

# EP4

---

Leandro Rodrigues nUSP: 10723944

**Testes empíricos:** Utilizando um texto de 137526 palavras com uma considerável distribuição entre elas, medindo o tempo com `time ./ep4` e executando 100.000 testes para cada estrutura e calculando uma média:

## **Vetor desordenado:**

*Imprimindo em ordem alfabética:* Tempo mínimo 2,231 segundos e máximo 2,513 segundos.

*Imprimindo em ordem de ocorrências:* Tempo mínimo 3,001 segundos e máximo 3,072 segundos.

## **Vetor ordenado:**

*Imprimindo em ordem alfabética:* Tempo mínimo 2,200 segundos e máximo 2,248 segundos.

*Imprimindo em ordem de ocorrências:* Tempo mínimo 10,102 segundos e máximo 12,228 segundos.

## **Lista ligada desordenada:**

*Imprimindo em ordem alfabética:* Tempo mínimo 2,361 segundos e máximo 2,548 segundos.

*Imprimindo em ordem de ocorrências:* Tempo mínimo 2,420 segundos e máximo 2,668 segundos.

## **Lista ligada ordenada:**

*Imprimindo em ordem alfabética:* Tempo mínimo 4,725 segundos e máximo 4,836 segundos.

*Imprimindo em ordem de ocorrências:* Tempo mínimo de 3,721 segundos e máximo de 3,859 segundos.

## **Árvore de busca binária:**

*Imprimindo em ordem alfabética:* Tempo mínimo de 1,880 segundos e máximo de 2,032 segundos.

*Imprimindo em ordem de ocorrências:* Tempo mínimo de 2,117 segundos e máximo de 2,210 segundos.

Para ter certeza dos resultados, aumentei o número de palavras e os resultados se mostraram consistentes. Alguns detalhes importantes foram notados e estarei comentando na conclusão.

## Conclusão

---

A árvore de busca binária com certeza foi a estrutura mais rápida, tanto para imprimir em ordem de ocorrências quanto em ordem alfabética. A lista ligada desordenada foi a estrutura mais lenta na média, enquanto que o vetor ordenado foi o que tomou mais tempo para impressão (no caso de ordenação por ocorrências). O motivo da demora do vetor desordenado foi que, por problemas que tive em montar uma função eficiente, tive que fazer uma implementação que utiliza de  $n$  quicksorts, tornando a complexidade da função maior que quadrática. Foram testados textos de outros idiomas. Textos em inglês, em média, apresentaram taxa de ocorrência repetida de palavras maior, devido aos artigos, pois em inglês todos acabam convergindo na palavra "the". Em textos menores, a diferença é quase insignificante, mudando apenas algumas casas decimais. Um fato curioso é que, quando a lista ligada ordenada é utilizada para imprimir em ordem de ocorrência, ela imprime as palavras com acento em uma ordem diferente das outras funções. Não tratei esse caso pois foi dito que não devemos nos preocupar com as palavras acentuadas.