Relatório do Trabalho de MPEI

**Classe Bloom**

Nesta classe o construtor inicia o counting bloom filter com tamanho passado pelo utilizador e com o número de hash functions que serão utilizadas.

A classe possui 5 métodos principais:

* **string2hash()**: responsável pela conversão de uma string para um inteiro. Neste método foi usada a função djb2.
* **insert()**: este método, conjuntamente com o método string2hash(), é responsável pela inserção de elementos no counting bloom filter.
* **delete()**: contrariamente ao método insert(), o método delete() tem como função eliminar um elemento do counting bloom filter.
* **count()**: o método count() devolve o número de vezes que um elemento foi inserido no counting bloom filter.
* **check()**: responsável pela verificação se um elemento se encontra no counting bloom filter.

**Classe MinHash**

A classe MinHash possui 5 métodos principais:

* **iniciarHash()**: este método é chamado para que sejam gerados valores aleatórios de *a* e *b*  que serão mais tarde utilizados na *universal hash function*.
* **string2hash()**: responsável pela conversão de uma string para um inteiro. Neste método foi usada uma *universal hash function*.
* **minhashing()**: este método, conjuntamente com o método string2hash(), é responsável pela inserção num array bidimensional dos valores de hash mínimos correspondentes a uma key.
* **jaccard()**: responsável pela análise do array bidimensional devolvido pelo método minhashing() e pelo cálculo das distâncias de jaccard entre cada valor desse array.
* **similar()**: este método analisa todas as distâncias de jaccard calculadas anteriormente e devolve os elementos similares (considerou-se como elementos similares elementos com uma distância de jaccard inferior a 0,15).

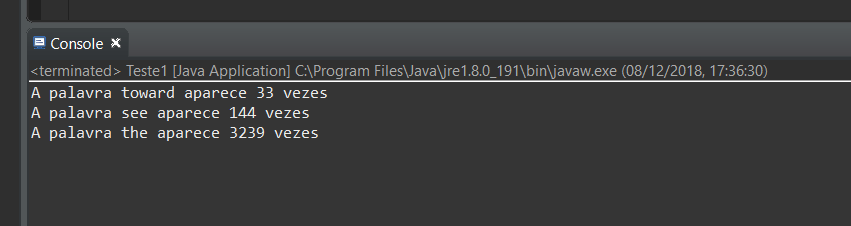
**Classe Contador**

A classe Contador faz uso de um contador estocástico para fazer a contagem de elementos inseridos no counting bloom filter.

**Testes**

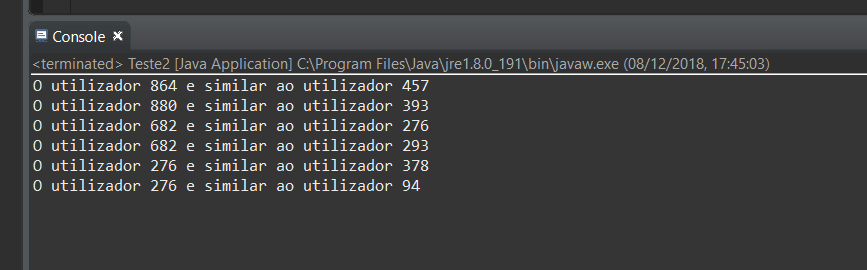
**Classe Teste1**

O teste ao counting bloom filter encontra-se na classe Teste1. Nesta classe é lido o ficheiro gutemberg.txt e são inseridas no counting bloom filter todas as palavras presentes nesse documento. Após a inserção é feito o teste para as palavras “toward”, “see” e “the”. A imagem seguinte mostra o resultado do teste.



**Classe Teste2**

O teste à classe Minhash encontra-se na classe Teste2. Nesta classe é lido o ficheiro u.data e é feita a comparação entre vários utilizadores que viram diferentes filmes. A imagem seguinte mostra o resultado do teste.



**Classe Main**

Para além das classes Teste1 e Teste2 existe a classe Main que faz o teste aos 3 módulos. Esta classe lê o ficheiro players.txt, que é constituído pelo nome de um jogador, a sua posição e equipa, e vai inserir linha a linha o nome do jogador ao counting bloom filter. Ao mesmo tempo é chamado o método Increment() da Classe contador para que este faça o incremento da variável count. No final temos uma estimativa de o número de jogadores que foram inseridos no counting bloom filter.

Durante os testes o utilizador pode verificar a existência de um jogador no counting bloom filter e também verificar os jogadores que têm a mesma equipa e posição.

As imagens seguintes mostram os resultados dos testes.

