

### Práctica 3. Entropía del Español como fuente de memoria nula.

#### Objetivo.

Esta práctica tiene por objeto estimar la probabilidad de las letras del alfabeto en el idioma español y la cantidad información media que cada una de estas letras proporciona. Dicha estimación posibilita la realización de códigos más eficientes aplicados a los textos escritos en este idioma.

#### Realización

El paso previo necesario para la realización de esta práctica consiste en la obtención de un fichero de texto escrito en español de longitud suficiente para asegurar una buena estimación de las probabilidades de las distintas letras en nuestro idioma.

Debe realizarse un programa, en lenguaje C (o C++) compilable en Linux, que lea el fichero anteriormente citado y calcule las probabilidades de aparición de cada una de las letras del alfabeto más el espacio, así como la información proporcionada por cada una de ellas. Los resultados obtenidos se presentarán resumidos en una tabla, junto con las características del fichero utilizado para generar dichos resultados, como la dimensión del mismo, el número de caracteres leídos o la fuente de la que se obtuvo (un diario, un libro, revista electrónica, etc.)

El mismo programa debe calcular los siguientes parámetros que definen el idioma español como fuente de memoria nula: su entropía, la entropía máxima, la redundancia y la eficiencia.

### Práctica 4. Compresión de un texto en español utilizando codificación Huffman.

#### Objetivo.

Esta práctica tiene por objeto el diseño de un codificador Huffman para la compresión de ficheros de texto escritos en español. Para ello se utilizarán los resultados obtenidos en la práctica anterior.

#### Realización

El alumno debe realizar un programa que, suministradas las probabilidades de las letras del alfabeto para el idioma español obtenidas en la práctica anterior, realice un código Huffman para dicho alfabeto fuente. Asimismo se calcularán las propiedades de dicho código, como la longitud media de las palabras o su eficacia. Comparar dichas magnitudes con las que se obtendrían para un código de longitud fija.

Una vez obtenido el código se realizará un programa que codifique/decodifique (comprima/descomprima) un fichero de texto ASCII utilizando dicho código. Debe tenerse en cuenta que en un texto normal aparecerán algunos caracteres diferentes a las letras del alfabeto y el espacio, como pueden ser los signos de puntuación o los retornos de carro, que no se han considerado en la práctica anterior. Para solventar esa situación, pueden considerarse como nuevos símbolos del alfabeto fuente o asignarles alguna codificación fija que permita distinguirlos de los demás. Debe comprobarse que el fichero comprimido es menor que el original y el factor de compresión que se obtiene.

## **Práctica 5. Compresión de un texto en español utilizando codificación LZW.**

### **Objetivo.**

Esta práctica tiene por objeto la implementación de un compresor basado en el algoritmo LZW para la compresión de ficheros de texto escritos en español.

### **Realización**

El alumno debe realizar un programa que comprima/descomprima un fichero de texto en español (se recomienda utilizar los mismos ficheros empleados en la práctica 4, Codificación Huffman). Se compararán los cocientes de compresión obtenidos con los conseguidos en la práctica anterior. Se presentarán los resultados y discusiones de dicha comparación.