

Introdução

A Árvore B foi criada por Bayer e McCreight[2] em 1972. É uma árvore m-ária, ou seja, seus nós possuem mais de um registro e também possuem mais de dois descendentes.

A dupla teve como guia uma descrição do segundo trabalho prático, a qual foi disponibilizada pela professora da disciplina. Nela há o requerimento da criação do código de uma Árvore B que armazena registros os quais devem ser inseridos, removidos e pesquisados dentro da árvore respeitando as regras da estrutura de árvore B para a execução de análises sobre o comportamento desta estrutura de dados.

Objetivo

Este relatório tem como objetivo analisar os dados obtidos ao inserir 10.000 registros em árvores-B de ordem 2, 20 e 40, compará-los e tirar conclusões sobre os mesmos.

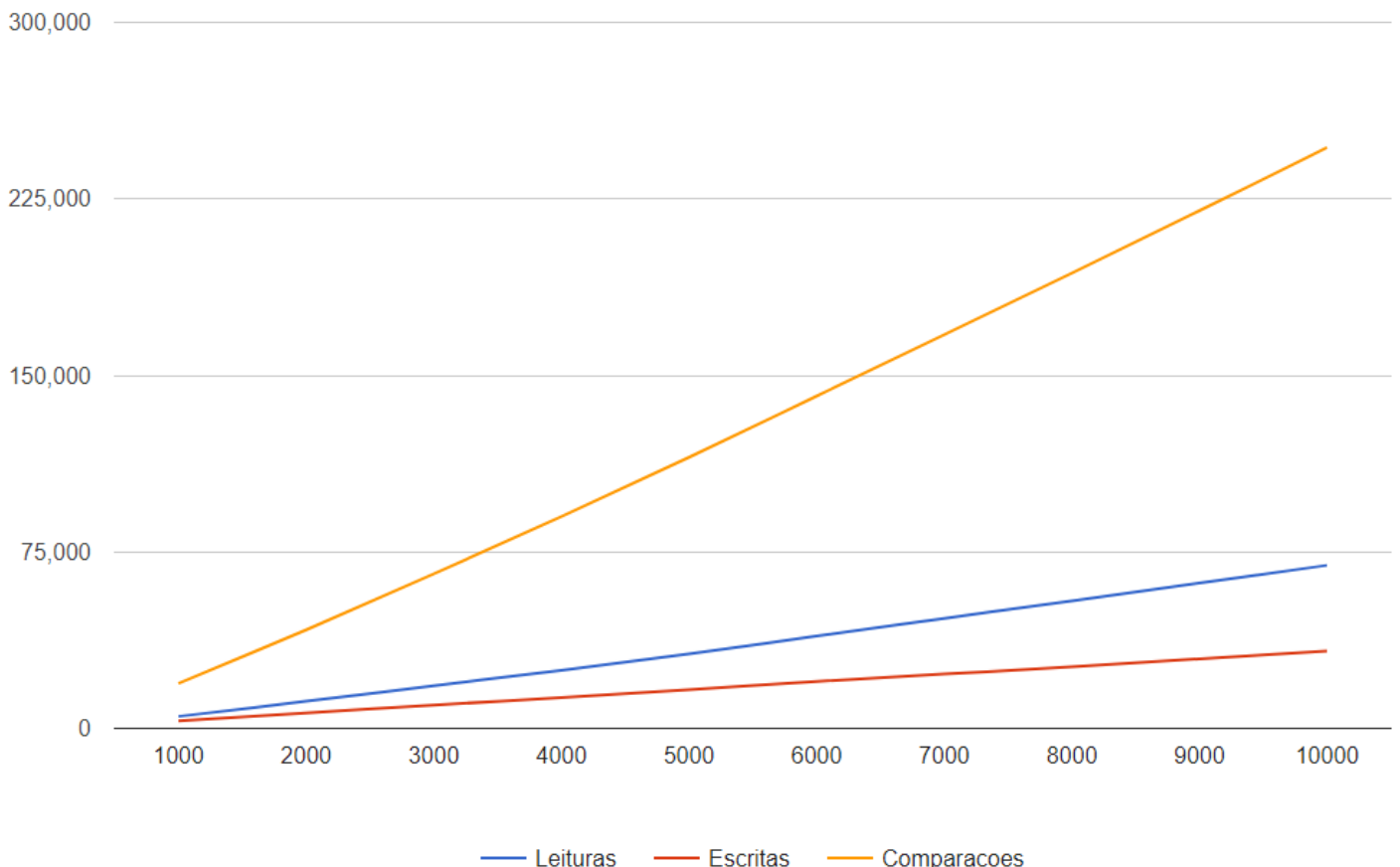
Análise Comparativa

- Testes comparativos com valores da ordem de árvore $M=2$, $M=20$ e $M=40$.
- Os gráficos abaixo podem ser gerados pelo programa principal (main.exe), nele foi implementada uma integração com o Google Charts (Para mais informações, acessar a documentação em anexo).
- Todos os gráficos abaixo podem ser encontrados em “\graph\generated”, eles são arquivos .html interativos mostrando os valores em cada ponto e estão com maior definição.

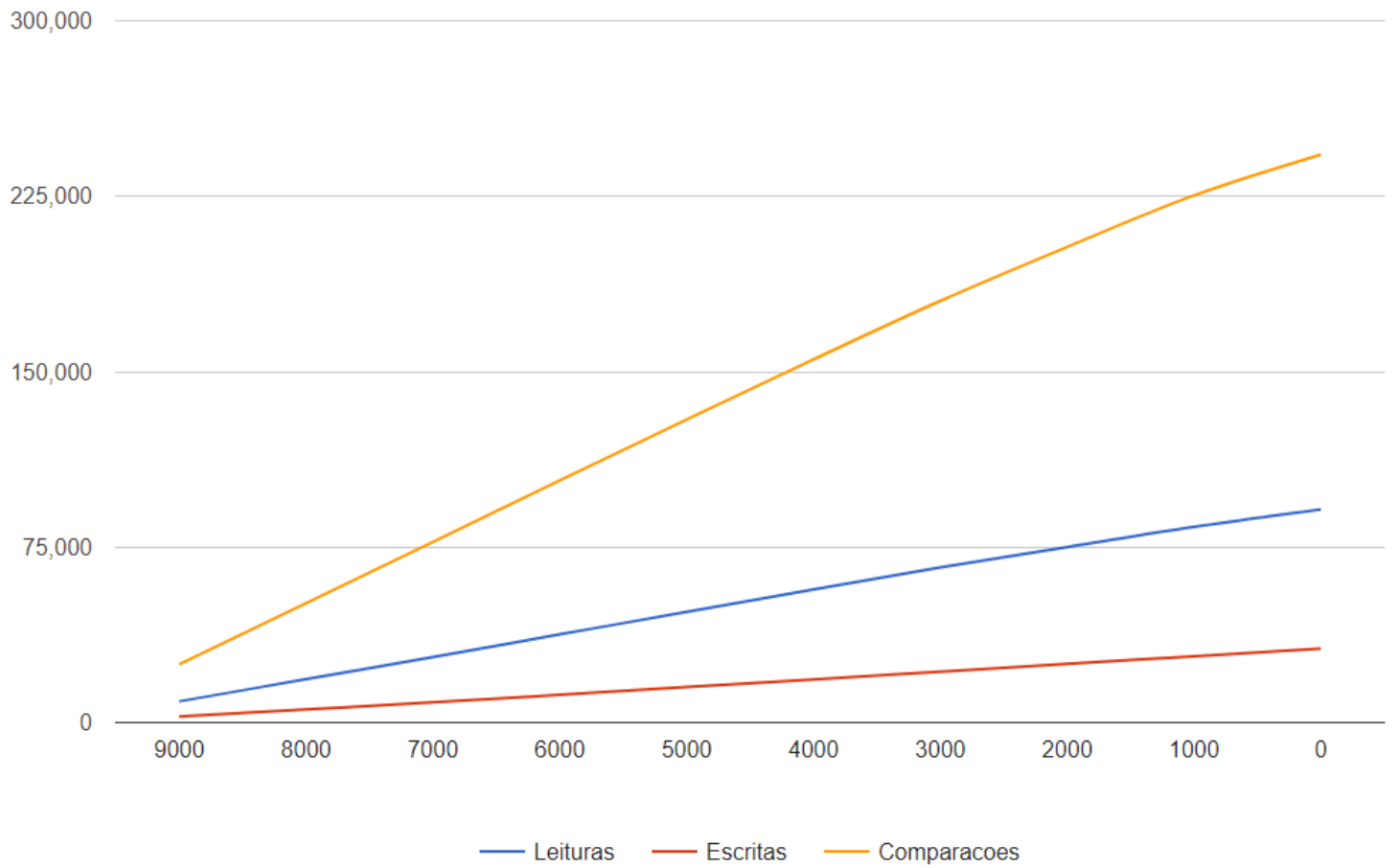
- **Para $M = 2$ e 10.000 registros:**

Altura: 7

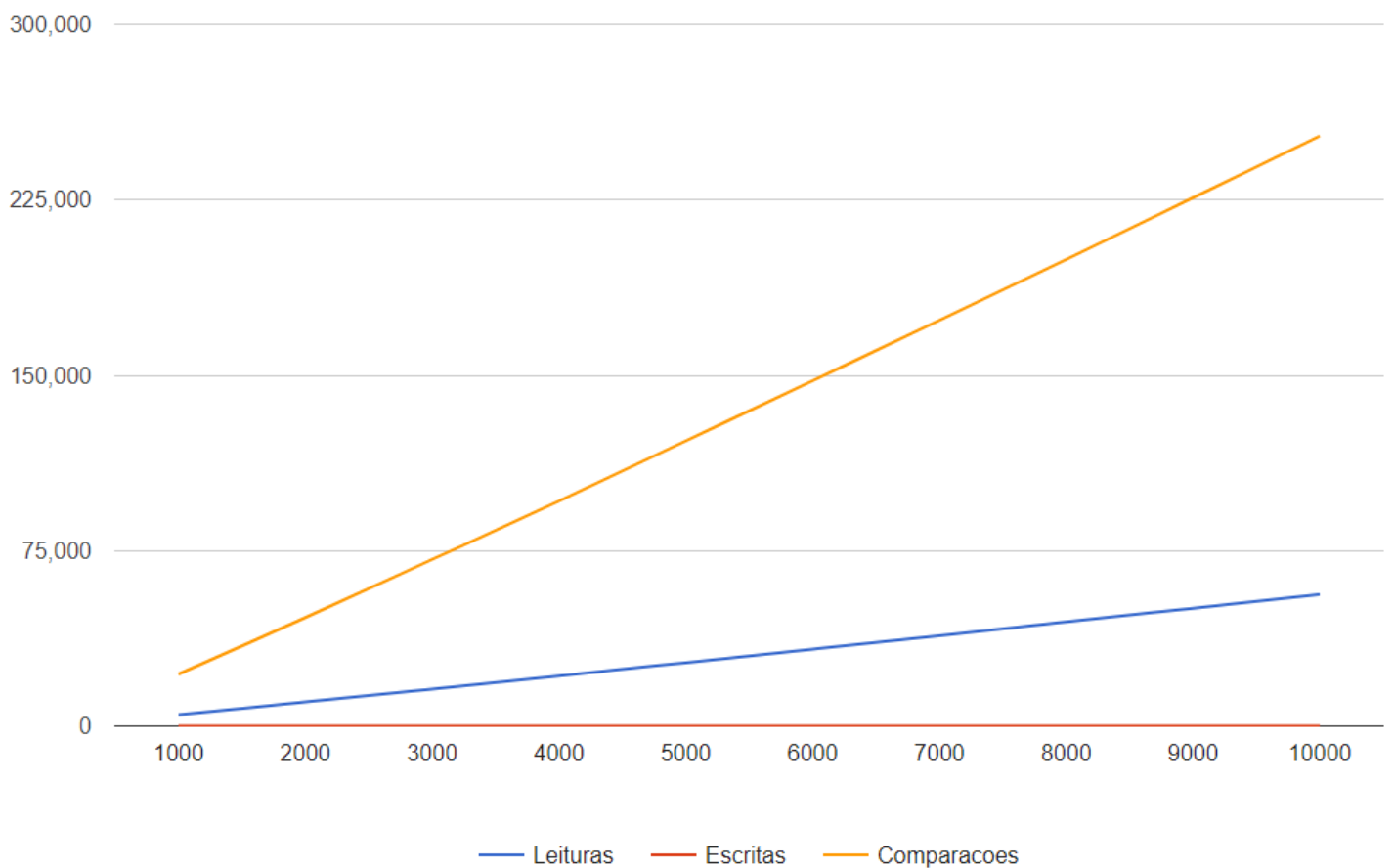
Insercao



Remocao



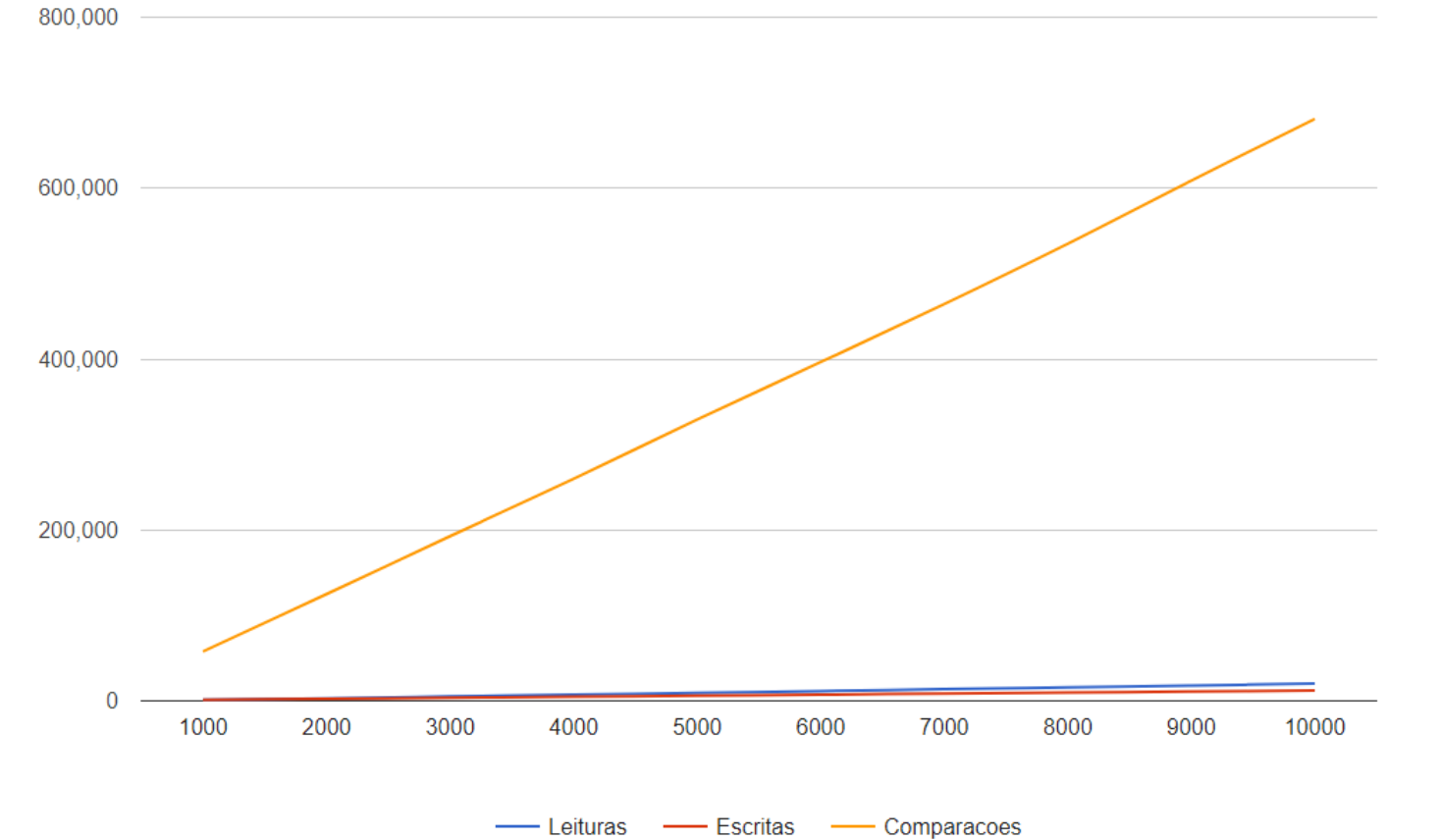
Pesquisa



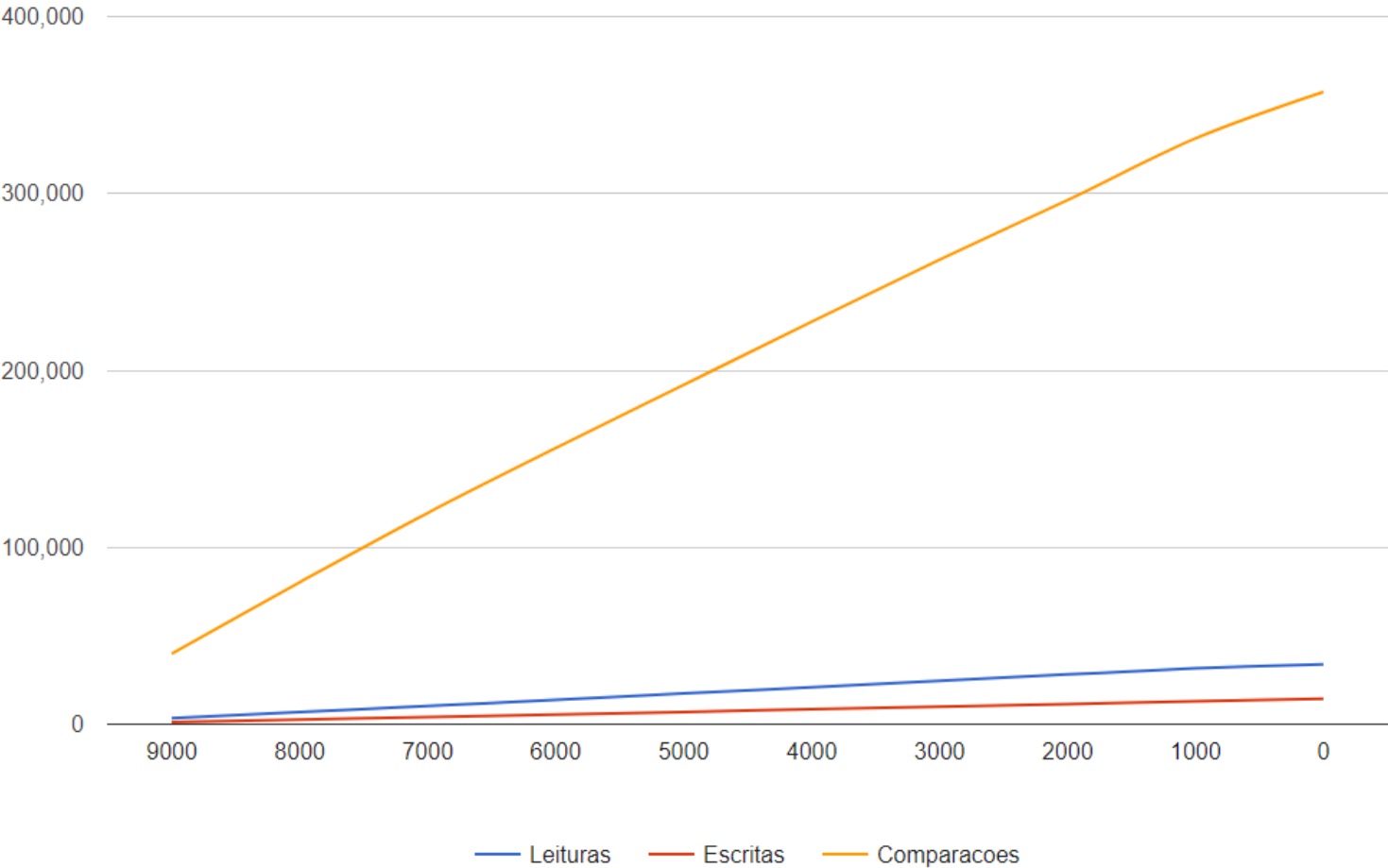
• Para M = 20 e 10.000 registros:

Altura: 3

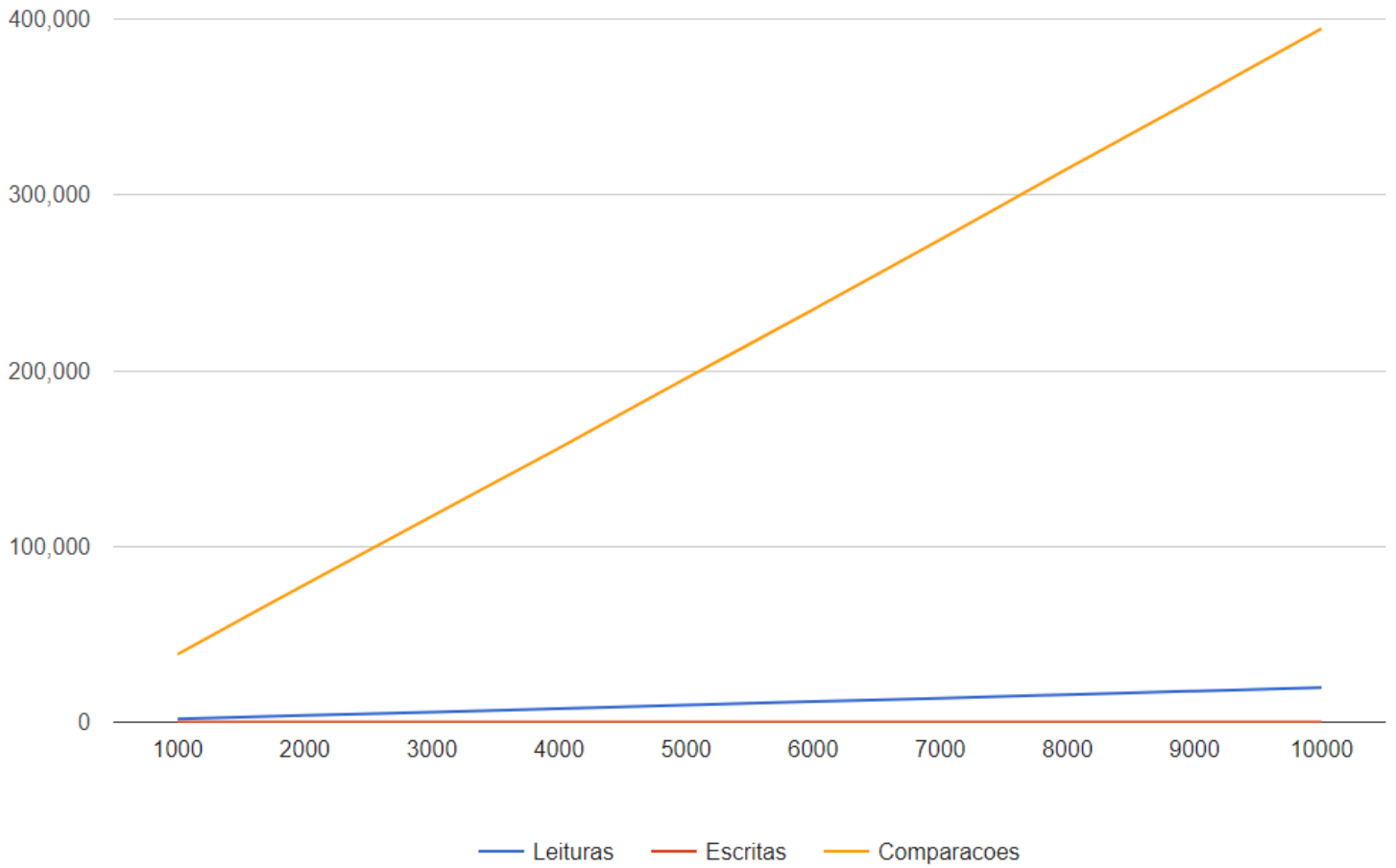
Insercao



Remocao



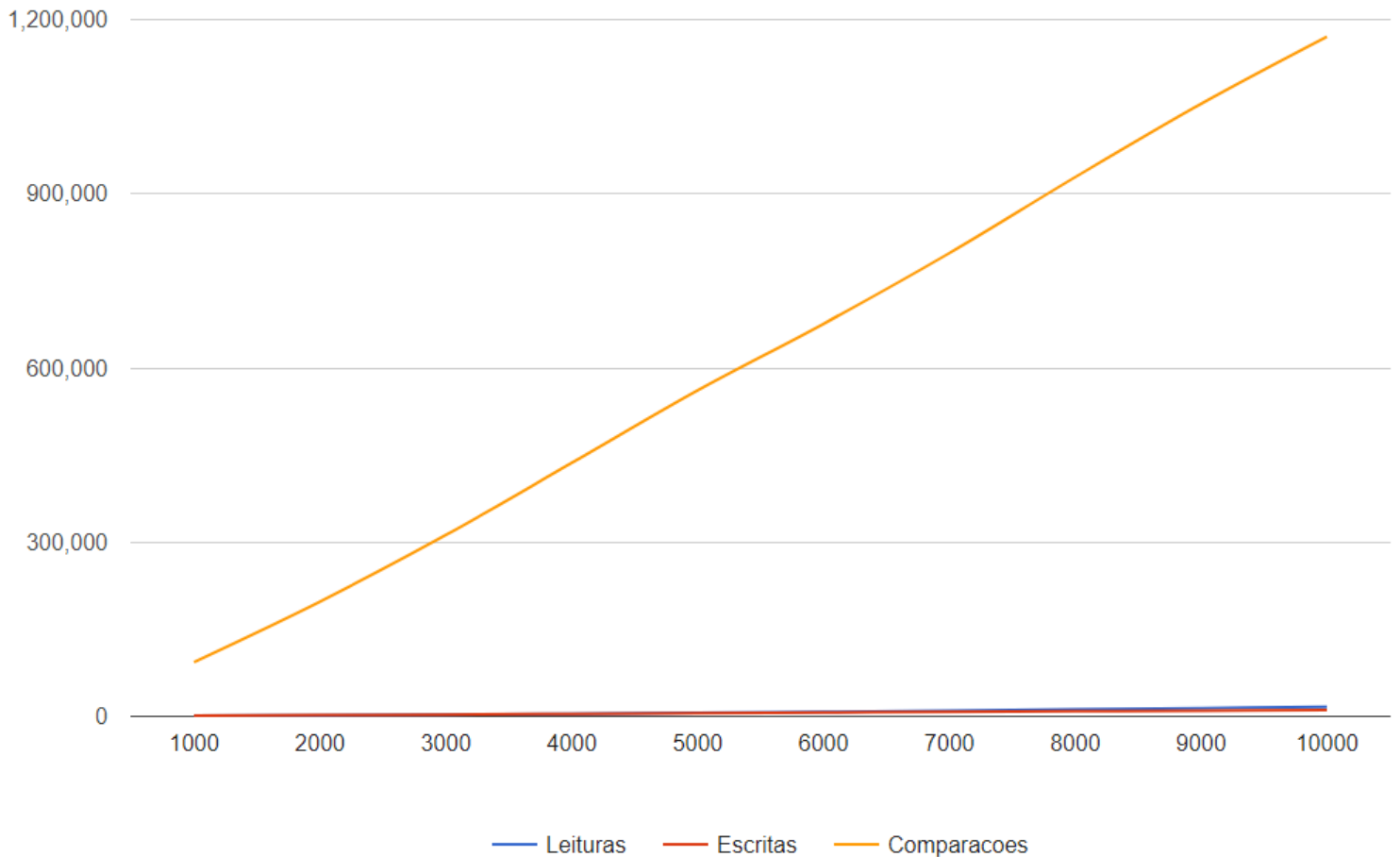
Pesquisa



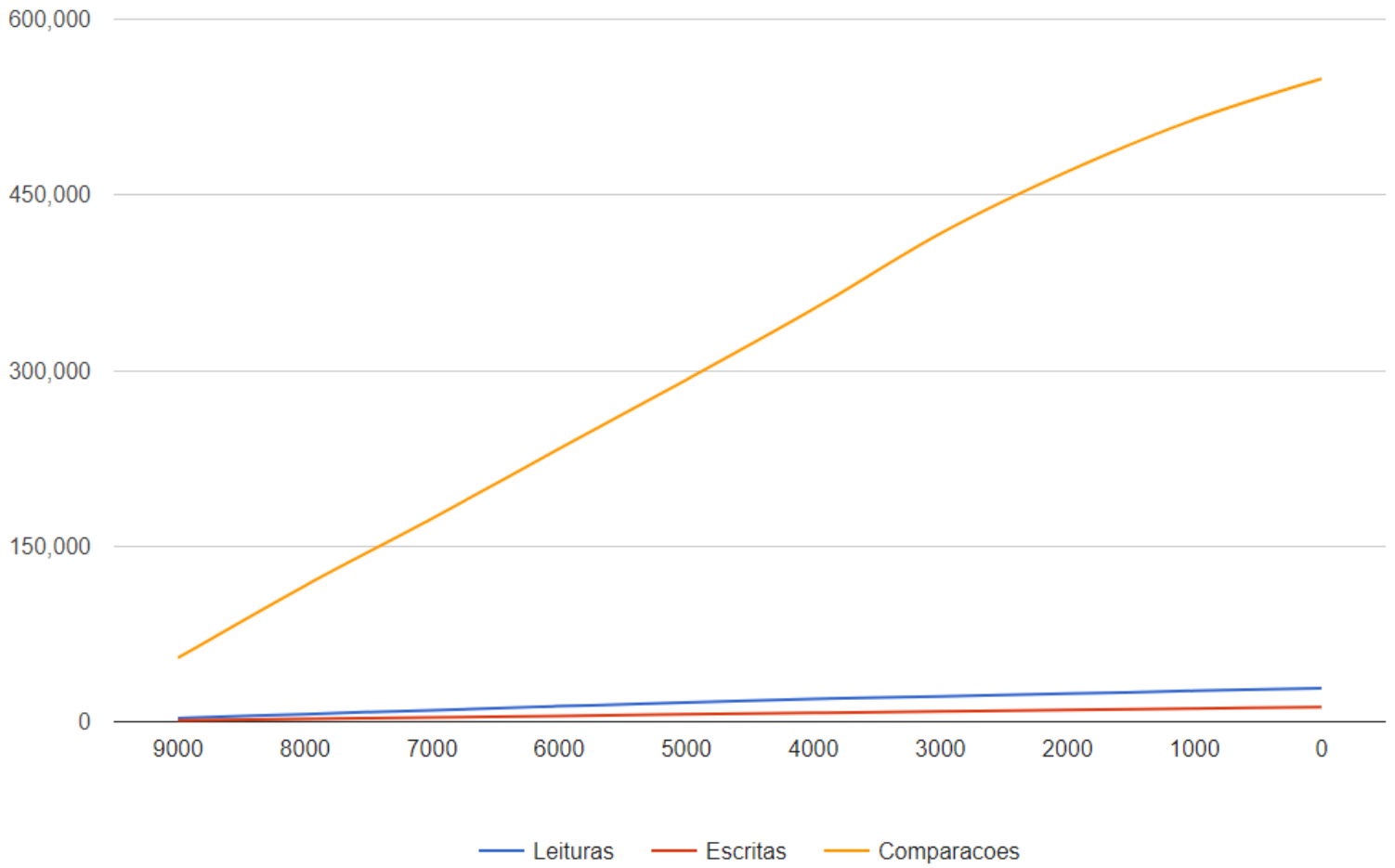
- Para $M = 40$ e 10.000 registros:

Altura: 3

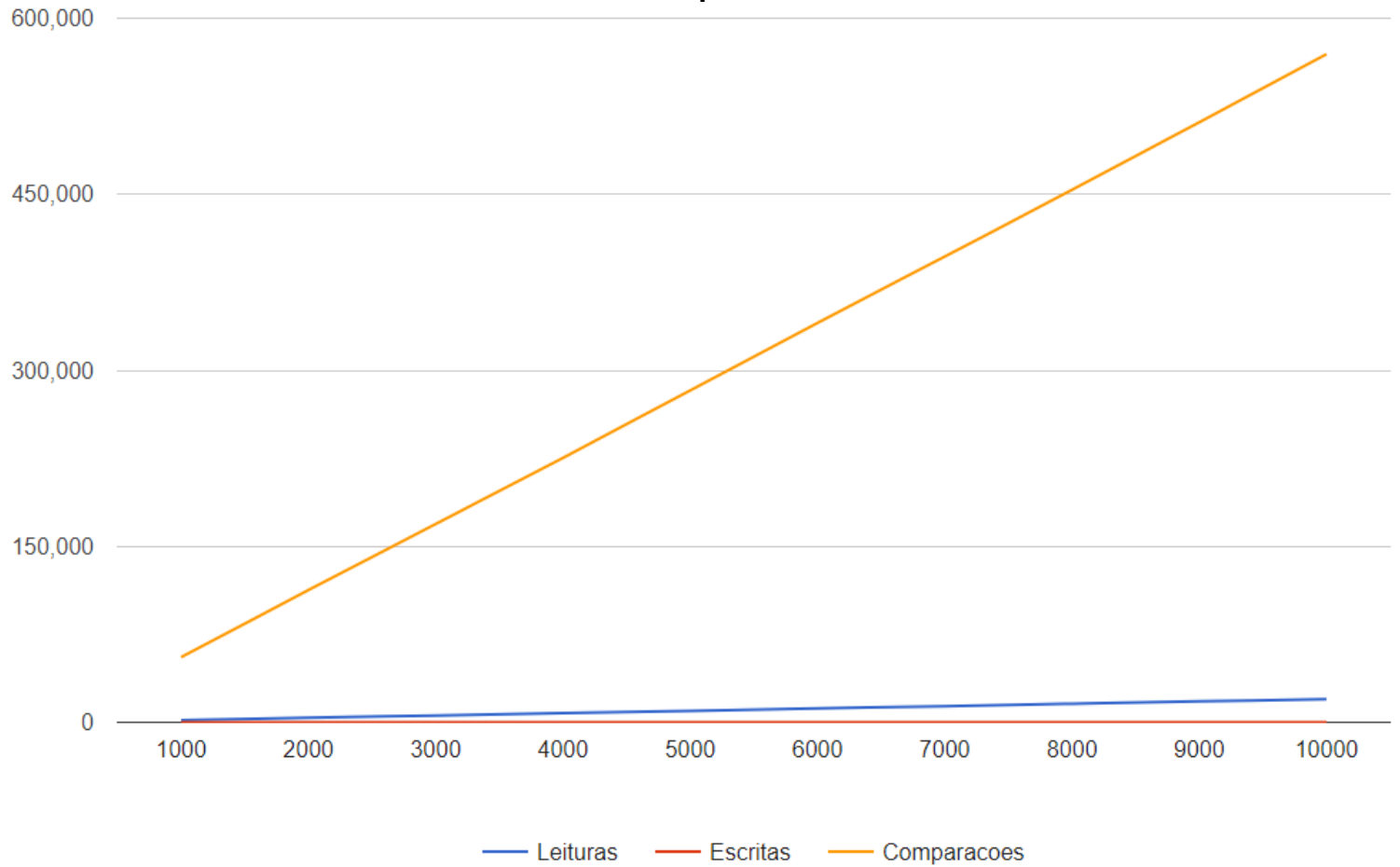
Insercao



Remocao



Pesquisa



Interpretação Exaustiva dos Dados Apresentados

Serão comparados exaustivamente os dados, no final haverá sua interpretação e inferência.

- **Ordem M=2 e Ordem M=20**

Inserção:

Comparações:

Árvore com ordem M=20 precisou de 2,76 vezes mais comparações.

Escritas:

Árvore com ordem M=2 precisou de 3,18 vezes mais acessos de escrita ao disco.

Leituras:

Árvore com ordem M=2 precisou de 3,44 vezes mais acessos de leitura ao disco.

Remoção:

Comparações:

Árvore com ordem M=20 precisou de 1,47 vezes mais comparações.

Escritas:

Árvore com ordem M=2 precisou de 2,2 vezes mais acessos de escrita ao disco.

Leituras:

Árvore com ordem M=2 precisou de 2,68 vezes mais acessos de leitura ao disco.

Pesquisa:

Comparações:

Árvore com ordem M=20 precisou de 1,56 vezes mais comparações.

Leituras:

Árvore com ordem M=2 precisou de 2,86 vezes mais acessos de leitura ao disco.

- **Ordem M=2 e Ordem M=40**

Inserção:

Comparações:

Árvore com ordem M=40 precisou de 4,74 vezes mais comparações.

Escritas:

Árvore com ordem M=2 precisou de 2,95 vezes mais acessos de escrita ao disco.

Leituras:

Árvore com ordem M=2 precisou de 4,29 vezes mais acessos de leitura ao disco.

Remoção:

Comparações:

Árvore com ordem $M=40$ precisou de 2,26 vezes mais comparações.

Escritas:

Árvore com ordem $M=2$ precisou de 2,51 vezes mais acessos de escrita ao disco.

Leituras:

Árvore com ordem $M=2$ precisou de 3,18 vezes mais acessos de leitura ao disco.

Pesquisa:

Comparações:

Árvore com ordem $M=40$ precisou de 2,26 vezes mais comparações.

Leituras:

Árvore com ordem $M=2$ precisou de 2,84 vezes mais acessos de leitura ao disco.

• **Ordem $M=20$ e Ordem $M=40$**

Inserção:

Comparações:

Árvore com ordem $M=40$ precisou de 1,72 vezes mais comparações.

Escritas:

Árvore com ordem $M=20$ precisou de 1,1 vezes mais acessos de escrita ao disco.

Leituras:

Árvore com ordem $M=20$ precisou de 1,24 vezes mais acessos de leitura ao disco.

Remoção:

Comparações:

Árvore com ordem $M=40$ precisou de 1,54 vezes mais comparações.

Escritas:

Árvore com ordem $M=20$ precisou de 1,14 vezes mais acessos de escrita ao disco.

Leituras:

Árvore com ordem $M=20$ precisou de 1,18 vezes mais acessos de leitura ao disco.

Pesquisa:

Comparações:

Árvore com ordem $M=40$ precisou de 1,44 vezes mais comparações.

Leituras:

Árvore com ordem $M=20$ precisou de aproximadamente o mesmo número de acesso de leitura ao disco do que a árvore de ordem $M=40$.

A árvore de ordem $M=2$ teve o menor número de comparações quando comparada com as de ordem $M=20$ e $M=40$.

A árvore de ordem $M=2$:

- Apresentou o maior número de acessos ao disco.
- Apresentou o menor número de comparações.

A árvore de ordem $M=40$:

- Apresentou o menor número de acessos ao disco.
- Apresentou o maior número de comparações.

A árvore de ordem $M=20$:

- Apresentou um número maior, porém próximo de acessos ao disco quando comparada à árvore de ordem $M=40$.
- Apresentou o menor que a árvore de ordem $M=40$.

Conclusão

Neste comparativo foi possível observar o comportamento das B-Trees, sua eficiência a lidar com dados em memória externa e a mudança dos resultados produzidos com a mudança de sua ordem.

Caso o objetivo seja unicamente diminuir o número de acessos ao disco a árvore de ordem $M=40$ é a melhor opção. Caso o número de comparações também seja uma preocupação, a árvore de ordem $M=20$ é uma opção intermediária que oferece menos comparações com um número de comparações próximo ao da árvore de ordem $M=40$.

Referências

[1] Comer, "The Ubiquitous B-Tree", 1971

[2] Bayer & McCreight, "Organization and Maintenance of Large Ordered Indexes", 1972

[3] Ronald Fisher and Frank Yates, "*Statistical tables for biological, agricultural and medical research.*", 1978

[4] G. T. de Assis. Pesquisa Externa. (Notas de Aula) UFOP. 2013. Disponível em< <http://www.decom.ufop.br/guilherme>>.