



UNIVERSIDAD TÉCNICA DE ORURO
FACULTAD NACIONAL DE INGENIERÍA
INGENIERÍA INFORMÁTICA



PROYECTO INF – 2610 A
“TEORIA DE LA INFORMACION”

COMPRESION Y DESCOMPRESION DE FICHEROS

Docente: Ing. Franz Chinche Imaña
Estudiante: Alex Baltazar Castellón

Oruro – Bolivia
2023

ÍNDICE GENERAL DE CONTENIDOS

1	ASPECTOS GENERALES	3
1.1	Objetivo	3
2	MARCO TEORICO.....	3
2.1	ALGORITMO DE HUFFMAN.....	3
2.2	CODIFICACION DE HUFFMAN.....	4
2.3	DISEÑO E IMPLEMENTACION	7
3	CONCLUSIONES.....	8
3.1	CONCLUSION.....	8
4	ANEXOS	9

1 ASPECTOS GENERALES

1.1 OBJETIVO

Diseñar un software (SW) que permita comprimir y descomprimir un archivo .txt aplicando codificación de Huffman.

2 MARCO TEORICO

2.1 ALGORITMO DE HUFFMAN

El algoritmo de Huffman es un algoritmo para la construcción de códigos de Huffman, desarrollado por David A. Huffman en 1952 y descrito en “A Method for the Construction of Minimum-Redundancy Codes”. Este algoritmo toma un alfabeto de n símbolos, junto con sus frecuencias de aparición asociadas, y produce un código de Huffman para ese alfabeto y esas frecuencias.

El algoritmo consiste en la creación de un árbol binario que tiene cada uno de los símbolos por hoja, y construido de tal forma que siguiéndolo desde la raíz a cada una de sus hojas se obtiene el código Huffman asociado.

1. Se crean varios árboles, uno por cada uno de los símbolos del alfabeto, consistiendo cada uno de los árboles en un nodo sin hijos, y etiquetado cada uno con su símbolo asociado y su frecuencia de aparición.
2. Se toman los dos árboles de menor frecuencia, y se unen creando un nuevo árbol. La etiqueta de la raíz será la suma de las frecuencias de las raíces de los dos árboles que se unen, y cada uno de estos árboles será un hijo del nuevo árbol. También se etiquetan las dos ramas del nuevo árbol: con un 0 la de la izquierda, y con un 1 la de la derecha.
3. Se repite el paso 2 hasta que sólo quede un árbol.

Con este árbol se puede conocer el código asociado a un símbolo, así como obtener el símbolo asociado a un determinado código.

Para obtener el código asociado a un símbolo se debe proceder del siguiente modo:

1. Comenzar con un código vacío
2. Iniciar el recorrido del árbol en la hoja asociada al símbolo
3. Comenzar un recorrido del árbol hacia arriba
4. Cada vez que se suba un nivel, añadir al código la etiqueta de la rama que se ha recorrido
5. Tras llegar a la raíz, invertir el código
6. El resultado es el código Huffman deseado

Para obtener un símbolo a partir de un código se debe hacer así:

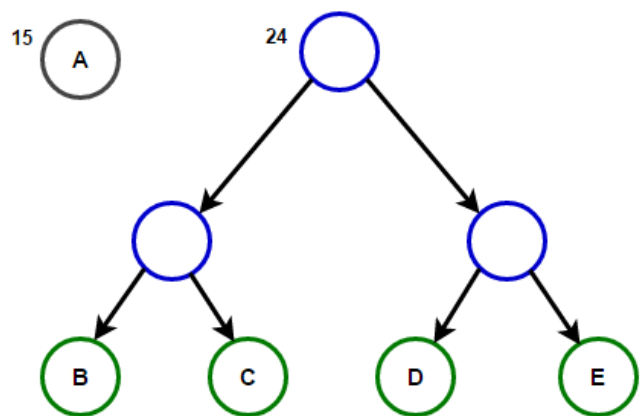
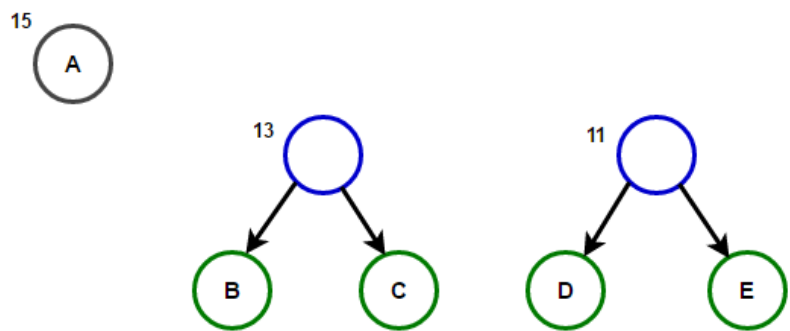
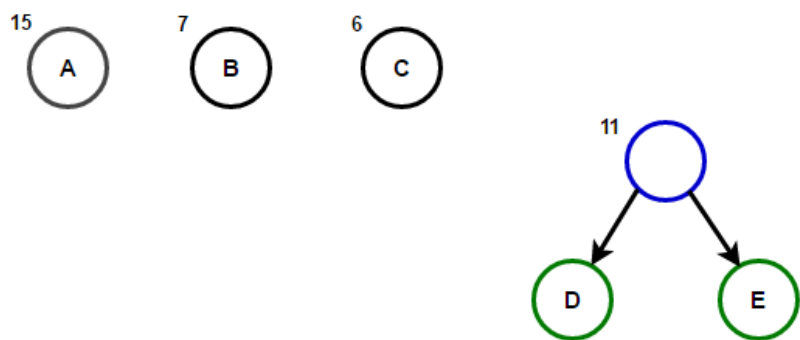
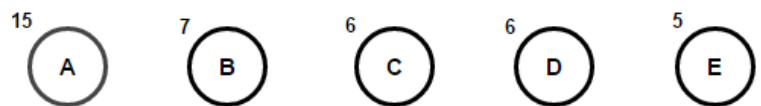
1. Comenzar el recorrido del árbol en la raíz de éste
2. Extraer el primer símbolo del código a decodificar
3. Descender por la rama etiquetada con ese símbolo
4. Volver al paso 2 hasta que se llegue a una hoja, que será el símbolo asociado al código

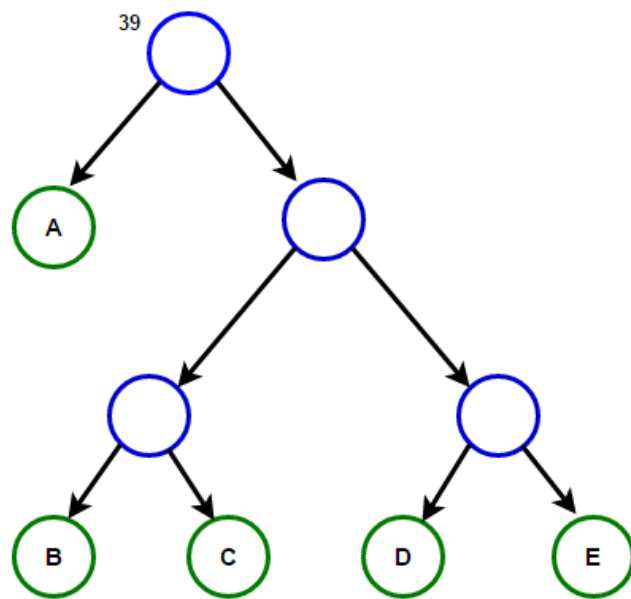
En la práctica, casi siempre se utiliza el árbol para obtener todos los códigos de una sola vez; luego se guardan en tablas y se descarta el árbol.

2.2 CODIFICACION DE HUFFMAN

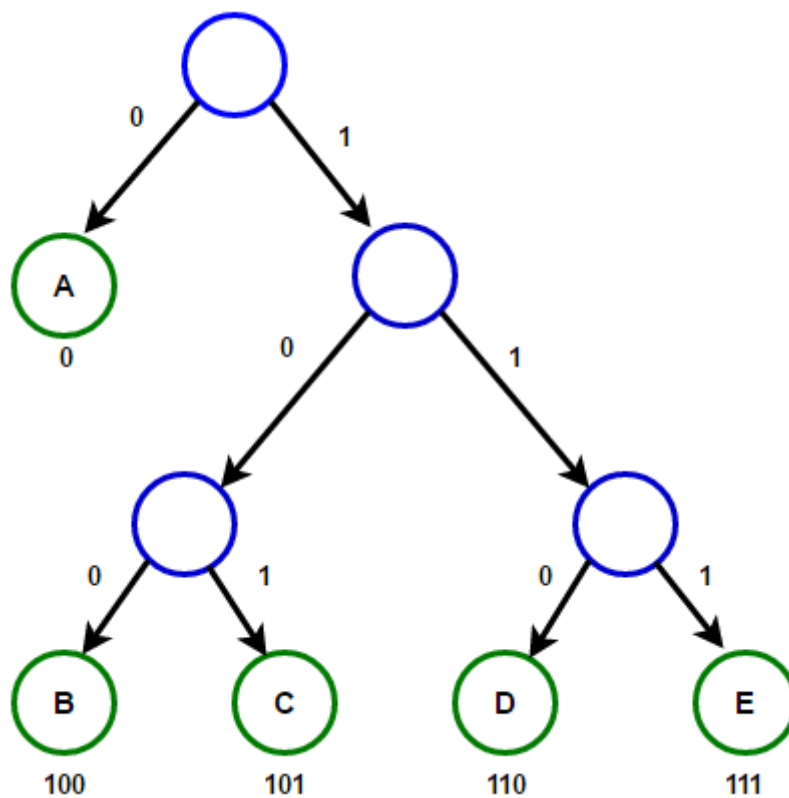
La codificación Huffman (también conocida como Codificación Huffman) es un algoritmo para realizar la compresión de datos y forma la idea básica detrás de la compresión de archivos. Esta publicación habla sobre la codificación de longitud fija y de longitud variable, los códigos decodificables únicos, las reglas de prefijo y la construcción del árbol de Huffman.

Considere un texto que consta de sólo 'A', 'B', 'C', 'D', y 'E' caracteres y sus frecuencias son 15, 7, 6, 6, 5, respectivamente. Las siguientes figuras ilustran los pasos seguidos por el algoritmo:





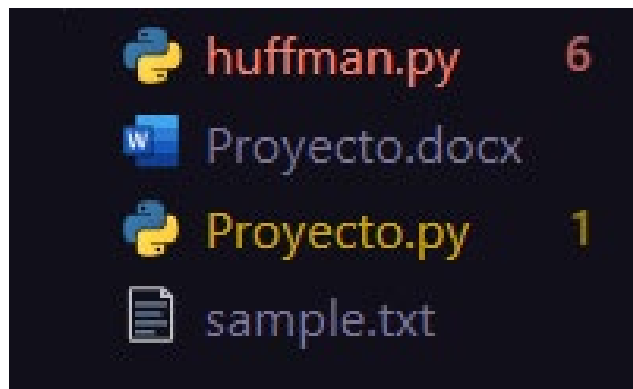
La ruta desde la raíz hasta cualquier nodo hoja almacena el código de prefijo óptimo (también llamado código Huffman) correspondiente al carácter asociado con ese nodo hoja.



2.3 DISEÑO E IMPLEMENTACION

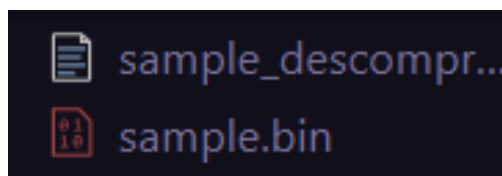
Dicho software lo implementaremos en lenguaje de programación Python, el proyecto contara con 3 archivos:

- ❖ Proyecto.py: Es donde estará el main del programa y llamara funciones.
- ❖ huffman.py: Contendra todas las clases, funciones para lograr la codificación Huffman.
- ❖ sample.txt: Sera un archivo de texto donde contiene el texto que deseamos comprimir (véase anexo2).



Al compilar y ejecutar el programa este creara dos archivos:

- ❖ sample.bin: Donde contendrá el código en binario (.bin) del texto en comprimido, lo cual el archivo estará para formato de grabar CD/DVD (véase anexo 3)
- ❖ sample_descomprimido.txt: Sera el archivo descomprimido.



3 CONCLUSIONES

3.1 CONCLUSION

Se logro diseñar con satisfacción el software (SW) el cual tiene la funcionalidad de comprimir y descomprimir ficheros, usando la codificación Huffman y aplicando el algoritmo de Colas de Montículo (heap) haciendo uso de las librerías correspondientes.

4 ANEXOS

ANEXO 1

EJECUCIÓN DEL PROGRAMA

```
PROBLEMAS 27 SALIDA CONSOLA DE DEPURACIÓN TERMINAL PUERTOS

[Done] exited with code=0 in 0.229 seconds

[Running] python -u "d:\INF 2610 - TEORIA DE LA INFORMACION\Proyecto\Proyecto.py"
Comprimido
El archivo comprimido esta en: sample.bin
Descomprimido
El archivo descomprimido esta en: sample_descomprimido.txt
|
[Done] exited with code=0 in 0.664 seconds
```

ANEXO 2

CONTENIDO DE SAMPLE.TXT

```
sample.txt
1 Los incendios forestales han dejado una huella desgarradora en Bolivia, con más de 3,3 millones de hectáreas quemadas hasta ahora.
2 Más de 3,3 millones de hectáreas de zonas boscosas, matorrales y pasturas secas se han quemado en los 160 incendios forestales registrados en
3 Bolivia en lo que va del año, según informó el ministro de Defensa, Edmundo Novillo.
4 Informe sobre las afectaciones causadas por los incendios
5 El ministro compareció ante los medios en la ciudad central de Cochabamba para dar un informe sobre las afectaciones causadas por los incendios
6 este año en el país y las acciones realizadas por las autoridades, así como la ayuda internacional recibida para combatir el fuego.
7 Cifras y daños ocasionados por los incendios
8 Detalló que hasta noviembre se han quemado 3.372.369,46 hectáreas y explicó que consideran que «los incendios han concluido», ya que no se
9 reportan más focos de calor y ha comenzado la época de lluvias en el país. De esa superficie dañada, 1,1 millones de hectáreas corresponden a
10 «cobertura arbórea» y 2,2 millones eran matorrales y pasturas secas. La región amazónica de Beni fue la más afectada con 1,9 millones de hectáreas
11 quemadas.
12 Historial histórico e impacto social
13 Según datos históricos presentados por el ministro, el año con mayores afectaciones fue 2019 con 5.305.512 hectáreas quemadas; mientras que en
14 2022 la superficie dañada fue de 4.466.540 hectáreas. Estas cifras son obtenidas mediante un control satelital a través del Sistema de Información
15 y Monitoreo de Bosques (SIMB), que verifica diariamente las áreas quemadas y proporciona informes anuales.
16 Familias damnificadas y ayuda humanitaria
17 Se estima preliminarmente que hay 18.194 familias damnificadas por los incendios y se reportaron 133 viviendas afectadas o destruidas por el
18 fuego. Durante el período de junio a noviembre, se registraron 160 incendios en todo el territorio boliviano, siendo las siete de las nueve
19 regiones del país las más afectadas. La región de Beni declaró «desastre departamental».
20 Acciones tomadas para combatir los incendios
21 El Gobierno distribuyó 225,75 toneladas de ayuda humanitaria que requirieron una inversión cercana a los 2 millones de bolivianos (más de 285.000
22 dólares). Además, se agradeció el envío de 2,77 toneladas de ayuda humanitaria desde Chile y la colaboración de bomberos forestales provenientes
23 de Venezuela y Francia.
24 El ministro destacó la participación activa de 5.605 bomberos forestales bolivianos y la movilización de 56 unidades militares para sofocar los
25 incendios. Se llevaron a cabo operaciones terrestres, aéreas y fluviales para reconocimiento y exploración, así como otras operaciones para
26 descargar agua y extinguir el fuego.
```

ANEXO 3

SAMPLE.BIN

```
sample.bin
1  ENQg+...A; EBb...ACKoFr...CB...yU+.../E...[{{ENZ,Qq...U...rACK...6^...yG...*2XP...SOH...~...yN!FSy}0C...5|
V2...i...RS...P...;gFFh[...NAK...ESC...9x...KY~J[...ESC...A&p^H|VT...w...ETS...bb\6.../]...Avr...P:iQ...G=9...VTw...7S...ACK
)|Z...S...rc...90...GS4...*...RS...;V|ENQ...NQ...DC4...SYN...US...4...R...+g)...DC2...R)...7S...cy...Y...dclN...T...Gn...Cu>...d...w...ESC
SUB...a...SYN)Z...{...|Oble...}FS...DC4]FS...9CQ...k...u+Sk...ENQ...9...Yt...9b...n...5...NAKTX...+.../...;...Su>...DEL
mTeF...2ETX...R...1...^...FS...;...SUBR?US...SI...cX...ETB...FP...DC2...v2U...r...P...5...BEL+q...~...DELhj
2  ...X...~
3  @...J...G..._...b...C...d...BEL+...BEL...RS...WDC3...c...NAK...FS...DEL...j...h...P...;w...SO^"R...NX......ESC...CANSYBEL...J7...e...K...70...SO...Y...J...5ETX...EM~X,G...
^...C...US...s...|m0...<9QM...ETB...{...n...9SI...~G...SUB...FS...SO...~...1...J...NUL...E...+SOH|[...XS...@...EOT...B...z7~...SI...[SUB...V...V...ENb
^9...FS...RS.../i...He)...~...h...n...j)...FS...>...|mRS...M...Z...ENCAN...ERACK...m...e)...X
4  ...b...dc2RS...>...[...h...~R...r.../ZrE...BEL...RV...~...0^...SI...NY...EM...EOT...r...F...BEL...STX...q.../i...a...t...SI.
5  ...BSRV...US...^...i...SYN...{4n...4%...Subly...8b...~...ENQ...:...STX...}...p...SOV...EMZ7...~...r...1...0...SI...N=...I...9[...NAK...NAK...F...S...sta...SI...SO...W...
6  ...FS...r.../Zq...;w...VBS...z...Eoth...F...D...D...k...Q...E...~...E...z...U...Q...GS...X...%...k+...DEL...NAK...4...X...or.../E;
gFFVr...X...b"G...~...~...ENCK...=...RS...DC4...q...F...WFS...f...X...<0a^6Zr...S...ESC...ETB...J
7  1...EX...DEL...V...~...~...ü:n?...X...p...GS...
8  ^2...X
9  >
10 ...SUB...+.../)...n...0...EOT...ë:...A...F7US...N...
11 ...g...~...âc...~...ACK...@...e...;gFFVr27...p...o...q...~...ETB...n...~...ENACK...W...ESC...~...ER...r.../Zq...~...~...CAN...e...~...M^2...FF...FFVOUS...SOH...SUB...UCAN...e|...5+;:c
C...^...S...~n...E09e...M^2...P...NR...~...l...j?...E...y|d...w...^...w...:EM9_
12 V25dcl...ENQY...~...8...ESC/G...E...9...DC4...O...~...90...~...n...~...J...V...RS...~...FF...~...~...ENP...l...Og...n...+...~...~...jUSI...SO...VTSONB...~...L...~...L...9z...NAK...r...y...+...3L...E
叁...:=...SYN...~...SUB...~...~...q...Q...5A...P...SO...~...~...ETB.../...T...?ACK...Zq...9...0...~...SYN...F^...~...SOH...t...BEL...7g/...~...m...^...U...G...N=...~...~...SOH...VT...~...~...X.../...j...~...~...US...~...n...+...
J...g...~...ETX...~...~...~...o...T...<...~...S...~...~...B...C...~...n...~...RS...~...Cf...~...BPK...n...5...US...~...P...RS...~...VTETX...~...FS...~...~...0mNY~q(2...+r...v...h...RS...~...VT...~...DC4...c...t...q...FW^
13 7DEL...~...~...~...6..._...dc4...~...~...ETB...~...Zr...~...S...~...ESC...V...r...~...FS...~...~...X'?...~...~...~...
```

ANEXO 5

SAMPLE.DESCOMPIMIDO

```
sample_descomprimido.txt X Proyecto.py 1 huffman.py 6 sample.txt
sample_descomprimido.txt
1 Los incendios forestales han dejado una huella desgarradora en Bolivia, con más de 3,3 millones de hectáreas quemadas hasta ahora.
2 Más de 3,3 millones de hectáreas de zonas boscosas, matorrales y pasturas secas se han quemado en los 160 incendios forestales registrados en
Bolivia en lo que va del año, según informó el ministro de Defensa, Edmundo Novillo.
3
4 Informe sobre las afectaciones causadas por los incendios
5 El ministro compareció ante los medios en la ciudad central de Cochabamba para dar un informe sobre las afectaciones causadas por los incendios
este año en el país y las acciones realizadas por las autoridades, así como la ayuda internacional recibida para combatir el fuego.
6
7 Cifras y daños ocasionados por los incendios
8 Detalló que hasta noviembre se han quemado 3.372.369,46 hectáreas y explicó que consideran que «los incendios han concluido», ya que no se
reportan más focos de calor y ha comenzado la época de lluvias en el país. De esa superficie dañada, 1,1 millones de hectáreas corresponden a
«cobertura arbórea» y 2,2 millones eran matorrales y pasturas secas. La región amazónica de Beni fue la más afectada con 1,9 millones de hectáreas
quemadas.
9
10 Historial histórico e impacto social
11 Según datos históricos presentados por el ministro, el año con mayores afectaciones fue 2019 con 5.305.512 hectáreas quemadas; mientras que en
2022 la superficie dañada fue de 4.466.540 hectáreas. Estas cifras son obtenidas mediante un control satelital a través del Sistema de Información
y Monitoreo de Bosques (SIMB), que verifica diariamente las áreas quemadas y proporciona informes anuales.
12
13 Familias damnificadas y ayuda humanitaria
14 Se estima preliminarmente que hay 18.194 familias damnificadas por los incendios y se reportaron 133 viviendas afectadas o destruidas por el
fuego. Durante el periodo de junio a noviembre, se registraron 160 incendios en todo el territorio boliviano, siendo las siete de las nueve
regiones del país las más afectadas. La región de Beni declaró «desastre departamental».
15
16 Acciones tomadas para combatir los incendios
17 El Gobierno distribuyó 225,75 toneladas de ayuda humanitaria que requirieron una inversión cercana a los 2 millones de bolivianos (más de 285.000
dólares). Además, se agradeció el envío de 2,77 toneladas de ayuda humanitaria desde Chile y la colaboración de bomberos forestales provenientes
```