

Algoritmos de compresión

Comprimir se refiere a convertir datos a una representación que necesite menos espacio que la representación original.

Existen dos tipos de compresión:

- Sin pérdida (lossless)

Comprime mediante la identificación y eliminación de redundancias calculando mediante alguna fórmula estadística

Con pérdida (lossy)

- Elimina información innecesaria. Hay pérdida de datos. Son algoritmos muy complejos. Especialmente usados para almacenar archivos multimedia.

- Se pierde información del texto/dato original y no hay forma de recuperarlo

- MP3 / AAC / OGG / JPEG

- Los algoritmos lossless comprimen 20% - 90% dependiendo de las características de dato original

=> utilizar información estadística

- Huffman Codes

- Desarrollado en 1950 por David Huffman cuando era estudiante en el MIT

- Se utiliza en muchos formatos de compresión

La idea es asignar un código binario lo más corto posible a los símbolos que aparecen en los datos por comprimir

↑ frecuencia => símbolo más corto

- Construye un árbol binario que se utiliza para asignar dichos símbolos

comprimir : mississippi-river

① Contar las apariciones de las letras

m: 1

s: 4

i: 1

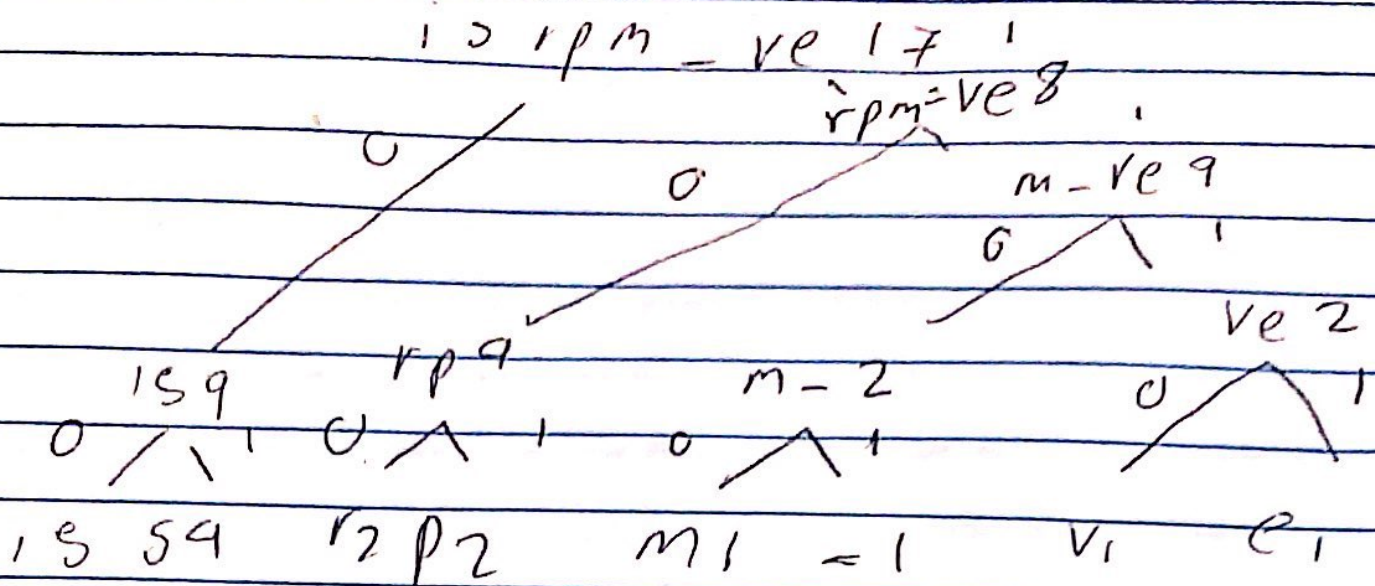
v: 1

i: 3

p: 2

r: 2

e: 1



| | |
|----------------------|------|
| $m : 1$ | 1100 |
| $i : 5$ | 00 |
| $s : 9\ \cancel{22}$ | 01 |
| $p : 2\ \cancel{22}$ | 101 |
| $: 1\ \cancel{22}$ | 1101 |
| $r : 2\ \cancel{22}$ | 100 |
| $v : 1\ \cancel{21}$ | 1110 |
| $e : 1$ | 1111 |