

## Práctica 1: Fuente de Información Asociada a un Texto

21 de febrero - 22 de febrero - 23 de febrero

**AVISO:** Las soluciones de la práctica deben darse con números decimales de cuatro cifras decimales (aproximaciones por redondeo).

En el capítulo 1 del libro “PLATERO Y YO” de Juan Ramón Jiménez podemos encontrar el siguiente texto:

La noche cae, brumosa ya y morada.  
Vagas claridades malvas y verdes perduran tras la torre de la iglesia.  
El camino sube, lleno de sombras, de campanillas, de fragancia de hierba,  
de canciones, de cansancio y de anhelo.

*(texto disponible en el fichero datos\_1, dentro de la carpeta práctica 1 de la moodle)*

Usando como referencia los símbolos que aparecen en el texto anterior (diferenciando mayúsculas de minúsculas y tomando el cambio de línea como **dos espacios**) y el número de veces que aparece cada símbolo, calcula:

1. La fuente de información  $\mathcal{F}$  asociada al texto y apunta la frecuencia absoluta y la probabilidad para el símbolo “d”.

frecuencia	probabilidad
13	0.0586

2. Calcula la entropía en bits de la fuente de información  $\mathcal{F}$ .

entropía en bits de $\mathcal{F}$
4.0347

3. Ordena la fuente de información de mayor a menor probabilidad, indicando las cuatro primeras entradas de la fuente ordenada (tanto el símbolo, como la frecuencia absoluta, como la probabilidad).

fuente ordenada

símbolo	frecuencia	probabilidad
□	40	0.1802
a	30	0.1351
e	21	0.0946
s	13	0.0586

**Extra:** Supongamos ahora que queremos obtener la fuente asociada al texto anterior, considerando como símbolos de la fuente las secuencias de longitud dos que van apareciendo en dicho texto. En nuestro caso: “La”, “ n”, “oc” y así sucesivamente. Ordena esta fuente de información de mayor a menor probabilidad, indicando las cuatro primeras entradas de la fuente ordenada (tanto la secuencia, como la frecuencia absoluta, como la probabilidad).

fuente extra ordenada

símbolo	frecuencia	probabilidad
,□	5	0.0450
de	5	0.0450
e□	4	0.0360
□c	3	0.0270

Calcula la entropía en bits de la fuente de información considerada.

entropía en bits

6.0107