**7.a**

Os seguintes valores foram calculados usando uma versão modificada do programa do exercício 1

|  |  |  |
| --- | --- | --- |
| Ficheiro | Nº médio de bits/símbolo do ficheiro inteiro | Nº médio de bits/símbolo mínimo necessário numa subsequência de 128 bits |
| cna21\_1f\_resultados.xls | 4,98 | 3,5 |
| cna21\_1f\_resultados.xlsx | 7,86 | 6,5 |
| FPessoa.bmp | 6,96 | 3,5 - 6,5 |
| FPessoa.jpg | 7,97 | 6,5 |
| FPessoa.png | 7,99 | 6,5 |
| FPessoa-g.bmp | 6,47 | 0,0 - 5,5 |
| gota.mp4 | 7,90 | 6,5 |
| gota.wav | 7,39 | 6,0 |
| gota8bm.wav | 5,95 | 5,0 |
| kodim10.bmp | 7,20 | 5,0 |
| kodim10.png | 7,98 | 6,5 |
| lenag.bmp | 7,46 | 6,0 |
| lz4io.c | 5,01 | 4,5 |
| M1F1-Alaw-AFsp.wav | 6,20 | 2,5 - 6,0 |
| rfc7932.html | 4,70 | 3,5 |
| 23961-8.txt | 5,24 | 4,5 |
| 28779-0.txt | 4,83 | 4,5 |
| O Ano da Morte de Ricardo Reis.srt | 5,13 | 4,5 |
| rfc7932.txt | 4,50 | 3,5 |

Com este método de geração de mensagens, a mensagem terá sempre um nº médio de bits mínimo necessário para codificar cada símbolo menor ou igual ao do ficheiro utilizado, isto porque está-se a retirar uma subsequência de símbolos de uma sequência maior. Esta subsequência tem menos símbolos e na maioria dos casos menos complexidade, oque resulta em menos entropia/informação.

No entanto há certos ficheiros em que uma subsequência aleatória tem uma alta probabilidade de ter uma entropia próxima da entropia inteira. Outros em que a entropia varia muito dependendo de que parte do ficheiro a sequencia foi retirada, podendo ser muito alta ou baixa. E também há ficheiros onde a entropia de uma subsequência é quase sempre muito menor que a original, sem contar com certos outliers.

Nos ficheiros “FPessoa.bmp”, “FPessoa-g.bmp” os valores têm uma alta variância, podendo às vezes chegar perto do valor original. Nos ficheiros com extensão .bmp (bitmap) isto é devido a maneira como bitmaps guardam a informação de uma imagem, que cria zonas separadas algumas com muita complexidade e outras com menos.

Nos ficheiros “FPessoa.jpg”, “FPessoa.png”, “gota.mp4”, “gota.wav”, “gota8bm.wav”, “kodim10.png”, “23961-8.txt”, “28779-0.txt” e “O Ano da Morte de Ricardo Reis.srt” os valores tem muito menos variância.

A diferença entre .png (portable network graphics) e .jpg (joint photographic experts group) é que PNG’s são *lossless* e JPEG’s são *lossy*. Neste caso não se nota discrepância no valores entre “FPessoa.jpg” e “FPessoa.png”, apesar de a PNG ter o dobro do tamanho, apesar de JPEG ser um formato *lossy* a complexidade do ficheiro não altera muito.

“gota.mp4”, “gota.wav”, “gota8bm.wav” mostram valores muito parecidos, no entanto “M1F1-Alaw-AFsp.wav” tem uma grande disparidade de complexidades apesar de também ser um ficheiro .wav (waveform áudio file format). Ficheiros .wav são *lossless*, ou seja não tem perdas de informação, a informação e guardada como ondas de som, por isso áudio com varias mudanças de tom e sons sobrepostos tem muitas zonas com entropias diferentes.

Em ficheiros .txt não há grande discrepância de complexidade entre subsequências, isto porque linguagem humana tem símbolos e sequencias de símbolos muito frequentes. Nota-se que o texto grego mostra menos complexidade por inteiro que o inglês, e quando se retira um excerto a complexidade não muda muito no grego. Conclui-se que grego tem menos disparidade, e mais símbolos frequentes que o inglês.

O resto dos ficheiros também mostra sempre entropias semelhantes entre subsequência sem contar com alguns outliers, que provavelmente vêm de partes específicas do ficheiro, como o prefacio e posfácio.