MC-202 Tabela de Espalhamento

Rafael C. S. Schouery rafael@ic.unicamp.br

Universidade Estadual de Campinas

 2° semestre/2019

Introdução

Queremos contar o número de ocorrências de cada palavra da biblioteca

— Não há norte. O encontro da cuase espanición, ou a expansão da dusa formas, pode determinar a supressão de não há morte ha supressão de não há morte ha vida, porque a supressão de uma e a condição da sobrevivência da costra, a a destruição não atingo o principio universal a communicação de moderna de communicação de compos de adatatas e desa tribo familias. An bastatas apenas compos de batatas e desa tribo familias. An bastata a penas compos de batatas e de casa de contribuição de compos d

Introdução

Queremos contar o número de ocorrências de cada palavra da biblioteca



• no idioma, há cerca de milhares de palavras (≈ 435.000)

Introdução

Queremos contar o número de ocorrências de cada palavra da biblioteca



- no idioma, há cerca de milhares de palavras (≈ 435.000)
- mas no total, há milhões de ocorrências!

dia: 6 ocorrências

escola: 13 ocorrências

gratuito: 1 ocorrência

ilha: 8 ocorrências

jeito: 5 ocorrências

lata: 2 ocorrências

dia: 6 ocorrências

escola: 13 ocorrências

gratuito: 1 ocorrência

ilha: 8 ocorrências

jeito: 5 ocorrências

lata: 2 ocorrências

Queremos acessar uma palavra como se fosse um vetor:

ocorrencias["ilha"] = 8

dia: 6 ocorrências

escola: 13 ocorrências

gratuito: 1 ocorrência

ilha: 8 ocorrências jeito: 5 ocorrências

2 ocorrências lata:

Queremos acessar uma palavra como se fosse um vetor:

ocorrencias["ilha"] = 8

dia: 6 ocorrências escola: 13 ocorrências

gratuito: 1 ocorrência ilha: 8 ocorrências jeito: 5 ocorrências

lata: 2 ocorrências

Queremos acessar uma palavra como se fosse um vetor:

Primeiras opções:

ullet Vetor - acesso/escrita em $\mathrm{O}(n)$

dia: 6 ocorrências escola: 13 ocorrências gratuito: 1 ocorrência ilha: 8 ocorrências

jeito: 5 ocorrências

lata: 2 ocorrências

Queremos acessar uma palavra como se fosse um vetor:

- Vetor acesso/escrita em O(n)
 - inserir uma nova palavra leva $\mathrm{O}(1)$

dia: 6 ocorrências escola: 13 ocorrências gratuito: 1 ocorrência ilha: 8 ocorrências jeito: 5 ocorrências

Queremos acessar uma palavra como se fosse um vetor:

lata:

2 ocorrências

- Vetor acesso/escrita em $\mathrm{O}(n)$
 - inserir uma nova palavra leva $\mathrm{O}(1)$
- ullet Vetor ordenado acesso/escrita em $\mathrm{O}(\lg n)$

dia: 6 ocorrências escola: 13 ocorrências gratuito: 1 ocorrência ilha: 8 ocorrências jeito: 5 ocorrências

2 ocorrências

Queremos acessar uma palavra como se fosse um vetor:

lata:

- Vetor acesso/escrita em O(n)
 - inserir uma nova palavra leva $\mathrm{O}(1)$
- Vetor ordenado acesso/escrita em $O(\lg n)$
 - inserir uma nova palavra leva $\mathrm{O}(n)$

dia: 6 ocorrências escola: 13 ocorrências gratuito: 1 ocorrência ilha: 8 ocorrências jeito: 5 ocorrências 2 ocorrências

Queremos acessar uma palavra como se fosse um vetor:

```
ocorrencias["ilha"] = 8
```

- Vetor acesso/escrita em O(n)
 - inserir uma nova palavra leva $\mathrm{O}(1)$
- Vetor ordenado acesso/escrita em $O(\lg n)$
 - inserir uma nova palavra leva O(n)
- ullet ABB balanceada acesso/escrita/inserção em $\mathrm{O}(\lg n)$

dia: 6 ocorrências escola: 13 ocorrências gratuito: 1 ocorrência ilha: 8 ocorrências jeito: 5 ocorrências

lata: 2 ocorrências

Queremos acessar uma palavra como se fosse um vetor:

```
ocorrencias["ilha"] = 8
```

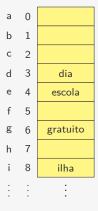
Primeiras opções:

- Vetor acesso/escrita em O(n)
 - inserir uma nova palavra leva $\mathrm{O}(1)$
- Vetor ordenado acesso/escrita em $O(\lg n)$
 - inserir uma nova palavra leva O(n)
- ullet ABB balanceada acesso/escrita/inserção em $\mathrm{O}(\lg n)$

Conseguimos fazer em O(1)?

Caso fácil

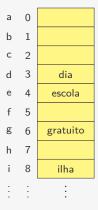
Se tivéssemos apenas uma palavra começando com cada letra era fácil



Caso fácil

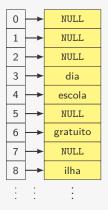
Se tivéssemos apenas uma palavra começando com cada letra era fácil

• bastaria ter um vetor de 26 posições

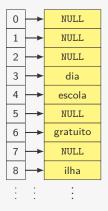


4



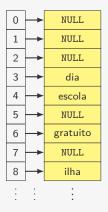


Ideia:



Ideia:

• uma lista ligada para cada letra



Ideia:

- uma lista ligada para cada letra
- guardamos os ponteiros para as listas em um vetor





Inserindo "bala":



Inserindo "bala":

• descobrimos a posição pela primeira letra



Inserindo "bala":

- descobrimos a posição pela primeira letra
- atualizamos o vetor para apontar para o nó de "bala"



Inserindo "bala":

- descobrimos a posição pela primeira letra
- atualizamos o vetor para apontar para o nó de "bala"





Inserindo "bela":



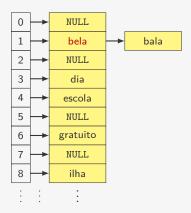
Inserindo "bela":

• descobrimos a posição pela primeira letra



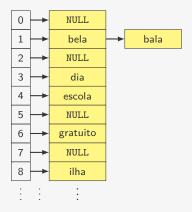
Inserindo "bela":

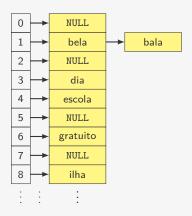
- descobrimos a posição pela primeira letra
- temos uma colisão com "bala"



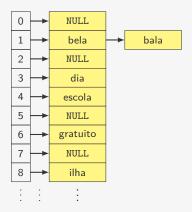
Inserindo "bela":

- descobrimos a posição pela primeira letra
- temos uma colisão com "bala"
- inserimos no começo da lista da letra b



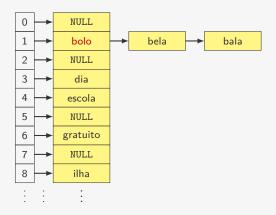


Após a inserção de várias palavras começando com b:



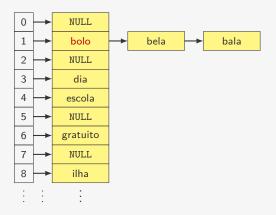
Após a inserção de várias palavras começando com b:

• inserimos "bolo",



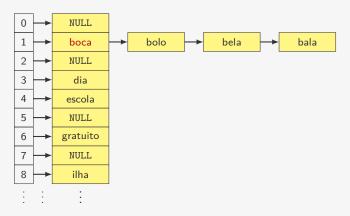
Após a inserção de várias palavras começando com b:

• inserimos "bolo",



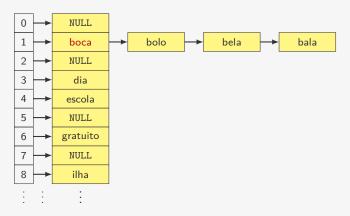
Após a inserção de várias palavras começando com b:

• inserimos "bolo", "boca",



Após a inserção de várias palavras começando com b:

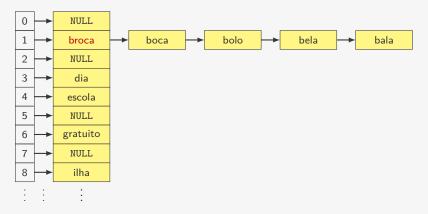
• inserimos "bolo", "boca",



Após a inserção de várias palavras começando com b:

• inserimos "bolo", "boca", "broca"

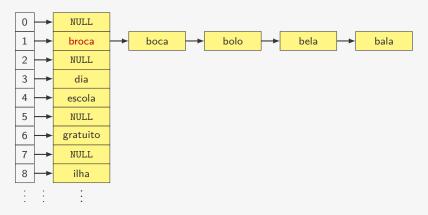
Palavras que começam com a mesma letra



Após a inserção de várias palavras começando com b:

• inserimos "bolo", "boca", "broca"

Palavras que começam com a mesma letra



Após a inserção de várias palavras começando com b:

- inserimos "bolo", "boca", "broca"
- a tabela ficou degenerada em lista

broca

boca

bolo

bela

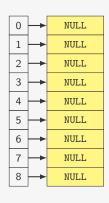
bala

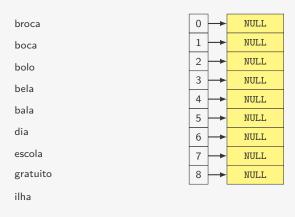
dia

escola

gratuito

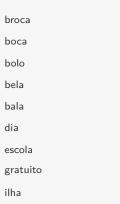
ilha

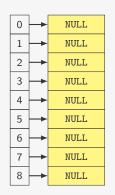




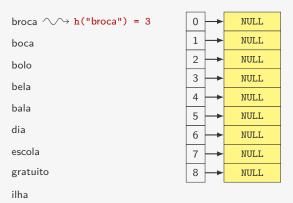
Corrigindo:

• vamos tentar espalhar melhor

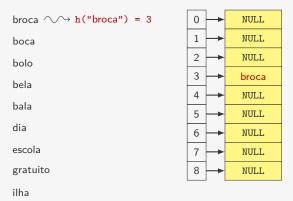




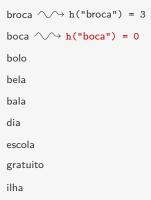
- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)

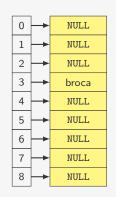


- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)



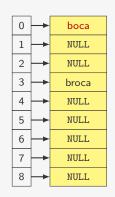
- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)





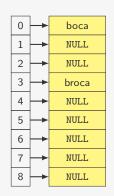
- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)

```
broca \rightsquigarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo
bela
bala
dia
escola
gratuito
ilha
```



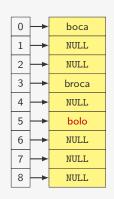
- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)

```
broca \rightsquigarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo \rightsquigarrow h("bolo") = 5
bela
bala
dia
escola
gratuito
ilha
```



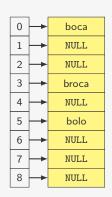
- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)

```
broca \rightsquigarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo \rightsquigarrow h("bolo") = 5
bela
bala
dia
escola
gratuito
ilha
```



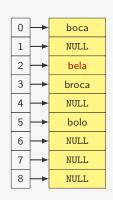
- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)

```
broca \rightsquigarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo \wedge \rightarrow h("bolo") = 5
bela \rightsquigarrow h("bela") = 2
bala
dia
escola
gratuito
ilha
```



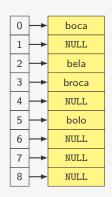
- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)

```
broca \rightsquigarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo \wedge \rightarrow h("bolo") = 5
bela \rightsquigarrow h("bela") = 2
bala
dia
escola
gratuito
ilha
```

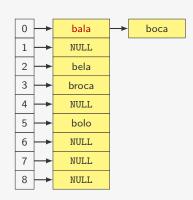


- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)

```
broca \land \land \Rightarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo \wedge \rightarrow h("bolo") = 5
bela \rightsquigarrow h("bela") = 2
bala \rightsquigarrow h("bala") = 0
dia
escola
gratuito
ilha
```

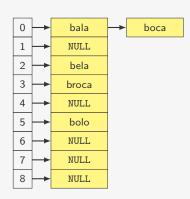


- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)

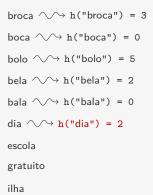


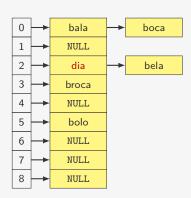
- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)

broca
$$\wedge \rightarrow h("broca") = 3$$
boca $\wedge \rightarrow h("boca") = 0$
bolo $\wedge \rightarrow h("bolo") = 5$
bela $\wedge \rightarrow h("bela") = 2$
bala $\wedge \rightarrow h("bela") = 0$
dia $\wedge \rightarrow h("dia") = 2$
escola
gratuito
ilha

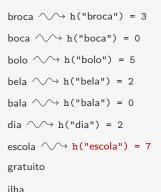


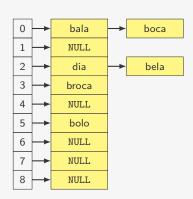
- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)



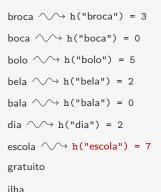


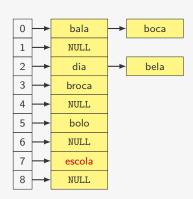
- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)



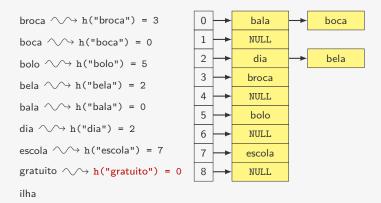


- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)

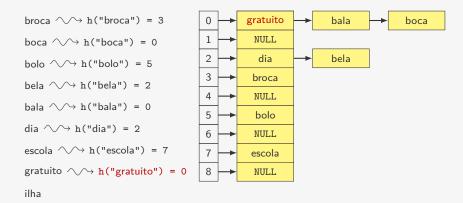




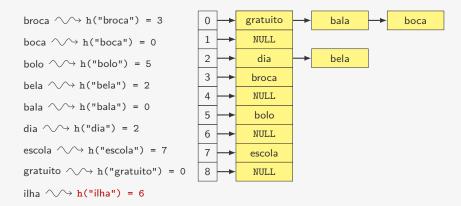
- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)



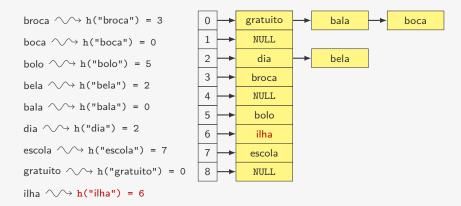
- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)



- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)



- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)



- vamos tentar espalhar melhor
- usamos um hash da chave (palavra)
- vamos associar a chave a um número inteiro (entre 0 e 8)

Uma função de hashing associa

Uma função de hashing associa

• um elemento de um conjunto (strings, números, etc.)

Uma função de hashing associa

- um elemento de um conjunto (strings, números, etc.)
- a um número inteiro de tamanho conhecido

Uma função de hashing associa

- um elemento de um conjunto (strings, números, etc.)
- a um número inteiro de tamanho conhecido

Uma tabela de espalhamento é um TAD para conjuntos dinâmicos com certas propriedades:

Uma função de hashing associa

- um elemento de um conjunto (strings, números, etc.)
- a um número inteiro de tamanho conhecido

Uma tabela de espalhamento é um TAD para conjuntos dinâmicos com certas propriedades:

 os dados são acessado por meio de um vetor de tamanho conhecido

Uma função de hashing associa

- um elemento de um conjunto (strings, números, etc.)
- a um número inteiro de tamanho conhecido

Uma tabela de espalhamento é um TAD para conjuntos dinâmicos com certas propriedades:

- os dados são acessado por meio de um vetor de tamanho conhecido
- a posição do vetor é calculada por uma função de hashing

Restrições

Restrições

estimativa do tamanho do conjunto de dados deve ser conhecida

Restrições

- estimativa do tamanho do conjunto de dados deve ser conhecida
- bits da chave devem estar disponíveis

Restrições

- estimativa do tamanho do conjunto de dados deve ser conhecida
- bits da chave devem estar disponíveis
 - em uma ABB, basta uma função de comparação

Restrições

- estimativa do tamanho do conjunto de dados deve ser conhecida
- bits da chave devem estar disponíveis
 - em uma ABB, basta uma função de comparação

Tempo das operações

Restrições

- estimativa do tamanho do conjunto de dados deve ser conhecida
- bits da chave devem estar disponíveis
 - em uma ABB, basta uma função de comparação

Tempo das operações

• depende principalmente da função de hashing escolhida

Restrições

- estimativa do tamanho do conjunto de dados deve ser conhecida
- bits da chave devem estar disponíveis
 - em uma ABB, basta uma função de comparação

Tempo das operações

- depende principalmente da função de hashing escolhida
- ullet chaves bem espalhadas: tempo "quase" $\mathrm{O}(1)$

Restrições

- estimativa do tamanho do conjunto de dados deve ser conhecida
- bits da chave devem estar disponíveis
 - em uma ABB, basta uma função de comparação

- depende principalmente da função de hashing escolhida
- chaves bem espalhadas: tempo "quase" O(1)
 - se temos n itens

Restrições

- estimativa do tamanho do conjunto de dados deve ser conhecida
- bits da chave devem estar disponíveis
 - em uma ABB, basta uma função de comparação

- depende principalmente da função de hashing escolhida
- chaves bem espalhadas: tempo "quase" O(1)
 - se temos n itens
 - uma tabela de tamanho M

Restrições

- estimativa do tamanho do conjunto de dados deve ser conhecida
- bits da chave devem estar disponíveis
 - em uma ABB, basta uma função de comparação

- depende principalmente da função de hashing escolhida
- chaves bem espalhadas: tempo "quase" O(1)
 - se temos n itens
 - uma tabela de tamanho M
 - tempo de acesso é o tempo de calcular a função de hashing mais O(n/M)

Restrições

- estimativa do tamanho do conjunto de dados deve ser conhecida
- bits da chave devem estar disponíveis
 - em uma ABB, basta uma função de comparação

- depende principalmente da função de hashing escolhida
- chaves bem espalhadas: tempo "quase" O(1)
 - se temos n itens
 - uma tabela de tamanho M
 - tempo de acesso é o tempo de calcular a função de hashing mais O(n/M)
- chaves muito agrupadas: pior caso de tempo O(n)

Restrições

- estimativa do tamanho do conjunto de dados deve ser conhecida
- bits da chave devem estar disponíveis
 - em uma ABB, basta uma função de comparação

- depende principalmente da função de hashing escolhida
- chaves bem espalhadas: tempo "quase" O(1)
 - se temos n itens
 - uma tabela de tamanho M
 - tempo de acesso é o tempo de calcular a função de hashing mais O(n/M)
- chaves muito agrupadas: pior caso de tempo O(n)
 - vira uma lista ligada com todos os elementos

Uma boa função de hashing deve espalhar bem:

Uma boa função de hashing deve espalhar bem:

• A probabilidade de uma chave ter um hash específico é (aproximadamente) 1/M

Uma boa função de hashing deve espalhar bem:

- A probabilidade de uma chave ter um hash específico é (aproximadamente) 1/M
- Ou seja, esperamos que cada lista tenha n/M elementos

Uma boa função de hashing deve espalhar bem:

- A probabilidade de uma chave ter um hash específico é (aproximadamente) 1/M
- Ou seja, esperamos que cada lista tenha n/M elementos

Métodos genéricos (que funcionam bem na prática):

Uma boa função de hashing deve espalhar bem:

- A probabilidade de uma chave ter um hash específico é (aproximadamente) 1/M
- Ou seja, esperamos que cada lista tenha n/M elementos

Métodos genéricos (que funcionam bem na prática):

1. Método da divisão

Uma boa função de hashing deve espalhar bem:

- A probabilidade de uma chave ter um hash específico é (aproximadamente) 1/M
- Ou seja, esperamos que cada lista tenha n/M elementos

Métodos genéricos (que funcionam bem na prática):

- 1. Método da divisão
- 2. Método da multiplicação

Uma boa função de hashing deve espalhar bem:

- A probabilidade de uma chave ter um hash específico é (aproximadamente) 1/M
- Ou seja, esperamos que cada lista tenha n/M elementos

Métodos genéricos (que funcionam bem na prática):

- 1. Método da divisão
- 2. Método da multiplicação

Hashing perfeito: Se conhecermos todos as chaves a priori, é possível encontrar uma função de hashing injetora

Uma boa função de hashing deve espalhar bem:

- A probabilidade de uma chave ter um hash específico é (aproximadamente) 1/M
- Ou seja, esperamos que cada lista tenha n/M elementos

Métodos genéricos (que funcionam bem na prática):

- 1. Método da divisão
- 2. Método da multiplicação

Hashing perfeito: Se conhecermos todos as chaves a priori, é possível encontrar uma função de hashing injetora

• isto é, não temos colisões

Uma boa função de hashing deve espalhar bem:

- A probabilidade de uma chave ter um hash específico é (aproximadamente) 1/M
- Ou seja, esperamos que cada lista tenha n/M elementos

Métodos genéricos (que funcionam bem na prática):

- 1. Método da divisão
- 2. Método da multiplicação

Hashing perfeito: Se conhecermos todos as chaves a priori, é possível encontrar uma função de hashing injetora

- isto é, não temos colisões
- tais funções podem ser difíceis de encontrar

Pressupomos que as chaves são números inteiros

Pressupomos que as chaves são números inteiros

E se não forem?

Pressupomos que as chaves são números inteiros

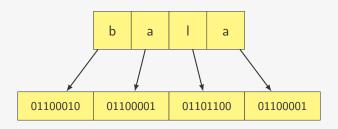
E se não forem?

• Reinterpretamos a chave como uma sequência de bits

Pressupomos que as chaves são números inteiros

E se não forem?

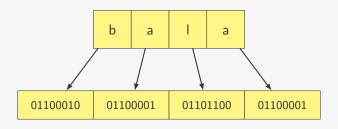
• Reinterpretamos a chave como uma sequência de bits



Pressupomos que as chaves são números inteiros

E se não forem?

Reinterpretamos a chave como uma sequência de bits

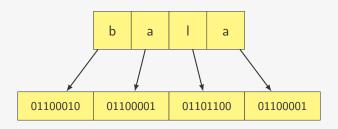


Assim, "bala" se torna o número 1.650.551.905

Pressupomos que as chaves são números inteiros

E se não forem?

Reinterpretamos a chave como uma sequência de bits



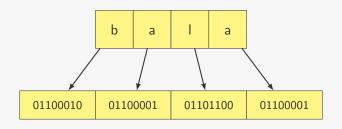
Assim, "bala" se torna o número 1.650.551.905

• Esse número pode explodir rapidamente

Pressupomos que as chaves são números inteiros

E se não forem?

Reinterpretamos a chave como uma sequência de bits



Assim, "bala" se torna o número 1.650.551.905

- Esse número pode explodir rapidamente
- Veremos como contornar isso para strings...

ullet obtemos o resto da divisão pelo tamanho M do hashing

ullet obtemos o resto da divisão pelo tamanho M do hashing

$$h(x) = x \bmod M$$

ullet obtemos o resto da divisão pelo tamanho M do hashing

$$h(x) = x \bmod M$$

Exemplo:

$$h(\text{"bala"}) = 1.650.551.905 \text{ mod } 1783 = 277$$

ullet obtemos o resto da divisão pelo tamanho M do hashing

$$h(x) = x \bmod M$$

Exemplo:

$$h(\text{``bala''}) = 1.650.551.905 \ \mathrm{mod} \ 1783 = 277$$

Escolhendo *M*:

ullet obtemos o resto da divisão pelo tamanho M do hashing

$$h(x) = x \bmod M$$

Exemplo:

$$h(\text{``bala''}) = 1.650.551.905 \ \mathrm{mod} \ 1783 = 277$$

Escolhendo M:

ullet escolher M como uma potência de 2 não é uma boa ideia

ullet obtemos o resto da divisão pelo tamanho M do hashing

$$h(x) = x \bmod M$$

Exemplo:

$$h(\text{"bala"}) = 1.650.551.905 \text{ mod } 1783 = 277$$

Escolhendo M:

- ullet escolher M como uma potência de 2 não é uma boa ideia
 - considera apenas os bits menos significativos

ullet obtemos o resto da divisão pelo tamanho M do hashing

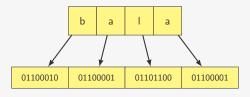
$$h(x) = x \bmod M$$

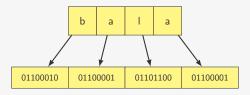
Exemplo:

$$h(\text{"bala"}) = 1.650.551.905 \text{ mod } 1783 = 277$$

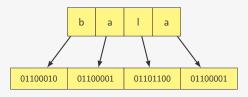
Escolhendo *M*:

- ullet escolher M como uma potência de 2 não é uma boa ideia
 - considera apenas os bits menos significativos
- ullet normalmente escolhemos M como um número primo longe de uma potência de 2

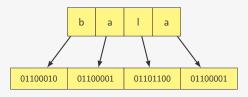




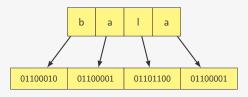
•
$$x = 'b' \cdot 256^3 + 'a' \cdot 256^2 + '1' \cdot 256^1 + 'a' \cdot 256^0$$



•
$$x=$$
 'b' $\cdot 256^3+$ 'a' $\cdot 256^2+$ 'l' $\cdot 256^1+$ 'a' $\cdot 256^0$ que pode ser rescrito como



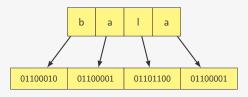
- x= 'b' $\cdot 256^3+$ 'a' $\cdot 256^2+$ 'l' $\cdot 256^1+$ 'a' $\cdot 256^0$ que pode ser rescrito como
 - $x = ((('b') \cdot 256 + 'a') \cdot 256 + '1') \cdot 256 + 'a'$



Como podemos calcular o número x que representa "bala"?

- x= 'b' $\cdot 256^3+$ 'a' $\cdot 256^2+$ 'l' $\cdot 256^1+$ 'a' $\cdot 256^0$ que pode ser rescrito como
 - $\bullet \ x = (((\texttt{'b'}) \cdot 256 + \texttt{'a'}) \cdot 256 + \texttt{'1'}) \cdot 256 + \texttt{'a'}$

Mas x poderia ser muito grande e estourar um int...

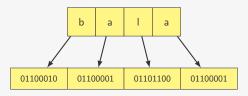


Como podemos calcular o número x que representa "bala"?

- x= 'b' $\cdot 256^3+$ 'a' $\cdot 256^2+$ 'l' $\cdot 256^1+$ 'a' $\cdot 256^0$ que pode ser rescrito como
 - $x = ((('b') \cdot 256 + 'a') \cdot 256 + '1') \cdot 256 + 'a'$

Mas x poderia ser muito grande e estourar um int...

Ao invés de calcular $x \mod M$, calculamos



Como podemos calcular o número x que representa "bala"?

- x= 'b' $\cdot 256^3+$ 'a' $\cdot 256^2+$ '1' $\cdot 256^1+$ 'a' $\cdot 256^0$ que pode ser rescrito como
 - $x = ((('b') \cdot 256 + 'a') \cdot 256 + '1') \cdot 256 + 'a'$

Mas x poderia ser muito grande e estourar um int...

Ao invés de calcular $x \mod M$, calculamos

$$(((\verb""b" \bmod M) \cdot 256 + \verb"a" \bmod M) \cdot 256 + \verb"1" \bmod M) \cdot 256 + \verb"a" \bmod M)$$

Método da multiplicação

ullet multiplicamos por um certo valor real A e obtemos a parte fracionária

- ullet multiplicamos por um certo valor real A e obtemos a parte fracionária
- escolhemos A conveniente, por exemplo $A = (\sqrt{5} 1)/2$

- ullet multiplicamos por um certo valor real A e obtemos a parte fracionária
- escolhemos A conveniente, por exemplo $A = (\sqrt{5} 1)/2$
- ${\color{blue} \bullet}$ posição relativa no vetor não depende de M (pode ser M=1024)

- ullet multiplicamos por um certo valor real A e obtemos a parte fracionária
- escolhemos A conveniente, por exemplo $A = (\sqrt{5} 1)/2$
- ullet posição relativa no vetor **não** depende de M (pode ser M=1024)

$$h(x) = \lfloor M \left(A \cdot x \bmod 1 \right) \rfloor$$

- ullet multiplicamos por um certo valor real A e obtemos a parte fracionária
- escolhemos A conveniente, por exemplo $A = (\sqrt{5} 1)/2$
- ullet posição relativa no vetor **não** depende de M (pode ser M=1024)

$$h(x) = \lfloor M \left(A \cdot x \bmod 1 \right) \rfloor$$

- ullet multiplicamos por um certo valor real A e obtemos a parte fracionária
- escolhemos A conveniente, por exemplo $A = (\sqrt{5} 1)/2$
- ullet posição relativa no vetor **não** depende de M (pode ser M=1024)

$$h(x) = \lfloor M \left(A \cdot x \bmod 1 \right) \rfloor$$

- ullet multiplicamos por um certo valor real A e obtemos a parte fracionária
- escolhemos A conveniente, por exemplo $A = (\sqrt{5} 1)/2$
- ullet posição relativa no vetor **não** depende de M (pode ser M=1024)

$$h(x) = \lfloor M \left(A \cdot x \bmod 1 \right) \rfloor$$

$$h(\text{``bala''}) = \lfloor 1024 \cdot \left[((\sqrt{5}-1)/2 \cdot 1.650.551.905) \text{ mod } 1 \right] \rfloor$$

- ullet multiplicamos por um certo valor real A e obtemos a parte fracionária
- escolhemos A conveniente, por exemplo $A = (\sqrt{5} 1)/2$
- posição relativa no vetor não depende de M (pode ser M=1024)

$$h(x) = \lfloor M \left(A \cdot x \bmod 1 \right) \rfloor$$

$$\begin{split} h(\text{``bala''}) &= \lfloor 1024 \cdot [((\sqrt{5}-1)/2 \cdot 1.650.551.905) \text{ mod } 1] \rfloor \\ &= \lfloor 1024 \cdot [1020097177,4858876 \text{ mod } 1] \rfloor \end{split}$$

- ullet multiplicamos por um certo valor real A e obtemos a parte fracionária
- escolhemos A conveniente, por exemplo $A = (\sqrt{5} 1)/2$
- ullet posição relativa no vetor **não** depende de M (pode ser M=1024)

$$h(x) = \lfloor M \left(A \cdot x \bmod 1 \right) \rfloor$$

$$\begin{split} h(\text{``bala''}) &= \lfloor 1024 \cdot [((\sqrt{5}-1)/2 \cdot 1.650.551.905) \text{ mod } 1] \rfloor \\ &= \lfloor 1024 \cdot [1020097177, 4858876 \text{ mod } 1] \rfloor \\ &= \lfloor 1024 \cdot 0.4858876 \rfloor \end{split}$$

- multiplicamos por um certo valor real A e obtemos a parte fracionária
- escolhemos A conveniente, por exemplo $A = (\sqrt{5} 1)/2$
- ullet posição relativa no vetor **não** depende de M (pode ser M=1024)

$$h(x) = \lfloor M \left(A \cdot x \bmod 1 \right) \rfloor$$

$$\begin{split} h(\text{``bala''}) &= \lfloor 1024 \cdot [((\sqrt{5}-1)/2 \cdot 1.650.551.905) \text{ mod } 1] \rfloor \\ &= \lfloor 1024 \cdot [1020097177, 4858876 \text{ mod } 1] \rfloor \\ &= \lfloor 1024 \cdot 0.4858876 \rfloor \\ &= \lfloor 497.5489024 \rfloor \end{split}$$

- multiplicamos por um certo valor real A e obtemos a parte fracionária
- escolhemos A conveniente, por exemplo $A = (\sqrt{5} 1)/2$
- ullet posição relativa no vetor **não** depende de M (pode ser M=1024)

$$h(x) = \lfloor M \left(A \cdot x \bmod 1 \right) \rfloor$$

$$\begin{split} h(\text{``bala''}) &= \lfloor 1024 \cdot [((\sqrt{5}-1)/2 \cdot 1.650.551.905) \text{ mod } 1] \rfloor \\ &= \lfloor 1024 \cdot [1020097177, 4858876 \text{ mod } 1] \rfloor \\ &= \lfloor 1024 \cdot 0.4858876 \rfloor \\ &= \lfloor 497.5489024 \rfloor = 497 \end{split}$$

- ullet multiplicamos por um certo valor real A e obtemos a parte fracionária
- escolhemos A conveniente, por exemplo $A = (\sqrt{5} 1)/2$
- ullet posição relativa no vetor **não** depende de M (pode ser M=1024)

$$h(x) = \lfloor M \left(A \cdot x \bmod 1 \right) \rfloor$$

Exemplo:

$$\begin{split} h(\text{``bala''}) &= \lfloor 1024 \cdot [((\sqrt{5}-1)/2 \cdot 1.650.551.905) \text{ mod } 1] \rfloor \\ &= \lfloor 1024 \cdot [1020097177,4858876 \text{ mod } 1] \rfloor \\ &= \lfloor 1024 \cdot 0,4858876 \rfloor \\ &= \lfloor 497,5489024 \rfloor = 497 \end{split}$$

O uso da razão áurea como valor de A é sugestão de Knuth

Interface do TAD

```
1 #define MAX 1783
2
3 typedef struct {
4    char chave[10];
5    int dado;
6    No * prox;
7 } No;
```

Interface do TAD

```
1 #define MAX 1783
2
3 typedef struct {
4   char chave[10];
5   int dado;
6   No * prox;
7 } No;
8
9 typedef No * p_no;
10
11 typedef struct {
12   p_no vetor[MAX];
13 } Hash;
```

Interface do TAD

```
1 #define MAX 1783
3 typedef struct {
4 char chave[10];
5 int dado:
6 No * prox;
7 } No:
8
9 typedef No * p no;
10
11 typedef struct {
12  p_no vetor[MAX];
13 } Hash;
14
15 typedef Hash * p_hash;
16
17 p_hash criar_hash();
18
19 void destruir_hash(p_hash t);
20
21 void inserir(p_hash t, char *chave, int dado);
22
23 void remover(p_hash t, char *chave);
24
25 p_no buscar(p_hash t, char *chave);
```

```
1 int hash(char *chave) {
```

```
1 int hash(char *chave) {
2  int i, n = 0;
```

```
1 int hash(char *chave) {
2   int i, n = 0;
3   for (i = 0; i < strlen(chave); i++)
4    n = (256 * n + chave[i]) % MAX;</pre>
```

```
1 int hash(char *chave) {
2   int i, n = 0;
3   for (i = 0; i < strlen(chave); i++)
4    n = (256 * n + chave[i]) % MAX;
5   return n;
6 }</pre>
```

```
1 int hash(char *chave) {
2   int i, n = 0;
3   for (i = 0; i < strlen(chave); i++)
4    n = (256 * n + chave[i]) % MAX;
5   return n;
6 }
7
8 void inserir(p_hash t, char *chave, int dado) {</pre>
```

```
int hash(char *chave) {
  int i, n = 0;
  for (i = 0; i < strlen(chave); i++)
    n = (256 * n + chave[i]) % MAX;
  return n;
  }
  void inserir(p_hash t, char *chave, int dado) {
    int n = hash(chave);
  t->vetor[n] = inserir_lista(t->vetor[n], chave, dado);
}
```

```
1 int hash(char *chave) {
2   int i, n = 0;
3   for (i = 0; i < strlen(chave); i++)
4    n = (256 * n + chave[i]) % MAX;
5   return n;
6 }
7
8 void inserir(p_hash t, char *chave, int dado) {
9   int n = hash(chave);
10   t->vetor[n] = inserir_lista(t->vetor[n], chave, dado);
11 }
12
13 void remover(p_hash t, char *chave) {
```

```
1 int hash(char *chave) {
2 int i, n = 0:
3 for (i = 0; i < strlen(chave); i++)</pre>
   n = (256 * n + chave[i]) % MAX;
5
    return n:
6 }
7
8 void inserir(p_hash t, char *chave, int dado) {
  int n = hash(chave);
    t->vetor[n] = inserir_lista(t->vetor[n], chave, dado);
10
11 }
12
13 void remover(p_hash t, char *chave) {
int n = hash(chave);
t ->vetor[n] = remover_lista(t->vetor[n], chave);
16 }
```

Sabendo a função de hashing, podemos prejudicar o programa:

Sabendo a função de hashing, podemos prejudicar o programa:

• insira muitos elementos com o mesmo hash

Sabendo a função de hashing, podemos prejudicar o programa:

• insira muitos elementos com o mesmo hash

Como nos proteger de um adversário malicioso?

Sabendo a função de hashing, podemos prejudicar o programa:

• insira muitos elementos com o mesmo hash

Como nos proteger de um adversário malicioso?

• Podemos escolher a função de hashing aleatoriamente

Sabendo a função de hashing, podemos prejudicar o programa:

• insira muitos elementos com o mesmo hash

Como nos proteger de um adversário malicioso?

Podemos escolher a função de hashing aleatoriamente

Sabendo a função de hashing, podemos prejudicar o programa:

• insira muitos elementos com o mesmo hash

Como nos proteger de um adversário malicioso?

Podemos escolher a função de hashing aleatoriamente

Uma boa função de hashing aleatória:

ullet fixe p um primo maior do que M

Sabendo a função de hashing, podemos prejudicar o programa:

insira muitos elementos com o mesmo hash

Como nos proteger de um adversário malicioso?

• Podemos escolher a função de hashing aleatoriamente

- ullet fixe p um primo maior do que M
- ullet escolha ${\color{blue}a} \in \{1,\dots,p\}$ e ${\color{blue}b} \in \{0,\dots,p\}$ uniform. ao acaso

Sabendo a função de hashing, podemos prejudicar o programa:

insira muitos elementos com o mesmo hash

Como nos proteger de um adversário malicioso?

Podemos escolher a função de hashing aleatoriamente

- fixe p um primo maior do que M
- escolha $\mathbf{a} \in \{1, \dots, p\}$ e $\mathbf{b} \in \{0, \dots, p\}$ uniform. ao acaso
- defina $h_{a,b}(k) = ((ak+b) \bmod p) \bmod M$

Sabendo a função de hashing, podemos prejudicar o programa:

insira muitos elementos com o mesmo hash

Como nos proteger de um adversário malicioso?

Podemos escolher a função de hashing aleatoriamente

- fixe p um primo maior do que M
- escolha $\mathbf{a} \in \{1, \dots, p\}$ e $\mathbf{b} \in \{0, \dots, p\}$ uniform. ao acaso
- defina $h_{a,b}(k) = ((ak+b) \bmod p) \bmod M$
- sabemos que essa função espalha bem

Sabendo a função de hashing, podemos prejudicar o programa:

insira muitos elementos com o mesmo hash

Como nos proteger de um adversário malicioso?

Podemos escolher a função de hashing aleatoriamente

- fixe p um primo maior do que M
- escolha $\mathbf{a} \in \{1, \dots, p\}$ e $\mathbf{b} \in \{0, \dots, p\}$ uniform. ao acaso
- defina $h_{a,b}(k) = ((ak+b) \bmod p) \bmod M$
- sabemos que essa função espalha bem
 - a probabilidade de colisão é no máximo 1/M

Sabendo a função de hashing, podemos prejudicar o programa:

insira muitos elementos com o mesmo hash

Como nos proteger de um adversário malicioso?

Podemos escolher a função de hashing aleatoriamente

- fixe p um primo maior do que M
- escolha $\mathbf{a} \in \{1, \dots, p\}$ e $\mathbf{b} \in \{0, \dots, p\}$ uniform. ao acaso
- defina $h_{a,b}(k) = ((ak+b) \bmod p) \bmod M$
- sabemos que essa função espalha bem
 - a probabilidade de colisão é no máximo 1/M
 - é um hashing universal

Endereçamento aberto

Existe uma alternativa para a implementação de tabela de espalhamento

Endereçamento aberto

Existe uma alternativa para a implementação de tabela de espalhamento

Endereçamento aberto:

Endereçamento aberto

Existe uma alternativa para a implementação de tabela de espalhamento

Endereçamento aberto:

• os dados são guardados no próprio vetor

Existe uma alternativa para a implementação de tabela de espalhamento

Endereçamento aberto:

- os dados são guardados no próprio vetor
- colisões são colocadas em posições livres da tabela

Existe uma alternativa para a implementação de tabela de espalhamento

Endereçamento aberto:

- os dados são guardados no próprio vetor
- colisões são colocadas em posições livres da tabela

Existe uma alternativa para a implementação de tabela de espalhamento

Endereçamento aberto:

- os dados são guardados no próprio vetor
- colisões são colocadas em posições livres da tabela

Características:

 evita percorrer usando ponteiros e alocação e deslocação de memória (malloc e free)

Existe uma alternativa para a implementação de tabela de espalhamento

Endereçamento aberto:

- os dados são guardados no próprio vetor
- colisões são colocadas em posições livres da tabela

- evita percorrer usando ponteiros e alocação e deslocação de memória (malloc e free)
- se a tabela encher, deve recriar uma tabela maior

Existe uma alternativa para a implementação de tabela de espalhamento

Endereçamento aberto:

- os dados são guardados no próprio vetor
- colisões são colocadas em posições livres da tabela

- evita percorrer usando ponteiros e alocação e deslocação de memória (malloc e free)
- se a tabela encher, deve recriar uma tabela maior
 - e mudar a função de hashing

Existe uma alternativa para a implementação de tabela de espalhamento

Endereçamento aberto:

- os dados são guardados no próprio vetor
- colisões são colocadas em posições livres da tabela

- evita percorrer usando ponteiros e alocação e deslocação de memória (malloc e free)
- se a tabela encher, deve recriar uma tabela maior
 - e mudar a função de hashing
- remoção é mais complicada

broca

boca

bolo

bela

bala

dia

escola

gratuito

ilha

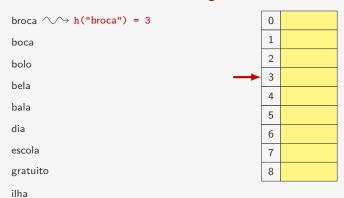
0	
1	
2	
3	
4	
5	
6	
7	
8	

```
broca \rightsquigarrow h("broca") = 3
boca
bolo
bela
bala
dia
escola
gratuito
ilha
```

0	
1	
2	
3	
4	
5	
6	
7	
8	

Inserindo:

• procuramos posição



- procuramos posição
- se houver espaço, guardamos

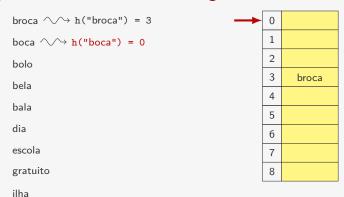
```
broca \rightsquigarrow h("broca") = 3
boca
bolo
bela
bala
dia
escola
gratuito
ilha
```

0	
1	
2	
3	broca
4	
5	
6	
7	
8	

- procuramos posição
- se houver espaço, guardamos

0	
1	
2	
3	broca
4	
5	
6	
7	
8	

- procuramos posição
- se houver espaço, guardamos



- procuramos posição
- se houver espaço, guardamos

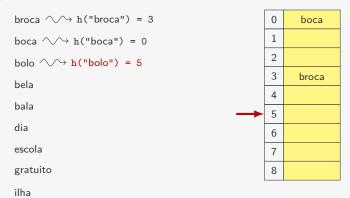
0	boca
1	
2	
3	broca
4	
5	
6	
7	
8	

- procuramos posição
- se houver espaço, guardamos

```
broca \rightsquigarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo \rightsquigarrow h("bolo") = 5
bela
bala
dia
escola
gratuito
ilha
```

boca
broca

- procuramos posição
- se houver espaço, guardamos



- procuramos posição
- se houver espaço, guardamos

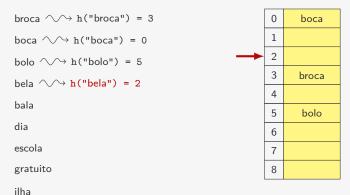
```
broca \rightsquigarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo \rightsquigarrow h("bolo") = 5
bela
bala
dia
escola
gratuito
ilha
```

0	boca
1	
2	
3	broca
4	
5	bolo
6	
7	
8	

- procuramos posição
- se houver espaço, guardamos

0	boca
1	
2	
3	broca
4	
5	bolo
6	
7	
8	
8	

- procuramos posição
- se houver espaço, guardamos

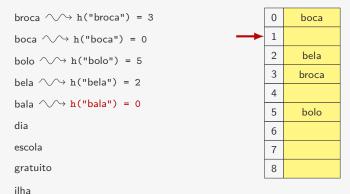


- procuramos posição
- se houver espaço, guardamos

0	boca
1	
2	bela
3	broca
4	
5	bolo
6	
7	
8	

- procuramos posição
- se houver espaço, guardamos

- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre $(m\acute{o}dulo\ M)$



- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre $(m\acute{o}dulo\ M)$

```
broca \rightsquigarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo \rightsquigarrow h("bolo") = 5
bela \wedge \rightarrow h("bela") = 2
bala \wedge \wedge \rightarrow h("bala") = 0
dia
escola
gratuito
ilha
```

0	boca
1	bala
2	bela
3	broca
4	
5	bolo
6	
7	
8	

- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre (m'odulo M)

```
broca \rightsquigarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo \rightsquigarrow h("bolo") = 5
bela \rightsquigarrow h("bela") = 2
bala \wedge \wedge \rightarrow h("bala") = 0
dia
escola
gratuito
ilha
```

0	boca
1	bala
2	bela
3	broca
4	
5	bolo
6	
7	
8	

- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre $(m\acute{o}dulo\ M)$

```
broca \rightsquigarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo \rightsquigarrow h("bolo") = 5
bela \rightsquigarrow h("bela") = 2
bala \wedge \wedge \rightarrow h("bala") = 0
dia \land \land \Rightarrow h("dia") = 2
escola
gratuito
ilha
```

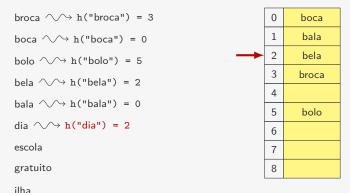
0	boca
1	bala
2	bela
3	broca
4	
5	bolo
6	
7	
8	

- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre $(m\acute{o}dulo\ M)$

```
broca \rightsquigarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo \rightsquigarrow h("bolo") = 5
bela \rightsquigarrow h("bela") = 2
bala \wedge \wedge \rightarrow h("bala") = 0
dia \land \land \Rightarrow h("dia") = 2
escola
gratuito
ilha
```

0	boca
1	bala
2	bela
3	broca
4	
5	bolo
6	
7	
8	

- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre $(m\acute{o}dulo\ M)$



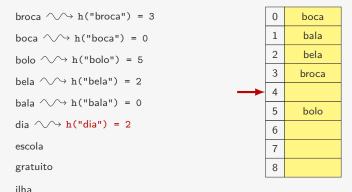
- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre $(m\acute{o}dulo\ M)$

broca \rightsquigarrow h("broca") = 3	0	boca
boca \rightsquigarrow h("boca") = 0	1	bala
bolo	2	bela
bela \rightsquigarrow h("bela") = 2	3	broca
	4	
bala $\wedge \rightarrow h("bala") = 0$	5	bolo
dia ∕∕→ h("dia") = 2	6	
escola	7	
gratuito	8	

Inserindo:

ilha

- procuramos posição
- se houver espaço, guardamos
- ullet se não houver espaço, procuramos a próxima posição livre (módulo M)



- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre $(m\'odulo\ M)$

```
broca \rightsquigarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo \rightsquigarrow h("bolo") = 5
bela \rightsquigarrow h("bela") = 2
bala \wedge \wedge \rightarrow h("bala") = 0
dia \land \land \Rightarrow h("dia") = 2
escola
gratuito
ilha
```

0	boca
1	bala
2	bela
3	broca
4	dia
5	bolo
6	
7	
8	

- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre $(m\acute{o}dulo\ M)$

```
broca \rightsquigarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo \rightsquigarrow h("bolo") = 5
bela \rightsquigarrow h("bela") = 2
bala \rightsquigarrow h("bala") = 0
dia \sim h("dia") = 2
escola \wedge \wedge \rightarrow h("escola") = 7
gratuito
ilha
```

0	boca
1	bala
2	bela
3	broca
4	dia
5	bolo
6	
7	
8	

- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre $(m\acute{o}dulo\ M)$

broca \rightsquigarrow h("broca") = 3	0	boca
boca \rightsquigarrow h("boca") = 0	1	bala
bolo \square h("bolo") = 5	2	bela
bela ∕∕→ h("bela") = 2	3	broca
bala \rightsquigarrow h("bala") = 0	4	dia
pala · 🗸 → II(bala) = 0	5	bolo
dia ∕∕→ h("dia") = 2	6	
escola $\wedge \rightarrow h("escola") = 7$	7	
gratuito	8	
ilha		

- procuramos posição
- se houver espaço, guardamos
- ullet se não houver espaço, procuramos a próxima posição livre (módulo M)

```
broca \rightsquigarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo \rightsquigarrow h("bolo") = 5
bela \rightsquigarrow h("bela") = 2
bala \rightsquigarrow h("bala") = 0
dia \sim h("dia") = 2
escola \wedge \wedge \rightarrow h("escola") = 7
gratuito
ilha
```

0	boca
1	bala
2	bela
3	broca
4	dia
5	bolo
6	
7	escola
8	

- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre (m'odulo M)

```
broca \rightsquigarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo \rightsquigarrow h("bolo") = 5
bela \rightsquigarrow h("bela") = 2
bala \wedge \wedge \rightarrow h("bala") = 0
dia \sim h("dia") = 2
escola \land \land \Rightarrow h("escola") = 7
gratuito \wedge \wedge \rightarrow h("gratuito") = 0
ilha
```

0	boca
1	bala
2	bela
3	broca
4	dia
5	bolo
6	
7	escola
8	

- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre $(m\acute{o}dulo\ M)$

broca ∕∕→ h("broca") = 3	0	
boca \rightsquigarrow h("boca") = 0	1	
bolo ∕∕→ h("bolo") = 5	2	
bela $\wedge \wedge \rightarrow h("bela") = 2$	3	
$bala \longrightarrow h("bala") = 0$	4	
	5	
dia ∕∕→ h("dia") = 2	6	
escola $\wedge \rightarrow h("escola") = 7$	7	
gratuito $\wedge \rightarrow h("gratuito") = 0$	8	
ilha		

boca
bala
bela
broca
dia
bolo

escola

- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre (módulo M)

broca \rightsquigarrow h("broca") = 3	0	boca
boca \rightsquigarrow h("boca") = 0	1	bala
bolo	2	bela
bela ∕ → h("bela") = 2	3	broca
, , , , , , ,	4	dia
$bala \wedge \wedge h("bala") = 0$	5	bolo
dia ∕∕→ h("dia") = 2	6	
escola $\wedge \rightarrow h("escola") = 7$	7	escola
gratuito $\wedge \rightarrow h("gratuito") = 0$	8	
ilha		

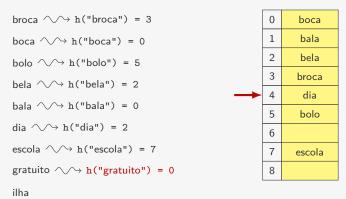
- procuramos posição
- se houver espaço, guardamos
- ullet se não houver espaço, procuramos a próxima posição livre (módulo M)

broca \rightsquigarrow h("broca") = 3	0	boca
boca $\sim h("boca") = 0$	1	bala
bolo	2	bela
bela ∕∕→ h("bela") = 2	3	broca
bala $\wedge \rightarrow h("bala") = 0$	4	dia
	5	bolo
dia ∕∕→ h("dia") = 2	6	
escola $\wedge \rightarrow h("escola") = 7$	7	escola
gratuito $\wedge h$ h("gratuito") = 0	8	
ilha		

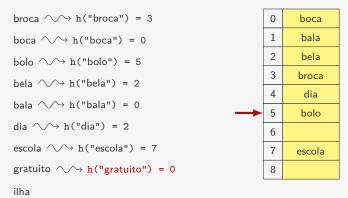
- procuramos posição
- se houver espaço, guardamos
- ullet se não houver espaço, procuramos a próxima posição livre (módulo M)

broca $\wedge \rightarrow$ h("broca") = 3	0	boca
boca \rightsquigarrow h("boca") = 0	1	bala
bolo	2	bela
bela	3	broca
bala $\wedge \rightarrow h$ ("bala") = 0	4	dia
	5	bolo
dia ∕∕→ h("dia") = 2	6	
escola $\wedge \rightarrow h("escola") = 7$	7	escola
gratuito $\wedge h$ h("gratuito") = 0	8	
ilha		

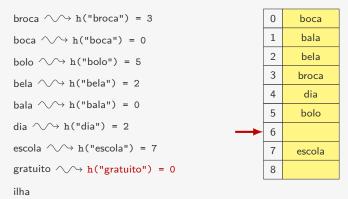
- procuramos posição
- se houver espaço, guardamos
- ullet se não houver espaço, procuramos a próxima posição livre (módulo M)



- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre $(m\acute{o}dulo\ M)$



- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre $(m\acute{o}dulo\ M)$



- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre $(m\acute{o}dulo\ M)$

```
broca \rightsquigarrow h("broca") = 3
boca \rightsquigarrow h("boca") = 0
bolo \rightsquigarrow h("bolo") = 5
bela \rightsquigarrow h("bela") = 2
bala \wedge \wedge \rightarrow h("bala") = 0
dia \sim h("dia") = 2
escola \land \land \Rightarrow h("escola") = 7
gratuito \wedge \wedge \rightarrow h("gratuito") = 0
ilha
```

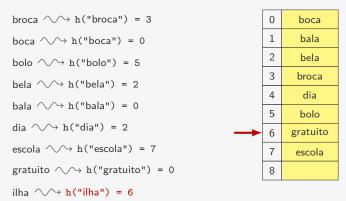
0	boca
1	bala
2	bela
3	broca
4	dia
5	bolo
6	gratuito
7	escola
8	

- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre $(m\acute{o}dulo\ M)$

```
broca \rightsquigarrow h("broca") = 3
boca \wedge \wedge \rightarrow h("boca") = 0
bolo \wedge \rightarrow h("bolo") = 5
bela \rightsquigarrow h("bela") = 2
bala \wedge \wedge \rightarrow h("bala") = 0
dia \sim h("dia") = 2
escola \wedge \wedge \rightarrow h("escola") = 7
gratuito \wedge \rightarrow h("gratuito") = 0
ilha \wedge \wedge \rightarrow h("ilha") = 6
```

0	boca
1	bala
2	bela
3	broca
4	dia
5	bolo
6	gratuito
7	escola
8	

- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre (módulo M)



- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre (m'odulo M)

broca $\wedge \rightarrow$ h("broca") = 3	0	boca
boca \rightsquigarrow h("boca") = 0	1	bala
bolo ∕√→ h("bolo") = 5	2	bela
bela ∕ → h("bela") = 2	3	broca
	4	dia
$bala \wedge h("bala") = 0$	5	bolo
dia ∕∕→ h("dia") = 2	6	gratuito
escola $\wedge \rightarrow h("escola") = 7$	7	escola
gratuito $\wedge \rightarrow h("gratuito") = 0$	8	
ilha $\rightsquigarrow h("ilha") = 6$		

- procuramos posição
- se houver espaço, guardamos
- ullet se não houver espaço, procuramos a próxima posição livre (módulo M)

broca $\wedge \rightarrow$ h("broca") = 3	0	boca
boca \rightsquigarrow h("boca") = 0	1	bala
bolo	2	bela
bela ∕	3	broca
	4	dia
bala ∕∕→ h("bala") = 0	5	bolo
dia ∕∕→ h("dia") = 2	6	gratuito
escola $\wedge \wedge h$ ("escola") = 7	7	escola
gratuito $\wedge \rightarrow$ h("gratuito") = 0	8	
$ilha \land \land \Rightarrow h("ilha") = 6$		

- procuramos posição
- se houver espaço, guardamos
- ullet se não houver espaço, procuramos a próxima posição livre (módulo M)

```
broca \rightsquigarrow h("broca") = 3
boca \wedge \wedge \rightarrow h("boca") = 0
bolo \rightsquigarrow h("bolo") = 5
bela \rightsquigarrow h("bela") = 2
bala \wedge \wedge \rightarrow h("bala") = 0
dia \sim h("dia") = 2
escola \wedge \wedge \rightarrow h("escola") = 7
gratuito \wedge \rightarrow h("gratuito") = 0
ilha \wedge \wedge \rightarrow h("ilha") = 6
```

0	boca
1	bala
2	bela
3	broca
4	dia
5	bolo
6	gratuito
7	escola
8	ilha

- procuramos posição
- se houver espaço, guardamos
- se não houver espaço, procuramos a próxima posição livre (m'odulo M)

Como fazer uma busca com endereçamento aberto?

• Basta simular a inserção:

- Basta simular a inserção:
 - Calcule a função de hashing

- Basta simular a inserção:
 - Calcule a função de hashing
 - Percorra a tabela em sequência procurando pela chave

- Basta simular a inserção:
 - Calcule a função de hashing
 - Percorra a tabela em sequência procurando pela chave
 - Se encontrar a chave, devolva o item correspondente

- Basta simular a inserção:
 - Calcule a função de hashing
 - Percorra a tabela em sequência procurando pela chave
 - Se encontrar a chave, devolva o item correspondente
 - Se encontrar um espaço vazio, devolva NULL

Como fazer uma busca com endereçamento aberto?

- Basta simular a inserção:
 - Calcule a função de hashing
 - Percorra a tabela em sequência procurando pela chave
 - Se encontrar a chave, devolva o item correspondente
 - Se encontrar um espaço vazio, devolva NULL

Como fazer uma busca com endereçamento aberto?

- Basta simular a inserção:
 - Calcule a função de hashing
 - Percorra a tabela em sequência procurando pela chave
 - Se encontrar a chave, devolva o item correspondente
 - Se encontrar um espaço vazio, devolva NULL

O que é um espaço vazio em um vetor?

Se for um vetor de ponteiros, pode ser NULL

Como fazer uma busca com endereçamento aberto?

- Basta simular a inserção:
 - Calcule a função de hashing
 - Percorra a tabela em sequência procurando pela chave
 - Se encontrar a chave, devolva o item correspondente
 - Se encontrar um espaço vazio, devolva NULL

- Se for um vetor de ponteiros, pode ser NULL
- Se n\u00e3o for, precisa ser um elemento dummy

Como fazer uma busca com endereçamento aberto?

- Basta simular a inserção:
 - Calcule a função de hashing
 - Percorra a tabela em sequência procurando pela chave
 - Se encontrar a chave, devolva o item correspondente
 - Se encontrar um espaço vazio, devolva NULL

- Se for um vetor de ponteiros, pode ser NULL
- Se não for, precisa ser um elemento dummy
 - ou um valor que nunca será usado

Como fazer uma busca com endereçamento aberto?

- Basta simular a inserção:
 - Calcule a função de hashing
 - Percorra a tabela em sequência procurando pela chave
 - Se encontrar a chave, devolva o item correspondente
 - Se encontrar um espaço vazio, devolva NULL

- Se for um vetor de ponteiros, pode ser NULL
- Se não for, precisa ser um elemento dummy
 - ou um valor que nunca será usado
 - ou ter um campo indicando que é dummy

Como fazer a remoção com endereçamento aberto?

• Não podemos apenas remover os elementos da tabela

- Não podemos apenas remover os elementos da tabela
 - Por quê?

- Não podemos apenas remover os elementos da tabela
 - Por quê? Quebraria a busca...

- Não podemos apenas remover os elementos da tabela
 - Por quê? Quebraria a busca...
- Opção 1: rehash dos elementos seguintes do bloco

- Não podemos apenas remover os elementos da tabela
 - Por quê? Quebraria a busca...
- Opção 1: rehash dos elementos seguintes do bloco
 - removemos os elementos até a próxima posição vazia

- Não podemos apenas remover os elementos da tabela
 - Por quê? Quebraria a busca...
- Opção 1: rehash dos elementos seguintes do bloco
 - removemos os elementos até a próxima posição vazia
 - recalculamos o hash de cada um e reinserimos na tabela

- Não podemos apenas remover os elementos da tabela
 - Por quê? Quebraria a busca...
- Opção 1: rehash dos elementos seguintes do bloco
 - removemos os elementos até a próxima posição vazia
 - recalculamos o hash de cada um e reinserimos na tabela
 - pode ser custoso e difícil de implementar

- Não podemos apenas remover os elementos da tabela
 - Por quê? Quebraria a busca...
- Opção 1: rehash dos elementos seguintes do bloco
 - removemos os elementos até a próxima posição vazia
 - recalculamos o hash de cada um e reinserimos na tabela
 - pode ser custoso e difícil de implementar
- Opção 2: trocamos por um valor dummy

- Não podemos apenas remover os elementos da tabela
 - Por quê? Quebraria a busca...
- Opção 1: rehash dos elementos seguintes do bloco
 - removemos os elementos até a próxima posição vazia
 - recalculamos o hash de cada um e reinserimos na tabela
 - pode ser custoso e difícil de implementar
- Opção 2: trocamos por um valor dummy
 - valor indica que o item foi removido

- Não podemos apenas remover os elementos da tabela
 - Por quê? Quebraria a busca...
- Opção 1: rehash dos elementos seguintes do bloco
 - removemos os elementos até a próxima posição vazia
 - recalculamos o hash de cada um e reinserimos na tabela
 - pode ser custoso e difícil de implementar
- Opção 2: trocamos por um valor dummy
 - valor indica que o item foi removido
 - mas não pode ser o mesmo que indica espaço vazio

- Não podemos apenas remover os elementos da tabela
 - Por quê? Quebraria a busca...
- Opção 1: rehash dos elementos seguintes do bloco
 - removemos os elementos até a próxima posição vazia
 - recalculamos o hash de cada um e reinserimos na tabela
 - pode ser custoso e difícil de implementar
- Opção 2: trocamos por um valor dummy
 - valor indica que o item foi removido
 - mas não pode ser o mesmo que indica espaço vazio
- Opção 3: marcamos o item como removido

- Não podemos apenas remover os elementos da tabela
 - Por quê? Quebraria a busca...
- Opção 1: rehash dos elementos seguintes do bloco
 - removemos os elementos até a próxima posição vazia
 - recalculamos o hash de cada um e reinserimos na tabela
 - pode ser custoso e difícil de implementar
- Opção 2: trocamos por um valor dummy
 - valor indica que o item foi removido
 - mas não pode ser o mesmo que indica espaço vazio
- Opção 3: marcamos o item como removido
 - usamos um campo adicional

Inserção e Busca Revisitadas

Se fizermos a remoção marcando o item como removido, precisamos mudar a inserção e a busca

Inserção e Busca Revisitadas

Se fizermos a remoção marcando o item como removido, precisamos mudar a inserção e a busca

Inserção:

Se fizermos a remoção marcando o item como removido, precisamos mudar a inserção e a busca

Inserção:

ullet Calculamos a função hashing e temos um resultado h

Se fizermos a remoção marcando o item como removido, precisamos mudar a inserção e a busca

Inserção:

- Calculamos a função hashing e temos um resultado h
- Inserimos na primeira posição vazia ou com item removido a partir de h

Se fizermos a remoção marcando o item como removido, precisamos mudar a inserção e a busca

Inserção:

- Calculamos a função hashing e temos um resultado h
- Inserimos na primeira posição vazia ou com item removido a partir de h

Se fizermos a remoção marcando o item como removido, precisamos mudar a inserção e a busca

Inserção:

- Calculamos a função hashing e temos um resultado h
- Inserimos na primeira posição vazia ou com item removido a partir de h

Busca:

ullet Calculamos a função hashing e temos um resultado h

Se fizermos a remoção marcando o item como removido, precisamos mudar a inserção e a busca

Inserção:

- Calculamos a função hashing e temos um resultado h
- Inserimos na primeira posição vazia ou com item removido a partir de h

- Calculamos a função hashing e temos um resultado h
- Procuramos cada posição a partir de h em sequência

Se fizermos a remoção marcando o item como removido, precisamos mudar a inserção e a busca

Inserção:

- Calculamos a função hashing e temos um resultado h
- Inserimos na primeira posição vazia ou com item removido a partir de h

- Calculamos a função hashing e temos um resultado h
- Procuramos cada posição a partir de h em sequência
 - Se encontrar o item, verifique se ele não foi removido

Se fizermos a remoção marcando o item como removido, precisamos mudar a inserção e a busca

Inserção:

- Calculamos a função hashing e temos um resultado h
- Inserimos na primeira posição vazia ou com item removido a partir de h

- Calculamos a função hashing e temos um resultado h
- Procuramos cada posição a partir de h em sequência
 - Se encontrar o item, verifique se ele n\u00e3o foi removido
 - Passe por cima de itens removidos

Se fizermos a remoção marcando o item como removido, precisamos mudar a inserção e a busca

Inserção:

- Calculamos a função hashing e temos um resultado h
- Inserimos na primeira posição vazia ou com item removido a partir de h

- Calculamos a função hashing e temos um resultado h
- Procuramos cada posição a partir de h em sequência
 - Se encontrar o item, verifique se ele não foi removido
 - Passe por cima de itens removidos
 - Pare ao encontrar uma posição vazia

Se fizermos a remoção marcando o item como removido, precisamos mudar a inserção e a busca

Inserção:

- Calculamos a função hashing e temos um resultado h
- Inserimos na primeira posição vazia ou com item removido a partir de h

- ullet Calculamos a função hashing e temos um resultado h
- Procuramos cada posição a partir de h em sequência
 - Se encontrar o item, verifique se ele n\u00e3o foi removido
 - Passe por cima de itens removidos
 - Pare ao encontrar uma posição vazia
- Cuidado para não ciclar...

É como a sondagem linear:

É como a sondagem linear:

• Estratégia mais geral para lidar com conflitos

É como a sondagem linear:

- Estratégia mais geral para lidar com conflitos
- Ao invés de saltarmos sempre de 1, saltamos de $hash_2(k)$

É como a sondagem linear:

- Estratégia mais geral para lidar com conflitos
- Ao invés de saltarmos sempre de 1, saltamos de $hash_2(k)$
- Onde *hash*₂ é uma segunda função de hashing

É como a sondagem linear:

- Estratégia mais geral para lidar com conflitos
- Ao invés de saltarmos sempre de 1, saltamos de $hash_2(k)$
- Onde hash₂ é uma segunda função de hashing

Isso é,

$$h(k,i) = (hash_1(k) + i \cdot hash_2(k)) \mod M$$

É como a sondagem linear:

- Estratégia mais geral para lidar com conflitos
- Ao invés de saltarmos sempre de 1, saltamos de $hash_2(k)$
- Onde hash₂ é uma segunda função de hashing

Isso é,

$$h(k,i) = (hash_1(k) + i \cdot hash_2(k)) \mod M$$

Cuidados:

É como a sondagem linear:

- Estratégia mais geral para lidar com conflitos
- Ao invés de saltarmos sempre de 1, saltamos de $hash_2(k)$
- Onde hash₂ é uma segunda função de hashing

Isso é,

$$h(k,i) = (hash_1(k) + i \cdot hash_2(k)) \mod M$$

Cuidados:

• $hash_2(k)$ nunca pode ser zero

É como a sondagem linear:

- Estratégia mais geral para lidar com conflitos
- Ao invés de saltarmos sempre de 1, saltamos de $hash_2(k)$
- Onde hash₂ é uma segunda função de hashing

Isso é,

$$h(k,i) = (hash_1(k) + i \cdot hash_2(k)) \mod M$$

Cuidados:

- $hash_2(k)$ nunca pode ser zero
- $hash_2(k)$ precisa ser coprimo com M

É como a sondagem linear:

- Estratégia mais geral para lidar com conflitos
- Ao invés de saltarmos sempre de 1, saltamos de $hash_2(k)$
- Onde hash₂ é uma segunda função de hashing

Isso é,

$$h(k,i) = (hash_1(k) + i \cdot hash_2(k)) \mod M$$

Cuidados:

- $hash_2(k)$ nunca pode ser zero
- $hash_2(k)$ precisa ser coprimo com M
 - garante que as sequências são longas

É como a sondagem linear:

- Estratégia mais geral para lidar com conflitos
- Ao invés de saltarmos sempre de 1, saltamos de $hash_2(k)$
- Onde hash₂ é uma segunda função de hashing

Isso é,

$$h(k,i) = (hash_1(k) + i \cdot hash_2(k)) \mod M$$

Cuidados:

- $hash_2(k)$ nunca pode ser zero
- $hash_2(k)$ precisa ser coprimo com M
 - garante que as sequências são longas

Exemplos:

É como a sondagem linear:

- Estratégia mais geral para lidar com conflitos
- Ao invés de saltarmos sempre de 1, saltamos de $hash_2(k)$
- Onde hash₂ é uma segunda função de hashing

Isso é,

$$h(k,i) = (hash_1(k) + i \cdot hash_2(k)) \mod M$$

Cuidados:

- $hash_2(k)$ nunca pode ser zero
- $hash_2(k)$ precisa ser coprimo com M
 - garante que as sequências são longas

Exemplos:

• Escolha M como uma potência de ${\bf 2}$ e faça que $hash_2(k)$ seja sempre ímpar

É como a sondagem linear:

- Estratégia mais geral para lidar com conflitos
- Ao invés de saltarmos sempre de 1, saltamos de $hash_2(k)$
- Onde *hash*₂ é uma segunda função de hashing

Isso é,

$$h(k,i) = (hash_1(k) + i \cdot hash_2(k)) \mod M$$

Cuidados:

- $hash_2(k)$ nunca pode ser zero
- $hash_2(k)$ precisa ser coprimo com M
 - garante que as sequências são longas

Exemplos:

- Escolha M como uma potência de ${\bf 2}$ e faça que $hash_2(k)$ seja sempre ímpar
- Escolha M como um número primo e faça que $hash_2(k) < M$

Sondagem linear - número de acessos médio por busca

n/M	1/2	2/3	3/4	9/10
com sucesso	1.5	2.0	3.0	5.5
sem sucesso	2.5	5.0	8.5	55.5

26

 $^{^{1}}$ Baseado em Sedgewick, R. Algorithms in C, third edition, Addison-Wesley. 1998.

Sondagem linear - número de acessos médio por busca

n/M	1/2	2/3	3/4	9/10
com sucesso	1.5	2.0	3.0	5.5
sem sucesso	2.5	5.0	8.5	55.5

Hashing duplo - número de acessos médio por busca

n/M	1/2	2/3	3/4	9/10
com sucesso	1.4	1.6	1.8	2.6
sem sucesso	1.5	2.0	3.0	5.5

 $^{^{1}}$ Baseado em Sedgewick, R. Algorithms in C, third edition, Addison-Wesley. 1998.

Sondagem linear - número de acessos médio por busca

n/M	1/2	2/3	3/4	9/10
com sucesso	1.5	2.0	3.0	5.5
sem sucesso	2.5	5.0	8.5	55.5

Hashing duplo - número de acessos médio por busca

n/M	1/2	2/3	3/4	9/10
com sucesso	1.4	1.6	1.8	2.6
sem sucesso	1.5	2.0	3.0	5.5

De qualquer forma, é muito importante não deixar a tabela encher muito:

 $^{^{1}}$ Baseado em Sedgewick, R. Algorithms in C, third edition, Addison-Wesley. 1998.

Sondagem linear - número de acessos médio por busca

n/M	1/2	2/3	3/4	9/10
com sucesso	1.5	2.0	3.0	5.5
sem sucesso	2.5	5.0	8.5	55.5

Hashing duplo - número de acessos médio por busca

n/M	1/2	2/3	3/4	9/10
com sucesso	1.4	1.6	1.8	2.6
sem sucesso	1.5	2.0	3.0	5.5

De qualquer forma, é muito importante não deixar a tabela encher muito:

• Você pode aumentar o tamanho da tabela dinamicamente

 $^{^{1}\}mbox{Baseado}$ em Sedgewick, R. Algorithms in C, third edition, Addison-Wesley. 1998.

Sondagem linear - número de acessos médio por busca

n/M	1/2	2/3	3/4	9/10
com sucesso	1.5	2.0	3.0	5.5
sem sucesso	2.5	5.0	8.5	55.5

Hashing duplo - número de acessos médio por busca

n/M	1/2	2/3	3/4	9/10
com sucesso	1.4	1.6	1.8	2.6
sem sucesso	1.5	2.0	3.0	5.5

De qualquer forma, é muito importante não deixar a tabela encher muito:

- Você pode aumentar o tamanho da tabela dinamicamente
- Porém, precisa fazer um rehash de cada elemento para a nova tabela

26

¹Baseado em Sedgewick, R. Algorithms in C, third edition, Addison-Wesley. 1998.

Hashing é uma boa estrutura de dados para

Hashing é uma boa estrutura de dados para

• inserir, remover e buscar dados pela sua chave rapidamente

Hashing é uma boa estrutura de dados para

- inserir, remover e buscar dados pela sua chave rapidamente
- com uma boa função de hashing, essas operações levam tempo $\mathrm{O}(1)$

Hashing é uma boa estrutura de dados para

- inserir, remover e buscar dados pela sua chave rapidamente
- com uma boa função de hashing, essas operações levam tempo $\mathrm{O}(1)$
- mas não é boa se quisermos fazer operação relacionadas a ordem das chaves

Hashing é uma boa estrutura de dados para

- inserir, remover e buscar dados pela sua chave rapidamente
- com uma boa função de hashing, essas operações levam tempo $\mathrm{O}(1)$
- mas não é boa se quisermos fazer operação relacionadas a ordem das chaves

Hashing é uma boa estrutura de dados para

- inserir, remover e buscar dados pela sua chave rapidamente
- com uma boa função de hashing, essas operações levam tempo $\mathrm{O}(1)$
- mas não é boa se quisermos fazer operação relacionadas a ordem das chaves

Escolhendo a implementação:

Sondagem linear é o mais rápido se a tabela for esparsa

Hashing é uma boa estrutura de dados para

- inserir, remover e buscar dados pela sua chave rapidamente
- com uma boa função de hashing, essas operações levam tempo $\mathrm{O}(1)$
- mas não é boa se quisermos fazer operação relacionadas a ordem das chaves

- Sondagem linear é o mais rápido se a tabela for esparsa
- Hashing duplo usa melhor a memória

Hashing é uma boa estrutura de dados para

- inserir, remover e buscar dados pela sua chave rapidamente
- com uma boa função de hashing, essas operações levam tempo $\mathrm{O}(1)$
- mas não é boa se quisermos fazer operação relacionadas a ordem das chaves

- Sondagem linear é o mais rápido se a tabela for esparsa
- Hashing duplo usa melhor a memória
 - mas gasta mais tempo para computar a segunda função de hash

Hashing é uma boa estrutura de dados para

- inserir, remover e buscar dados pela sua chave rapidamente
- com uma boa função de hashing, essas operações levam tempo O(1)
- mas não é boa se quisermos fazer operação relacionadas a ordem das chaves

- Sondagem linear é o mais rápido se a tabela for esparsa
- Hashing duplo usa melhor a memória
 - mas gasta mais tempo para computar a segunda função de hash
- Encadeamento separado é mais fácil de implementar

Hashing é uma boa estrutura de dados para

- inserir, remover e buscar dados pela sua chave rapidamente
- com uma boa função de hashing, essas operações levam tempo $\mathrm{O}(1)$
- mas não é boa se quisermos fazer operação relacionadas a ordem das chaves

- Sondagem linear é o mais rápido se a tabela for esparsa
- Hashing duplo usa melhor a memória
 - mas gasta mais tempo para computar a segunda função de hash
- Encadeamento separado é mais fácil de implementar
 - Usa memória a mais para os ponteiros

Além disso, funções de hashing têm várias outras aplicações, ex:

• Para evitar erros de transmissão, podemos, além de informar uma chave, transmitir o resultado da função de hashing.

- Para evitar erros de transmissão, podemos, além de informar uma chave, transmitir o resultado da função de hashing.
 - dígitos verificadores

- Para evitar erros de transmissão, podemos, além de informar uma chave, transmitir o resultado da função de hashing.
 - dígitos verificadores
 - sequências de verificação para arquivos (MD5 e SHA)

- Para evitar erros de transmissão, podemos, além de informar uma chave, transmitir o resultado da função de hashing.
 - dígitos verificadores
 - sequências de verificação para arquivos (MD5 e SHA)
- Guardamos o hash de uma senha no banco de dados ao invés da senha em si

- Para evitar erros de transmissão, podemos, além de informar uma chave, transmitir o resultado da função de hashing.
 - dígitos verificadores
 - sequências de verificação para arquivos (MD5 e SHA)
- Guardamos o hash de uma senha no banco de dados ao invés da senha em si
 - evitamos vazamento de informação em caso de ataque

- Para evitar erros de transmissão, podemos, além de informar uma chave, transmitir o resultado da função de hashing.
 - dígitos verificadores
 - sequências de verificação para arquivos (MD5 e SHA)
- Guardamos o hash de uma senha no banco de dados ao invés da senha em si
 - evitamos vazamento de informação em caso de ataque
 - mas temos que garantir que a probabilidade de duas senhas terem o mesmo hash seja ínfima...