Assignment 6: Web Crawler Report

Abdullah Hasani - AHH190004

Knowledge Base

To build the knowledge base, I initially tried using the scraped data from the .txt files, and any line or sentence that contained one of the top 10 words, I would add to the dictionary. However, the resulting data was very cluttered and at times repetitive, and synonyms or other forms of the words were often found to contain some information while the word I had chosen contained other information.

For instance, for my chosen topic "Formula One", there were some articles that primarily used "F1", some that used "Formula 1" and some that used a combination of both. In the end, I just went to ChatGPT and inputted the list of words, and had it generate facts based on the word in the context of Formula One. I then had it format the output into a JSON (or python dictionary) format, which I wrote to a .txt file and picked to serve as the knowledge base.

My top ten terms were:
'formula one', 'race', 'season', 'championship', 'team', 'driver', 'grand prix', 'car', 'points', and 'fia'

The screenshot below shows part of the dictionary in JSON format in the .txt file:

This dictionary was then pickled to serve as a knowledge base for a future project.

Sample Dialog for a Chatbot

User: What is F1?

F1 Bot: F1 stands for Formula One, which is a type of open-wheel, single-seat racing that is considered the pinnacle of motorsports.

User: Where does it take place?

F1 Bot: Formula One races are held all over the world, on circuits that range from traditional tracks to purpose-built street courses.

User: Tell me more about the races.

F1 Bot: Each race is typically held on a purpose-built race track, although some races take place on public roads that have been temporarily closed for the event.

User: Tell me more.

F1 Bot: The length of a race varies depending on the track and the specific race, but most races are around 300 kilometers in length and last between 1.5 and 2 hours. The winner of a race is the driver who completes the set number of laps or distance in the shortest amount of time, with the top three drivers awarded points towards the championship.