

PRÁCTICA 1 - GEOMETRÍA COMPUTACIONAL - 2024

24 de enero – 7 de febrero

CÓDIGO HUFFMAN Y PRIMER TEOREMA DE SHANNON

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

- “GCOM2024_pract1_auxiliar_eng.txt” con la muestra en inglés.
- “GCOM2024_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, así como los espacios y otros signos.

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 *Lorentz* para ambas lenguas. Comprobar la eficiencia de longitud frente al código binario usual. (0.50 puntos)

iii) Realiza un programa para **decodificar** cualquier palabra y comprueba que funciona con el resultado del apartado anterior para ambos idiomas

Ayuda SÓLO para quien lo necesite (¡no es obligatorio!):

Ver plantilla “GCOM2024-practical1_plantilla.py” (puede contener líneas no convenientes),

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** (no en *zip*), 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**.