

Felipe Gallois

Comparação entre Árvores Binárias de Pesquisa e Árvores de Adelson-Velsky

Universidade Federal do Rio Grande do Sul – UFRGS

Disciplina de Estruturas de Dados

Porto Alegre

4 de Abril de 2022

Sumário

1	INTRODUÇÃO	3
2	METODOLOGIA	5
2.1	Testes com textos-fonte diferentes	5
2.1.1	Capítulo primeiro de <i>O Alienista</i>	5
2.1.2	Capítulo terceiro de <i>O Alienista</i>	7
2.2	Testes com dicionários diferentes	8
2.2.1	Capítulo primeiro de <i>O Alienista</i>	8
2.2.2	Capítulo terceiro de <i>O Alienista</i>	11
3	CONCLUSÃO	13
4	REFERÊNCIAS	15

1 Introdução

Os objetivos deste relatório são expor um experimento criado para comparar o desempenho de Árvores Binárias de Pesquisa (ABPs) e Árvores de Adelson-Velsky (AVLs) em termos de inserções e consultas de nodos e analisar em quais situações a utilização de cada modelo seria mais recomendável, visando a performance. Foram desenvolvidas duas versões de uma aplicação capaz de, se suprida de um texto de entrada e um texto contendo pares de palavras sinônimas, produzir um arquivo com o conteúdo de entrada parafraseado. Para tanto, no início da sua execução, são carregadas em uma ABP ou uma AVL, dependendo da versão do programa, as palavras do dicionário de sinônimos, que, após isso, são acessadas apenas através da estrutura. Ao término da rotina, após realizar todas as operações, o programa informará dados importantes para a contraposição dos modelos, tais como total de comparações e rotações. Testando em condições diversas, pôde-se verificar, na prática, características de ambos tipos de árvores que serviram para determinar os seus melhores usos.

2 Metodologia

O programa desenvolvido inicializa carregando o arquivo de dicionário na árvore. Em seguida, mede a altura da estrutura e calcula o seu número de nodos. Após isso, percorre o texto de entrada do início ao fim, copiando suas palavras literalmente ou realizando as substituições indicadas pela árvore. Finalmente, imprime em linha de comando as seguintes informações obtidas após a operação: número de nodos da árvore, sua altura, quantidade de rotações realizadas e número de comparações. Para selecionar entre os modos ABP e AVL, deve-se informar o compilador da escolha no momento de geração do arquivo executável.

2.1 Testes com textos-fonte diferentes

2.1.1 Capítulo primeiro de *O Alienista*

Compiladas duas versões do programa, que faziam uso de ABP e de AVL, seguiu-se com os testes. Primeiramente, cada uma das aplicações foi testada com um dicionário de 20.000 palavras desordenadas, que formavam 10.000 pares de sinônimos, e com o capítulo primeiro da novela de Machado de Assis, *O alienista*^[1]. Adiante estão gráficos com os resultados do primeiro experimento. O gráfico representando número de nodos foi omitido por ser constante durante todo o experimento.

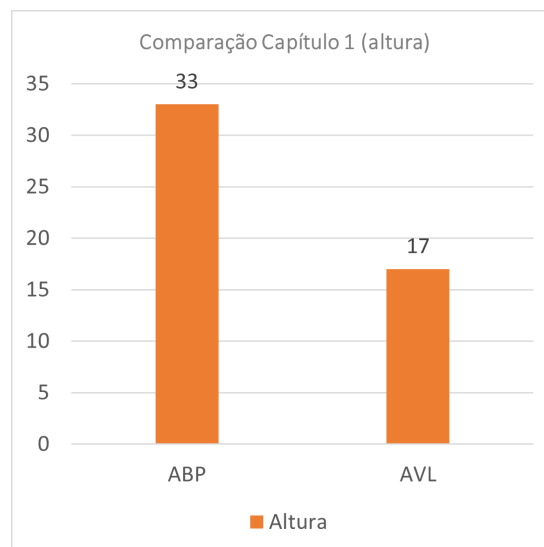


Figura 1 – Altura das árvores.

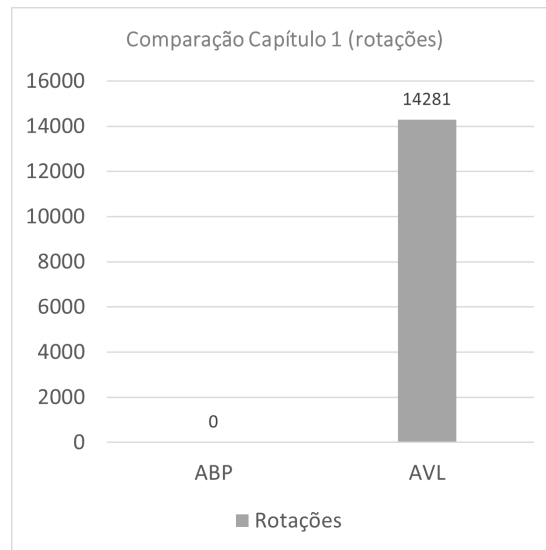


Figura 2 – Número de rotações realizadas.

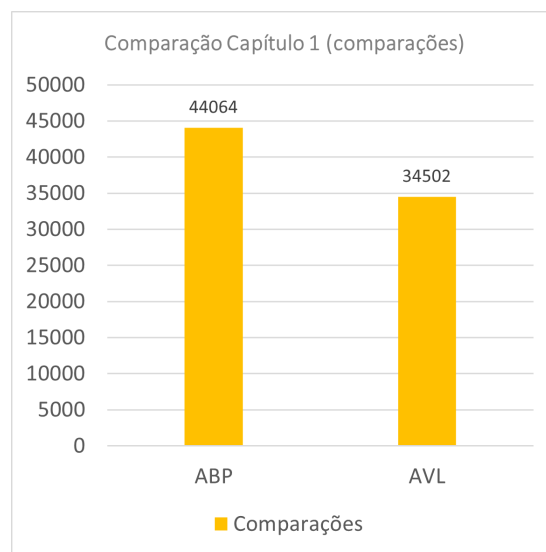


Figura 3 – Número de comparações realizadas.

Como é possível observar, a ABP era muito mais alta que a AVL, fato explicado pelo balanceamento por rotações realizado pela última e ilustrado na figura 2. Já em termos de número de comparações, a ABP realizou aproximadamente 27,71% mais comparações que a outra.

Uma função recursiva extra foi utilizada para calcular a média de consultas que se empregaria pesquisando por um nodo aleatório nas árvores e obteve os seguintes resultados: aproximadamente 17,9 para a ABP e 13,59 para a AVL. Isso corresponde a aproximadamente 31,71% mais consultas feitas pela primeira, próximo dos 27,71% encontrados no

experimento. A imprecisão pode ser atribuída ao fato de que a amostra possuía apenas 1.212 palavras, enquanto que o dicionário continha 20.000.

2.1.2 Capítulo terceiro de *O Alienista*

Foi realizado um segundo experimento, semelhante ao primeiro, porém com o capítulo terceiro de *O Alienista*, no qual observou-se uma situação semelhante:

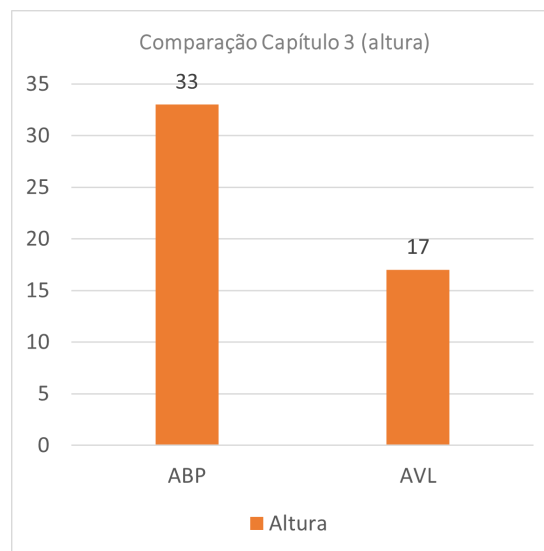


Figura 4 – Altura das árvores.

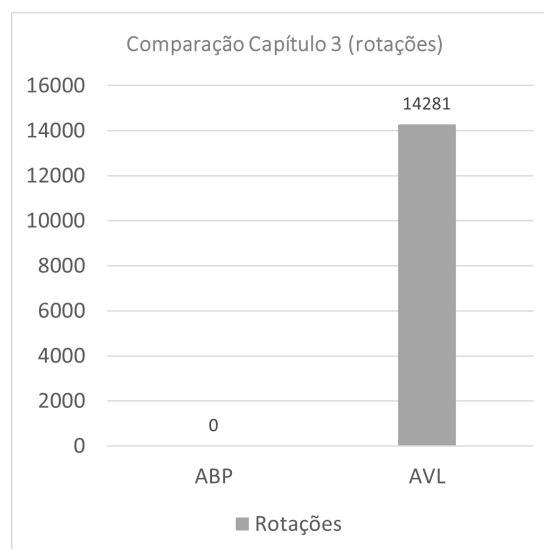


Figura 5 – Número de rotações realizadas.

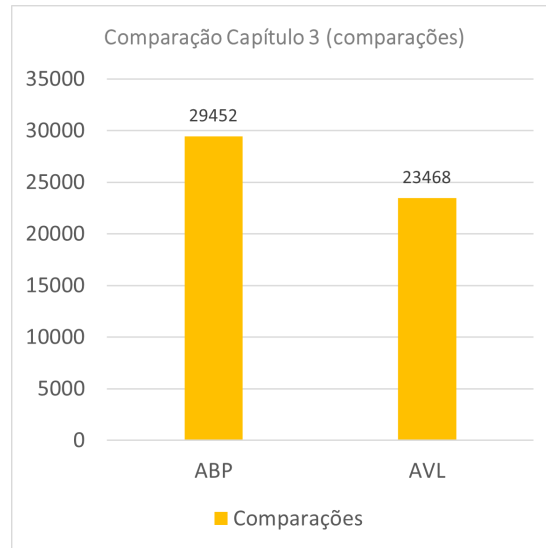


Figura 6 – Número de comparações realizadas.

As alturas e os números de rotações das árvores permaneceram iguais, pois o dicionário não foi alterado; contudo, o número de comparações mudou, devido ao texto dessa vez possuir apenas 825 palavras. Assim, o programa com ABP fez aproximadamente 25,50% mais comparações do que o com AVL, valor próximo do observado na situação anterior.

2.2 Testes com dicionários diferentes

2.2.1 Capítulo primeiro de *O Alienista*

Na sequência, testou-se o texto do capítulo 1 com um dicionário com as mesmas palavras do anterior, mas em ordem alfabética. Eis os resultados:

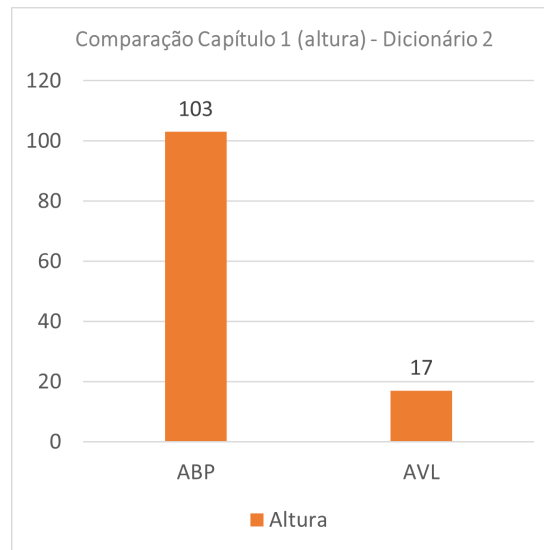


Figura 7 – Altura das árvores.

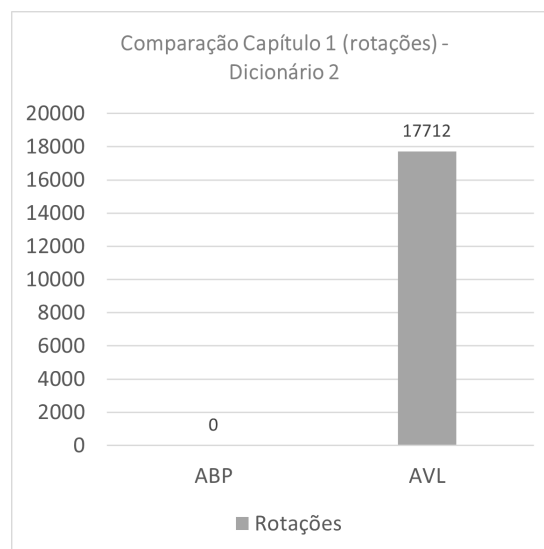


Figura 8 – Número de rotações realizadas.

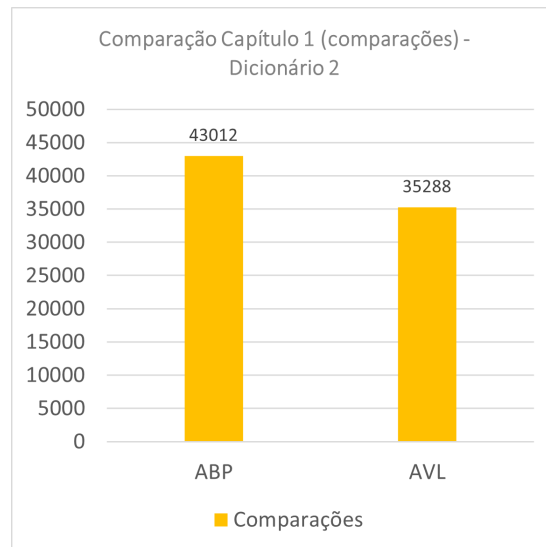


Figura 9 – Número de comparações realizadas.

A diferença de altura entre as árvores aumentou muito, visto que, pelo fato de as palavras do dicionário terem estado em ordem alfabética, mais nodos foram inseridos desproporcionalmente nas subárvores direitas da ABP. Todavia, ao analisarmos o gráfico, não verificamos um aumento proporcional do número de comparações na ABP, consequência esperada do aumento da altura. Ao contrário, a porcentagem de comparações adicionais em relação à AVL reduziu para 21,89%. Por efeito de comparação, a média de consultas adicionais para encontrar um nodo qualquer nessa árvore deveria ser 46,62%. O que poderia explicar essa diferença?

Após momentos de investigação, chegou-se a uma possível causa para o fenômeno: as chaves das palavras do texto eram pequenas, em média. Outro programa auxiliar foi criado, dessa vez para calcular a média dos códigos ASCII dos primeiros caracteres das palavras do texto procuradas nas árvores, o qual determinou que o valor que melhor aproxima o capítulo um era a letra *f*, a sexta letra dentre as 26 do alfabeto. Isso indicou que, para uma palavra qualquer do texto, já nas primeiras comparações da ABP, era comum que uma consulta fosse realizada em uma subárvore da esquerda, o que aumentava as chances de se atingir um nodo folha rapidamente e, por consequência, de se encerrar a sequência de comparações.

A grande vantagem que uma AVL tem em relação a uma ABP é diminuir o número de comparações realizadas para encontrar um nodo, especialmente em situações em que seriam necessárias inúmeras para que a última acabasse de pesquisar. Contudo, no caso específico do dicionário ordenado, situações assim devem ter sido menos frequentes. Como pudemos verificar, as chaves das palavras do capítulo primeiro eram muito pequenas, o que combinado com a estrutura da ABP do segundo dicionário, resultou em um número mais baixo de comparações por palavra nesta, apesar da sua altura.

2.2.2 Capítulo terceiro de *O Alienista*

Por fim, foi realizado um último teste, com o dicionário ordenado e o capítulo terceiro, o qual produziu os seguintes resultados:

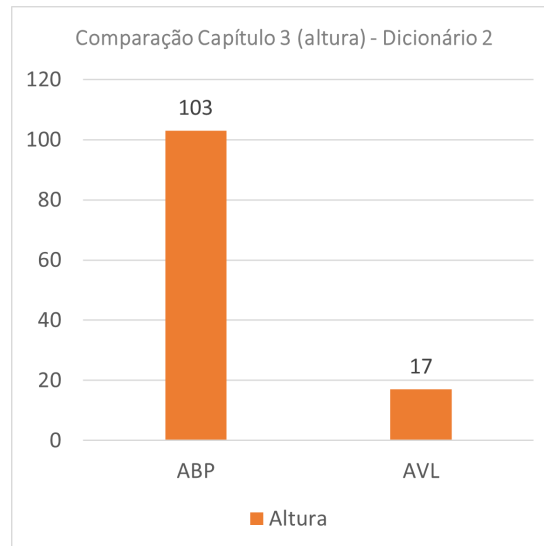


Figura 10 – Altura das árvores.

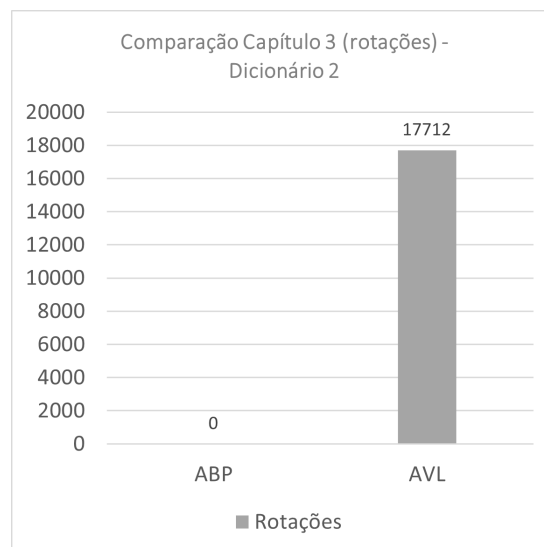


Figura 11 – Número de rotações realizadas.

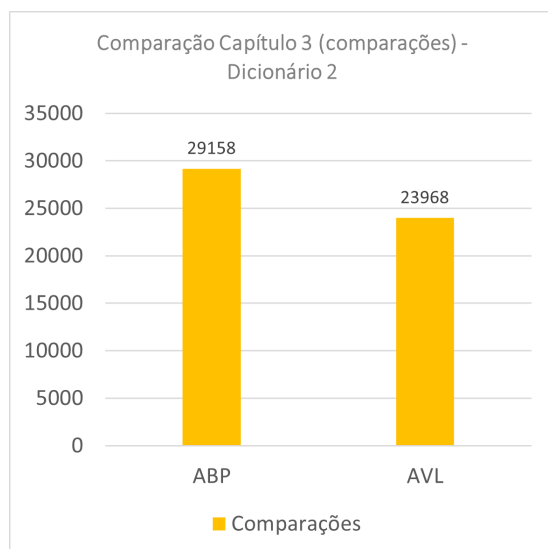


Figura 12 – Número de comparações realizadas.

Exceto pelo gráfico de comparações, devido à menor quantidade de palavras, esse caso foi idêntico ao anterior; portanto, dispensa-se explicações.

3 Conclusão

Tanto ABPs quanto AVLs apresentam vantagens e desvantagens. Uma ABP é mais rápida para fazer inserções de nodos, visto que não realiza nenhuma rotação; todavia, é potencialmente ineficaz em consultas, especialmente quando o nodo procurado se localiza em um nível muito elevado. Uma AVL, por outro lado, é quase sempre mais eficiente que a outra em se tratando de consultas, não obstante requer que sua estrutura seja modificada a cada inserção de um novo nodo. A determinação do modelo para utilizar em um programa, portanto, deve levar em consideração o mínimo impacto de performance, o que se traduz em ponderar ambas opções no contexto específico da aplicação.

Por outro lado, é de se ressaltar a dificuldade técnica de quantificar precisamente os benefícios de cada modelo em um determinado contexto, em vista das variáveis que podem alterar significativamente tais valores. Como foi mencionado no terceiro caso de teste, uma expectativa de 46,62% de desvantagem da ABP em relação à AVL foi frustrada pelos 21,89% medidos, simplesmente por causa das palavras utilizadas no texto do capítulo 1. Ademais, somente a diferente ordenação das palavras do dicionário do terceiro caso em relação ao primeiro causou a redução da diferença de comparações de 27,71% para 21,89%. Em conclusão, não há escapatória para o fato de que para cada aplicação devam ser individualmente testados ambos modelos com uma ampla amostra de valores representativos de utilizações normais do programa para determinar qual seria o mais indicado para o caso específico. Há diferentes ferramentas disponíveis ao desenvolvedor que medem o tempo decorrido para executar algum trecho de código e que podem servir para auxiliar a tomar a decisão mais sábia. Em situações em que performance não representa um fator crítico, porém, a intuição deve bastar para determinar o procedimento correto.

4 Referências

1. FUNDAÇÃO BIBLIOTECA NACIONAL. O alienista, 2002. Arquivo em PDF da novela. Disponível em: <<http://www.dominiopublico.gov.br/download/texto/bn000012.pdf>>. Acesso em: 04 de abr. de 2022.