

Relatório sobre Funções Hash

Vinicius Bittencourt Chinoli

November 23, 2023

<https://github.com/ViniciusBittencourtChinoli/hash>

1 Introdução

Uma função hash é uma ferramenta fundamental em ciência da computação, sendo utilizada em diversas aplicações, como bancos de dados, tabelas hash e criptografia. Este relatório abordará três funções hash comuns: resto da divisão, multiplicação e dobramento com giro.

2 Resto da Divisão

A função hash de resto da divisão é uma abordagem simples, dividindo a entrada de dados por um número inteiro primo e retornando o resto da divisão. Embora fácil de implementar, essa função pode gerar colisões com frequência, especialmente para entradas pequenas. Exemplificando com a entrada "abc" e o número primo 13:

$$\text{hash}("abc", 13) = (\text{ord}("a") + \text{ord}("b") + \text{ord}("c")) \% 13$$

O algoritmo de Resto da Divisão demonstra um comportamento consistente em relação ao tempo médio de inserção, número médio de colisões, tempo médio de busca e número médio de comparações. Como esperado, o tempo de inserção aumenta gradualmente com o aumento do conjunto de dados. O número médio de colisões também cresce proporcionalmente. O tempo e as comparações de busca mantêm-se constantes, já que a técnica de Resolução de Colisões utilizada não afeta essas operações. Este algoritmo é eficaz para tamanhos pequenos a moderados da tabela hash.

3 Multiplicação

A função hash de multiplicação é mais eficiente, multiplicando a entrada por um número inteiro primo e dividindo o resultado por um número maior. Essa abordagem reduz a probabilidade de colisões, mas ainda pode ocorrer, principalmente para entradas pequenas. Usando a entrada "abc" com os números primos 13 e 17 como exemplo:

$$\text{hash}("abc", 13, 17) = (\text{ord}("a") \times 13 + \text{ord}("b") \times 13^2 + \text{ord}("c") \times 13^3) \% 17$$

O algoritmo de Multiplicação exibe um desempenho semelhante ao de Resto da Divisão. O tempo de inserção, o número médio de colisões, o tempo de busca e as comparações seguem uma tendência

linear com o aumento do conjunto de dados. Ambos os algoritmos, Resto da Divisão e Multiplicação, são simples e eficientes, mas podem enfrentar desafios com tamanhos muito grandes de conjuntos de dados.

4 Dobramento com Giro

A função hash de dobramento com giro é eficiente e menos propensa a colisões. Ela divide a entrada em blocos, girando cada bloco para a esquerda e combinando os resultados. Embora mais complexa, é vantajosa para reduzir colisões. Exemplificando com "abc" e blocos de tamanho 2:

$$\text{hash}("abc", 2) = (\text{ord}("a") \ll 1 | \text{ord}("b") \gg 1) \oplus \text{ord}("c")$$

O algoritmo de Dobramento com Giro mostra uma eficácia notável, especialmente para tamanhos pequenos de tabela hash. O tempo médio de inserção e o número médio de colisões são comparativamente menores em relação aos outros dois algoritmos para tamanhos pequenos. No entanto, à medida que o conjunto de dados aumenta, o desempenho pode não ser tão eficiente quanto o esperado. O tempo de busca e o número médio de comparações também crescem, tornando-o menos eficaz para tamanhos grandes de conjunto de dados.

5 Conclusão

Cada função hash apresenta vantagens e desvantagens. O resto da divisão é simples, mas menos eficiente. A multiplicação é mais eficiente, mas pode gerar colisões. O dobramento com giro é eficiente e reduz as colisões, mas é mais complexo. A escolha depende de fatores como o tamanho da entrada, a probabilidade de colisões e a eficiência desejada. Este relatório fornece uma base para a tomada de decisões sobre qual função hash utilizar em diferentes cenários. Ao comparar os três algoritmos, é importante considerar o equilíbrio entre tempo de inserção, número de colisões, tempo de busca e comparações. O algoritmo de Resto da Divisão e Multiplicação é mais estável para uma variedade de tamanhos de conjunto de dados, enquanto o Dobramento com Giro destaca-se para tamanhos pequenos de tabela hash. A escolha do algoritmo deve depender das características específicas do conjunto de dados e dos requisitos de desempenho.

6 Tabelas e graficos

Resto da divisao

Tempo Médio de Inserção	Número Médio de Colisões	Tempo Médio de Busca	Número Médio de Comparações
697860	19980	487440	20
823980	99980	732460	20
4596480	499980	3788040	20
5633000	999980	3059620	20
28693120	4999980	15125360	20

Table 1: Resultados para Tamanho da Tabela Hash: 20

Table 2: Resultados para Tamanho da Tabela Hash: 200

Tempo Médio de Inserção	Número Médio de Colisões	Tempo Médio de Busca	Número Médio de Comparações
104680	19800	55560	200
528160	99800	279180	200
3377000	499800	1529940	200
5578320	999800	4703260	201
30485640	4999800	16059920	201

Table 3: Resultados para Tamanho da Tabela Hash: 2000

Tempo Médio de Inserção	Número Médio de Colisões	Tempo Médio de Busca	Número Médio de Comparações
137360	18000	80280	2000
581720	98000	314880	2001
2871760	498000	1580380	2000
5723480	998000	3111360	2006
27899780	4998000	15435200	2012

Table 4: Resultados para Tamanho da Tabela Hash: 20000

Tempo Médio de Inserção	Número Médio de Colisões	Tempo Médio de Busca	Número Médio de Comparações
206640	7395	172680	12605
794860	80140	543220	19862
3069160	480000	1735100	20013
5975580	980000	3334960	20020
30679120	4980000	15673040	20094

Table 5: Resultados para Tamanho da Tabela Hash: 200000

Tempo Médio de Inserção	Número Médio de Colisões	Tempo Médio de Busca	Número Médio de Comparações
204540	964	94580	19036
1125380	21375	813180	78628
6145680	316365	4400860	183696
9883920	801351	6712000	198796
38367480	4800000	22984420	200938

Table 6: Resultados para Algoritmo de Multiplicação com Tamanho da Tabela Hash: 20

Tempo Médio de Inserção	Número Médio de Colisões	Tempo Médio de Busca	Número Médio de Comparações
3399100	19980	3174520	20
14698800	99980	15677880	20
78501640	499980	74655960	20
152748820	999980	146273700	20
786944960	4999980	763896500	20

Table 7: **Multiplicacao**

Resultados para Algoritmo de Multiplicação com Tamanho da Tabela Hash: 200

Tempo Médio de Inserção	Número Médio de Colisões	Tempo Médio de Busca	Número Médio de Comparações
2979300	19800	2897300	200
16226080	99800	14681680	200
75997240	499800	78840640	200
151850500	999800	154078440	200
757260400	4999800	745892840	201

Table 8: Resultados para Algoritmo de Multiplicação com Tamanho da Tabela Hash: 2000

Tempo Médio de Inserção	Número Médio de Colisões	Tempo Médio de Busca	Número Médio de Comparações
2910920	18000	2831960	2000
18118780	98000	15962920	2001
81713120	498000	73142080	2001
150689620	998000	149449040	2004
759634280	4998000	738599560	2010

Table 9: Resultados para Algoritmo de Multiplicação com Tamanho da Tabela Hash: 20000

Tempo Médio de Inserção	Número Médio de Colisões	Tempo Médio de Busca	Número Médio de Comparações
2947720	7579	2909060	12421
15577840	80196	15934640	19807
75064780	480000	74395960	20011
155395700	980000	149752140	20017
757679460	4980000	743661060	20101

Table 10: Resultados para Algoritmo de Multiplicação com Tamanho da Tabela Hash: 200000

Tempo Médio de Inserção	Número Médio de Colisões	Tempo Médio de Busca	Número Médio de Comparações
2978980	3781	2874480	16219
15009480	58658	15813740	41345
80149540	452516	72697060	47505
153729280	952389	154264540	47650
761662160	4952333	747525280	47924

Table 11: **Dobramento com Giro**

Resultados para Algoritmo de Dobramento com Giro com Tamanho da Tabela Hash: 20

Tempo Médio de Inserção	Número Médio de Colisões	Tempo Médio de Busca	Número Médio de Comparações
1367460	19980	1034740	20
2092940	99980	2210800	20
12152920	499980	9859940	20
18693820	999980	21835080	20
117359840	4999980	78593020	20

Table 12: Resultados para Algoritmo de Dobramento com Giro com Tamanho da Tabela Hash: 200

Tempo Médio de Inserção	Número Médio de Colisões	Tempo Médio de Busca	Número Médio de Comparações
529400	19800	479960	200
2649720	99800	2406740	200
10844440	499800	9011400	200
25190060	999800	22910100	201
108800900	4999800	121710920	201

Table 13: Resultados para Algoritmo de Dobramento com Giro com Tamanho da Tabela Hash: 2000

Tempo Médio de Inserção	Número Médio de Colisões	Tempo Médio de Busca	Número Médio de Comparações
418880	18000	338080	2000
1944740	98000	1620640	2001
9819500	498000	8174260	2001
20034940	998000	25161660	2005
115300020	4998000	113257060	2014

Table 14: Resultados para Algoritmo de Dobramento com Giro com Tamanho da Tabela Hash: 20000

Tempo Médio de Inserção	Número Médio de Colisões	Tempo Médio de Busca	Número Médio de Comparações
787280	7345	691780	12655
3699060	80156	3110840	19846
17322260	480000	14800560	20013
38859860	980000	24916480	20021
184269780	4980000	119129900	20101

Table 15: Resultados para Algoritmo de Dobramento com Giro com Tamanho da Tabela Hash: 200000

Tempo Médio de Inserção	Número Médio de Colisões	Tempo Médio de Busca	Número Médio de Comparações
1039560	958	792180	19042
4885500	21321	4334940	78682
24691040	316322	21708740	183736
46872680	801315	34647600	198834
197014000	4800000	182812120	200961





