How I created the knowledge base
1. Scraped the web for good links, from the starting url
2. Extracted the text from 15 of the good links
3. Cleaned up the text that was extracted
4. Extract 40 top important terms for the 15 texts using tf-idf
5. Manually determined the top 10 terms from the 40 terms using my domain knowledge
6. Built the knowledge base off of the top 10 terms and the texts
   a. Knowledge base is a dictionary where the keys are the top 10 terms
   b. The values are lists of sentences where that term occurred (extracted from the 15 cleaned texts)

My top 10 terms:

```
jump
manga
dawn
romance
oda
shonen
ending
treasure
luffy
pirate
```

Screenshot from knowledge base

```
C:\Python311\python.exe D:\source\nlp_portfolio\Web-Crawler\5_create_knowledge_base.py
jump : ['oda began serializing the manga in weekly shonen jump in 1997, and the manga has been published in more than 30 countries.', "shueisha released the manga's 91st compiled
manga : ['manga creator eiichiro oda has said that chiisana viking vicke asserted a strong influence on his one piece pirate manga and anime.', 'manga creator eiichiro oda has sa
dawn : ['this year\'s 32nd issue of shueisha\'s weekly shonen jump magazine has confirmed on july 7 that the jump super anime tour will animate for the first time romance dawn, o
romance : ['this year\'s 32nd issue of shueisha\'s weekly shonen jump magazine has confirmed on july 7 that the jump super anime tour will animate for the first time romance daw
oda : ['manga creator eiichiro oda has said that chiisana viking vicke asserted a strong influence on his one piece pirate manga and anime.', 'manga creator eiichiro oda has sai
shonen : ['oda began serializing the manga in weekly shonen jump in 1997, and the manga has been published in more than 30 countries.', 'oda began serializing the manga in weekl
ending : ["he later becomes a higher-up in blackbeard's crew, just as jean lafitte worked his way up the maritime ladder, starting low and ending up a captain.", 'she led severa
treasure : ['join monkey d. luffy and his swashbuckling crew in their search for the ultimate treasure, the one piece.', 'years later, luffy sets off in search of the one piece,
luffy : ['join monkey d. luffy and his swashbuckling crew in their search for the ultimate treasure, the one piece.', 'as a child, monkey d. luffy dreamed of becoming king of th
pirate : ['as a child, monkey d. luffy dreamed of becoming king of the pirates.', 'as a child, monkey d. luffy dreamed of becoming king of the pirates.', 'years later, luffy set

Process finished with exit code 0
```

Sample Dialog I would like to create with a chatbot based on my knowledge base

Chatbot: "Hello! I am One Bot, a One Piece fan bot, what is your name?"

Me: "Hello One Bot, my name is Rigre."

Chatbot: "Hello Rigre, are you a one piece fan?"

Me: "Yes! My favorite character is luffy."

Chatbot: "Did you know that as a child, monkey d. luffy dreamed of becoming king of the pirates."

Me: "Yes I did know that. Can you tell me something more interesting?"

Chatbot: "Did you know that manga creator eiichiro oda has said that chiisana viking vicke asserted a strong influence on his one piece pirate manga and anime."

Me: "That's pretty cool actually. Any news about One Piece?"

Chatbot: "I know right! One Piece is awesome. Here is one piece of news related to one piece romance dawn: this year's 32nd issue of shueisha's weekly shonen jump magazine has confirmed on july 7 that the jump super anime tour will animate for the first time romance dawn, eiichiro oda's "prototype" manga that eventually evolved into his hit one piece series."

Me: "Thank you, you've been very helpful!"

Chatbot: "Of course! Let me know if I can do anything else for you.'