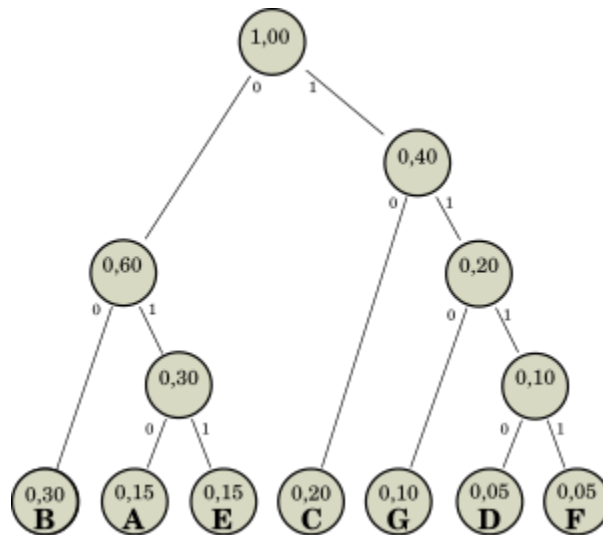


Árbol de Huffman

Consulta realizada por las estudiantes Ana María Fernández y Alejandra Díaz Navarro.

Los árboles de Huffman son una estructura de datos utilizada en informática y teoría de la información para realizar compresión de datos sin pérdida. Su propósito es representar información de manera más eficiente, asignando códigos binarios más cortos a los símbolos más frecuentes y códigos más largos a los menos frecuentes. Esta técnica fue desarrollada por David Huffman en 1952 y desde entonces se ha convertido en uno de los métodos de compresión más conocidos y utilizados.



Tomada de Wikipedia

La base de esta compresión se encuentra en la codificación de longitud variable. A diferencia de otros métodos como el código ASCII, que asigna el mismo número de bits a cada carácter (normalmente 8), la codificación de Huffman adapta la longitud del código en función de la frecuencia con la que aparece cada símbolo. Esto permite reducir significativamente el tamaño total del archivo cuando ciertos símbolos aparecen muchas veces.

Para construir un árbol de Huffman, primero se debe analizar el conjunto de datos y contar la frecuencia de aparición de cada símbolo. Luego, se crea un nodo para cada símbolo con su frecuencia correspondiente. Estos nodos se organizan en una estructura de árbol binario combinando iterativamente los dos nodos de menor frecuencia en uno nuevo, sumando sus frecuencias. Este proceso se repite hasta formar un solo árbol que contenga todos los símbolos.

Una vez que el árbol ha sido construido, se genera el código binario de cada símbolo recorriendo el árbol desde la raíz hasta cada hoja. Al bajar por la izquierda se agrega un 0 y por la derecha un 1. El camino que se forma hasta llegar al símbolo en cuestión es su código Huffman. Como resultado, los símbolos más comunes tienden a estar más cerca de la raíz, por lo tanto tienen códigos más cortos, lo que mejora la eficiencia de la compresión.

El algoritmo de Huffman es muy útil en aplicaciones prácticas como los formatos de compresión de archivos (.zip, .rar), imágenes (JPEG), audios (MP3), y en general en cualquier contexto donde se necesite reducir el tamaño de los datos sin perder información. Su eficiencia se debe a que se adapta automáticamente a las características del contenido que se quiere comprimir, y garantiza una codificación óptima para una fuente de datos conocida.

En resumen, los árboles de Huffman permiten una compresión eficiente al aprovechar las diferencias en frecuencia entre los distintos símbolos de un mensaje. Su implementación no solo reduce el uso de espacio de almacenamiento, sino también mejora la velocidad de transmisión de datos. Gracias a su sencillez y eficacia, sigue siendo una herramienta fundamental en el campo de la computación y las telecomunicaciones.

Ejemplo:

Supongamos que queremos codificar la palabra “ABRACADABRA” usando un árbol de Huffman.

1. Contamos la frecuencia de cada letra:

A: 5 veces

B: 2 veces

R: 2 veces

C: 1 vez

D: 1 vez

2. Creamos nodos para cada letra y los combinamos según su frecuencia:

C y D ($1 + 1 = 2$)

Este nuevo nodo se combina con B o R (todos tienen frecuencia 2)

Repetimos hasta formar un solo árbol con frecuencia total de 1.

3. Asignamos códigos:

A, al ser el más frecuente, tendrá el código más corto (por ejemplo: 0)

Las otras letras recibirán códigos más largos como 10, 110, 1110, etc.

4. Codificamos “ABRACADABRA”:

Si A = 0, B = 110, R = 111, C = 10, D = 101

Entonces la palabra quedaría codificada como:

0 110 111 0 10 0 101 0 110 111 0

(sin espacios: 01101110 10010101101110)

Esto genera una versión comprimida de la palabra original usando menos bits que una codificación fija como ASCII.

Referencias.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). Introduction to Algorithms (3rd ed.). MIT Press.

Colaboradores de Wikipedia. (2024). *Algoritmo de Huffman* [Fotografía]. Wikipedia, la Enciclopedia Libre. https://es.wikipedia.org/wiki/Algoritmo_de_Huffman

Huffman, D. A. (1952). A Method for the Construction of Minimum-Redundancy Codes. Proceedings of the IRE, 40(9), 1098–1101. <https://doi.org/10.1109/JRPROC.1952.273898>