

Speedup é um indicador de desempenho comparativo entre duas versões de um algoritmo, geralmente a sequencial e a paralela. Ele é amplamente utilizado em computação paralela para avaliar o ganho de desempenho obtido com a adição de recursos, como múltiplos processadores ou núcleos. Acesse nosso material para saber mais.

Há uma ampla variedade de problemas que poderíamos fazer a nossa primeira atividade de implementação. Como preciso escolher uma, escolho a seguinte tarefa:

### Problema

Já ouviu falar do projeto [The Project Gutenberg](#)? Ele oferece mais de 60.000 e-books gratuitos, principalmente literatura clássica e histórica. É útil para treinamento de modelos de linguagem e análise de texto, permitindo uma ampla gama de aplicações educacionais e de pesquisa.

Para esta tarefa, escolhi [“Clarissa Harlowe: or the history of a young lady” de Samuel Richardson](#). A versão completa, contendo os 9 volumes, é considerada um dos livros mais longos já escritos na língua inglesa.

Peço que vocês implementem um programa que busque na obra completa as palavras abaixo, são palavras de grande ocorrência no texto. Ao final, seu programa deve dizer o total de ocorrências de cada palavra buscada (não considere partes de palavras na sua busca).

clarissa  
lovelace  
letter  
dear  
miss  
virtue

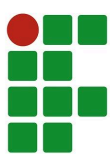
### Solicitações de implementação

- 1) Faça duas estratégias de implementação:
  - a) uma sequencial (programa A);
  - b) uma paralela utilizando threads (programa B).
- 2) Faça um programa testador que irá compor nosso *benchmark*. Esse testador será o responsável por disparar as execuções de cada estratégia (programa A e B), bem como aferir os tempos de execução. O testador deve fornecer dados brutos de desempenho em arquivo para permitir posterior análise estatística.

Os dados brutos esperados para cada bateria de testes são: todos os tempos coletados nos experimentos, tempos excluídos (*warm-up* e *outliers*) e tempos médios.

Os cenários de coleta são:

- a) Programa A - 30 execuções;
- b) Programa B - 30 execuções em cada um dos cenários abaixo:
  - i) 2 threads;
  - ii) 4 threads;



iii) 8 threads.

Ao fim dos experimentos, você deve calcular os *speedups* do programa A em relação a cada uma das versões do programa B.

- 3) Agora escolha a versão mais rápida do algoritmo B. Refaça o mesmo experimento (comparar programa A com a versão mais rápida do programa B) usando as seguintes palavras que possuem nenhuma ou pouca ocorrência no texto:

eita  
bacana  
vixe  
forbidden  
indignation  
oppression

4) REFLITA e ANOTE:

- a) Por que obteve esses resultados?
- b) A relação do tempo do *speedup* coletado com os itens abaixo:
  - i) Qual o hardware usado? Isso influencia?
  - ii) Qual o sistema operacional usado? Isso influencia?
  - iii) Qual a linguagem e recursos da linguagem usados? Haveria diferença se as linguagens fossem diferentes?
  - iv) O conjunto de dados influenciou nos resultados (palavras mais frequentes e mais raras)?
  - v) Por que determinada abordagem obteve melhor desempenho em relação a outra (paralelo x sequencial)?
  - vi) Onde foi possível aumentar a vazão na sua proposta?

Peço que se organizem em trio. A avaliação, contudo, será individual. A apresentação da solução e resultados deverá ser feita em sala de aula.

Lembrem-se que o objetivo desse exercício é que vocês comparem algoritmos em relação ao tempo de execução. Sendo assim, se é desejado aferir o tempo de execução do processo, é necessário levar em consideração que outros processos estão concorrendo por esse tempo de processamento na máquina.

**A APRESENTAÇÃO DO TRABALHO SERÁ NO DIA 06/05 NA AULA. Vocês devem montar o ambiente em máquina e disponibilizar TODAS as informações pedidas para a apresentação. Toda a equipe deve estar presente na apresentação e responder igualmente pelo trabalho.T**