

# Filtro de vídeos - Projeto de MPEI

Francisco Xavier Santos Petronilho - 89241 , José Pedro Silva - 89195

1

## 1. Módulos

- Counting Bloom Filter
- MinHash
- Contador Estocástico
- Classes teste
- Class main

## 2. Execução da aplicação

A aplicação executa-se a partir do ficheiro ProjetoMain.java. Este faz uso de todas as classes que compõe o programa.

Os outros ficheiros executaveis (que contêm funções main) correspondem aos testes dos módulos desenvolvidos e são inicializados com valores teóricos semelhantes aos usados no programa main.

## 3. Descrição dos módulos

### 3.1. Counting Bloom Filter

O bloom filter é composto pelas operações addElem, contains, remove, count, genSeeds e getInds. Usamos um array de inteiros para representar os vários índices do filter.

Para introduzir objetos no filtro é passada como argumento uma string que resulta da concatenação do id do vídeo e o hashcode do título do vídeo.

É usada a função getInds que passa o objeto que queremos introduzir pelas várias hash functions e retorna um array com os índices desse objeto no filter. Esta função é usada para adicionar, remove e verificar se o objeto se encontra no filtro.

Depois serem determinados os índices do filtro, as funções addElem e remove incrementam/decrementam os inteiros do filtro que se encontram nos índices determinados anteriormente. A função contains verifica se o valor encontrado nesses índices é diferente de zero, retornando true nesse caso. Esta função pode levar a falsos positivos caso haja colisões dos índices dos elementos, que leva a que estes sejam a um inteiro maior que zero mesmo tendo sido removido.

```
Percentagem de falsos negativos: 0%  
Percentagem de falsos positivos: 36.6578%  
  
Número de elementos: 1000000
```

Figura 1. Teste BloomFilter

### 3.2. Contador Estocástico

O contador estocástico tem apenas uma função que quando chamada tem a probabilidade de 0.5 de incrementar por dois uma variável estática.

```
ContadorEstocástico c= new ContadorEstocástico();  
for (int i = 0; i < 1000000; i++) {  
    c.counter();  
}  
out.println("Número de elementos: " +c.getCount());
```

Figura 2. Teste do Contador

```
Número de elementos: 1000196
```

Figura 3. Resultados do Teste do Contador

### 3.3. MinHash

A MinHash recebe dois Sets de strings e o número de hash functions a ser utilizadas. Estes Sets passam pelas hash functions e é guardado o valor mínimo dos hashes obtidos.

Este módulo é usado na procura de vídeos através de uma ou mais palavras-chave e também na identificação de vídeos com o mesmo conjunto de categorias.

```
Set<String> set1= new HashSet<>();  
set1.add("aa"); set1.add("bb"); set1.add("cc"); set1.add("dd");  
  
Set<String> set2= new HashSet<>();  
set2.add("aa"); set2.add("bb"); set2.add("cc"); set2.add("dd");  
  
Set<String> set3= new HashSet<>();  
set3.add("aa"); set3.add("bb"); set3.add("cc"); set3.add("xd");  
  
MinHash m= new MinHash();  
System.out.println("Similaridade set1 e set2: "+m.similarity(set1,set2, numHash: 20));  
System.out.println("Similaridade set1 e set3: "+m.similarity(set1,set3, numHash: 20));
```

Figura 4. Teste da MinHash

```
Similaridade set1 e set2: 1.0  
Similaridade set1 e set3: 0.35
```

Figura 5. Resultados do Teste da MinHash

#### 4. Aplicação

A aplicação inicialmente dá a opção de procurar vídeos, listar os vídeos ou os canais que estão em ficheiros de texto que foram gerados automaticamente através de um programa desenvolvido para termos uma grande quantidade de vídeos na aplicação.

```
Numero Videos: 1010276
1-Pesquisar por nome.
2-Listar Videos.
3-Listar Canais.
4-Sair.
```

Na opção de procurar vídeos é pedido para inserir um título, aqui pode se adicionar uma ou mais palavras.

```
Insira um titulo:
asd
```

De seguida irá aparecer na consola todos os vídeos encontrados e estes estão ordenados segundo o número de visualizações de cada vídeo.

```
Visualizacoes: 904023 Nome: rrtxxy asd Canal: ajaavxjqbg Categorias: Comedia Animais
Visualizacoes: 805907 Nome: asd ycliw qpi Canal: cocjugmwmk Categorias: Tecnologia Entretenimento
Visualizacoes: 727539 Nome: asd vlcbvsn kwtsxyh Canal: fpxuoarvj Categorias: Acao
Visualizacoes: 581582 Nome: n xiiate uwhogru asd Canal: obmvk Categorias: Desporto Musica Tecnologia
Visualizacoes: 444923 Nome: sx asd hlga Canal: lhcjt Categorias: Tecnologia Moda
Visualizacoes: 345650 Nome: nhwmjivgmg kxmf asd Canal: xtlk Categorias: Videojogos Animais
Visualizacoes: 339604 Nome: asd xtwavq Canal: xunug Categorias: Politica
Visualizacoes: 317113 Nome: mkbarx avcfduz asd sjeuisyhnj k Canal: arwhdl Categorias: Musica
Visualizacoes: 120040 Nome: asd iujjpkddhp Canal: axxgl Categorias: Entretenimento Animais Comedia
```

A seguir à lista ser apresentada surge a opção de filtrar os vídeos por categoria e as categorias existentes.

```
Filtrar por:
1-Categorias.
2-Sair.

1
1-Comedia.
2-Acao.
3-Videojogos.
4-Entretenimento.
5-Animais.
6-Desporto.
7-Musica.
8-Politica.
9-Automoveis.
10-Moda.
11-Tecnologia.
12-Sair.
```

Visualizacoes: 727539 Nome: asd vlcbsvn kwtsxyh Canal: fpxuoarvjj Categorias: Acao

O contador estocástico conta o número de vídeos na aplicação, este número é apresentado assim que o programa termina de ler os ficheiros. O BloomFilter armazena o id dos vídeos e o seu nome. A MinHash é utilizada na procura de vídeos por nome e, também quando se filtra os vídeos por categoria.

A criação do gerador de vídeos(gera aleatoriamente o nome de um vídeo, as suas categorias e o número de visualizações) ajudou-nos bastante nos testes devido ao grande número de dados que tínhamos para usar, sendo por isso uma grande vantagem da nossa aplicação apesar de não estar relacionada com os objetivos do trabalho.

## 5. Conclusão

Inicialmente a aplicação tinha como objetivo ver o histórico de um utilizador e sugerir vídeos consoante o seu histórico, devido a várias complicações não conseguimos atingir esse objetivo. Como não atingimos os nossos objetivos a aplicação acabou por ficar muito limitada e simples.

Francisco Petronilho-50%

Desenvolvimento da aplicação e MinHash.

José Silva(89195)-50%

Desenvolvimento do BloomFilter e Contador Estocástico.