Codage de Huffman

Le **codage de Huffman** est un [algorithme](https://fr.wikipedia.org/wiki/Algorithmique) de [compression de données sans perte](https://fr.wikipedia.org/wiki/Compression_de_donn%C3%A9es_sans_perte). Il a été développé en 1952 par [David A.Huffman](http://boowiki.info/art/informatique-americain/david-a-huffman.html), doctorant au MIT, et publié en Une méthode pour la construction de codes minimum Redondance.

Le codage de Huffman consiste à remplacer un caractère (ou symbole) par une suite de bits de longueur variable. Le code est déterminé à partir d'une estimation des probabilités d'apparition des symboles de source. L’idée est de coder ce qui est fréquent sur peu de bits et au contraire ce qui est rare sur des séquences de bits plus longues.

Un code de Huffman est optimal au sens de la plus courte longueur pour un codage par symbole, et une distribution de probabilité connue.

Il existe trois variantes du codage de Huffman : statique, semi-adaptatif et adaptatif.

## Aperçue historique

Claude Shannon, inventeur de théorie de l'information essayait de répondre à deux questions fondamentales: quelles sont les ressources nécessaires à la transmission de l'information, et quelle est la quantité d'informations que l'on peut transmettre de façon fiable.

En 1951 David Huffman et son collègue au cours du MIT en théorie de l'information a eu le choix entre une thèse écrite ou un examen. L'enseignant, Robert M. Fano, assigna une thèse sur le problème de trouver le code binaire le plus efficace. Huffman, incapable de démontrer tout code qui a été le plus efficace, démissionnait à l'idée d'avoir à étudier pour l'examen, quand il a eu l'idée d'utiliser un arbre binaire trié par fréquence, et si vite prouvé que cette méthode était la plus efficace.

Une idée similaire avait été utilisée auparavant dans [Codage Shannon-Fano](http://boowiki.info/art/theorie/codage-shannon-fano.html), mais Huffman a disposé son plus grand écart en construisant un arbre ascendant au lieu de descendre un.