

Código Huffman y Primer Teorema de Shannon

Consideramos los alfabetos del inglés y el español, respectivamente. Utilizaremos una pequeña muestra de cada idioma en los siguientes archivos:

- `pract1_auxiliar_eng.txt` con la muestra en inglés.
- `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.

1. A partir de las muestras dadas, hallar el código Huffman binario de S_{Eng} y S_{Esp} , y sus longitudes medias $L(S_{\text{Eng}})$ y $L(S_{\text{Esp}})$. Comprobar que se satisface el Primer Teorema de Shannon.
2. 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.
3. Realizar un programa para decodificar cualquier palabra y comprobar que funciona con el resultado del apartado anterior para ambos idiomas.