TDs de Théorie de l'information. Feuille 2

Exercice 1.— On considère un ensemble fini de mots sur un alphabet à deux lettres. Donnez un algorithme permettant de tester si cet ensemble de mots est un code préfixe. Quelles sont les complexités en temps et espace ?

Exercice 2.— Etant donné un alphabet A, un *automate* sur A est ici un couple (Q, F), où Q est un ensemble fini appelé l'ensemble des *états* (qui sont généralement représentés par des entiers naturels), et $F \subset Q \times A \times Q$ est l'ensemble des *flèches*, c'est-à-dire un ensemble de triplets (p, a, q), où p et q sont dans Q et a est dans A. La flèche (p, a, q) est représentée aussi par $p \xrightarrow{a} q$.
A tout ensemble fini C constitué de mots non vides sur un alphabet A, on associe un automate appelé l'*automate en pétales* de C comme suit.
Son ensemble d'états est formé d'un état distingué noté 0 et, pour chaque mot $u \in C$, de $r-1$ états (u, p_i), où $r = |u|$ et p_i est le préfixe propre non vide de longueur i de u pour $i = 1, \ldots, r-1$.
Chaque mot $u \in C$ contribue $r = |u|$ flèches à l'automate. Si $r = 1$, c'est la flèche $0 \xrightarrow{u} 0$; si $u = a_1 \cdots a_r$ où les a_i sont des lettres et $r > 1$, ce sont les r flèches $0 \xrightarrow{a_1} (u, p_1)$, $(u, p_{i-1}) \xrightarrow{a_i} (u, p_i)$ pour $1 \le i \le r-2$ et $(u, p_{r-1}) \xrightarrow{a_r} 0$.

Exemple. On considère l'ensemble $C = \{a, ab, ac\}$ sur l'alphabet $A = \{a, b, c\}$. L'automate en pétales de C est donné dans la figure 1.
Le terme d'automate en pétales provient du fait que ce dessin ressemble à une fleur, chaque pétale correspondant à un mot de C. Il existe par exemple un chemin d'étiquette $abac$ allant de l'état 0 à l'état 0.

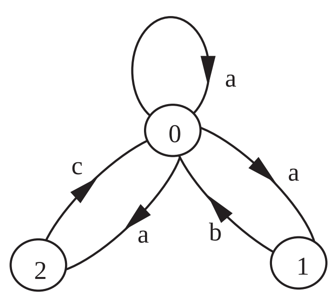

Fig. 1: Automate en pétales de C

Etant donné un automate $\mathcal{A} = (Q, F)$ sur un alphabet A, on définit, sur le même alphabet, deux nouveaux automates associés à $\mathcal{A}$, l'un dit *automate des couples*, l'autre dit *automate des paires*. L'automate des couples est l'automate $(Q \times Q, G)$, où $G = \{((p, q), a, (r, s)) \mid (p, a, r) \in F \text{ et } (q, a, s) \in F\}$. L'automate des paires est obtenu à partir de l'automate des couples en identifiant les états (p, q) et (q, p) lorsque p et q sont distincts.

Exemple. L'automate en pétales de la figure 1 admet comme automate des paires l'automate de la figure 2. Seuls sont représentés les états sur lesquels arrivent des flèches.

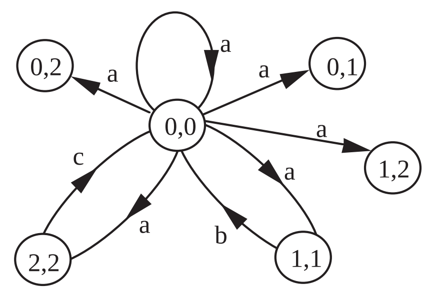

Fig. 2: Automate des paires de l'automate en pétales de la figure 1.

Montrer qu'un ensemble fini C est un code si et seulement son automate des paires a la propriété suivante : il n'existe pas de circuit passant par l'état $(0, 0)$ et passant par un état (p, q) avec $p \neq q$. Indiquer la complexité algorithmique de ce test. Faire tourner cet algorithme sur $\{b, ab, abb, abaa, aaaa\}$, $\{b, ab, aab, abba\}$, $\{aa, ba, bb, baa, bba\}$, $\{ab, ababa, baa\}$.

Exercice 3.— Un ensemble fini C de mots non vides est un *code circulaire* si on a la propriété suivante : si $u_1, \ldots, u_m$ et $v_1, \ldots, v_n$ sont des éléments de C, si p et s sont deux mots avec s non vides, tels que $su_2 \cdots u_m p = v_1 \cdots v_n$ et $u_1 = ps$, alors $m = n$, p est le mot vide et $u_i = v_i$, pour tout i.
Cette propriété signifie que tout mot «vu sur un cercle» s'écrit d'au plus une façon comme produit de mots du code circulaire. Tout code circulaire est un code mais l'inverse n'est pas vrai.

Exemple. L'ensemble $C = \{a, ab, ac\}$ est un code circulaire. L'ensemble $C = \{ab, ba\}$ est un code mais n'est pas un code circulaire. On donne dans la figure 3 deux «décompositions circulaires» distinctes du mot $ababab$.

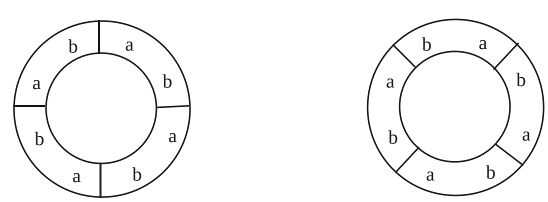

Fig. 3: Exemple d'un code non circulaire

Montrer qu'un code fini est circulaire ssi son automate des couples n'a pas de circuit passant par un état $(0, x)$ *et* par un état $(y, 0)$, avec $x \neq 0$ ou (non exclusif) $y \neq 0$. Indiquer la complexité algorithmique de ce test. Faire tourner cet algorithme sur $\{b, ab, abb, abaa, aaaa\}$, $\{b, ab, aab, abba\}$, $\{aa, ba, bb, baa, bba\}$, $\{ab, ababa, baa\}$.