

Compression & Algo du Texte

– M1 –
2014-2015

Feuille 2 : Théorie de l'information & LZ77

► **Exercice 1** ◀ On considère une source S sur l'alphabet $A = \{a, b, c\}$. On suppose que les symboles sont émis (sans mémoire) avec probabilités $p_a = \frac{1}{2}$, $p_b = \frac{1}{4}$ et $p_c = \frac{1}{4}$.

- Calculez l'entropie de la source.
- Quel est le taux de compression du message par Huffman, en prenant les probabilités comme nombre d'occurrences (sans encoder l'arbre).
- Même question si on regroupe les symboles par blocs de 2 : on crée un nouvel alphabet $\{A, B, C, D, E, F, G, H, I\}$ avec $A = aa$, $B = ab$, ..., $I = cc$.

► **Exercice 2** ◀ Représentez le graphe de la chaîne de Markov d'ordre 1 associé au message

ananassassin

► **Exercice 3** ◀ On a une probabilité p sur l'ensemble fini $E = \{1, \dots, n\}$, qui nous est donné par une liste qui contient, dans l'ordre, p_1, p_2, \dots, p_n .

En supposant qu'on a un générateur aléatoire `uniform()` qui retourne un flottant de $[0, 1[$, écrire (en pseudo-code) un algorithme efficace pour tirer au sort un élément de E en respectant les probabilités.

► **Exercice 4** ◀ On considère le message suivant :

aacaacabcaaaacbab

- Encodez le message avec la méthode de LZ77. On utilisera un buffer et une fenêtre de taille 8 tous les deux.
- Appliquez la variante LZSS avec un seuil pour garder les couples fixé à 2.

► **Exercice 5** ◀ Détaillé ce qui se passe quand on applique l'algorithme LZ77 à un mot de la forme a^n ? On supposera que la fenêtre est de taille 256, le buffer de taille 8, et que n est de la forme $n = 8k + 2$.