COSC 310 - Individual Project

For my individual project, which has been done in Java, the user takes on the role of an interviewer and the agent takes on the role of an interviewee for a software engineering job. The user prompts the agent by asking questions related to their history, experience, etc. and the agent responds with a relevant answer. The software we developed utilizes Google Translate API, Wikipedia API (MediaWiki), Named Entity Recognition, Coreference Resolution, SpellCheck, and POS Tagging to identify various languages, subjects, and structures within user's input phrases to respond appropriately. The software utilizes a minimalistic User Interface that allows users to easily enter interview-based phrases and receive responses that are cohesive to an interviewee.

Added Features

Google Translate API

The Google Translate API takes in the user's prompt and detects the language the user is using. If the prompt is detected as English then the original prompt is returned. If the prompt is a different language, then the API returns an English translation of the prompt. The user can prompt the chatbot in multiple languages, but will always get an answer back in English.

Wikipedia API (MediaWiki)

The Wikipedia API is used after the chatbot fails to identify anything inside the CSV prompt files. It then takes the last word in the prompt (which we assume is the subject), puts it through the API query, and retrieves a JSON file for the chatbot to extract the definition of that subject, and returns an excerpt from the Wikipedia page to print back out to answer the user's prompt. If the user wants the chatbot to look up a multiple-worded subject title to the API, they need to use '_' (underscores) instead of spaces.
E.g. 'Software Engineering' needs to be written as 'Software_Engineering'.

Features

Misspelled Word matcher

If a potential keyword doesn't match any known keywords, then the program will try to match it against known words. If the misspelled word nearly matches a keyword, the program will recognize it as the keyword.

POS Tagging

Without the POS tagger, the program must loop through every word in the input string until it finds a matching keyword. By adding the POS tagger, the number of words the program must search is reduced. The POS tagger takes the input string, tags each word, and returns a list of potential keywords based on their tag. So instead of searching the entire string, a list of two to three words is searched instead.

Named Entity Recognition

The named entity recognition improves the responses our chatbot gives to particular questions without those responses needing to be hardcoded. The named entity recognition allows the program to recognize names, and use those names in its responses. For example, if the interviewer asks if the chatbot has ever traveled to a particular place, the chatbot can recognize the place name and use that name in its response for a more natural-sounding conversation.

Coreference Resolution

The coreference toolkit enhances our chatbot by improving the flow of conversation for the interviewer. Instead of explicitly stating the topic each time, the chatbot is now able to match a previous topic to the current conversation. For example, the interviewer can ask the chatbot about work experience and then ask the chatbot to talk more about it. Even though a particular topic is not named, the chatbot can recognize what the topic is from a previous conversation.

Compiling and Running

CSV Path Change

After getting all of the files from the GitHub repository the CSVs location must be changed. This can either be absolute or relative and is found within ChatBot.java in the search function.

JavaFX

After this JavaFX must be installed onto the machine.

VSCode: * Update VSCode to 1.49.3 or above. * Install JavaFX support extension in VSCode. * Download and install open JDK and configure Java runtime in VSCode (may need to add to system variables). * Download JavaFX SDK, extract, and copy path of lib folder. * Reference all jar files within JavaFX SDK to the Java project. * Edit run configuration within launch.json by adding:

"vmArgs": "--module-path {Your path}/javafx-sdk-11.0.2/lib --add-modules javafx.controls,javafx.fxml",

Eclipse: * Follow this tutorial: https://www.youtube.com/watch?v=bC4XB6JAaoU

CoreNLP

After setting up JavaFX. The next step is to properly reference the Stanford CoreNLP jars.

Generalized method: * Download the CoreNLP zip file from: https://stanfordnlp.github.io/CoreNLP/ and extract * Reference all of the jars to the classpath of the Java project