the length of matching characters in the user input to the length of the key to get a ratio of a match. If this match was above a certain threshold, then the key would be identified and the data would be printed. This greatly improved my program's ability to recognize misspellings and similar words (property vs properties). My key system was also reused by both Jacob Stoll and CodyMiller.

PA5

This PA simply outputs recorded stats from precious chat-sessions. To do this, I first had to modify PA4 to create logs of chat-sessions that included when it started, when it ended, how many times the bot spoke, how many times the user spoke, and the actual chat-logs themselves. I then made a series of commands to do things such as show a summary for a specific chat-session or print the whole chat of a specific session.

Integration

Integrating these five components was not too difficult. Projects 1-4 built on each other and were already implemented together. The most difficult part was creating a new series of regex patterns to search for PA5 commands inside of PA4. However, this did not take too long, especially since I reused the necessary regex patterns from Cody Miller.

Evaluation

I would overall describe this chatbot as decent. In order to properly interact with it, you need to already have some knowledge of the key system, which is not ideal. This is due to a lack of search keys. Testing yielded a precision of 10/13, a recall of 10/12, and an overall F1 score of 0.8.

Discussion

This chat bot was significant in my learning process. It taught me alot about language processing and how chatbots work in general. The purpose of the chatbot is also significant as it allows for more open-access into the financial data of two companies that are important in the lives of millions of people globally. I found the overall experience building the chatbot to be frustrating but rewarding. This is true in essentially all learning processes, where at first the challenge seems daunting and tedious, but whenever you overcome whatever obstacles block you, the success results in a rush of self-gratitude. In terms of collaboration, or the re-use of

others' code, I found this to be hard. Everyone has different coding tendencies and ways in which they approach problems. This makes integrating just a part of someone else's solution very difficult as you seek to put a round puzzle piece into a sometimes square hole. In the future, I would like to try making this bot in another language such as C++ and maybe compare their processing speeds. I would also want to dive more into natural language processing and possibly even look into some AI usage for better language processing. I was inspired by this after seeing a group at the 2024 Cocky Hackathon use an AI model to map their chatbot queries to keys regarding the University of South Carolina's website.

Conclusion

The goal of this project was to create a chat-bot that allowed users to access financial data from two different companies, which in my case was Wendys and Restaurant Brands International. This included creating a series of projects, and integrating them to create a fully functional program. This ended up becoming a decent success that I hope to expand on in the future.