

MOMENTO DE AVALIAÇÃO

Curso(s): Licenciatura em Informática
Unidade Curricular: Estrutura de Dados e Algoritmos
Avaliação Continuada
Docente: André Monteiro

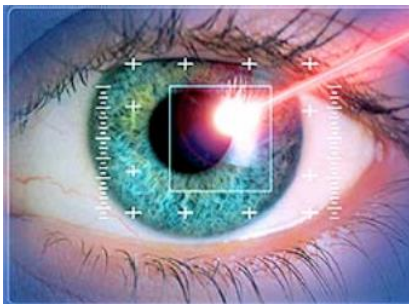
Duração: Até 07-01-2017

Data: 22-11-2016

Observações: O objetivo do projeto final é desenvolver e aplicar algumas das técnicas de conceção e análise de algoritmos abordadas nas aulas aplicadas à resolução de um problema real de biométrica.

CONTEXTO

Os sistemas biométricos visam identificar univocamente um determinado indivíduo com base na sua assinatura biométrica (íris, impressão digital, face,...) Uma assinatura biométrica consiste num conjunto de valores numéricos (características) que são únicos para cada indivíduo.



Suponha que uma polícia forense internacional guarda as assinaturas biométricas de determinado indivíduos considerados perigosos na base de dados DANGER_DB. O objetivo é que de cada vez que se registam os dados biométricos de alguém, seja gerada uma assinatura biométrica e comparada com as assinaturas existentes na DANGER_GB, sendo devolvida uma lista dos “n” potenciais indivíduos com assinatura mais parecida aos dados registados, isto é, que correspondem às identidades mais prováveis.

O trabalho, a implementar em linguagem Java, consiste na leitura do ficheiro de texto com a informação dos indivíduos e assinaturas registados na DANGER_DB, carregamento em memória, pesquisa dos elementos mais parecidos a uma assinatura e devolução dos “top-n matches”:

- n=1, top-1: o mais parecido
- n=2, top-2: os dois mais parecidos

Para este trabalho vamos supor que as características registadas são valores binários (0's e 1's) apenas. Exemplo: 0,1,1,0,0,1,1,1,0,0,1,1,0,1,0,0,1,0,1,1,0...

Exemplo de implementação:

```
//////////////////////////////////////////
// Programa de gestão biométrica      //
// 1 - Ler ficheiro                    //
// 2 - Pesquisar assinatura algoritmo1 //
// 3 - Pesquisar assinatura algoritmo2 //
// 4 - Ordenar sequência algoritmo1   //
// 5 - Ordenar sequência algoritmo2   //
// 0 - Sair                           //
// Opção?                             //
> 2
> Indique a assinatura biométrica (ou parte) a pesquisar:
> 1010010100101
> Introduza o top-n desejado:
> 3
> As identidades mais prováveis são:
BI: 12345678, Distancia 3
BI: 11111222, Distancia 7
BI: 12120303, Distancia 9
Demorou 3600ms
```

PROJETO FINAL

Dentro deste contexto, o objetivo do projeto final é utilizar algoritmos que pesquisem padrões específicos dentro de assinaturas biométricas e forneçam o BI correspondente e os n-elementos mais parecidos da lista. Também é necessária a ordenação, uma experiência puramente acadêmica pois a ordenação corrompe a assinatura biométrica.

O projeto final será composto de três partes:

1. Elaboração de uma proposta de projeto com uma breve descrição dos algoritmos utilizados;
2. Implementação, teste e validação dos algoritmos;
3. Redação de um relatório sucinto que descreva o problema, algoritmos usados, análise dos algoritmos usados e resultados experimentais incluindo o tempo de execução e gráficos.

Como o propósito principal deste projeto é verificar a eficiência de algoritmos, deverão ser implementados pelo menos dois métodos, sendo um considerado "menos bom", ou seja, com um tempo de execução não adequado à magnitude do problema tratado e um outro método que forneça "o melhor" tempo de execução possível.

DADOS

Para o desenvolvimento do trabalho, são fornecidos 2 ficheiros de teste que contêm diferentes sequências de pares BI/Assinatura:

- DANGER_DB_SMALL.txt
 - Linhas= 1001
- DANGER_DB_LARGE.txt
 - Linhas= 100001

Os ficheiros compactados (ZIP) estão disponíveis para download no Moodle.

Notas:

- As assinaturas biométricas em DANGER_DB têm dimensão fixa (N=2048).
- As assinaturas a procurar podem ter dimensão variável (K, $1 < K < 2048$).
- Devem ser procurados não só assinaturas completas, como excertos de assinaturas.

Ex.: Procurar "0110010".

Neste caso deverão ser devolvidos os indivíduos em DANGER_DB que contenham sequências mais próximas da pedida em qualquer parte da sua assinatura:

- 01000011001000101010... (Distância 0)
- 011001001011100101010 (Distância 0)
- 111001001011100101010 (Distância 1)

Exemplo do Ficheiro DANGER_DB.txt

- 12345678 \t 1001001101001001101001001101001001101001001101001001101001001101...
- 11004562 \t 0101010010011010010011010010011000100100110100100110110...

RECOMENDAÇÕES

O projeto final poderá ser elaborado individualmente ou em grupo. Pode utilizar e reutilizar código, algoritmos, funções, bibliotecas e classes de outras fontes desde que sejam referenciadas. Deve pesquisar um algoritmo que resolva seu problema, analisá-lo (funcionamento e tempo de execução) e então implementá-lo utilizando a linguagem de programação Java.

IMPLEMENTAÇÃO

Uma vez definidos os algoritmos a serem utilizados, deverá implementá-los utilizando a linguagem de programação Java.

O programa deverá ler uma assinatura ("01000101011") de tamanho qualquer **k** de um ficheiro cujo padrão seja similar ao dos ficheiros fornecidos e receber do teclado um **n**, cuja presença se vai verificar as assinaturas que apenas diferem em **n** elementos. Além disso, deve ter a opção de escolher entre o algoritmo "mau" e o "bom".

No ecrã de saída deve ser indicado:

1. as top-n assinaturas mais parecidas com a introduzida, na pesquisa;
2. o BI/CC correspondente à assinatura, na pesquisa;
3. tempo de execução, ambos para a pesquisa e ordenação