

# **Trabalho 2:**

# **Tabelas Hash**

Antônio Pedro Avanzi Nunes	RA 407852
Guilherme Koiti Hoçoya	RA 407887
Thales Eduardo Adair Menato	RA 407976

## Introdução

Tabela hash é uma estrutura utilizada na pesquisa de dados. Apesar de seu método de busca ter custo baixo, o tratamento de colisões surge como um novo problema. Dentre as várias maneiras de se tratar colisões, a sondagem linear é a mais simples. O desempenho da tabela então depende da quantidade média de sondagens realizadas.

## Objetivos

Será feita uma análise do desempenho de uma tabela hash a partir do número de sondagens em diferentes contextos. Para tal objetivo, será implementada uma tabela hash com as operações de inserção e busca. Após os testes, dados serão coletados, analisados e discutidos.

## Descrição dos Experimentos

Foram realizados testes para os seguintes tamanhos de tabela:

- 13 chaves
- 31 chaves
- 283 chaves
- 599 chaves
- 1069 chaves
- 10007 chaves

Nota-se que todos os valores são primos.

Para cada tamanho de tabela, serão realizados 5 testes.

Cada teste foi realizado em três situações:

- Chaves ocupando 60% do tamanho máximo da tabela
- Chaves ocupando 75% do tamanho máximo da tabela
- Chaves ocupando 90% do tamanho máximo da tabela

Os testes então seguiram na seguinte ordem:

1. Cria-se uma tabela vazia, com seu tamanho máximo sendo um número primo
2. São realizadas inserções até que o número de chaves para o teste seja atingido
3. É feito o levantamento do número médio de sondagens em buscas bem sucedidas
4. É feito o levantamento do número médio de sondagens em buscas mal sucedidas

Para buscas bem sucedidas, foram buscados todos os valores existentes na tabela.

Para as buscas mal sucedidas, foram buscados valores diferentes dos já existentes, em um número equivalente a 5 vezes o número de valores presentes na tabela.

A função de sondagem utilizada foi o incremento de 1 (a partir da posição  $k$ , são realizadas sondagens nas posições  $k + 1$ ,  $k + 2$ ,  $k + 3$  e assim em diante).

A função hash utilizada foi o módulo simples ( $h(\text{chave}) = \text{chave} \% \text{tamanho da tabela}$ ).

A tabela só aceita chaves de valor inteiro e positivo.

A população de uma tabela recém criada é feita com números aleatórios entre 1 e o dobro do tamanho máximo da tabela.

Após os experimentos, foram criados gráficos correlacionando os dados obtidos nas diferentes situações.

## Resultados Obtidos

A seguir encontram-se as tabelas referentes aos testes do número de sondagens. Cada tabela apresenta cinco testes para cada tamanho de tabela-hash. As consultas foram testadas em seis tamanhos de tabelas-hash.

Tabela I: Testes das sondagens bem sucedidas com 60% de ocupação para tabelas de tamanho 13, 31, 283, 599, 1069 e 10007.

Tamanho da tabela	Teste 1	Teste 2	Teste 3	Teste 4	Teste 5	Média
13	0,25	0,375	0,125	0,875	0,5	0,425
31	0,58	0,26	0,47	0,1	0,26	0,334
283	0,36	0,38	0,43	0,45	0,39	0,402
599	0,42	0,43	0,35	0,37	0,36	0,386
1069	0,36	0,41	0,29	0,33	0,41	0,36
10007	0,39	0,37	0,4	0,38	0,4	0,388

Tabela II: Testes das sondagens mal sucedidas com 60% de ocupação para tabelas de tamanho 13, 31, 283, 599, 1069 e 10007.

Tamanho da tabela	Teste 1	Teste 2	Teste 3	Teste 4	Teste 5	Média
13	0,65	1,15	0,375	0,825	1,575	0,915
31	1,13	0,76	1,35	1,28	0,94	1,092
283	1,23	1,8	1,15	1,93	1,62	1,546
599	1,54	1,33	1,41	1,21	1,11	1,32
1069	1,08	1,29	1,17	1,32	1,39	1,25
10007	1,3	1,33	1,38	1,29	1,29	1,318

Tabela III: Testes das sondagens bem sucedidas com 75% de ocupação para tabelas de tamanho 13, 31, 283, 599, 1069 e 10007.

<b>Tamanho da tabela</b>	<b>Teste 1</b>	<b>Teste 2</b>	<b>Teste 3</b>	<b>Teste 4</b>	<b>Teste 5</b>	<b>Média</b>
13	0,1	0,2	1,1	0,6	0,2	0,44
31	0,54	0,83	0,5	1,08	0,25	0,64
283	0,6	0,66	0,5	1,05	0,68	0,698
599	0,74	0,73	0,65	0,7	0,6	0,684
1069	0,74	0,65	0,75	0,65	0,77	0,712
10007	0,73	0,73	0,77	0,78	0,74	0,75

Tabela IV: Testes para sondagens mal sucedidas com 75% de ocupação para tabelas de tamanho 13, 31, 283, 599, 1069 e 10007.

<b>Tamanho da tabela</b>	<b>Teste 1</b>	<b>Teste 2</b>	<b>Teste 3</b>	<b>Teste 4</b>	<b>Teste 5</b>	<b>Média</b>
13	1,42	2,08	1,36	3,04	1,48	1,876
31	5,67	2,08	2,61	1,94	2,71	3,002
283	3,6	2,46	2,51	2,99	2,96	2,904
599	3,15	2,72	3,34	3,05	2,95	3,042
1069	3,34	3,18	4,49	3,3	3,45	3,552
10007	4,01	3,69	3,96	3,76	3,38	3,76

Tabela V: Testes para sondagens bem sucedidas com 90% de ocupação para tabelas de tamanho 13, 31, 283, 599, 1069 e 10007.

<b>Tamanho da tabela</b>	<b>Teste 1</b>	<b>Teste 2</b>	<b>Teste 3</b>	<b>Teste 4</b>	<b>Teste 5</b>	<b>Média</b>
13	0,25	1,5	1,17	0,84	0,58	0,868
31	0,68	0,32	1,07	0,46	1,18	0,742
283	2,62	2,04	1,31	1,56	2,85	2,076
599	1,62	1,7	2,17	2,48	1,72	1,938
1069	2,15	1,7	1,96	1,6	1,76	1,834
10007	2,22	2,33	2,33	2,12	2,22	2,244

Tabela VI: Testes para sondagens mal sucedidas com 90% de ocupação para tabelas de tamanho 13, 31, 283, 599, 1069 e 10007.

<b>Tamanho da tabela</b>	<b>Teste 1</b>	<b>Teste 2</b>	<b>Teste 3</b>	<b>Teste 4</b>	<b>Teste 5</b>	<b>Média</b>
13	5,03	3,63	5,12	5,34	5,6	4,944
31	5,67	4,36	5,81	10,57	8,64	7,01
283	19,9	10,89	10,78	13,18	17,16	14,382
599	15,8	17,35	19	14,9	18,1	17,03
1069	16,46	19,71	16,61	13,69	14,71	16,236
10007	27,46	24,84	29,39	22,5	22,19	25,276

Os primeiros testes realizados com tabelas de tamanho 13 e 31 apresentaram alguns valores de sondagens com grandes diferenças em relação às médias, isso ocorreu devido aos seus tamanhos reduzidos.

#### **Tabelas com 60% de ocupação:**

A partir da tabela-hash de tamanho 283, observamos que os resultados obtidos para o número de sondagens médias de uma tabela-hash com 60% de ocupação se mantém próximos.

Desconsiderando as duas tabelas-hash com tamanhos menores, observamos na tabela I que a maior média de sondagens bem sucedidas obtida foi de 0,402, na tabela-hash de tamanho 283, enquanto a menor média foi de 0,36, na tabela-hash de tamanho 1069. Nota-se que a média encontrada nos testes é muito próxima da média linear apresentada no gráfico 1.

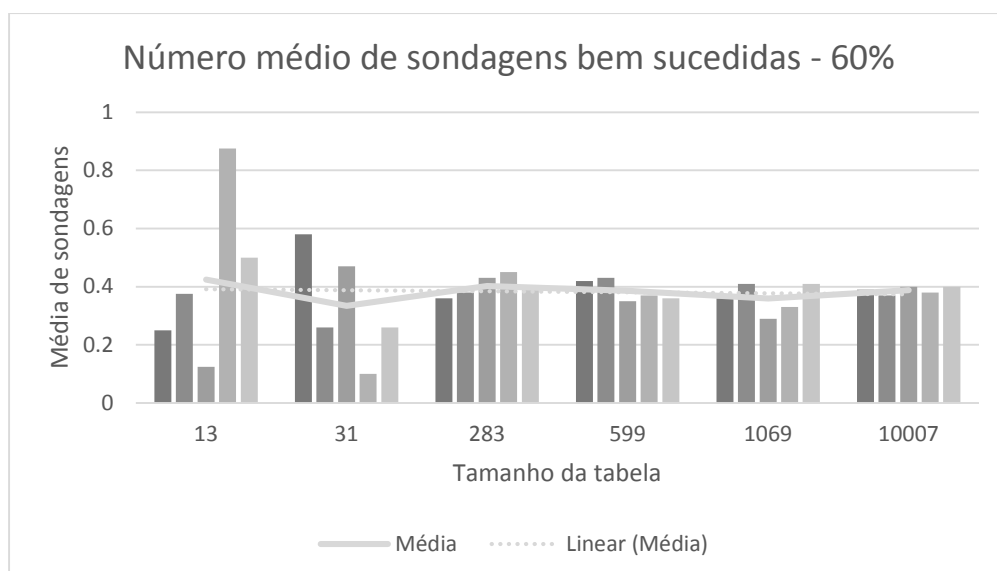


Gráfico 1: Gráfico de Média de sondagens bem sucedidas em relação ao tamanho das tabelas, com 60% de ocupação. Cada barra vertical aponta um teste individual. A média linear corresponde à uma linha de tendência das médias.

No gráfico de médias mal sucedidas encontramos comportamento semelhante, com pouca variação da maior média de 1,546 para a menor média de 1,25 (novamente desconsiderando as duas primeiras tabelas-hash com tamanhos inferiores).

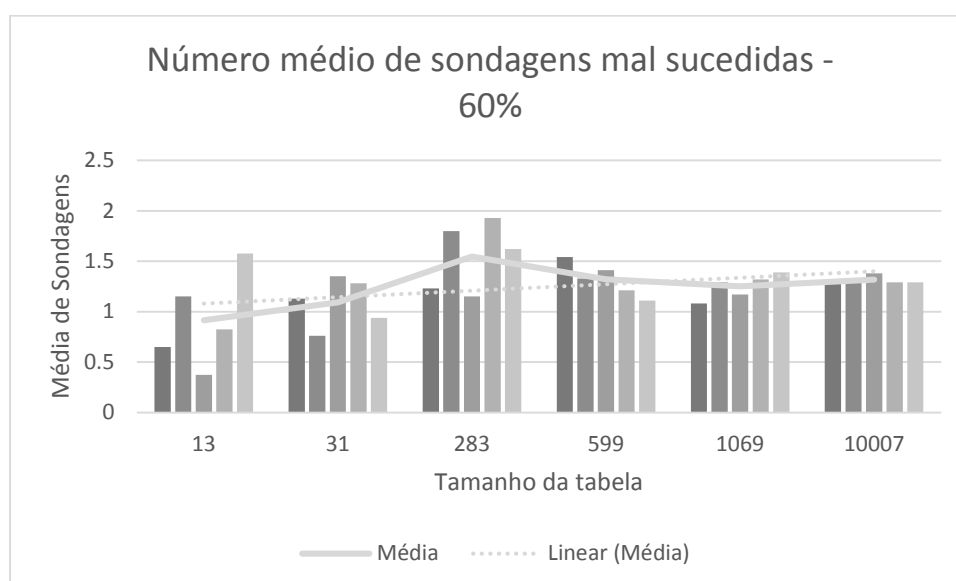


Gráfico 2: Gráfico de Média de sondagens mal sucedidas em relação ao tamanho das tabelas, com 60% de ocupação. Cada barra vertical aponta um teste individual. A média linear corresponde à uma linha de tendência das médias.

### Tabelas com 75% de ocupação:

Para as tabelas-hash com preenchimento de 75%, podemos observar um crescimento gradativo na média de sondagens em relação ao tamanho das tabelas-hash. Isso ocorre no número médio de sondagens bem sucedidas e mal sucedidas, conforme apresentado nos gráficos 3 e 4.

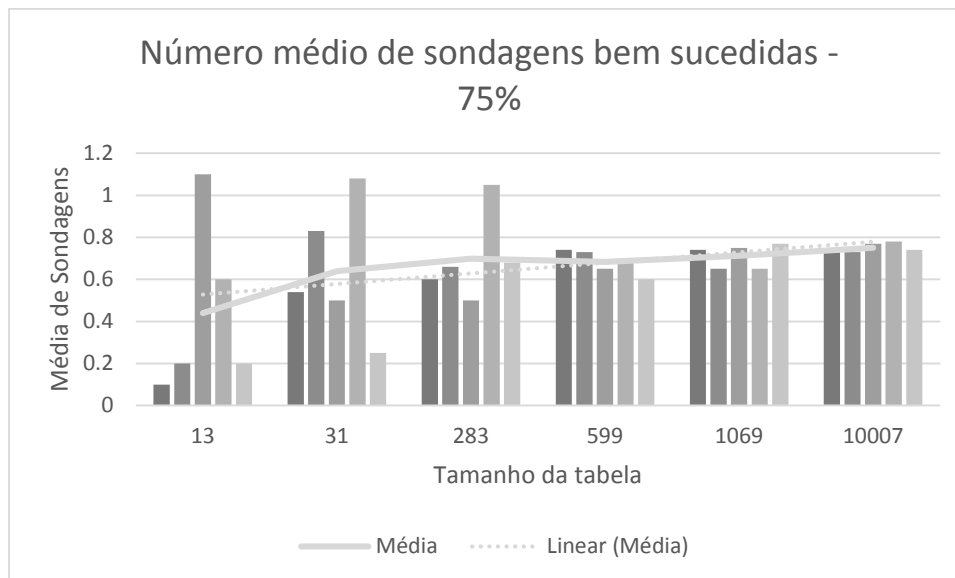


Gráfico 3: Gráfico de Média de sondagens bem sucedidas em relação ao tamanho das tabelas, com 75% de ocupação. Cada barra vertical aponta um teste individual. A média linear corresponde à uma linha de tendência das médias.

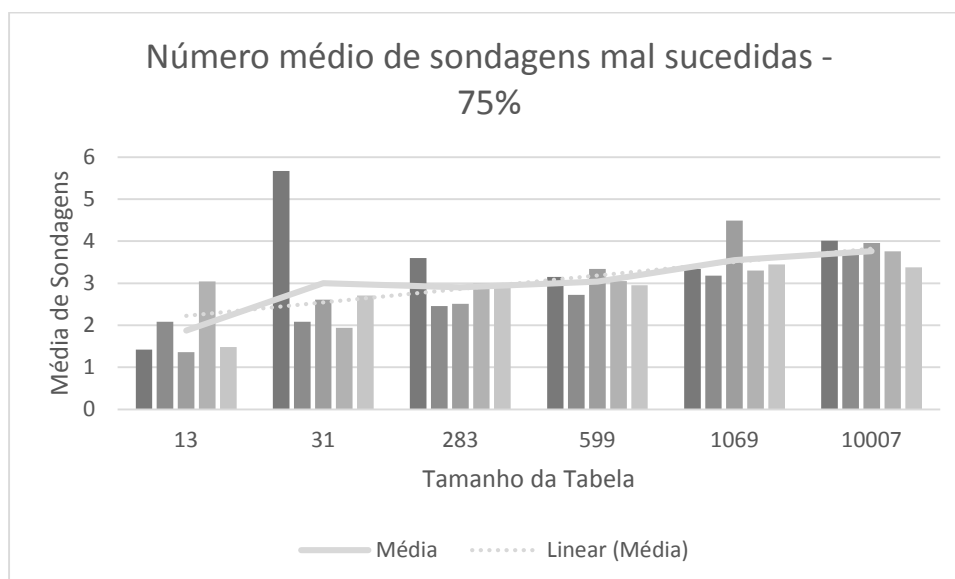


Gráfico 4: Gráfico de Média de sondagens mal sucedidas em relação ao tamanho das tabelas, com 75% de ocupação. Cada barra vertical aponta um teste individual. A média linear corresponde à uma linha de tendência das médias.

### Tabelas com 90% de ocupação:

Se desconsiderarmos as duas tabelas-hash com tamanho inferior, veremos na tabela V uma média de sondagens bem sucedidas com pouca variação, entre 1,834 e 2,244.

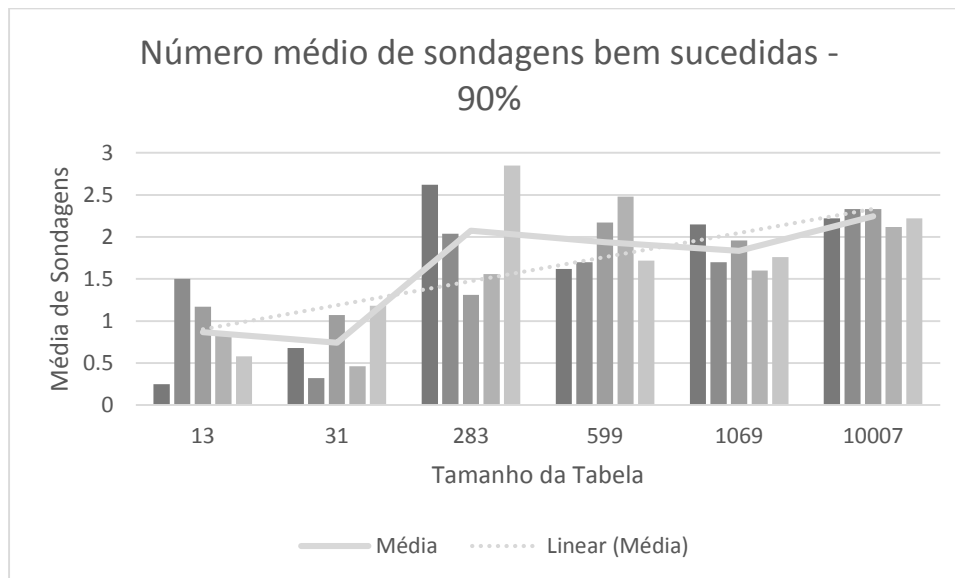


Gráfico 5: Gráfico de Média de sondagens bem sucedidas em relação ao tamanho das tabelas, com 90% de ocupação. Cada barra vertical aponta um teste individual. A média linear corresponde à uma linha de tendência das médias.

No gráfico 6, observamos que o número médio de sondagens mal sucedidas apresentou grande aumento a cada teste em tabelas maiores. Como mostra a tabela VI, a média da tabela-hash de tamanho 10007 é mais do que cinco vezes maior do que a média da tabela de tamanho 13.



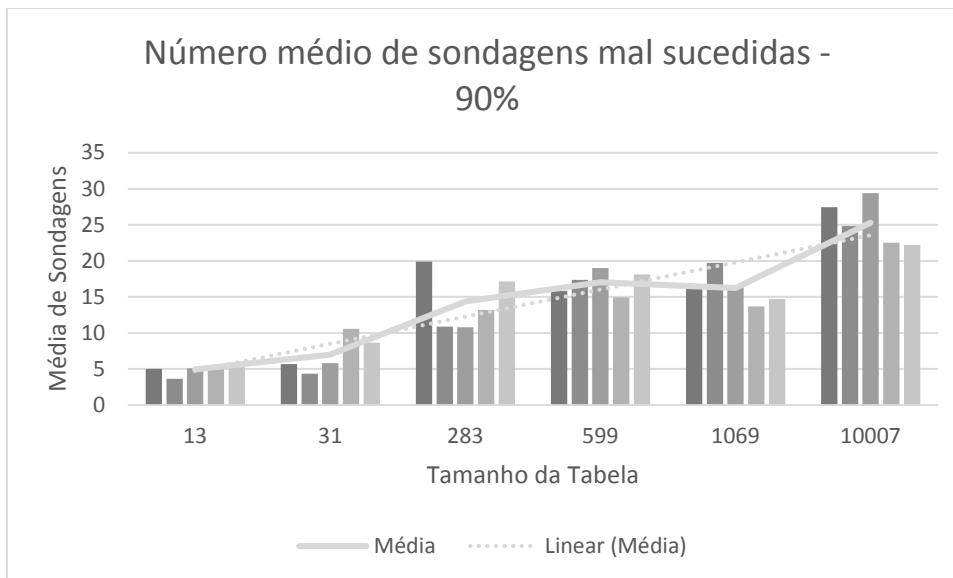


Gráfico 6: Gráfico de Média de sondagens mal sucedidas em relação ao tamanho das tabelas, com 90% de ocupação. Cada barra vertical aponta um teste individual. A média linear corresponde à uma linha de tendência das médias

## Conclusão

Como mostram os testes nas tabelas-hash com ocupação de 60%, os resultados das sondagens bem sucedidas são satisfatórios, com uma média abaixo de 0,4. Nas sondagens mal sucedidas, as médias são um pouco superiores, porém ainda estáveis.

Nas tabelas-hash com ocupação de 75%, existe um pequeno aumento nas médias das sondagens com tabelas maiores, porém ainda apresentam desempenho satisfatório para sondagens bem sucedidas, com média abaixo de 0,7, e para sondagens mal sucedidas com média abaixo de 3,1.

No entanto, fica claro a queda de desempenho nos casos onde a tabela possui ocupação igual ou superior a 90%. Isso ocorre devido aos agrupamentos primários (primary clustering). Por meio do gráfico 6 podemos comprovar o aumento no número médio de sondagens, principalmente nos casos mal sucedidos, onde alcançou uma média 25,276 sondagens para uma tabela de 10007 posições, como mostra a tabela VI. Os agrupamentos longos prejudicam o desempenho da hash-table, aproximando a performance a  $O(n)$ .

## **Bibliografia Consultada**

Jander, Moreira (2011). **Armazenamento e Recuperação da Informação.**

Drozdek, Adam (2001). **Algorithms and Data Structures in C++, Second Edition.**

**Hash Table.** Disponível em [https://en.wikipedia.org/wiki/Hash\\_table](https://en.wikipedia.org/wiki/Hash_table)

**Linear Probing.** Disponível em [https://en.wikipedia.org/wiki/Linear\\_probing](https://en.wikipedia.org/wiki/Linear_probing)

**Primary Clustering.** Disponível em [http://en.wikipedia.org/wiki/Primary\\_clustering](http://en.wikipedia.org/wiki/Primary_clustering)