

Pair Exercise #3

Prepared by Wallace Chipidza

June 2023

Due: Write your code in 2 files named **pe4a.py** and **pe4b.py** and push them to a Github repository named **303 summer _23**. Submit a link to the Github repository by 3:59PM on Monday, 7/10/23. Link on Canvas. The exercise will be graded out of 50.

1 Concurrency

The **wikipedia** package enables access to and extraction of Wikipedia content using Python. More details on how the package works can be found here: <https://www.thepythoncode.com/article/access-wikipedia-python>. You can install the package using pip as follows **pip install wikipedia**

1.1 Downloading Wikipedia pages sequentially

You will first create code in **pe4a.py** that extracts content related to the topic “generative artificial intelligence”

- a) Use the method **wikipedia.search** to return a list of topics related to generative artificial intelligence.
- b) Iterate over the list of related topics using a for loop. In the loop do the following for each topic:
 - i) Create a file with the title of the topic and a .txt extension. The title of the topic is returned using **page.title**, and the page is retrieved using **wikipedia.page**. For example, the topic “Music and artificial intelligence” will have an associated file named “Music and artificial intelligence.txt”
The page method may complain that it cannot find the matching content, so make sure the autosuggest flag is false when you call it.
 - ii) Write the contents of the Wikipedia page (retrieved using **page.content** to the file you created in [i]) above.
- c) Print to the console the amount of time it took for your program to execute.

```
Programs/Python/Python310/python.exe c:/Users/jeffr/OneDrive/School_Work/MA/Summer_Semester_2/Software_Development/Apps/303-summer-2
3/pe4a.py
Execution Time: 3.772793769836426 seconds
PS C:\Users\jeffr\OneDrive\School_Work\MA\Summer_Semester_2\Software_Development\Apps\303-summer-23>
```

1.2 Downloading Wikipedia pages concurrently

You will now create code in **pe4b.py** that concurrently extracts content related to the topic “generative artificial intelligence.”

- a) Create a function **download wiki page(topic)** that:
 - i) Creates a file with the title of the topic and a .txt extension.
 - ii) Writes the contents of the Wikipedia page (retrieved using **page.content** to the file you created in [i]) above.
- c) Using the **ThreadPoolExecutor**, concurrently execute code to write files corresponding to all the topics related to “generative artificial intelligence.”
- d) Print to the console the amount of time it took for your program to execute.

```
Programs/Python/Python310/python.exe c:/Users/jeffr/OneDrive/School_Work/MA/Summer_Semester_2/Software_Development/Apps/303-summer-2/Exercise_4/pe4b.py
Execution time: 0.8811652660369873 seconds
```

