

Knowledge Base

In the program, my initial url was a link to a Wikipedia page on the Bible. After going through the page and getting 15 links to various other pages, I collected all the information from those various pages and saved them in 15 different files. After that, I used NLTK to separate out all the various sentences for each file and saved those in 15 other files. Then, I looked through all the files and figured out which words appeared the most often. Of the top 40 most important words, I chose out 10: ["books", "testament", "new", "bible", "canon", "canonical", "church", "hebrew", "christian", "apocrypha"]

To create the knowledge base, I made a dictionary where the key was one of the 10 words listed above. For the value, it was a long string that is a concatenation of all the sentences from all the files that contain that word. So the value associated with “books” is a string that has all the sentences from all the files that contain the word “books”

Sample Dialog

Me: “What is an example of an apocryphal book?”

Bot: “Tobit is an example of an apocryphal book”

Me: “What is the Bible about?”

Bot: “The Bible is a collection of the sacred texts of the Christian religion. The central focus is the teaching, life, suffering, death, burial, and resurrection of Jesus, believed to be the Son of God and the Messiah of the Old Testament”