Justin Hardy

JEH180008

CS 4395.001

Dr. Mazidi

Web Crawler Overview

For this assignment, we were tasked with creating a knowledge base using a web crawler to scrape information from a webpage, and any webpages they link to, of our choosing. We'd chosen to use a poker-related webpage as the root page. I went about creating the knowledge base by using the ten most frequent, non-stopword terms in each of the fifteen web pages we'd scraped. The keys point to a list of each sentence they occur in, out of all sentences from the webpages we'd scraped. It isn't perfect, as it may think a sentence starts at a strange point due to the way the text was sometimes displayed. But most of the sentence content is very sensible. You can see an example of a key from the knowledge base in the screenshot below.

If we were to create a chatbot out of this information, I'd like for it to be able to answer a question such as, "What different types of pots are there?", using the key term "pot."