# C:\Users\Admin\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Logo-IG2I-couleur-RVB.PNG

# TP\_5

**Compression par la méthode d’Huffman**

**Introduction**

Cette partie porte sur l’implémentation d’une méthode de compression de données sans perte, inventée par David Huffman en 1959

Cette méthode s’appuie sur un codage des symboles du document initial, d’autant plus court que ceux-ci sont fréquents dans ce document.

Le principe de cette méthode repose sur l’utilisation d’un **arbre de codage** et sur un **minimier** qui améliore l’efficacité de sa construction. Cet arbre de codage permet de produire un **codage préfixe**.

Des informations complémentaires sur cette méthode, les structures de données et le codage préfixe sont présentes dans le document d’accompagnement – *poly de cours* –.

**La méthode**

La méthode de compression de Huffman procède de la façon suivante :

1. On cherche tout d’abord le nombre d’occurrences de chaque caractère à traiter
2. Puis, on construit l’arbre de codage d’Huffman en partant des feuilles qui, associe aux caractères, portant comme information leur nombre d’occurrences. On associe ensuite les deux nœuds de plus faible nombre d’occurrences pour former un nouveau nœud interne dont la valeur est la somme des valeurs fils. On réitère ce processus avec les feuilles et les nœuds internes restants jusqu’à ne plus en avoir qu’un seul, la racine
3. L’arbre de codage étant créé, le codage d’un caractère est déterminé par le chemin depuis la racine, permettant d’atteindre la feuille qui lui est associée. A la branche conduisant au fils de plus grand nombre d’occurrences on associe un bit à 1, l’autre un bit à 0 (ou inversement, car peu importe)
4. On utilise le codage ainsi obtenu pour coder chaque caractère du document à traiter.

Toute cette démarche est présentée dans le cours avec l’exemple pas à pas.

**Exemple**

Le programme devra réaliser un codage de huffman et afficher sur sa sortie standard la table de codage générée, le texte compressé ainsi qu’un bilan affichant le ratio de compression.

Par exemple, lors de l’encodage de la chaîne "algorithme de huffman pour la compression de chaines", l’arbre de codage devra être affiché sous la forme suivante :

car : occ | long | bits

----+-----+------+--------

' ' :   7 |    3 | 010

'a' :   4 |    4 | 0011

'c' :   2 |    5 | 00000

'd' :   2 |    5 | 01110

'e' :   5 |    3 | 111

'f' :   2 |    5 | 00001

'g' :   1 |    5 | 11000

'h' :   3 |    4 | 1010

'i' :   3 |    4 | 1001

'l' :   2 |    5 | 00011

'm' :   3 |    4 | 1000

'n' :   3 |    4 | 1011

'o' :   4 |    4 | 0010

'p' :   2 |    5 | 00010

'r' :   3 |    4 | 1101

's' :   3 |    4 | 0110

't' :   1 |    5 | 11001

'u' :   2 |    5 | 01111

Le code généré devra être affiché sous cette forme :

001100011110000010110110011100110101000111010011101110101010011110000100001100000111011010000100010011111101010000110011010000000010100000010110111101100110100100101011010011101110100000010100011100110111110110

La conclusion devra être affichée sous cette forme :

Longueur du code binaire : 416 bits

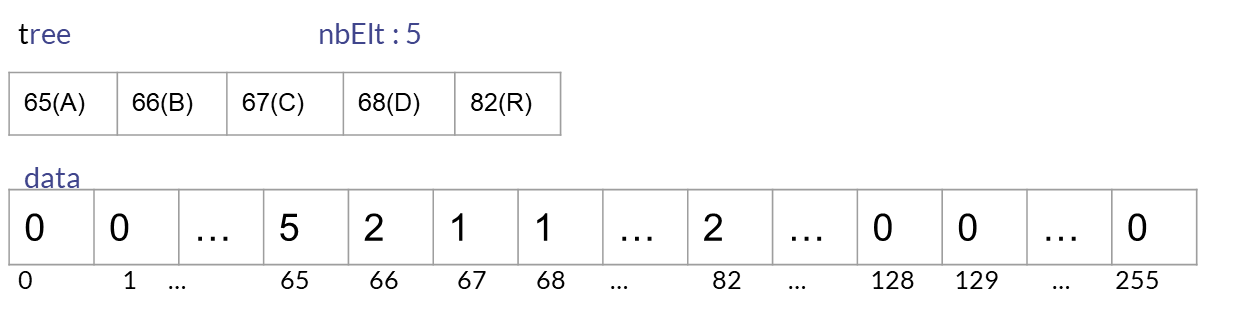
Longueur du code de huffman : 210 bits

Ratio de compression : 50.48%

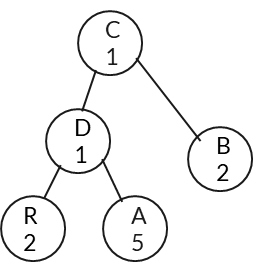
Le programme devra respecter les contraintes suivantes :

* Au fur et à mesure de son exécution, le programme génère des fichiers permettant de visualiser l’arbre partiellement ordonné et l’arbre de codage de manière graphique, à l’aide de l’outil graphviz.

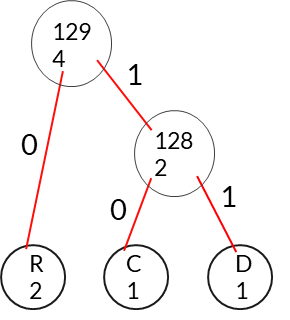
Par exemple, le minimier indirect présenté dans le poly du cours :



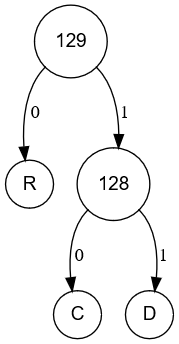
devra être présenté sous la forme suivante :



L’arbre de codage présenté ainsi :



devra être présenté sans les occurrences et peut être obtenu par le code graphviz :



# TNE

Produire un programme prenant un ou deux paramètres :

* Lorsque le programme prend deux paramètres, il doit compresser un texte et produire un fichier contenant le texte compressé précédé de l’entête de huffman permettant de le décompresser
  + paramètre 1 : chemin du fichier à compresser
  + paramètre 2 : chemin du fichier à produire

Exemple d’appel : huffman.exe ./fichierACompresser ./fichierResultat

* Lorsque le programme prend un seul paramètre, il s’agit du chemin d’un fichier à décompresser. Le contenu du document décompressé doit s’afficher sur la sortie standard.

Exemple d’appel : huffman.exe ./fichierADecompresser

Bien entendu, votre programme doit pouvoir décompresser les fichiers qu’il a lui-même compressé.