

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:

```
{
  "formula one": [
    "F1 stands for Formula One, which is a type of open-wheel, single-seat racing that is considered the pinnacle of motorsports.",
    "Formula One races are held all over the world, on circuits that range from traditional tracks to purpose-built street courses.",
    "Formula One cars are highly advanced, with cutting-edge technology and materials used in their design and construction.",
    "F1 cars are powered by hybrid power units that combine a turbocharged internal combustion engine with an electric motor, producing over 1000 horsepower.",
    "F1 is governed by the FIA (Fédération Internationale de l'Automobile), which is responsible for setting technical regulations, safety standards, and overseeing the organization and management of each Grand Prix race.",
    "The F1 season typically runs from March to December each year, with races held all around the world.",
    "F1 races are watched by millions of fans worldwide, making it one of the most popular and prestigious forms of motorsport.",
    "F1 drivers are among the most skilled and talented athletes in the world, with years of training and experience required to compete at the highest level.",
    "F1 teams are highly competitive and constantly seeking ways to gain an advantage on the track, investing heavily in research and development to improve their cars and technologies.",
    "The history of Formula One is filled with legendary drivers, teams, and races, and continues to captivate and inspire fans around the world."
  ],
  "f1": [
    "F1 stands for Formula One, which is a type of open-wheel, single-seat racing that is considered the pinnacle of motorsports.",
    "Formula One races are held all over the world, on circuits that range from traditional tracks to purpose-built street courses.",
    "Formula One cars are highly advanced, with cutting-edge technology and materials used in their design and construction.",
    "F1 cars are powered by hybrid power units that combine a turbocharged internal combustion engine with an electric motor, producing over 1000 horsepower.",
    "F1 is governed by the FIA (Fédération Internationale de l'Automobile), which is responsible for setting technical regulations, safety standards, and overseeing the organization and management of each Grand Prix race.",
    "The F1 season typically runs from March to December each year, with races held all around the world.",
    "F1 races are watched by millions of fans worldwide, making it one of the most popular and prestigious forms of motorsport.",
    "F1 drivers are among the most skilled and talented athletes in the world, with years of training and experience required to compete at the highest level.",
    "F1 teams are highly competitive and constantly seeking ways to gain an advantage on the track, investing heavily in research and development to improve their cars and technologies.",
    "The history of Formula One is filled with legendary drivers, teams, and races, and continues to captivate and inspire fans around the world."
  ],
  "formula 1": [
    "F1 stands for Formula One, which is a type of open-wheel, single-seat racing that is considered the pinnacle of motorsports.",
    "Formula One races are held all over the world, on circuits that range from traditional tracks to purpose-built street courses.",
    "Formula One cars are highly advanced, with cutting-edge technology and materials used in their design and construction.",
    "F1 cars are powered by hybrid power units that combine a turbocharged internal combustion engine with an electric motor, producing over 1000 horsepower.",
    "F1 is governed by the FIA (Fédération Internationale de l'Automobile), which is responsible for setting technical regulations, safety standards, and overseeing the organization and management of each Grand Prix race.",
    "The F1 season typically runs from March to December each year, with races held all around the world.",
    "F1 races are watched by millions of fans worldwide, making it one of the most popular and prestigious forms of motorsport.",
    "F1 drivers are among the most skilled and talented athletes in the world, with years of training and experience required to compete at the highest level.",
    "F1 teams are highly competitive and constantly seeking ways to gain an advantage on the track, investing heavily in research and development to improve their cars and technologies.",
    "The history of Formula One is filled with legendary drivers, teams, and races, and continues to captivate and inspire fans around the world."
  ],
  "race": [
    "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.",
    "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.",
    "Races are high-speed, high-intensity events that require extreme focus, skill, and physical fitness from the drivers.",
    "In addition to the drivers, races also involve a team of engineers and mechanics who work together to fine-tune the car's performance and strategy.",
    "Races often involve pit stops, during which drivers can refuel and change tires, as well as make adjustments to the car.",
    "Weather conditions can play a significant role in the outcome of a race, with rain and other factors affecting track conditions and driver performance.",
    "Formula One races are broadcast on television around the world, and attract millions of viewers and fans who follow the sport closely.",
    "Winning a Formula One race is considered one of the greatest achievements in motorsport, and many legendary drivers have become household names thanks to their success on the track."
  ],
  "season": [
    "The Formula One season typically runs from March to December, with races held all around the world.",
    "The season is made up of a series of races, with drivers and teams competing for the title of World Champion.",
    "In the modern era, the season usually consists of around 20 races, although this can vary from year to year.",
    "Each race awards points towards the championship, with the driver and team with the most points at the end of the season being crowned World Champion.",
    "The Formula One season involves a significant amount of travel, with teams and drivers traveling to different continents and time zones to compete in races.",
    "The season often begins with pre-season testing, during which teams test their cars and prepare for the upcoming season.",
    "Teams and drivers may make changes to their cars throughout the season to improve performance, with updates and new parts being introduced at various stages of the year.",
    "The season finale, typically held in December, is often one of the most anticipated races of the year and can be crucial in determining the World Championship winner.",
    "The Formula One season is not without controversy, with disputes between teams, drivers, and officials sometimes making headlines and influencing the outcome of races.",
    "Formula One fans around the world follow the season closely, with many attending races in person and millions more tuning in to watch on television."
  ],
  "championship": [
    "The Formula One World Championship is the pinnacle of the sport, with drivers and teams competing for the title of World Champion.",
    "The championship is awarded to the driver and team that score the most points over the course of the season."
  ]
}
```

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.