Quelques rappels pour se (re)familiariser avec les notations.

Langages

Soit Σ un *alphabet* (ou *vocabulaire*) quelconque (un ensemble de caractères, ou symboles) :
- un *mot* x (de longueur n) sur Σ est une suite finie de caractères de Σ (de longueur n) ;
- ϵ est le mot ne contenant aucun caractère ;
- Σ^n est l'ensemble des mots de longueur n ;
- Σ^* signifie $\bigcup_{n\geq 0} \Sigma^n$, c'est l'ensemble de tous les mots sur Σ ;
- xy est la concaténation des mots x et y.

Un langage L sur Σ est un sous-ensemble de Σ : $L \in \mathcal{P}(\Sigma^*)$, $L \subseteq \Sigma^*$. Quelques opérations courantes sur les langages :
- $L_1 \cup L_2$: union de langages ;
- L_1L_2 signifie $\{xy|x \in L_1, y \in L_2\}$, concaténation de langages ;
- $\overline{L}$ signifie $\Sigma^* - L$, complément d'un langage ;
- L^n signifie $\{x_1 \ldots x_n | x_i \in L, 1 \leq i \leq n\}$;
- L^* signifie $\bigcup_{n\geq 0} L^n$, fermeture d'un langage.

Définition 1 (Langages réguliers) Un langage régulier sur Σ est défini récursivement :
- $\emptyset$, $\{\epsilon\}$ sont des langages réguliers ;
- pour tout $a \in \Sigma$, $\{a\}$ est un langage régulier ;
- si L_1 et L_2 sont des langages réguliers sur Σ alors $L_1 \cup L_2$, L_1L_2 et L_1^* sont aussi des langages réguliers sur Σ ;
- il n'y a pas d'autre langage régulier sur Σ.

Les langages réguliers sont clôts par union, intersection, concaténation, complémentaire, différence ensembliste, et il est facile de construire l'automate correspondant à partir des automates correspondant aux langages opérandes.

Expressions régulières

Définition 2 (Expressions régulières) Les expressions régulières (er) sur Σ et les langages réguliers qu'elles décrivent sont aussi décrits récursivement :
- $\emptyset$ est une er qui décrit le langage $\emptyset$;
- ϵ est une er qui décrit le langage $\{\epsilon\}$;
- $\mathtt{a}$ (pour $a \in \Sigma$) est une er qui décrit le langage $\{a\}$.

Si $\mathtt{e}_1$ et $\mathtt{e}_2$ sont des er sur Σ, décrivant les langages $L(e_1)$ et $L(e_2)$, alors les expressions suivantes sont aussi des er :
- $\mathtt{e}_1 \,|\, \mathtt{e}_2$ décrit le langage $L(e_1) \cup L(e_2)$;
- $\mathtt{e}_1\mathtt{e}_2$ décrit le langage $L(e_1)L(e_2)$;
- $\mathtt{e}_1{}^*$ décrit le langage $L(e_1)^*$.

Il n'y a pas d'autre expression régulière.

Automates à nombre fini d'états (AF)

Définition 3 (Automate fini non déterministe, AFND) Un automate fini non déterministe (AFND) est un tuple $A = (\Sigma, Q, \Delta, q_0, F)$ dans lequel :
- Σ est l'alphabet d'entrée ;
- Q est un nombre fini d'états ;
- $q_0 \in Q$ est l'état initial ;
- $F \subseteq Q$ est l'ensemble des états finaux ;
- $\Delta \subseteq Q \times (\Sigma \cup \{\epsilon\}) \times Q$ est la relation de transition.

Un AFND admet des transitions étiquetées par ϵ et des états dont plus d'une transition sortante ont la même étiquette.

Définition 4 (Configuration, langage accepté) Soit $A = (\Sigma, Q, \Delta, q_0, F)$ un AFND :

- un couple (q, w) avec $q \in Q$ et $w \in \Sigma^*$ est une *configuration* de A ;
- (q_0, w) est une *configuration initiale* ;
- (q_f, ϵ) avec $q_f \in F$ est une *configuration finale*.

La *relation de dérivation* $\vdash_A \subseteq\ (Q \times \Sigma^*) \times (Q \times \Sigma^*)$ est telle que

$$(q, aw) \vdash_A (q', w) \text{ ssi } (q, a, q') \in \Delta \text{ avec } a \in \Sigma \cup \{\epsilon\}$$

Le *langage accepté* par A est

$$L(A) = \{w \in \Sigma^* | (q_0, w) \vdash_A^* (q_f, \epsilon) \text{ avec } q_f \in F\}$$

Théorème 1 Un langage est régulier ssi il est accepté par un AFND.

Définition 5 (AF Deterministe - AFD) Soit $A = (\Sigma, Q, \Delta, q_0, F)$ un AFND. A est un AFD si Δ est une fonction partielle $\delta : Q \times \Sigma \to Q$.

Dans un AFD il n'y a pas de transition sur ϵ, et pour tout couple (q, a) il existe au plus un état successeur.

Propriété 1 Si un langage L est accepté par un AFND alors il existe un AFD acceptant L.