

PRÁCTICA 2 - GEOMETRÍA COMPUTACIONAL - 2022

9 y 16 de febrero de 2022

CÓDIGO HUFFMAN

Sean $S_{Eng} = \{a, b, c, \dots, Z, !, ?, -, /, n, , 0, \dots, 9\}$ y $S_{Esp} = \{a, á, b, c, \dots, ñ, \dots, Z, i, !, ¿, ?, -, /, n, , 0, \dots, 9\}$, dos variables aleatorias, hemos tomado una pequeña muestra de cada población y las disponemos en los archivos “GCOM2022_pract2_auxiliar_eng.txt” y “GCOM2022_pract2_auxiliar_esp.txt”, respectivamente. Si suponemos que las mayúsculas y minúsculas son distintos estados de las variables, e incluso con o sin tilde, se pide:

- i) Hallar el código Huffman binario de S_{Eng} y S_{Esp} , sus longitudes medias $L(S_{Eng})$ y $L(S_{Esp})$, y comprobar que se satisface el Primer Teorema de Shannon. (0.50 puntos):
- ii) Codificar con dicho código la palabra cognada $X = \text{“medieval”}$ para ambas lenguas, y comprobar la eficiencia de longitud comparada con el código binario usual. (0.25 puntos)
- iii) Decodifica la siguiente palabra del inglés (0.25 puntos):
“10111101101110110111011111”

Ayuda para quien lo necesite (no es obligatorio):

- Utiliza la plantilla “GCOM2022-Practica2_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).

El **código fuente debe incluirse** como archivo ‘.py’ independiente. **No** subir archivos comprimidos.