EE 551 – Voice Assistant for Wikipedia Summarization

Jayden Pereira
jpereir2@stevens.edu

Problem

With the likes of Alexa, Okay Google, Siri, and more, there are multitudes of options for voice assistants – each with its own advantages and downsides. While directly interfacing with APIs that companies like Amazon provide would be simple, I decided to take a look at replicating what it's like to process and provide meaningful data to queries posed through spoken word.

Method

In order to actually process a meaningful outcome from a spoken query, there needed to be an outline of what exactly is to be delivered to the User. For this project, while the scope may be small, I thought procuring and summarizing Wikipedia articles for the subject of the query would be the widest gamut for which my Voice Assistant would be deemed useful without compromising on the complexity of the switch cases needed to operate various different query types. In this way, I can process every query alike and try to match it as best as possible. This entire process can be outlined as such:

1. Get spoken query from User and generate string from Google Speech Recognition
2. Generate an array of Google Search results from User query
3. Iterate through array to find a Wikipedia article associated to the topic
4. Take Wikipedia summary and generate a temporary Voice Assistant audio file through Google Text-to-Speech
5. Play audio file for User to receive result from query

Since Google Search queries often result in Wikipedia articles, even when structured in various syntactically similar phrasing, I utilized the service as a funnel to trace a related Wikipedia article that could possibly answer the User query. Although this does not answer every single case (especially in geolocation specific queries such as "How is the weather today?"), the Google Search interface I used does actually generate relevant search results that could be processed further. Although, as aforementioned this would complicate the code base and pigeonhole specific use cases.

Conclusion

For what is a project that I used to primarily discover and research how these established interfaces could work together to perform a task that most of our phones and any smart speakers that we might have can do, it was really interesting to understand how complex it is to actually generate meaningful results from a Voice Assistant without limiting its scope or becoming too general.