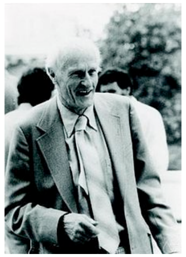
2 - Théorème de Kleene



Stephen C. Kleene
Mathématicien et logicien américain
1909-1994

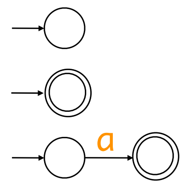
Théorème de Kleene (1968)

Théorème : Un langage sur un alphabet Σ est rationnel si et seulement si il est reconnu par un automate fini.

<u>Premier sens</u> * (preuve par induction)

- on cherche des A.F.N pour chaque langage rationnel dénoté par les expressions régulières de base

Expressions régulières
Définition inductive (B) $\varnothing \in$ ER $\varepsilon \in$ ER $a \in$ ER, pour tout $a \in \Sigma$ (I) ...

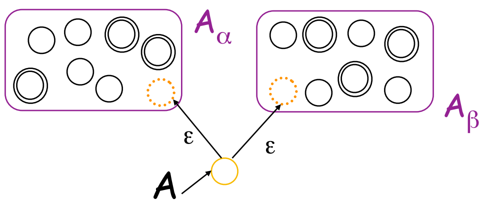


* Inspiré de l'algorithme de Thompson pour passer d'une E.R. à un A.F.N.

Suite preuve (opération +)

- <u>par hypothèse d'induction</u>,
on suppose qu'il existe deux A.F.N. A_α et A_β tels que
L(A_α) est le langage décrit par l'expression régulière α et
L(A_β) est le langage décrit par l'expression régulière β.

Expressions régulières
Définition inductive (B) ... (I) si $\alpha, \beta \in$ ER alors $(\alpha + \beta) \in$ ER ...



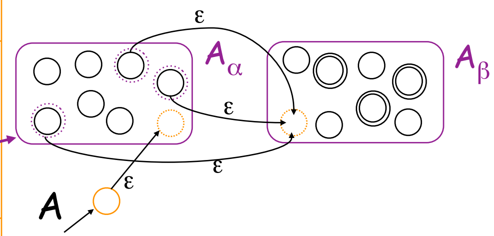
(nouvel état initial raccordé à ceux de A_α et A_β)

- A reconnaît le langage décrit par $(\alpha + \beta)$

Suite preuve (opération .)

- A reconnaît le langage décrit par $(\alpha . \beta)$

Expressions régulières
Définition inductive (B) ... (I) si $\alpha, \beta \in$ ER alors ... $(\alpha . \beta) \in$ ER ...



(le nouvel état initial est raccordé à celui de A_α , les anciens états finals de A_α sont raccordés à l'ancien état initial de A_β , les états finals de A_β sont les seuls états finals dans A)

Suite preuve (opération *)

✦ A reconnaît le langage décrit par α^*

Expressions régulières

Définition inductive

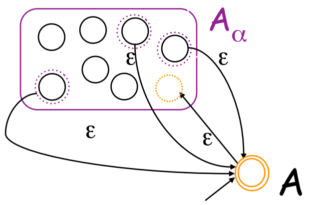
(B) ...

(I) si $\alpha, \beta \in$ ER alors

...

...

$\alpha^* \in$ ER

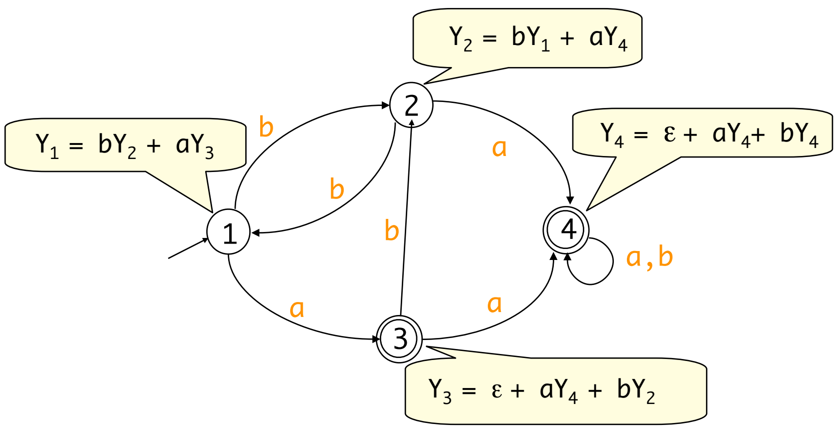


(le nouvel état initial de A est raccordé à l'ancien de A_α , les anciens états finals de A_α sont raccordés au nouvel état initial de A , l'état initial de A est le seul état final de A)

Passer d'un A.F.D. à une E.R.

- Réciproquement : à partir d'un A.F.D, peut-on construire une expression régulière qui décrive le langage reconnu par A ?
- Les systèmes d'équations linéaire permettent de ramener le calcul d'une expression régulière à la résolution d'un système d'équations symboliques.
- A chaque état q accessible et co-accessible on associe le langage $L_q(A)$ qui serait le langage reconnu par A si q était son seul état initial.
- On cherche alors une expression régulière Y_q dénotant le langage $L_q(A)$.
- On obtient un système d'équations dont les inconnues sont des expressions régulières dénotant des langages (rationnels donc).

Exemple



Y_q décrit le langage des mots w tels que $\delta^*(q,w) \in F$

δ^* : prolongement de δ aux mots

Exemple

$$\begin{cases} Y_1 = bY_2 + aY_3 \\ Y_2 = bY_1 + aY_4 \\ Y_3 = \varepsilon + aY_4 + bY_2 \\ Y_4 = \varepsilon + (a+b)\,Y_4 \end{cases}$$

Résoudre un tel système revient ici à calculer Y_1 car il est associé à l'état initial 1.

On procède par substitutions et c'est le lemme d'Arden qui va nous permettre de résoudre l'équation :

$$Y = A\,Y + B$$

Preuve du lemme d'Arden

Soit à résoudre l'équation $Y = AY + B$:

- $A^* B$ est bien solution :

  $A^*B = A\,A^*B + B = A^+ B + B = A^*B$

- $A^* B$ est une solution minimale :

  si Y est une solution, alors $A^* B \subseteq Y$

  Récurrence sur la hauteur d'étoile :
  - si $i=0$, $A^0 B = B \subseteq Y$ car $Y = AY + B$
  - hyp. réc. pour $i=n$: $A^n B \subseteq Y$
  - pour $n+1$: $A^{n+1} B = A\,A^n B \subseteq A\,Y \subseteq A\,Y + B = Y$ cqfd

Preuve du lemme d'Arden (suite)

- si $\varepsilon \notin A$ alors A^*B est l'unique solution.

  On suppose la non-unicité de la solution A^*B.
  Soit X une autre solution
  et soit un mot w de longueur minimale tel que $w \in X \setminus A^*B$.

  $w \in X = AX + B$ et $w \notin B$ donc $w = u\,v$ avec $u \in A$ et $v \in X$.
  Or $v \notin A^*B$ (sinon w aussi) donc $v \in X \setminus A^*B$.
  Contradiction : la longueur de w était supposée minimale dans $X \setminus A^*B$!

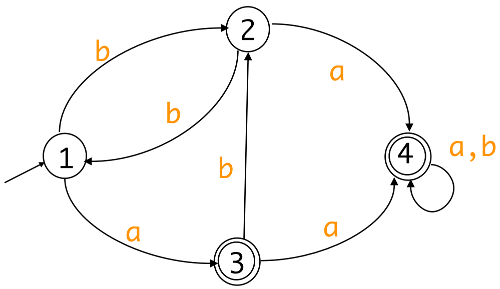
- si $\varepsilon \in A$ alors pour tout $C \subseteq \Sigma^*$, $Y = A^*B + A^*C$ est aussi solution :

  $A^*B + A^*C = A\,(A^*B + A^*C) + B = A^+B + A^+C + B = A^*B + A^*C$

- cela dit, si l'automate de départ est déterministe,
  on a toujours : $\varepsilon \notin A$.

Exemple

On garde toujours à l'esprit que si $\varepsilon \notin A$ alors A^*B est l'unique solution de $Y=AY+B$.



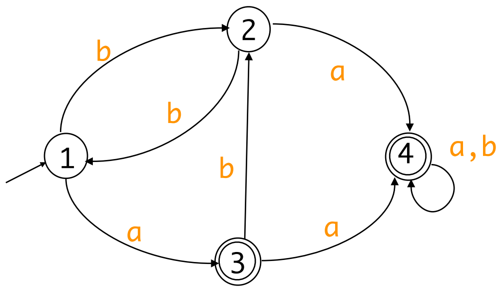
$$\begin{cases} Y_1 = bY_2 + aY_3 \\ Y_2 = bY_1 + aY_4 \\ Y_3 = \varepsilon + aY_4 + bY_2 \\ Y_4 = \varepsilon + (a+b)Y_4 \end{cases}$$

$\Rightarrow$

$Y_4 =$???

Exemple

On garde toujours à l'esprit que si $\varepsilon \notin A$ alors A^*B est l'unique solution de $Y=AY+B$.



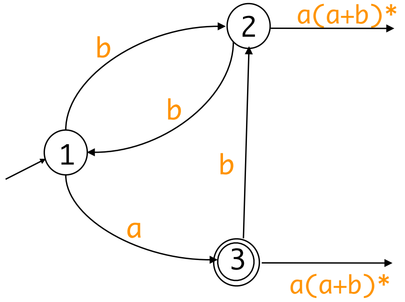
$$\begin{cases} Y_1 = bY_2 + aY_3 \\ Y_2 = bY_1 + aY_4 \\ Y_3 = \varepsilon + aY_4 + bY_2 \\ Y_4 = \varepsilon + (a+b)Y_4 \end{cases}$$

$\Rightarrow$

$Y_4 = (a+b)^*$

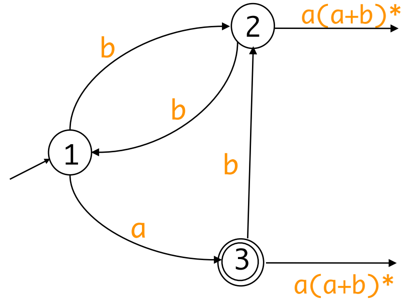
(Y_4 est alors éliminé)

Exemple



$$\begin{cases} Y_1 = bY_2 + aY_3 \\ Y_2 = bY_1 + a(a+b)^* \\ Y_3 = \varepsilon + a(a+b)^* + bY_2 \end{cases}$$

Exemple



$$\begin{cases} Y_1 = bY_2 + aY_3 \\ Y_2 = bY_1 + a(a+b)^* \\ Y_3 = \varepsilon + a(a+b)^* + bY_2 \end{cases}$$

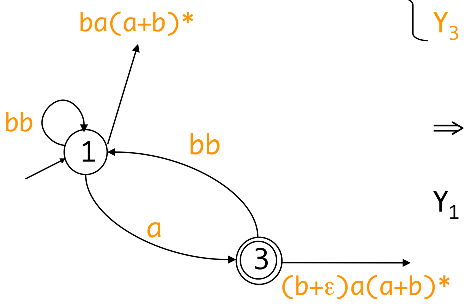
$$\Rightarrow$$

$$\begin{cases} Y_1 = bbY_1 + ba(a+b)^* + aY_3 \\ Y_3 = \varepsilon + a(a+b)^* + bbY_1 + ba(a+b)^* \end{cases}$$

(Y_2 est alors éliminé)

Exemple

On garde encore à l'esprit que si $\varepsilon \notin A$ alors A^*B est l'unique solution de $Y = AY + B$.



$$\begin{cases} Y_1 = bbY_1 + ba(a+b)^* + aY_3 \\ Y_3 = \varepsilon + bbY_1 + (b+\varepsilon)a(a+b)^* \end{cases}$$

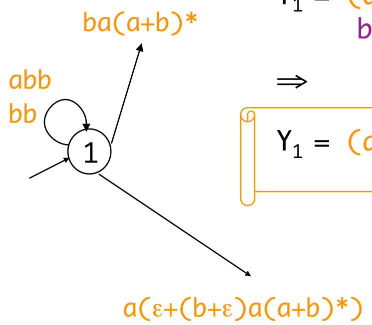
$$\Rightarrow$$

$$Y_1 = bbY_1 + ba(a+b)^* + a\varepsilon + abbY_1 + a(b+\varepsilon)a(a+b)^*$$

(Y_3 est alors éliminé)

Exemple

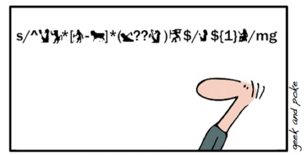
On garde encore et toujours à l'esprit que si $\varepsilon \notin A$ alors A^*B est l'unique solution de $Y = AY + B$.



$$Y_1 = (abb+bb)\, Y_1 + ba(a+b)^* + a(\varepsilon+(b+\varepsilon)a(a+b)^*)$$

$$\Rightarrow$$

$$Y_1 = (abb+bb)^*(ba(a+b)^* + a(\varepsilon+(b+\varepsilon)a(a+b)^*))$$



ANCIENT EGYPTIAN REGEXP

Dénombrabilité

RAPPEL !!!

- un alphabet Σ est un ensemble fini (ou dénombrable) de symboles
- l'ensemble des mots Σ^* sur l'alphabet Σ
  - ➡ (infini) dénombrable
- l'ensemble des langages rationnels sur l'alphabet Σ
  - ➡ (infini) dénombrable (cf. TD 2)
- l'ensemble des langages sur l'alphabet Σ
  - ➡ non-dénombrable

donc il existe des langages non rationnels ...

(in fine, toutes les fonctions ne seront pas calculables ...)

Preuve de non-dénombrabilité

(argument de la diagonale de Cantor)

	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8	σ_9	σ_{10}	σ_{11}	σ_{12}	σ_{13}	σ_{14}	...
L_0	$\boxed{1}$	0	0	0	1	0	0	1	0	0	1	0	0	1	0	...
L_1	1	$\boxed{0}$	0	0	0	0	0	1	0	1	0	0	0	0	0	...
L_2	0	1	$\boxed{0}$	0	1	1	1	0	0	0	0	1	0	0	0	...
L_3	0	1	0	$\boxed{1}$	0	0	0	0	0	0	0	0	0	1	0	...
L_4	0	0	1	0	$\boxed{0}$	0	1	0	0	1	0	0	0	0	0	...
L_5	1	0	1	0	1	$\boxed{1}$	0	0	0	0	0	1	0	1	0	...
$\vdots$																

- ✦ soit L le langage tel que :

$$\sigma_i \in L \text{ ssi } \sigma_i \notin L_i$$

- ✦ ici $L = \{\sigma_1,\ \sigma_2,\ \sigma_4,\ \dots\}$
- ✦ L ne figure pas dans cette tentative de lister les langages en les numérotant en indice par les entiers donc cette liste est incomplète
- ✦ pour n'importe quelle liste, on pourrait exhiber un langage qui n'y appartiendrait pas

$\Rightarrow$ l'ensemble des langages sur Σ n'est pas dénombrable.