



UNIVERSIDAD
DE GRANADA

Este documento está protegido por la Ley de Propiedad Intelectual ([Real Decreto Ley 1/1996 de 12 de abril](#)).

Queda expresamente prohibido su uso o distribución sin autorización del autor.

Teoría de la Información y la Codificación

4º Grado en Ingeniería Informática

Tarea Tema 3

Compresión de datos con el método de Huffman

1. Objetivo.....	2
2. Materiales.....	2
3. Descripción de la tarea.....	2
4. Entrega de la práctica.....	3

© Prof. Manuel Pegalajar Cuéllar
Dpto. Ciencias de la Computación e I. A.
Universidad de Granada



DECSAI

**Departamento de Ciencias de la
Computación e Inteligencia Artificial**

Compresión de datos con el método de Huffman

1. Objetivo

El objetivo de la tarea consiste en que el estudiante se familiarice con los métodos de generación de códigos óptimos mediante el algoritmo de Huffman. En particular, se utilizará una aplicación de este algoritmo para compresión de datos de texto. **La tarea debe realizarse de forma individual.**

2. Materiales

Se proporciona al estudiante los siguientes materiales:

- **Video: 3.ImplementacionHuffman. Duración: 19' 43"**. Introduce una explicación sobre cómo abordar la tarea de la semana, junto con el cuaderno de entrega de la práctica.
- **Cuaderno Jupyter: Huffman.ipynb**. Contiene el esqueleto y formato de entrega de la tarea, para que sirva como base al estudiante para la realización del trabajo.
- **Fichero "quijote.txt"**, con un fragmento de "El Ingenioso Hidalgo Don Quijote de la Mancha" sin tildes, ñ's, etc., con un alfabeto formado por los símbolos "ABCDEFGHIJKLMNOPQRSTUVWXYZ.,;:".
- **Fichero "TextoAComprimir.txt"**, con un ejemplo de texto que se desea comprimir, procedente de un alfabeto como el descrito en la línea anterior.

3. Descripción de la tarea

La tarea consiste en utilizar el método de Huffman como mecanismo de compresión de ficheros de datos sin pérdida. Para ello, se deberá:

1. Partir de la base de un fichero de texto que contenga texto en castellano que represente significativamente al idioma, en el sentido de que el uso de cada símbolo del alfabeto, y de puntuación, sea significativo como para poder calcular la probabilidad de uso de cada símbolo con fiabilidad. Se restringirá al conjunto de símbolos (mensajes de la fuente) "ABCDEFGHIJKLMNOPQRSTUVWXYZ.,;:". Se añadirá un símbolo final '\0' que se usará para delimitar el fin del mensaje. Se proporciona al estudiante una función para leer un fichero de texto y devolver un diccionario con pares *{símbolo : probabilidad}*, denominada **LeerFicheroDiccionario**.
2. Crear el árbol de codificación de acuerdo a los pares *{símbolo : probabilidad}* facilitados en el punto anterior. Esta función Huffman tendrá como entrada un diccionario, y como salida un árbol de codificación binario con el método de Huffman.
3. Se proporciona una función para codificar una secuencia de texto en una cadena con símbolos '0' y '1', codificados con el árbol del código generado por la función anterior. Se permite modificar esta función para que tenga un funcionamiento correcto de acuerdo a la función Huffman implementada por el estudiante.
4. Implementar una función para decodificar una secuencia de texto partiendo desde una cadena con símbolos '0' y '1', codificados con el árbol del código generado por la función anterior. La función deberá parar de decodificar (y por tanto devolver su resultado) bien cuando se hayan procesado todos sus símbolos, o bien cuando se haya decodificado el

símbolo de fin de mensaje '\0'.

5. Realizar pruebas de codificación de ficheros de texto en ficheros binarios con la técnica de codificación de Huffman, y **realizar un análisis** de la ganancia obtenida con respecto a la compresión de datos realizada, justificando los resultados.

La entrega de la tarea tiene dos partes: Una teórica y otra práctica. Con respecto a la parte teórica, se deberá indicar:

- Algoritmo de Huffman. Diseño y ejemplo de funcionamiento. Explicar paso a paso dicho ejemplo.
- Describir un ejemplo que ilustre cómo codificar, y cómo decodificar, un texto con el árbol generado a partir del algoritmo de Huffman.

La parte práctica consiste en implementar en Python diferentes funciones que calculen lo siguiente (cuaderno Jupyter **Huffman.ipynb**):

- Una función Huffman que tenga como entrada un diccionario de símbolos de un alfabeto y la probabilidad de ocurrencia de cada símbolo en este alfabeto. La función deberá devolver un árbol de codificación generado mediante el método de Huffman.
- Dado un árbol de Huffman generado mediante la función anterior, y un símbolo (carácter válido) a codificar, la función genere una cadena de caracteres con una secuencia de 0's y 1's que codifiquen al símbolo de entrada. Se proporciona al estudiante un código que realiza esta funcionalidad, pero que se permitirá modificar para adaptarlo a la implementación del propio estudiante en caso de ser necesario.
- Dado un árbol de Huffman generado mediante la función Huffman, y una cadena de caracteres conteniendo una secuencia de 0's y 1's que representan secuencias de mensajes codificados, devuelva una secuencia de mensajes (caracteres) decodificados (sin incluir caracteres de fin de mensaje). La función dejará de decodificar bien cuando ya no queden símbolos 0's o 1's en la cadena de caracteres de entrada, o bien cuando se decodifique un símbolo de fin de mensaje.
- Realizar pruebas de las funciones implementadas para compresión de datos contenidos en ficheros de texto. Hacer uso de las funciones auxiliares facilitadas por el profesor.

4. Entrega de la práctica

Se deberá entregar un fichero cuaderno Jupyter (**Huffman.ipynb**) con:

1. Varias celdas Markdown que contengan las soluciones a los conceptos teóricos, asociadas a cada función (código Python) a implementar.
2. Las implementaciones requeridas, en celdas de código separadas.
3. Una sección o varias (celda o celdas) donde se realicen las pruebas, y que sea fácilmente modificable por el profesor. **Cada prueba deberá ser analizada y sus resultados justificados por el estudiante. Este análisis crítico será una parte crucial en la evaluación, y su inexistencia puede hacer que la valoración de la parte teórica sea de 0.**

El fichero del cuaderno Jupyter (.ipynb) con la solución deberá estar suficientemente documentado como para que cualquier persona con conocimientos de los conceptos teóricos pueda entenderlo claramente.

La tarea se valorará de 0 a 10, siendo equitativo 50% para la parte de teoría, 50% para la parte práctica de la tarea. **Deberán entregarse ambas partes para que la tarea sea evaluada.** De otra forma, la calificación será 0.

El profesor podrá realizar, en horario de clase o tutorías, **en cualquier momento del curso**, una auditoría de las tareas realizadas por cualquier estudiante, debiendo responder adecuadamente a las cuestiones que se le planteen. **Es responsabilidad del estudiante llevar un estudio diario de todo el material entregado a lo largo del curso**, La auditoría podrá variar la calificación obtenida si el estudiante no es capaz de responder adecuadamente a las consultas planteadas por el profesor.