

# Comunicaciones Digitales

## Práctica 1

Curso 2009/2010

### 1. Introducción

El objetivo de esta práctica es realizar un programa en MATLAB que determine el código Huffman de un fichero de texto. Los caracteres posibles que pueden formar el texto son las letras del alfabeto (distinguiendo mayúsculas de minúsculas), los números del 0 al 9 y los siguientes signos de puntuación: '.,', '!', ':', ';', '?', ' '.

### 2. Cálculo del código Huffman

Se tomará un fichero de texto como entrada lo suficientemente largo para obtener resultados aceptables. Se calculará el número de veces que aparece cada carácter en el fichero de texto y, posteriormente, se calculará la probabilidad de cada carácter.

Se implementará el algoritmo de codificación de Huffman visto en clase:

Paso 1: Ordenar los símbolos según sus probabilidades.

Paso 2: Localizar los dos símbolos,  $a_1$  y  $a_2$ , con menor probabilidad,  $p_1$  y  $p_2$  ( $p_1 < p_2$ ).

Paso 3: Asignar 0 al símbolo  $a_2$  (mayor probabilidad) y 1 a  $a_1$ .

Paso 4: Crear un nuevo símbolo  $a_{12}$  con probabilidad  $p_1 + p_2$ .

Repetir los pasos 2 – 4 hasta que quede un único símbolo.

Paso 5: Recorrer el árbol en orden inverso concatenando los símbolos asignados.

Para construir el código Huffman puede ser necesario utilizar los comandos: *sort()*, *max()*, *min()* y *find()*. La salida del programa debe tener la forma  $[l, C]$ , donde:

- $l$  es un vector  $L \times 1$  cuya entrada  $i$ -ésima,  $l(i)$ , representa la longitud de la palabra código para el carácter  $i$ .
- $C$  es una matriz cuya  $i$ -ésima fila representa el código asociado al carácter  $i$ .

Una vez generado el código compare la longitud de cada palabra código con la cantidad de información teórica,  $-\log_2(p_i)$ , donde  $p_i$  es la probabilidad del  $i$ -ésimo carácter. Calcule también la eficiencia y redundancia del código generado.

### 3. Codificación de textos en general

En este apartado, utilizaremos el código Huffman obtenido en el apartado anterior para codificar distintos textos. Para ello,

1. Lea un fichero de texto distinto al utilizado en el apartado anterior y calcule la probabilidad de cada carácter.
2. Calcule cuál sería la eficiencia y la redundancia del código Huffman calculado anteriormente si se utilizase para codificar este segundo texto.
3. Compare el resultado anterior con la eficiencia y la redundancia que se obtendría codificando el texto con un código de longitud fija.