

Série 11

1 Un peu de magie noire

a) Calculez l'entropie de la séquence de lettres suivante (sans l'espace):

AVADA KEDAVRA

- en écrivant d'abord le résultat sous la forme $a + b \log_2(3) + c \log_2(5)$, où a, b, c sont des fractions de nombres entiers;

- puis en calculant le résultat numérique ou en l'approximant avec $\log_2(3) \simeq 1.58$ et $\log_2(5) \simeq 2.32$.

b) Créez un dictionnaire pour cette même séquence de lettres à l'aide de l'algorithme de Shannon-Fano. Combien de bits utilisez-vous pour représenter la séquence? Essayez différentes versions de l'algorithme et comparez.

c) Créez ensuite un dictionnaire à l'aide de l'algorithme de Huffman. Combien de bits utilisez-vous pour représenter la séquence? A nouveau, essayez différentes versions de l'algorithme et comparez.

d) Comparez les résultats obtenus aux points a), b) et c). Ceci est-il cohérent avec ce que vous avez appris au cours?

2 De bien mauvais dicos

Pour encoder le mot "BANANA" sous forme de 0 et de 1, six minions ont proposé les dictionnaires suivants:

a)

A	N	B
1	11	111

c)

A	N	B
0	10	110

e)

A	N	B
00	01	1

b)

A	N	B
0	1	11

d)

A	N	B
00	01	10

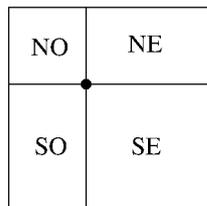
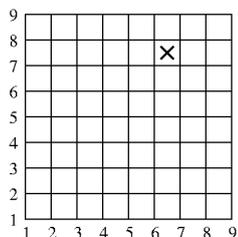
f)

A	N	B
0	01	001

Mais Grou n'est satisfait d'aucun d'entre eux. Pouvez-vous expliquer le(s) défaut(s) de chacun? Et aussi proposer un meilleur dictionnaire?

3 A la recherche d'un trésor

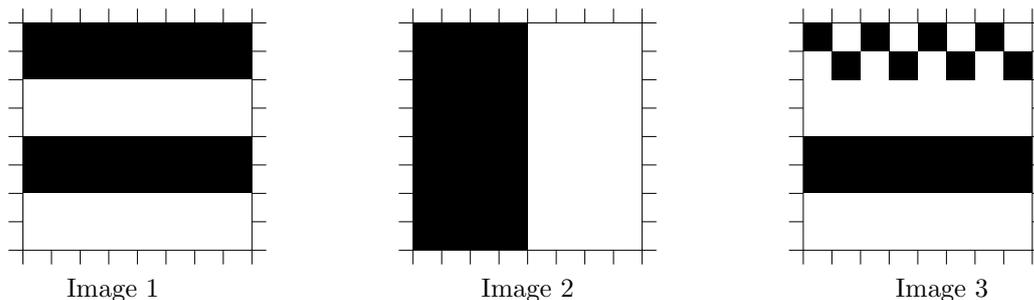
Vous êtes à la recherche d'un trésor, dont la position est une case choisie au hasard et indiquée par une croix sur le plan à gauche ci-dessous. Pour trouver la position du trésor, vous devez choisir un point sur la grille et poser la question: dans quelle direction se trouve le trésor? Un oracle vous indiquera alors s'il est au nord-est (NE), au nord-ouest (NO), au sud-est (SE) ou au sud-ouest (SO), comme indiqué par exemple à droite ci-dessous. Bien sûr, il vous faudra interroger l'oracle en plusieurs points de la grille pour localiser le trésor.



prière de tourner la page %

- a) Si vous utilisez une stratégie optimale, combien de questions faudra-t-il pour localiser le trésor?
- b) Dans l'exemple ci-dessus à gauche, quelle est la suite des points où vous interrogerez l'oracle avec votre stratégie optimale? (utilisez la notation (a, b) pour un point, avec a, b allant de 1 à 9, a dénotant l'abscisse et b l'ordonnée du point).
- c) Proposez un système *efficace* d'encodage avec des bits pour transmettre l'information de l'emplacement du trésor à un ami. En se basant sur votre système, quelle sera la séquence de bits que vous transmettez à votre ami dans l'exemple ci-dessus à gauche?

4 Pour le plaisir: codage par plages (run-length encoding)*



Dans cet exercice, on considère une autre type de compression, utilisé principalement pour les images en noir et blanc comme celles présentées ci-dessus. Il s'agit du *codage par plages* ou run-length encoding (RLE) en anglais. L'idée est la suivante: dans une image en noir et blanc, chaque pixel est représenté par un 1 (noir) ou un 0 (blanc). Pour compresser une image avec $8 \times 8 = 64$ pixels, on transforme tout d'abord celle-ci en une séquence de 64 bits, en "lisant" l'image ligne par ligne. Ainsi l'image 1 ci-dessus est représentée de manière "brute" par la séquence de bits:

```
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

Vu le grand nombre de 1 et de 0 consécutifs dans cette séquence, il semble qu'on peut économiser de l'espace-mémoire en procédant comme suit: on divise la séquence en paquets de 4 bits de longueur; dans chaque paquet, le premier bit symbolise la couleur (0 ou 1) de la suite de pixels et les 3 bits suivants indiquent en binaire le nombre de pixels consécutifs de cette couleur, moins 1. Exemples:

0010 signifie "3 pixels consécutifs de couleur 0"

1101 signifie "6 pixels consécutifs de couleur 1"

- a) Quel est le codage RLE des trois images ci-dessus, et quelle est la longueur de ce code pour chaque image (i.e. combien épargne-t-on de bits par rapport à la taille originale de l'image qui est de 64 bits)?
- b) Si vous avez résolu correctement la partie a), vous aurez constaté que le codage RLE de la seconde image est bien plus long que celui de la première image, alors que ces images sont très similaires par nature. Pouvez-vous proposer quelque chose pour remédier à ce problème?
- c) Si vous avez résolu correctement la partie a), vous aurez également constaté que le codage RLE de la troisième image pose problème. A nouveau, pouvez-vous imaginer une solution pour remédier à ce problème?

Indication: Cette fois-ci, nous vous proposons de modifier la taille de paquet (sachant que cette taille reste utilisée pour toute l'image) et le rôle d'un ou de plusieurs bits de chaque paquet. Le but est d'introduire une plus grande flexibilité dans le codage pour pouvoir aussi bien coder des plages uniformes et des zones comme le début de l'image 3.