Knowledge Base

Jonathan Yu

October 10, 2022

In order to make my knowledge base, I began my web crawl on a ESPN article on a hard hit that occurred in the NFL. I then crawled through all the links and recorded the text on each page. From there, I then tokenized the text. I noticed that the top 50 terms was mostly composed of legal words, so I realized that my web crawler was picking up the EULA at the bottom of the page. Because of this, I just added the entire sentence to a filter, along with stop words and punctuation.

Once I found the top 10 most important words I built the knowledge base. I first find all occurrences of the given word. Then, I find the last period in the substring before the occurrence, along with the first period after the occurrence. I then take a substring between the two, giving me a sentence describing the word. My top ten words are game, required, digital, dispute, york, south, tagovailoa, concussion, united, and subject. For each word, I simply append these sentences to my knowledge base. To look up an item in the knowledge base, we index the word we want and select a sentence.

```
In [36]: print(f"Fun fact: {knowledge_base['game'][2]}!\n")
         print(f"Fun fact: {knowledge_base['tagovailoa'][0]}\n")
         print(f"Fun fact: {knowledge_base['concussion'][1]}\n")

         Fun fact: a comprehensive listing of the elements that have been added to our
         fantasy game for the upcoming season!

         Fun fact: paul finebaum and keyshawn johnson break down why the dolphins will
         move on from tua tagovailoa if a better option at quarterback comes up

         Fun fact: bridgewater was put into the protocol after the booth atc spotter r
         uled him a "no-go" after he took a hit on the dolphins' opening offensive dri
         ve, in compliance with the nfl's amended concussion protocol
```

Example:
Me: ”Hi, I just got a concussion!”
Chat bot: ”Bridgewater was put into the protocol after the booth atc spotter ruled him a ”no-go” after he took a hit on the dolphins’ opening offensive drive, in compliance with the NFL’s amended concussion protocol”