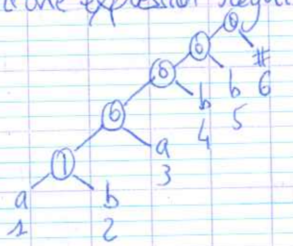
Kleene efficace

Réf : Aho, Lam, Sethi, Ullman : Compilers, principles, techniques & tools (dragon book)

En regardant la construction de l'automate déterministe à partir de l'automate avec ε-transitions, on se rend compte que certains états sont plus important que d'autre, à savoir ceux dont au moins une transition sortante n'est pas étiqueté par ε. Or, ces états importants proviennent des différentes positions dans l'expression régulière de départ.

• Déf : Arbre de syntaxe d'une expression régulière

ex $A = (a|b)^* abb$ ⇝



⇝ $(1|2)^* 3456 =: A'$

• des fonctions : À partir de cette arbre de syntaxe, on va pouvoir calculer efficacement des fonctions qui vont donner les infos nécessaires pour calculer le DFA final.

* $\varepsilon(n)$ est vraie pour n un sous-arbre ssi l'expression régulière représentée par le sous arbre contient ε dans son langage
* PREMIER(n) est l'ensemble des positions qui apparaissent comme première lettre du langage engendré par le sous-arbre (ie $\{w[1] / w \in \mathcal{L}(A'_{|n})\}$)
* DERNIER(n) idem
* SUIVANT(p) est l'ensemble des positions qui suivent p dans un mot du langage $\mathcal{L}(A')$

• Calcul des fonctions : par induction sur l'arbre pour $\varepsilon(n)$, PREMIER et DERNIER

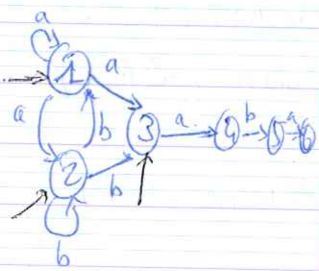
* $\varepsilon(n)$:
  - une feuille étiqueté ε : $\top$
  - une feuille étiqueté $\neq \varepsilon$: $\perp$
  - un noeud $n = c_1 | c_2$: $\varepsilon(c_1) \vee \varepsilon(c_2)$
  - un noeud $n = c_1 \circ c_2$: $\varepsilon(c_1) \wedge \varepsilon(c_2)$
  - un noeud $n = c_1^*$: $\top$

* PREMIER(n) :
- une feuille ε : $\emptyset$
- une feuille avec position i : $\{i\}$
- un noeud $n = c_1 | c_2$: PREMIER(c_1) $\cup$ PREMIER(c_2)
- un noeud $n = c_1 \circ c_2$: PREMIER(c_1) $\left(\bigcup_{\text{si } \varepsilon(c_1)} \right.$ PREMIER(c_2) $\left. \right)$
- un noeud $n = c_1^*$: PREMIER(c_1)

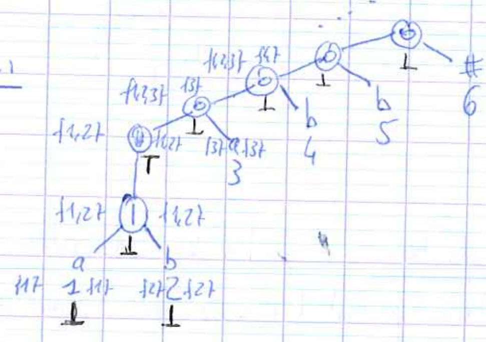
* DERNIER(n) : idem

* SUIVANT(p) : il n'y a que deux manières d'avoir deux positions qui se suivent :
  - si n est un noeud $c_1 \circ c_2$, alors pour tout $p \in$ DERNIER(c_1), toutes les positions dans PREMIER(c_2) appartiennent à SUIVANT(p)
  - si n est un noeud c_1^*, alors pour toute position $p \in$ DERNIER(c_1), toutes les positions dans PREMIER(c_1) sont dans SUIVANT(p).

Position p	SUIVANT(p)
1	$\{1,2,3\}$
2	$\{1,2,3\}$
3	$\{4\}$
4	$\{5\}$
5	$\{6\}$
6	$\emptyset$



• Retour à l'exemple :



• Construction d'un "graphe étiqueté" (automate non-dét) :
- états initial : positions dans PREMIER(racine)
- $\delta(i,a) = j$ ssi $j \in$ SUIVANT(i) et la position i correspond à la lettre a.

• Déterminisation classique avec les parties : le calcul de SUIVANT nous permet de ne pas avoir à parcourir l'automate.
Par exemple, pour trouver le successeur de l'état initial $\{1,2,3\}$ en lisant la lettre a, on garde d'abord les états correspondant à un a, soit $\{2,3\}$ et on fait l'union de leur SUIVANT, soit $\{1,2,3,4\}$.