

Rapport Structures de Données

Groupe : Mikaël Ferreira de Almeida et Jérémie Henrion

1. Fonctionnalités mise en oeuvre

Fonctionnalité 2	Fonctionnalité 3	Fonctionnalité 4	Fonctionnalité 5
✓	✓	✓	✓

Fonctionnalité bonus : Nous avons fait le choix d'ajouter une interface graphique permettant une plus grande facilité d'utilisation lors de l'encodage ou de décodage de fichiers.

2. Structure de données de l'arbre de Huffman

Pour construire l'arbre de Huffman nous avons utilisé un arbre binaire avec pour chaque nœud le caractère, sa fréquence et sa représentation binaire sous forme d'une chaîne de caractère de '0' et '1'.

```
public class BinaryTree {  
    private BinaryTree left;  
    private BinaryTree right;  
  
    private int freq;  
    private char car;  
    private String bytes;  
    ...  
}
```

3. Résultat

Pour comparer deux fichiers, nous avons utilisé Meld (cf : <https://meldmerge.org/>).

<i>Fichier</i>	<i>Taille initiale</i>	<i>Taille compressé</i>	<i>Taux de compression</i>	<i>Correspondance</i>	<i>Encodage</i>
fichier1.txt	1144 octets	1380 octets	-20,63 %*	100.0 %	A : 00100 B : 10111010 C : 1110010
fichier2.txt	17,2 Mo	7,2 Mo	56,16 %	99.66 %**	A : 1010111 B : 1010110001 C : 100110111
fichier3.txt	12 Mo	7,8 Mo	32,35 %	100.0 %	A : 1010001 B : 1101101 C : 1001000

- *Fichier1 :

Le taux de compression de fichier1.txt est négatif à cause de la méthode pour stocker les tableaux de conversions expliquée dans la partie suivante.

- *L'erreur que nous rencontrons est due aux caractère '❖' présent dans fichiers2.txt.

```

fichier2.txt 90
PLOUESCAT - PRIX A.P.G.O - 2❖ME EPREUVE DU CHALLENGE...
4663;CRITERIUM DES JEUNES ❖ 50%;20070513;05;81;5164...
3887;GRAND PRIX DE LYON (4❖ME ETAPE DU DEFI DU GAL...
;PRIX OMNIUM ANGLO-ARABES ❖ 25%;20070629;05;50;516...
;PRIX DE L'ASS. DES PROPRI❖TAIRES DE CHEVAUX DE COUR...
PLOUESCAT - PRIX A.P.G.O (3❖EP CHALLENGE A.P.G.O OBS...
LOUIS MAHEO - PRIX DES PROPRI❖TAIRES DE GALOP L'OU...
4209;CRITERIUM DES JEUNES ❖ 37.5%;20080518;05;81;516...
- CHALLENGE A.P.G.O. (1ERE ❖PREUVE);20080818;01;07;51...
;PRIX DU "PMU LE VALLION" ❖ SAINT JOUAN DES GUERETS...
3672;GRAND PRIX DES MEDAILL❖S DE LA JEUNESSE ET DE...
BOISSONS - (PRIX LAMAUGARNY❖);20090503;07;15;516692;...
LOUIS MAHEO - PRIX DES PROPRI❖TAIRES DE GALOP L'OU...
3963;CRITERIUM DES JEUNES ❖ 37.5%;20090510;05;81;516...
95227;20;0;4600;PRIX DE LA 20❖ME FOIRE AUX FROMAGES...
2009;99702;0;0;4812;PRIX DU 3❖ME ANNIVERSAIRE DU NO...
PECHE MIGNON BOULANGERIE ❖ CARNAC;20090920;05;02...
0;4547;PRIX DE CAMPEL - 5❖ME EPREUVE DU CHALLENGE ...
3680;CRITERIUM DES JEUNES ❖ 37.5%;20100503;05;81;516...
VILLENEUVE - FETE DES COURSES - 2❖ME EP. CHALLENGE A...
;0;0;4631;PRIX FUSION FM N❖1 SUR LES HITS (PRIX D'AUBI...
4333;PRIX APGO-EQUIDEOS (5❖ETAPE DU CHALLENGE AP...
20;0;3929;PRIX FUSION FM N❖1 SUR LES HITS (PRIX DE MO...
PRIX MITTELHAEUSER ELECTRICIT❖ G❖N❖RALE;20...

```

Ligne : 1698, 2120, 2984 etc.

Au total, on comptabilise 90 occurrences du caractère '❖'. Le problème vient certainement des caractères accentués point, tel que : é, è, etc.

4. Méthode pour stocker les encodings

```
final String ENC_SEPARATOR = "\0";  
final String CHAR_SEPARATOR = "\0\0";  
final String END_LINE = "\0\n\0";
```

Nous avons choisi de stocker les codes de chaque caractère en début de fichier en suivant le format suivant:

- 1 caractère
- le caractère '\0'
- Une chaîne de caractère représentant sa valeur binaire. ex : "001101"
- La chaîne "\0\0"

Une fois que tous les couples caractères - représentations binaires sont écrits on termine la ligne avec la chaîne "\0\n\0".

Nous avons choisi le caractère '\0' car il y a de très fortes chances qu'il ne soit pas présent dans un fichier texte à encoder.

Cette méthode n'est pas optimale car nous écrivons des valeurs binaires '0' et '1' sous forme de caractère. Certes il est plus facile de les lire, mais cela fait que pour des petits fichiers (cf fichier1.txt) la compression ne sera pas optimale.