

CÓDIGO HUFFMAN Y PRIMER TEOREMA DE SHANNON 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.
- 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.
- iii) Realiza un programa para decodificar cualquier palabra y comprueba que funciona con el resultado del apartado anterior para ambos idiomas.