

## Práctica 2 (Medida de la información)

1. El libro *Estudio lexicométrico del diario "El País"* tomó como muestra los ejemplares publicados durante una semana (526159 letras en total) para obtener la siguiente tabla de frecuencias de las letras en castellano:

altas		medias		bajas			
E	16.78	R	4.94	Y	1.54	J	0.30
A	11.96	U	4.80	Q	1.53	Ñ	0.29
O	8.69	I	4.15	B	0.92	Z	0.15
L	8.37	T	3.31	H	0.89	X	0.06
S	7.88	C	2.92	G	0.73	K	0.00
N	7.01	P	2.76	F	0.52	W	0.00
D	6.87	M	2.12	V	0.39		

Suponiendo que el alfabeto castellano, como fuente de información, no tiene memoria:

- (a) Calcular la cantidad de información (en bits) que contiene la palabra **codificación**.
  - (b) Calcular la cantidad de información (en bits) que contiene la palabra **informática**.
  - (c) Implementar un programa que, basándose en la tabla anterior, calcule la cantidad de información (en bits) de una palabra dada como input. <sup>1</sup>
2. Supongamos que una imagen de televisión está formada por un conjunto de puntos, dispuestos en 576 filas y 720 columnas, y que además cada punto puede tomar 256 colores con igual probabilidad. Calcular la cantidad de información o entropía (en bits/símbolo) de una imagen.
  3. Supongamos que un tertuliano de televisión tiene un vocabulario de 10 000 palabras que usa al azar con igual probabilidad. Calcular la cantidad de información o entropía (en bits/símbolo) del tertuliano.
  4. Decidir qué contiene más información, una imagen o mil palabras:
    - (a) En los supuestos anteriores.
    - (b) Suponiendo que el tertuliano es muy culto y conoce las 88 431 palabras de la 22<sup>a</sup> edición del diccionario de la R.A.E.

---

<sup>1</sup>El programa se presentará en un fichero ejecutable tal que en el inicio de la ejecución se muestren los datos del autor y la asignatura. A continuación, el programa pedirá una palabra y devolverá su cantidad de información.