

PRÁCTICA 1 - GEOMETRÍA COMPUTACIONAL - 2023

25 de enero – 8 de febrero

CÓDIGO HUFFMAN Y PRIMER TEOREMA DE SHANNON

Sean $S_{Eng} = \{a, b, c, \dots, Z, !, ?, -, /, n, , 0, \dots, 9\}$ y $S_{Esp} = \{a, á, b, c, \dots, ñ, \dots, Z, ¡, ¡, ?, -, /, n, , 0, \dots, 9\}$ los alfabetos del inglés y el español, respectivamente. Consideramos una pequeña muestra de cada idioma en los siguientes archivos:

- “GCOM2023_pract1_auxiliar_eng.txt” con la muestra en inglés.
- “GCOM2023_pract1_auxiliar_esp.txt” con la muestra en español.

Suponemos que las mayúsculas y minúsculas son distintos estados de las variables. Del mismo modo, las vocales con y sin tilde también se consideran estados distintos.

i) A partir de las muestras dadas, hallar el código Huffman binario de S_{Eng} y S_{Esp} , y sus longitudes medias $L(S_{Eng})$ y $L(S_{Esp})$. Comprobar que se satisface el Primer Teorema de Shannon. (1.50 puntos)

ii) Utilizando los códigos obtenidos en el apartado anterior, codificar la palabra cognada $X = \text{“dimension”}$ para ambas lenguas. Comprobar la eficiencia de longitud frente al código binario usual. (0.50 puntos)

iii) Decodificar la siguiente palabra del inglés:

“010101000110011100110111100010111110101010001110”. (0.50 puntos)

Ayuda para quien lo necesite (no es obligatorio):

Utiliza la plantilla “GCOM2023-practical1_plantilla.py”, especialmente para el último apartado.

Observaciones:

La memoria debe entregarse antes de que transcurran **21 días** desde el inicio de la práctica, salvo que se indique lo contrario.

La memoria, siempre en **pdf**, debe incluir **al menos** la siguiente información: (1) Introducción (motivación/objetivo de la práctica), (2) Material usado (método y datos), (3) Resultados, (4) Conclusión y (5) Anexo con el script/código utilizado.

La extensión máxima de la memoria **no superará las 2 páginas**, sin contar el código anexo (ilimitado). El total de la superficie de las figuras/tablas (si las hubiese) no podrán excederse del 50 % de la memoria (cuerpo principal).

Además, el **código fuente debe incluirse** en un archivo ‘.py’ independiente. **No** subir archivos **comprimidos**.