

Semana 7: Hashing - Tratamento de Colisões por Endereçamento Aberto com Sondagem Linear

d) Preencha a tabela abaixo usando as funções que você implementou. Para cada M, crie uma tabela hash com endereçamento aberto com sondagem linear e insira as chaves geradas aleatoriamente com a função `int* random_vector(int n, int max, int seed)` Use $n = 10000$, $\max = 10n$ e $\text{seed} = 42$.

M	N	α	Custo Mal-sucedidas	Custo Bem-sucedidas	Maior Cluster
9973	9481	0.950667	31.394039	0.500527	837
10000	9481	0.948100	29.262346	0.503480	864
11987	9481	0.790940	6.845488	0.522939	138
12000	9481	0.790083	6.850433	0.519774	156
13999	9481	0.677263	4.147419	0.542557	71
14000	9481	0.677214	4.134758	0.541818	80
15991	9481	0.592896	3.218262	0.562598	70
16000	9481	0.592562	3.221543	0.563389	58
17989	9481	0.527044	2.635075	0.586914	44
18000	9481	0.526722	2.630688	0.587177	36
19997	9481	0.474121	2.274712	0.607482	29
20000	9481	0.474050	2.329484	0.602683	25
219997	9481	0.431013	2.111111	0.624521	26
22000	9481	0.430955	2.092012	0.624680	27
23993	9481	0.395157	1.938062	0.644879	22

24000	9481	0.395042	1.919231	0.648308	19
25999	9481	0.364668	1.828191	0.658543	23
26000	9481	0.364654	1.8232269	0.660442	23
27997	9481	0.351187	1.783148	0.668147	22
28000	9481	0.338607	1.736765	0.675530	17
29989	9481	0.316149	1.666549	0.688189	13
30000	9481	0.316033	1.680432	0.685867	14

e. Com base nos dados da tabela, responda as questões a seguir.

i. Existe alguma relação entre o valor de α e o custo das buscas? Se sim, qual é?

R: Sim, quanto mais próximo de 1 é o fator de busca então maior é o custo de uma busca mal sucedida, e vale o contrário para a busca bem sucedida, quanto maior o fator de carga, o custo diminui.

ii. Existe alguma diferença entre o custo das buscas quando comparamos entre M primo e M composto? Caso exista, essa diferença se mantém para todos os valores de α ?

R: Sim , se observarmos os resultados veremos que para qualquer valor do fator de carga, o custo das buscas mal sucedidas com M primo é maior que com o M composto, com exceção de alguns casos em que essa diferença é menos significativa (ex: quando o M primo e o M composto são próximos), já para as buscas bem sucedidas, o custo não sofreu alteração.

iii. Suponha que depois da inserção de muitas chaves no início do seu programa você vai fazer muito mais consultas na sua tabela do que inserções de novas chaves. Você se preocuparia mais em escolher M para diminuir o custo de buscas bem-sucedidas ou mal-sucedidas?

R: Escolheria um M que diminuiria o custo das buscas mal sucedidas mesmo que o custo para as buscas bem sucedidas tivessem um aumento pequeno, já que o custo para as buscas bem sucedidas é significativamente menor que o custo para as buscas bem sucedidas .

iv. α é comumente utilizado como limiar para o redimensionamento da tabela conforme mais elementos são inseridos. Qual um limiar que você acha adequado

para os casos que você não conhece N de antemão? Por quê? (Estudaremos como implementar o redimensionamento na próxima semana.

R: Utilizaria o tamanho do maior cluster, pois quanto mais próximo de M estiver o valor do maior cluster então tem menos posições livres na tabela, poderia ser verificado também se o tamanho do cluster ultrapassou $m/2$ ou seja a metade do tamanho da tabela, para fazer o redimensionamento e ter posições livres para inserção.