# THIN - THÉORIE DE L'INFORMATION - TD

**Exercice 1 : Quantité d'information et entropie**

Dans un disque dur se trouvent deux dossiers : `desktop` et `home`.
Supposons que `desktop` contient 50 fichiers et `home` en contient 150.

- Quelle est la quantité d'information $I_1$ associée à l'événement $e_1$ "le fichier appartient au dossier `desktop`" ?

On suppose aussi que `desktop` contient 20 fichiers de type `.jpg` et 30 de type `.txt`, et que `home` contient 110 `.jpg` et 40 `.txt`.

- Quelle est la quantité d'information $I_2$ associée à l'événement $e_2$ "le fichier est de type `.jpg`"?
- Quelle est la quantité d'information $I_3$ associée à l'événement "le fichier est de type `.jpg` et se trouve dans le dossier `desktop`"?
- Est-ce que $I_3 = I_1 + I_2$ ?
- Peut-on en conclure que les deux événements $e_1$ et $e_2$ sont dépéndants ou indépendants?
- Calculer l'entropie de la source $A = \{\texttt{.jpg}, \texttt{.txt}\}$ et de la source $B = \{\texttt{desktop}, \texttt{home}\}$.

**Exercice 2 : Entropie conditionnelle**

Une certaine population a une probabilité de 2% d'attraper une certaine maladie. On appelera donc $A$ la source $A = \{a_1 = sain, a_2 = malade\}$.
Calculer l'entropie de $A$.
Supposons qu'il existe un test pour detecter la maladie, qui repond toujours "positif" pour un malade, mais qui repond souvent "positif" pour les sujets sains aussi. En particulier, on considère la source $B = \{b_1 = positif, b_2 = negatif\}$ et supposons de savoir que:

- la probabilité que le test soit positif est de 51% (i.e. $p(b_1) = 51/100$)
- la probabilité qu'un sujet positif soit vraiment malade est de 2/51 (i.e. $p(a_2|b_1) = 2/51$).

- Calculer tous les $p(a_i|b_j)$ et $p(a_i, b_j)$.
- Calculer l'entropie conditionnelle $H(A|B)$.
- Est-ce que l'entropie a augmenté ou diminué ?
- En se rappelant que l'entropie est la moyenne d'information est que l'information répresente l'incertitude d'un événement...est-ce que l'incertitude augmente ou diminue?

**Exercice 3 : Code de Huffman**

On suppose d'avoir une image composée par 10 couleurs, dont on va donner la fréquence dans le tableau ci-dessous :

| couleur | fréquence | couleur | fréquence |
|---|---|---|---|
| 1=noir | 0,05 | 6 = vert | 0,08 |
| 2 = blanc | 0,02 | 7 = violet | 0,12 |
| 3 = bleu | 0,03 | 8 = orange | 0,10 |
| 4 = jaune | 0,15 | 9 = marron | 0,20 |
| 5 = rouge | 0,20 | 10 = gris | 0,05 |

(TOURNEZ SVP)

*Date*: 14 janvier 2011.

- Calculer le code de Huffman associé à cette source (avec cette loi de probabilité) en dessinant aussi les arbres binaires aussociés.
- Calculer la longueur moyenne du code de Huffman ci-dessus.
- Calculer l'efficacité du code de Huffman et la comparer avec l'efficacité d'un code à longueur fixe pour cette source.
- Est-ce qu'on peut améliorer encore cette efficacité? Comment?

**Exercice 4 : LZ77**

(1) On considère la chaîne de caractères "toto_murmurait" ('_' indique le caractère espace).
On prend une fenêtre de 7 caractères dont 3 pour le tampon de recherche.
Appliquer l'algorithme de codage LZ77 et écrire la suite de triplets de la forme $(p, l, c)$ qu'on obtient à la sortie.
Si on suppose de coder $p$ et $l$ avec 2 bits chacun ($p, l \leq 3$) et chaque caractère avec 8 bits, quelle est le nombre de bits nécessaire pour le message codé? Et pour le message originaire?

(2) On prend maintenant une fenêtre de longueur 11 (avec tampon de recherche de longueur 5), décoder le message compressé suivant :
$(0, 0, i); (0, 0, l); (3, 1, c); (0, 0, a); (4, 1, c); (0, 0, u); (3, 1, e); (0, 0, \_); (3, 2, n); (0, 0, t); (3, 1, m); (5, 3, \emptyset)$.
(où $\emptyset$ répresente la fin du message).

**Exercice 5 : LZ78**

(1) Utiliser la méthode LZ78 pour compresser la chaîne de caractères suivante :
"ababcbcaaabcabbcc".
En supposant que un symbole soit representé par 8 bits, quelle est la taille originaire de la chaîne? Une fois compressée (en supposant qu'on utilise un code de longueur fixe minimale pour les entiers) quelle sera la taille?

(2) Décompresser le message suivant : $(0, t); (0, r); (0, a); (0, l); (3; l); (5, a); (4, e); (2, e)$.