

(1) describe how you created your knowledge base, include screen shots of the knowledge base, and indicate your top 10 terms;

I created the knowledge base by using a dictionary of lists. The keys in the dictionary are the top 10 terms that I chose, and the values are all of the sentences across all 40 urls that contain that term. Then, I pickled this large dictionary so I can read it in for part 2 of this project. The only issue is that when I scrape from the web pages sometimes code will be included in my text files (javascript and html). I was unable to figure out a good way to get rid of this code for all of the urls. I will have to take care of this before part 2 of the project. Here are the top terms I decided on:

- bukowski
- charles
- poetry
- poems
- stories
- work
- story
- life
- writing
- Time

Screenshots of output and knowledgebase

```
File Edit View Navigate Code Refactor Run Tools VCS Window Help NLP_Code - hw_5_amm180005.py - PyCharm
NLP_Code Homeworks HW-5 hw_5_amm180005.py hw_5_amm180005 (1) [Run] [Debug] [Test] [Search]

Project Structure
Project
  cleaned_url_text32.txt
  cleaned_url_text33.txt
  cleaned_url_text34.txt
  cleaned_url_text35.txt
  cleaned_url_text36.txt
  cleaned_url_text37.txt
  cleaned_url_text38.txt
  cleaned_url_text39.txt
  hw_5_amm180005.py
  knowledge_base.pickle
  url_text1.txt
  url_text2.txt
  url_text3.txt
  url_text4.txt
  url_text5.txt
  url_text6.txt

hw_5_amm180005.py
115 # pickle the knowledge base dict for next part of the project
116 with open('knowledge_base.pickle', 'wb') as handle:
117     pickle.dump(knowledge_base, handle)
118
119
120
121 # get urls to pass into functions
122 with open('urls.txt', 'r', encoding="utf-8") as f:
123     urls = f.read().splitlines()
124
125 #scrape_all_text(urls)
126 clean_text(urls)
127 get_important_words(urls)
128

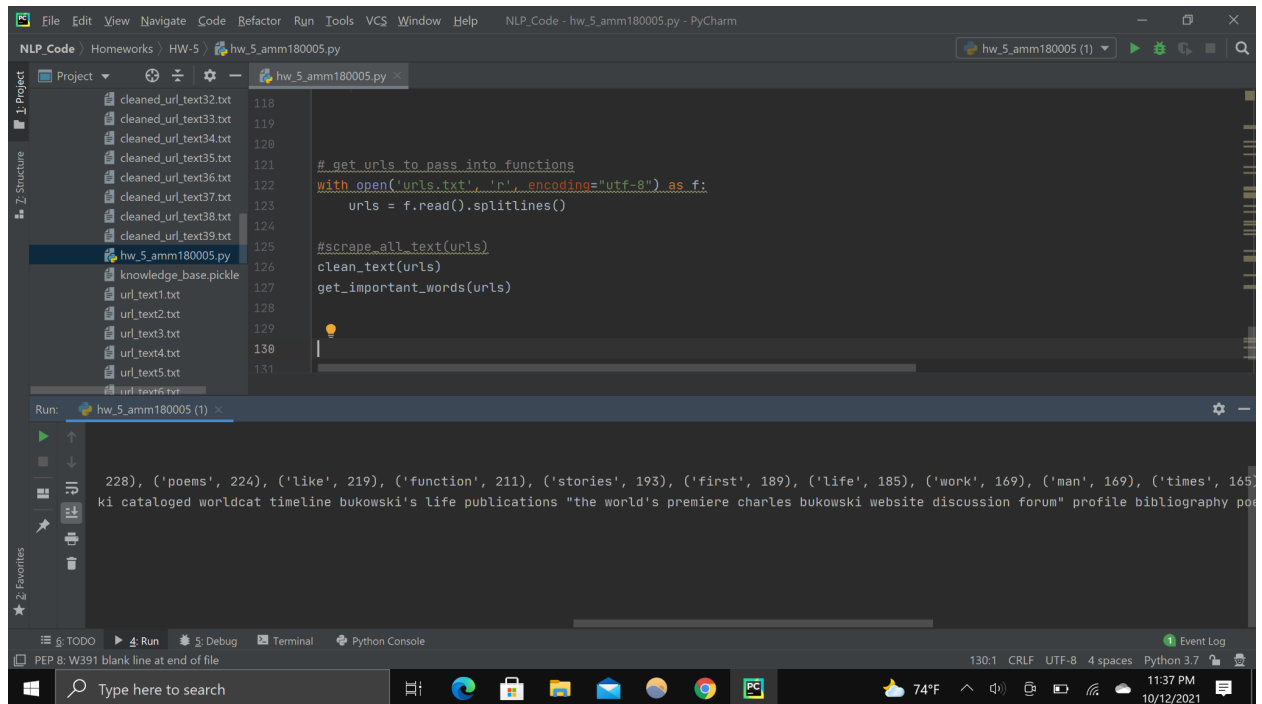
Run: hw_5_amm180005 (1)
end of crawler
Done cleaning original files

Important words: [('bukowski', 925), ('charles', 350), ('los', 314), ('angeles', 314), ('var', 314), ('new', 290), ('one', 276), ('solid', 250), ('b',
Knowledge base: "smashedthe pulp poetry charles bukowski" adam kirsch new yorker march 14 2005 harpercollins profile timeline resources archived fel

pickling the knowledge base for part 2 of the project

Process finished with exit code 0

PyCharm 2021.2.2 available: // Update... (today 8:45 PM)
119:1 CRLF UTF-8 4 spaces Python 3.7
74°F 11:39 PM 10/12/2021
```



(2) write up a sample dialog you would like to create with a chatbot based on your knowledge base

Hello I am BukowskiFanBoiBot, what is your name?

Anthony

Hi, Anthony. Are you a fan of Chalres Bukosi as well?

Yes

Yippie! What is your favorite poem by Bukowski?

[insert poem name]

That is a great one, display random quote from poem. Did you know [output random fact about Bukowsk's life]

No I did not know that

Oh you must be a fake fan. Did you know [output another random fact]

Yes

Okay, bye

Example functionality:

User: Tell me a bed time story

[output one of Bukowski's short stories]

