Feuille de TD n° 5 : Lemme de l'étoile vs. expressions régulières

Pour décider si un langage est reconnaissable par un automate, on procède le plus souvent de la manière suivante :
- soit on on veut montrer qu'il n'est pas reconnaissable, et on utilise le lemme de l'étoile
- soit on veut montrer qu'il est reconnaissable, et on exhibe un automate le reconnaissant OU une expression régulière le décrivant. On admettra pour ce TD qu'un langage décrit par une expression régulière est toujours reconnaissable.

Exercice 1 : reconnaissance de langages

On considère l'alphabet $\mathcal{A} = \{a, b\}$. Les langages suivants sont-ils reconnaissables ?
- $\mathcal{L}_1 = \{a^n b^n \mid n \geq 0\}$
- $\mathcal{L}_2 = \{u \in A^* \mid |u|_a = |u|_b\}$
- $\mathcal{L}_3 = \{u \in A^* \mid |u|_a < |u|_b\}$
- $\mathcal{L}_4 = \{baba^2ba^3 \dots ba^n \mid n \geq 0\}$
- $\mathcal{L}_5 = \mathcal{L}_2 \setminus A^*\{aa, bb\}A^*$
- $\mathcal{L}_6 = \{u \in A^* \mid \text{u est un palindrome}\}$
- $\mathcal{L}_7 = \{uu \mid u \in A^*\}$
- $\mathcal{L}_8 = \mathrm{Pref}(\mathcal{L}_1)$
- $\mathcal{L}_9 = \mathrm{Pref}(\mathcal{L}_2)$
- $\mathcal{L}_{10} = \{a^n \mid n \text{ est un nombre premier}\}$

où $\mathrm{Pref}(\mathcal{L})$ est l'ensemble des facteurs préfixes des mots du langage $\mathcal{L}$. On en profitera pour montrer que si $\mathcal{L}$ est reconnaissable alors $\mathrm{Pref}(\mathcal{L})$ est reconnaissable.

Exercice 2 : rappels sur les expressions régulières

- Soit $E_1 = (a + b)^*$. Les mots a^2, a^2b et ba^2 appartiennent-ils au langage décrit par E_1 ?
- Soit $E_2 = a^*b^*c^*$. Les mots a^2, ab, a^2b, cb et $aabc$ appartiennent-ils au langage décrit par E_2 ?
- Soit $E_3 = (a + b)^*abb$. Les mots a^2, a^2b, ab^2ab et a^3b^2 appartiennent-ils au langage décrit par E_3 ?

Donner des automates (pas nécessairement déterministes) reconnaissant ces langages.

Exercice 3 : langages décrits par une expression régulière

Donner des expressions régulières décrivant les langages ci-dessous :
- $\mathcal{L}_1 = \{u \in A^* : \text{toute occurrence de } b \text{ est immédiatement suivie de deux occurrences de } a\}$,
- $\mathcal{L}_2 = \{u \in A^* : u \text{ ne contient pas deux } a \text{ successifs}\}$,
- $\mathcal{L}_3 = \{u \in A^* : \text{le nombre d'occurrences de } a \text{ dans } u \text{ est pair}\}$,
- $\mathcal{L}_4 = \{u \in A^* : \text{les blocs de } a \text{ dans } u \text{ sont alternativement de longueur paire et impaire}\}$.

Exercice 4 : un peu plus dur...

Décrire (en français) les langages définis par les expressions régulières ci-dessous, puis donner des automates (pas nécessairement déterministes) les reconnaissant.
- $E_1 = (aa + b)^*(a + bb)^*$,
- $E_2 = (a + ba + bba)^*(\varepsilon + b + bb)$,
- $E_3 = (aa + bb + (ab + ba)(aa + bb)^*(ab + ba))^*$.

Exercice 5 : petite variation sur le lemme de l'étoile

On considère ici l'alphabet $B = \{a, b, c\}$.

On dit qu'un mot contient un carré si on peut l'écrire $uvvw$ avec $v \neq \epsilon$. On admettra (c'est loin d'être évident) qu'on peut construire des mots sans carré arbitrairement longs sur l'alphabet $\{a, b, c\}$. Montrer que le langage suivant n'est pas reconnaissable par un automate déterministe *complet* (cela implique donc qu'il n'est reconnaissable par aucun automate fini) :

$\mathcal{K} = \{u \in B \mid u \text{ contient un carré}\}$