BIRZEIT UNIVERSITY

Faculty of Engineering & Technology
Department of Electrical & Computer Engineering
ENCS3390: Operating System Concepts
First Semester, 2024/2025

Project 1

Assigned: 20/10/2024
Due: 20/11/2024

Write a program, in a language and platform of your choice, to list the top 10 most frequent words in the dataset enwik8, found at:
https://huggingface.co/datasets/LTCB/enwik8

Compare the following approaches. Measure the time it takes to complete the program in each case.

1- **Naive approach, a program that does not use any child processes or threads.**
2- **Multiprocessing approach: a program that uses multiple child processes running in parallel. Try different numbers of child processes and compare the outcome: 2, 4, 6, 8 children.**
3- **Multithreading approach: a program that uses multiple joinable threads running in parallel. Try different numbers of threads and compare the outcome: 2, 4, 6, 8 threads.**

In each of the approaches, measure the execution time.

Note: You must work on a computer that has at least 4 cores. If you use a virtual machine, make sure that at least 4 cores are allocated to the virtual machine.

You have to submit a report, along with your code, that discusses the following:

1- **Describe your environment: computer specs (Cores? Speed? Memory?), OS, programming language, and IDE tool, and whether you used a virtual machine.**
2- **How you achieved the multiprocessing and multithreading requirements, i.e. The API and functions that you used.**
3- **An analysis according to Amdahl's law. What percentage is the serial part of your code? What is the maximum speedup according to the available number of cores? What is the optimal number of child processes or threads?**
4- **A table that compares the performance of the 3 approaches.**
5- **Comment on the differences in performance, and**
6- **Conclusion.**