

UTFPR - Universidade Tecnológica Federal do Paraná
Campus Campo Mourão

Aluno: Jessé Pires Barbato Rocha

RA: 2149389

Disciplina: Algoritmos e Estruturas de Dados 2

6. Preencha a tabela abaixo usando as funções que você implementou. A tabela mostra a quantidade de colisões que aconteceram com cada função de sondagem. Para cada N, crie uma tabela hash com tamanho inicial $M = 5$ e insira as chaves geradas aleatoriamente com a função `int* random_vector(int n, int max, int seed)`. Use $\text{max} = 10n$ e $\text{seed} = 42$.

Colisões		N				
		100	1000	10000	100000	1000000
Função de Sondagem	Linear	11	201	3000	10986	183918
	Quadrático	11	194	2808	10693	176474
	Duplo	8	185	2593	10204	168346

7. Analisando os resultados obtidos no item 6:

i. Alguma função de sondagem resultou em consistentemente menos colisões para todos os N analisados? Se sim, qual?

R: a função de sondagem dupla se mostrou consistentemente menor para todos os N analisados. Isso se deve ao fato de os caminhos de sondagem que vão sendo criados na tabela hash com essa função são mais distintos e espalhados do que nas outras, gerando assim, menos colisões.

ii. Como os números de colisões se comparam com os N correspondentes? O que isso quer dizer para o custo da busca? O que aconteceria se permitíssemos que α ficasse maior antes de redimensionar a tabela?

R: a quantidade de colisões em todos os casos ficou abaixo da metade de N. Isso mostra que a quantidade de iterações necessárias para encontrar uma posição livre para inserção dos elementos é consideravelmente pequena em relação a N. Como a busca por um elemento segue uma estratégia semelhante à da inserção para encontrar uma posição válida (calcular a *função hash* incrementando k quantas vezes for necessário), a busca também terá um custo baixo. Caso fosse permitido que α ficasse maior que $\frac{m}{2}$, a quantidade de colisões, bem como o custo das buscas, aumentaria, uma vez que menos posições livres estariam disponíveis antes que a tabela fosse redimensionada.

iii. Quais outros padrões você identificou? Explique.

R: analisando os resultados expostos na tabela, nota-se que, conforme N aumenta, a diferença entre a quantidade de colisões de acordo com a função de sondagem aumenta também. Tal diferença aparece principalmente entre a sondagem linear e a dupla. Isso se dá em razão de a sondagem dupla criar caminhos mais espalhados entre os elementos do que a sondagem linear. Como o valor de M depende de N nesta implementação, quanto maior for N, maior será M. Isso resulta em uma maior possibilidade de caminhos criados pela sondagem dupla e consequentemente, menor quantidade de colisões.