Amanuel Shiferaw
CS 4376
Professor Karen Mazidi
12 October 2022

Portfolio Web Crawler

1) The way that I created the knowledge base was by first determining the most frequent words in all 15 URLs that were explored by the algorithm. After using a dictionary to calculate the frequency of all the words present in the URLs I retrieved the top 40 and manually determined the top 10 relevant words in all the texts. Furthermore, I have also been keeping track of all the sentences in all 15 URLs so I simple word-tokenized all the sentences and checked to see if a word from the top 10 words list is in the sentence. I simply appended the sentence to an array. This array is stored as a value in a dictionary where it can be easily referenced by the key which is a word from the top 10 keywords list. This was we have a list of all the sentences in which the word was used.

   a) Final 10 words on the topic of Muhammad Ali are:

['ali', 'muhammad', 'fight', 'vs', 'clay', 'boxing', 'frazier', 'heavyweight', 'title', 'greatest']

Screenshot of Knowledge Base [ali]



2) write up a sample dialog you would like to create with a chatbot based on your knowledge base

**Sample Conversation:**
Bot: Hey Do you know Muhammad?
User: Muhammad who?
Bot: Muhammad Ali, he was the greatest heavyweight fighter
User: Interesting, who did he fight?
Bot: His most famous fight was with Clay for the heavyweight title
User: Nice. What kind of fight was it?
Bot: It was a boxing fight with Clay for the heavyweight title
User: Was he good?
Bot: Mandela said he was the greatest of all time!