

# INF01124 - Classificação e Pesquisa de Dados - Exercício 4

## Professor João Comba

### 1 Tabelas Hash usando Endereçamento Fechado

Implemente uma tabela Hash com  $M$  entradas para armazenar nomes de uma lista de contatos. O tamanho da tabela  $M$  é variável e deve ser informado quando da construção da tabela. Para a implementação da Tabela Hash, realize as seguintes tarefas:

- Escolha uma função de hash para mapear um string para um inteiro entre 0 e  $M - 1$ ;
- Use o método para resolução de conflitos de endereçamento fechado;
- Implemente a operação de inserção de um elemento na tabela hash;
- Implemente a operação de busca de um nome na tabela de hash, que retorna -1 caso o nome não esteja na tabela, ou um número  $a > 0$  representando o número de acessos necessário para achar o nome;

### 2 Arquivos de Testes

A implementação da tabela Hash deve ser testada com dois arquivos. O primeiro arquivo, *nomes\_10000.txt*, contém 10,000 nomes de contato que devem ser inseridos na tabela Hash. O segundo arquivo, *consultas.txt*, contém listas de 50 nomes que devem ser usadas para acionar consultas na tabela Hash. Ambos arquivos contém em cada linha o nome de um contato no formato :

nome sobrenome

Um exemplo é dado abaixo:

Alfredo Derouen  
Janyce Koster  
Lesley Trabue  
Francesco Lauber  
Angel Ridgeway  
Buford Montez  
Coleen Harley  
Viviana Genova  
Veronique Zambrana  
Naida Meldrum  
Artie Smithers  
Milton Delgado  
Emory Casarez  
Debbie Musser  
Ines Teal  
Jamaal Duty

### 3 Experimentos

O objetivo deste laboratório é implementar e testar uma tabela Hash com tamanhos variáveis. Para testar sua implementação, iremos usar os arquivos de entrada com valores variáveis de tamanho de tabela: 503, 2503, 5003 e 7507. Para cada um destes tamanhos de tabela, deverá ser realizada um experimento que consiste em inserir todos os 10,000 nomes do arquivo *nomes\_10000.txt* na tabela Hash, e depois realizar as consultas descritas no arquivo *consultas.txt*.

Para cada experimento, gere um arquivo de nome *experimentoM.txt* (onde M possui os valores 503, 2503, 5003 e 7507) contendo estatísticas sobre a tabela hash construída e as consultas realizadas.

Para a tabela hash construída, as seguintes informações precisam ser retornadas: número de entradas não vazias, número mínimo de lista, número máximo de lista, média de tamanho de lista (para entradas não vazias).

Para as consultas, para cada nome deve se retornar HIT e o número de testes para encontrar o valor, ou simplesmente MISS se ele não for encontrado. Ao final, retornar o número mínimo, máximo e médio de testes para os nomes encontrados na tabela.

O arquivo de saída deve ter o seguinte formato:

```
PARTE1: ESTATISTICAS DA TABELA HASH
NUMERO DE ENTRADAS DA TABELA USADAS #NUMERO1
NÚMERO DE ENTRADAS DA TABELA VAZIAS #NUMERO2
TAXA DE OCUPACAO #NUMERO1/TAMANHODATABELA
MINIMO TAMANHO DE LISTA #MIN1
MAXIMO TAMANHO DE LISTA #MAX1
MEDIO TAMANHO DE LISTA #MED1

PARTE 2: ESTATISTICAS DAS CONSULTAS
NOME01 HIT #CONSULTAS1
NOME02 MISS
NOME03 MISS
NOME04 HIT #CONSULTAS4
...
NOME50 HIT #CONSULTAS3
MÍNIMO NUMERO DE TESTES POR NOME ENCONTRADO #MIN2
MAXIMO NUMERO DE TESTES POR NOME ENCONTRADO #MAX2
MEDIA NUMERO DE TESTES NOME ENCONTRADO #MED2
MEDIA #MÉDIA DAS CONSULTAS
```

### 4 Entrega

A solução deve ser enviada pelo Moodle dentro de um arquivo .zip, contendo os seguintes arquivos:

- **integrantes.txt**: coloque o nome dos integrantes do grupo (até 2 pessoas) , com um nome por linha
- **readme.txt**: explique a função de hash e a implementação do método de resolução de conflitos;
- **experimento503.txt**: estatísticas do experimento com tamanho de tabela 503
- **experimento2503.txt**: estatísticas do experimento com tamanho de tabela 2503
- **experimento5003.txt**: estatísticas do experimento com tamanho de tabela 5003
- **experimento7507.txt**: estatísticas do experimento com tamanho de tabela 7507
- código fonte correspondente a solução