

Análise de Algoritmos
Prof. Marcelo Keese Albertini
Faculdade de Computação - Universidade Federal de Uberlândia
Lista de exercícios 7: Hashing

1. Em qual situação não é possível trocar uma árvore de busca binária por uma tabela hash?
2. Em termos de custo de busca, vale a pena inserir árvores de busca binária em listas de colisões de uma tabela hash por encadeamento?
3. Há um resultado matemático surpreendente chamado “paradoxo do aniversário” que afirma que, se há mais de 23 pessoas em uma sala, há mais de 50% de chance de que duas pessoas façam aniversário no mesmo dia. Explique porque este paradoxo é um exemplo do maior problema do hash.
4. Descreva uma forma de criar uma tabela hashing em que o custo de comparações de igualdade em uma busca por uma chave é no máximo 1.
5. Porque a função hash $h(x) = x \bmod m$ não é boa?
6. Defina hashing uniforme e explique porque essa premissa não é realista.
7. Explique o que é a premissa do hashing fortemente uniforme.
8. Mostre que hashing multiplicativo primal é universal.
9. Se o fator de carga é 0.1, qual é a chance de uma lista de colisões na tabela hash conter no máximo 2 chaves?
10. Se o fator de carga é 0.1, qual é a chance de uma lista de colisões na tabela hash conter 3 chaves?
11. Escreva um código que converte dados de uma pessoa (um nome completo, CPF, data nascimento) em um número inteiro que será usado por uma função hashing.
12. Mostre que, com alta probabilidade, a maior lista de colisões de uma tabela hash com encadeamento contém $O(\lg N / \lg \lg N)$ chaves quando o número de posições na tabela for igual ao número de chaves inseridas.
13. Escreva os algoritmos de inserção e remoção de uma tabela hash com endereçamento e varredura linear usando marcadores de deleção.
14. Escreva os algoritmos de inserção e remoção de uma tabela hash com endereçamento e varredura linear SEM usar marcadores de deleção.
15. Explique o funcionamento do hashing Cuckoo.
16. Quais são os custos exatos para uma busca de uma chave que não está em uma tabela hashing implementada com: 1) encadeamento; 2) endereçamento aberto; 3) hashing Cuckoo.
17. Qual é o menor tamanho de uma tabela para garantir que o custo no pior caso médio seja, com alta probabilidade, no máximo 3 comparações de igualdade considerando uma família de hashing 2-quasi-uniforme e tratamento de colisões por encadeamento?
18. O que é clustering primário em hashing por endereçamento aberto?
19. Um programador quer comparar o custo de memória de implementar um banco de dados baseado em hashing (com encadeamento) com uma árvore-B. Supondo que a aridade média de um nó dessa árvore é 1000. Um nó pode ser com 1 seek. e que a leitura de uma lista de colisões exige 1 operação de seek a cada 1000 chaves. O programador esperar tem 10 bilhões de chaves no banco de dados. Discuta as vantagens e desvantagens para ele usar cada uma dessas estruturas de dados.
20. [CLRS] Considere uma tabela de hash com endereçamento aberto que utiliza uma função de hash que distribui os elementos da tabela de maneira uniforme. Calcule o limite superior para o número de sondagens necessárias para se determinar que um elemento não está presente na tabela e o limite superior para o número de sondagens necessárias quando o elemento está presente. Considere os casos de um fator de carga igual a $3/4$ e de um fator de carga $7/8$, ou seja, quando $n = 4m$ e $n = 8m$.

21. Discuta as diferenças entre as seguintes tabelas:

- Tabela de endereçamento direto
- Tabela de hash com tratamento de colisão por encadeamento
- Tabela de hash com tratamento de colisão por endereçamento aberto com sondagem linear
- Tabela de hash com tratamento de colisão por endereçamento aberto com sondagem linear quadrática
- Tabela de hash com tratamento de colisão por endereçamento aberto com espalhamento duplo

22. O que é e como funciona o filtro de Bloom?