

## Trabalho Prático

Os conteúdos multimédia (áudio e vídeo) representam hoje em dia uma parte significativa do volume de dados da Web.

Para este trabalho vamos analisar, independentemente do tipo de multimédia (áudio ou vídeo), a informação ao nível do bit.

Exemplo:

0,1,1,0,0,1,1,1,0,0,1,1,0,0,0,1,0,1,1,0...

Suponha que uma polícia forense possui uma base de dados de conteúdos considerados inapropriados e deseja implementar um sistema computacional que procure nessa base de dados por excertos binários.



O trabalho, a implementar em linguagem ANSI C, consiste na leitura do ficheiro de texto com a informação dos conteúdos, o respetivo carregamento em memória, e a 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
- ...

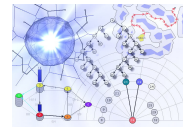
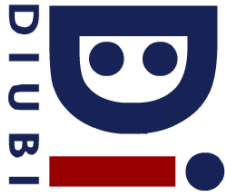
### Exemplo de funcionamento:

```

*****
MENU
*****
1:  Leitura de ficheiro
2:  Pesquisa de conteúdo
*****

```

### Exemplo de Pesquisa:



- > Introduza o valor de "n" desejado
- > 3
- > Introduza o conteúdo a procurar
- > 1010010100101

As conteúdos mais parecidos são:

ID: 12345678, Distancia 3  
ID: 11111222, Distancia 7  
ID: 12120303, Distancia 9

### Notas:

- Os segmentos disponíveis na base de dados têm dimensão fixa (N=2048).
- Os segmentos a procurar podem ter dimensão variável (K,  $1 < K < 2048$ ).
- Exemplo: Procurar "0110010".
  - Neste caso deverão ser devolvidos os conteúdos na base de dados que contenham sequências mais próximas da pedida em qualquer parte da sua assinatura:
    - ID1: 01000**0110010**00101010... (Distância 0)
    - ID2: **0110010**01011100101010 (Distância 0)
    - ID3: **1110010**01011100101010 (Distância 1)

### Exemplo do Ficheiro MEDIA.txt

```
12345678 \t 100100110...  
11004562 \t 0101000110...  
...
```