

Wiki Scraper Microservice

This program runs independently of your web app. To begin, make sure that **wiki_scraper.py** and its 3 helper text files (**output.txt**, **ready_check.txt**, and **wiki_url.txt**) are all in the same directory.

I recommend including some housekeeping code to make sure the text files are clear when you start your web app. Here is an example of what I did in Python:

```
if __name__ == '__main__':  
    # Housekeeping - Clear contents of all helper text files to begin  
    file = open('website/output.txt', 'r+', encoding="cp437")  
    file.truncate(0)  
    file.close()  
  
    file = open('website/ready_check.txt', 'r+', encoding="cp437")  
    file.truncate(0)  
    file.close()  
  
    file = open('website/wiki_url.txt', 'r+', encoding="cp437")  
    file.truncate(0)  
    file.close()  
  
    # Run web app  
    app.run(debug=True)
```

How it works:

The **wiki_scraper.py** microservice will always be running in a While loop as a separate process alongside your web app. Once your web app backend is ready to receive web-scraped text, you will have the backend write the URL to be scraped to **wiki_url.txt** and also write “ready” to **ready_check.txt**. The microservice will be constantly scanning the **ready_check.txt** file, checking if it contains the word “ready” and if so it proceeds to scrape the URL. Once the microservice checks **wiki_url.txt** for the URL and scrapes the URL, it will write the scraped text to **output.txt**. The microservice automatically deletes the contents of **ready_check.txt** and **wiki_url.txt** upon each performance, and **output.txt** contents are deleted before the scraped text is written to it, so that part is handled for you. Once the scraped text is written to **output.txt**, you are free to pass that into your web app as you see fit!

Below is an example in Python Flask of how to write “ready” and the URL to the text files in the backend, and then check the **output.txt** file and pass that text into the HTML page. It is recommended that your backend sleeps for at least 0.5 seconds between writing “ready” and checking for the scraped text from the output file. You may have to tweak the amount it sleeps as you see fit.

```

# Opens 'wiki_url.txt' file (used in wiki_scraper.py service) and adds url based on state capital
url_path = r"C:\Users\evanm\PycharmProjects\CS 361\website\wiki_url.txt"
url_file = open(url_path, 'w', encoding="utf-8")
url_link = f"https://en.wikipedia.org/wiki/{capital}_{state}"
url_file.write(url_link)
print(f"New URL: https://en.wikipedia.org/wiki/{capital}_{state}")
url_file.close()

# Open 'ready_check.txt' and write "ready" to let web scraping service know to run
ready_path = r"C:\Users\evanm\PycharmProjects\CS 361\website\ready_check.txt"
ready_file = open(ready_path, 'w', encoding="utf-8")
ready_file.write("ready")
ready_file.close()

time.sleep(0.5)

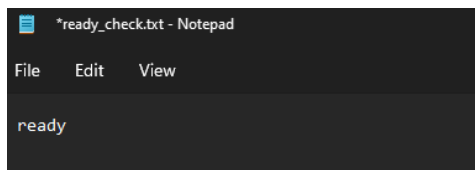
# Opens 'output.txt' file from wiki_scraper.py and adds contents to a string
path = r"C:\Users\evanm\PycharmProjects\CS 361\website\output.txt"
history_file = open(path, 'r', encoding="cp437")
history = history_file.read()
history_file.close()

return render_template("capital.html", capital=capital, state=state, weather=weather, gov_site=gov_site, pic=pic,
                       history=history url=url_link)

```

Examples of what should be written to each helper text file:

ready_check.txt

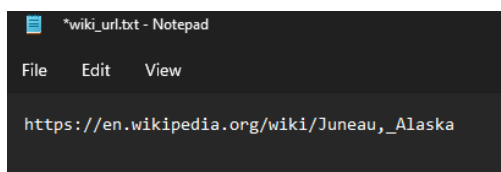


```

ready

```

wiki_url.txt



```

https://en.wikipedia.org/wiki/Juneau,_Alaska

```

output.txt

```
output.txt - Notepad
File Edit View

The City and Borough of Juneau, more commonly known simply as Juneau, is the capital city of the state of Alaska. Located in the City of Juneau, the downtown Juneau is nestled at the base of Mount Juneau and across the channel from Douglas Island. As of the 2020 census, the City of Juneau had a population of 3,343. The city is named after a gold prospector from Quebec, Joe Juneau, though the place was once called Rockwell and then Harrisburg. Juneau is unique among the 49 U.S. capitals on mainland North America in that there are no roads connecting the city to the rest of the state. The Alaska State Capitol in downtown Juneau was built as the Federal and Territorial Building in 1931. Prior to statehood, it housed the Alaska Territorial Legislature. Long before European settlement in the Americas, the Gastineau Channel was a fishing ground for the Auk and Taku tribes, who had lived in the area for centuries. Since the late 20th century, the A'akw Kwáan, together with the Sealaska Heritage Institute, have resisted European-American development of the area. Descendants of these indigenous cultures include the Tlingit people. Native cultures have rich artistic traditions expressed in carvings and totem poles. Although the Russians had a colony in the Alaska territory from 1784 to 1867, they did not settle in Juneau. They conducted extensive fur trading and mining operations in the area.
```