

TD n° 1

Expressions régulières

Exercice 1.1 Déterminer tous les mots de longueur maximale 4 qui appartiennent au langage dénoté par chacune des expressions régulières suivantes :

$$\begin{array}{llll} (i) & (b+ba)^* & (ii) & ab^*+b \\ (iii) & (a+b)^*abb & & \\ (iv) & (x+\varepsilon)^*dd^* & (v) & (xd+\varepsilon)^*d^* \\ (vi) & a^*(b+c)d^* & & \end{array}$$

Exercice 1.2 Donner une description en français des langages donnés par les expressions régulières suivantes :

$$\begin{array}{llll} (i) & (a+b)^* & (ii) & a(a+b)^* \\ (iii) & (a+b)^*a & (iv) & (b+ab)^*(a+\varepsilon) \\ (v) & a^*+b^* & (vi) & (aa+b)^* \\ (vii) & (ab^*a+b)^* & & \end{array}$$

Exercice 1.3 Décrire sous la forme d'une expression régulière les langages suivants

- (i) le langage des mots de longueur 2.
- (ii) le langage des mots de longueur au plus 2.
- (iii) le langage des mots de longueur paire.
- (iv) le langage des mots ayant au moins une occurrence de a .
- (v) le langage des mots ayant au moins une occurrence du facteur ab .
- (vi) le langage des mots ayant au moins une occurrence de a puis ensuite une occurrence d'un b .
- (vii) le langage des mots commençant par a , finissant par b et n'ayant pas deux a ou deux b consécutifs.

Exercice 1.4 Prouver les équivalences suivantes :

1. $(a+b) + (a+b)(a+b)^* + \emptyset^* \equiv (a+b)^*$
2. $a(c^+ + \emptyset^*) + (a+b)(c^* + (c^*)^*) \equiv (a+b)c^*$
3. $(x+y)\emptyset + (x+y)\emptyset^* + ((x+y)^*\emptyset^*)^* \equiv (x+y)^*$
4. $0(\varepsilon+00)^*(1+01)+1 \equiv 0^*1$

Exercice 1.5 Pour chacun des langages suivants, donner une expression régulière représentant son complément : (i) $(a+b)^*b$; (ii) $((a+b)(a+b))^*$.

Exercice 1.6 1. Il y a-t-il toujours une infinité d'expressions régulières dénotant un langage régulier ?

2. Le miroir d'un langage régulier est-il régulier ?

Exercice 1.7 On rappelle que deux ensembles X et Y sont *équipotents* s'il existe une bijection de l'un vers l'autre. On note alors $X \sim Y$. En particulier, un ensemble équipotent à \mathbb{N} est dit *dénombrable*.

- (a) Prouver qu'un ensemble n'est jamais équipotent à l'ensemble de ses parties.
- (b) Soit Σ un alphabet. Montrer que l'ensemble des *mots* sur Σ est dénombrable alors que l'ensemble des *langages* sur Σ ne l'est pas.
- (c) En déduire qu'il existe des langages non réguliers.

Exercice 1.8 Pour chacune de ces affirmations, dire si elle est vraie ou fausse en argumentant brièvement.

- i) Tout langage régulier est infini.
- ii) Tout langage non régulier est infini.
- iii) Il y a une infinité de langages réguliers.
- iv) Il y a une infinité de langages non réguliers.
- v) Tout langage inclus dans un langage régulier est régulier.
- vi) Il y a toujours une infinité d'expressions régulières pour décrire un langage régulier.
- vii) si L^* est régulier alors L est régulier.
- viii) si $L_1 \cup L_2$ est régulier et $|L_1| < \infty$ alors L_2 est régulier.
- ix) si $L_1 \cup L_2$ est régulier et L_1 est régulier alors L_2 est régulier.

Exercice 1.9 Dans le système de numération des Romains, les valeurs inférieures à 3999 s'écrivent avec des "chiffres" dont les équivalents décimaux sont donnés dans le tableau suivant :

Chiffre romain	Equivalent décimal
M	1000
D	500
C	100
L	50
X	10
V	5
I	1

Les nombres en chiffres romains sont formés selon les règles suivantes :

Ordre des chiffres : Les chiffres romains sont écrits en ordre décroissant, et les valeurs des chiffres s'additionnent pour donner la valeur souhaitée, sauf dans le cas des préfixes, comme on le verra dans la règle **préfixes**.

Nombre de répétitions : Pour M, C, X et I, il est interdit de faire plus de 3 répétitions consécutives, et il ne peut y avoir qu'une seule occurrence de D, L ou V.

Préfixes : Les chiffres C, X et I peuvent s'utiliser en tant que préfixes d'un chiffre supérieur, rompant ainsi l'ordre décroissant. Leur valeur est alors soustraite et non additionnée pendant la conversion, avec les restrictions suivantes :

- Un seul C peut préfixer un M ou un D pour donner respectivement 900 et 400 ; le M préfixé s'écrit à la suite des autres M, comme dans MMMCM. Les chiffres situés après un M ou un D préfixé représentent une valeur qui ne peut dépasser 99.
- Un seul X peut préfixer un C ou un L pour donner respectivement 90 et 40 ; le C préfixé s'écrit à la suite des autres C. Les chiffres après un C ou un L préfixé représentent une valeur qui ne peut dépasser 9.
- un seul I peut préfixer un X ou un V pour donner respectivement 9 ou 4. Le chiffre préfixé ne peut apparaître qu'à la fin du nombre.

Quelques exemples :

Nombre décimal	Nombre en chiffres romains
1987	MCMLXXXVII
1999	MCMXCIX
339	CCCXXXIX

Décrire par une expression régulière le langage des nombres romains compris entre 1 et 3999.