Huf

# Descripción

Este documento es la memoria del trabajo de fin de grado sobre un compresor de textos en castellano usando el algoritmo de compresión de Huffman.

# Introducción

A partir de ahora, se describirá en el documento una introducción a qué es algoritmo de Huffman, el camino mediante el cual se llena a él y se demuestra su validez y eficiencia; también se hará una descripción del programa realizado para comprimir y descomprimir textos en castellano usándolo.

# Programa

Introducción

En el programa elaborado, podemos:

* Analizar la frecuencia de uso de caracteres en un archivo de texto en formato txt que preferiblemente use la codificación de texto ISO 8859-1[[1]](#footnote-1).
* Habiendo analizado un texto, crear un archivo diccionario para la compresión/descompresión de textos.
* Comprimir archivos txt utilizando un archivo diccionario.
* Descomprimir textos (comprimidos con el programa) utilizando un archivo diccionario.

Archivos

El programa hará uso de tres clases de archivos:

* Archivos .txt que contendrán los textos antes y después de ser comprimidos.
* Archivos .huf que contendrán el texto comprimido. Internamente un archivo .huf no es más que una cadena binaria formada a partir del diccionario y el texto.
* Archivos .dhu que contienen el árbol Huffman usado tanto para comprimir como descomprimir el texto.

Sobre el desarrollo

El programa ha sido desarrollado haciendo uso del lenguaje de programación **Java** junto al IDE[[2]](#footnote-2) Netbeans 8.1 debido a la facilidad de implementación, documentación, detección de errores, depuración de éstos y portabilidad del programa.

Además de por las características intrínsecas al lenguaje y el IDE, también Java cuenta con una comunidad extensa que proporciona muchas librerías y ayudas a los desarrolladores.

## Ayudas externas

**Owen Astrachan**: librerías bit(input/output)stream usadas para leer/escribir respectivamente archivos binarios bit a bit.

[**https://rosettacode.org/wiki/Huffman\_coding#Java**](https://rosettacode.org/wiki/Huffman_coding#Java): clase que contiene el algoritmo de huffman que ha sido modificado para que las entradas y salidas coincidan con las necesarias para el programa.

1. Norma de la ISO que define la codificación del alfabeto latino, incluyendo los diacríticos (como letras acentuadas, ñ, ç), y letras especiales (como ß, Ø), necesarios para la escritura de las siguientes lenguas originarias de Europa occidental: afrikáans, alemán, español, catalán, euskera, aragonés, asturiano, danés, escocés, feroés, finés, francés, gaélico, gallego, inglés, islandés, italiano, holandés, noruego, portugués y sueco. [↑](#footnote-ref-1)
2. Siglas en inglés de Entorno de Desarrollo Integrado [↑](#footnote-ref-2)