Petit cours sur les fonctions séquentielles

Jean-Éric Pin

LIAFA, CNRS et Université Denis Diderot

Mai 2013, Sainte-Marie de Ré

Outline

(1) Fonctions séquentielles

(2) Une caractérisation des transducteurs séquentiels

(3) Composition de transducteurs séquentiels

(4) Le principe du produit en couronne

Première partie I

Fonctions séquentielles

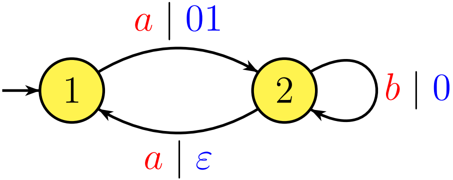
Définitions informelles

Un transducteur est un automate muni d'une fonction de sortie. Un transducteur calcule une relation sur $A^* \times B^*$.

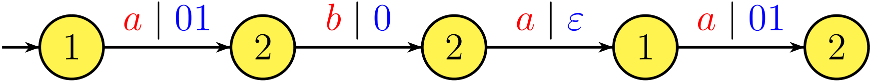
Un transducteur séquentiel est un transducteur dont l'automate sous-jacent est déterministe (mais pas nécessairement complet). Il calcule une fonction (partielle) de A^* dans B^*.

Un transducteur séquentiel pur calcule une fonction (partielle) φ préservant les préfixes : si u est un préfixe de v, alors $\varphi(u)$ est un préfixe de $\varphi(v)$.

An exemple de transducteur séquentiel pur



Sur l'entrée $abaa$, la sortie est 01001.

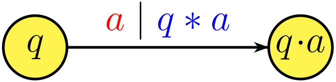


Transducteurs séquentiels purs

Un transducteur séquentiel pur est un 6-uplet

$$\mathcal{A} = (Q, A, B, i, \cdot, *)$$

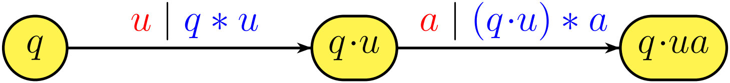
où la fonction d'entrée $(q, a) \to q{\cdot}a \in Q$ et la fonction de sortie $(q, a) \to q * a \in B^*$ sont définies sur le même domaine $D \subseteq Q \times A$.



Extensions des fonctions de transition et de sortie

La fonction de transition s'étend à $Q \times A^* \to Q$ en posant $q{\cdot}\varepsilon = q$ et, si $q{\cdot}u$ et $(q{\cdot}u){\cdot}a$ sont définies, $q{\cdot}(ua) = (q{\cdot}u){\cdot}a$.

La fonction de sortie s'étend à $Q \times A^* \to B^*$ en posant $q * \varepsilon = \varepsilon$ et, si $q * u$ et $(q{\cdot}u) * a$ sont définies, $q * (ua) = (q * u)((q{\cdot}u) * a)$.



Fonctions séquentielles pures

La fonction $\varphi\colon A^* \to B^*$ définie par

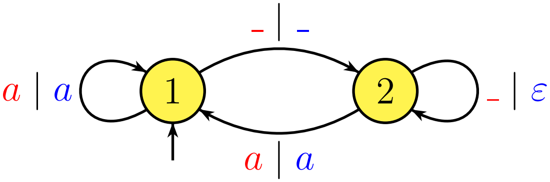
$$\varphi(u) = i * u$$

est appelée la fonction réalisée par $\mathcal{A}$.

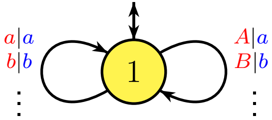
Une fonction est séquentielle pure si elle peut être réalisée par un transducteur séquentiel pur.

Exemples de fonctions séquentielles pures

Remplacer des espaces consécutifs par un seul espace :



Convertir des majuscules en minuscules :

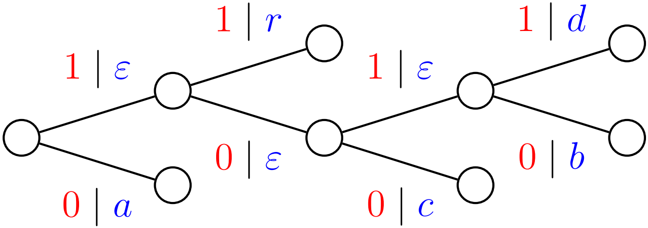


Codage et décodage

Pour le codage

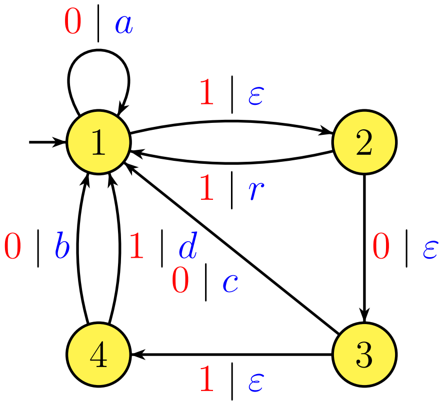
$$a \to 0 \quad b \to 1010 \quad c \to 100 \quad d \to 1011 \quad r \to 11$$

la fonction de décodage correspondante est



Décodage

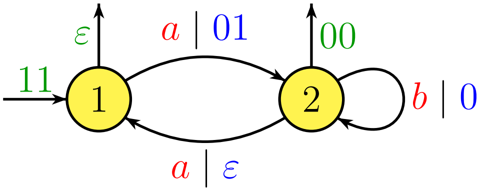
$a \to 0 \quad b \to 1010 \quad c \to 100 \quad d \to 1011 \quad r \to 11$



$01010110100010110101010110 \to abracadabra$

Transducteurs séquentiels : définition informelle

Un transducteur séquentiel est un transducteur dont l'automate sous-jacent est déterministe (mais pas nécessairement complet). Il y a un préfixe initial et une fonction terminale.



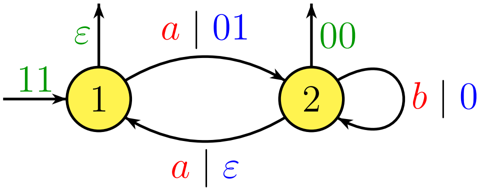
Pour l'entrée $abaa$, la sortie est 110100100.

Transducteurs séquentiels

Un transducteur séquentiel est un 8-uplet

$$\mathcal{A} = (Q, A, B, i, \cdot, *, m, \rho)$$

où $(Q, A, B, i, \cdot, *)$ est un transducteur séquentiel pur, $m \in B^*$ est le préfixe initial et $\rho \colon Q \to B^*$ est une fonction (partielle), appelée la fonction terminale.



Fonctions séquentielles

La fonction $\varphi\colon A^* \to B^*$ définie par

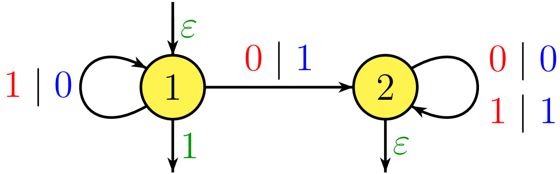
$$\varphi(u) = m(i * u)\rho(i{\cdot}u)$$

est appelée la fonction réalisée par $\mathcal{A}$.

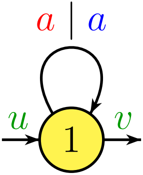
Une fonction est séquentielle si elle peut être réalisée par un transducteur séquentiel.

Quelques exemples de fonctions séquentielles

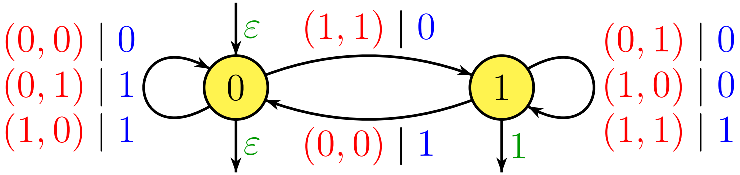
La fonction $x \to x + 1$ (en binaire renversé)



La fonction $\varphi : A^* \to A^*$ définie par $\varphi(x) = uxv$.



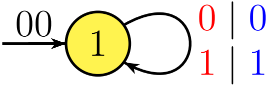
Addition (en binaire inversé)



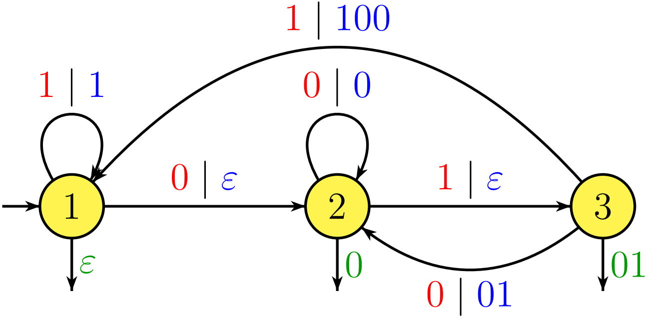
En binaire inversé, $22 = 2 + 4 + 16 \to 01101$ et $13 = 1 + 4 + 8 \to 10110$. En prenant comme entrée $(0,1)(1,0)(1,1)(0,1)(1,0)$, la sortie est 110001, la representation de $35 = 1 + 2 + 32$ en binaire inversé.

Autres exemples

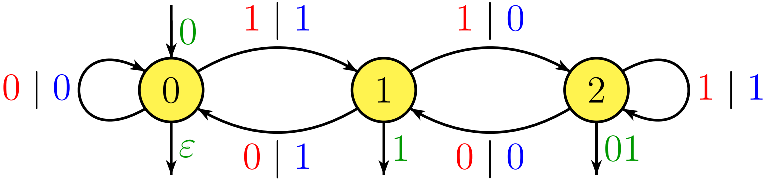
Multiplication par 4



Remplacer chaque occurrence de 011 par 100.

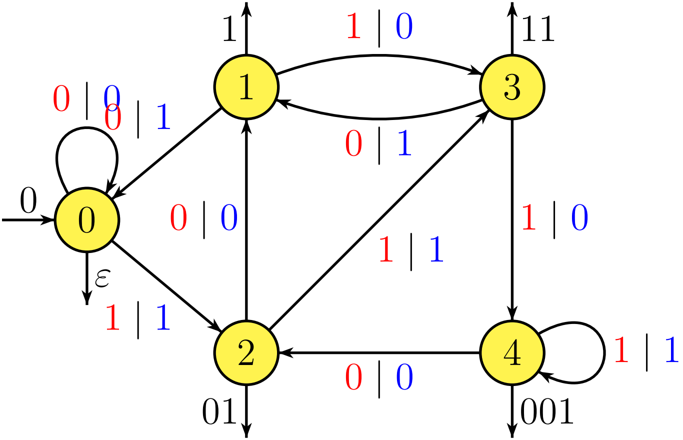


Multiplication par 6



$185 = 1 + 8 + 16 + 32 + 128$ et
$6 \times 185 = 1110 = 2 + 4 + 16 + 64 + 1024$.
Donc $\varphi(10011101) = 0110101001$

Multiplication par 10

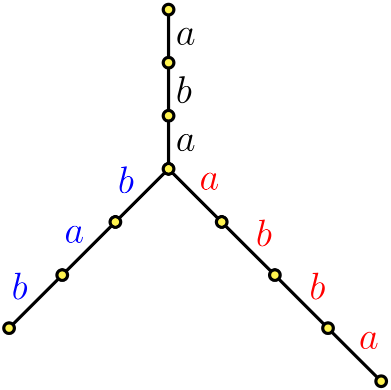


Deuxième partie II

Une caractérisation

La distance géodésique

La distance entre $ababab$ et $abaabba$ est 7.



La distance géodesique (2)

Notons $u \wedge v$ le plus long préfixe commun des mots u et v. Alors

$$d(u,v) = |u| + |v| - 2|u \wedge v|$$

Exemple : $d(ababab, abaabba) = 6 + 7 - 2 \times 3 = 7$.
On peut montrer que d est une distance :

(1) $d(u,v) = 0$ iff $u = v$,
(2) $d(u,v) = d(v,u)$,
(3) $d(u,v) \leqslant d(u,w) + d(w,v)$.

Une caractérisation des fonctions séquentielles

Une fonction $\varphi : A^* \to B^*$ est lipschitzienne s'il existe un $K > 0$ tel que, pour tout $u, v \in A^*$,

$$d(\varphi(u), \varphi(v)) \leqslant K d(u, v)$$

Théorème (Choffrut 1979)

Soit $\varphi : A^* \to B^*$ une fonction dont le domaine est préfixiel. Sont équivalents :

(1) *φ est séquentielle,*

(2) *φ est lipschitzienne, et φ^{-1} préserve les langages réguliers.*

Une caractérisation de fonctions séquentielles pures

Théorème (Ginsburg-Rose 1966)

Soit $\varphi : A^* \to B^*$ *une fonction dont le domaine est préfixiel. Sont équivalents :*

(1) φ *est une fonction séquentielle pure,*

(2) φ *est Lipschitzienne et préserve les préfixes, et* φ^{-1} *preserve les langages réguliers.*

Troisième partie III

Composition

Composition de transducteurs séquentiels purs

Théorème

Les fonctions séquentielles pures sont fermées par composition.

Soient σ et τ des fonctions séquentielles pures réalisées par les transducteurs

$$\mathcal{A} = (Q, A, B, q_0, \cdot, *) \text{ et } \mathcal{B} = (P, B, C, p_0, \cdot, *)$$

Le produit en couronne de $\mathcal{B}$ par $\mathcal{A}$ est obtenu en prenant pour entrée de $\mathcal{B}$ la sortie de $\mathcal{A}$. Ce transducteur calcule $\tau \circ \sigma$.

Produit en couronne de transducteurs séqu. purs

Le produit en couronne est défini par

$$\begin{aligned}
\mathcal{B} \circ \mathcal{A} &= (P \times Q, A, C, (p_0, q_0), \cdot, *) \\
(p,q) \cdot a &= (p \cdot (q * a), q \cdot a) \\
(p,q) * a &= p * (q * a)
\end{aligned}$$

$$(p,q) \xrightarrow{\; a \mid p * (q * a) \;} (p \cdot (q * a), q \cdot a)$$

Composition de deux transducteurs séquentiels

Théorème

Les fonctions séquentielles sont fermées par composition.

Soient deux fonctions séquentielles réalisées par les transducteurs $\mathcal{A}$ (muni du mot initial n et de la fonction terminale ρ) et $\mathcal{B}$ (muni du mot initial m et de la fonction terminale σ).

Le produit en couronne de $\mathcal{B}$ par $\mathcal{A}$ est obtenu en prenant $m(p_0 * n)$ comme mot initial et $\omega(p,q) = (p * \rho(q))\sigma(p \cdot \rho(q))$ comme fonction terminale.

Itération de fonctions séquentielles…

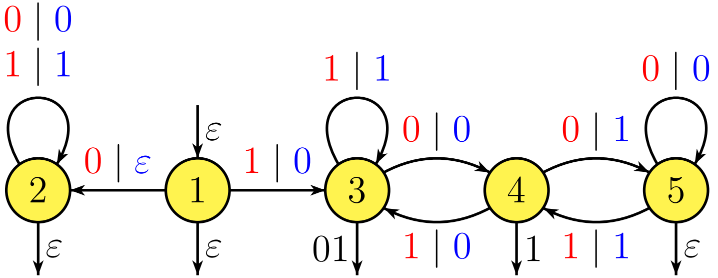
Elle peut conduire à des problèmes très difficiles …

Soit $f(n) = \begin{cases} 3n+1 & \text{si } n \text{ est impair} \\ n/2 & \text{si } n \text{ est pair} \end{cases}$

On conjecture que pour tout $n > 0$, il existe k tel que $f^k(n) = 1$. Le problème est toujours ouvert. La conjecture a été vérifiée pour $n \leqslant 5 \times 2^{60}$.

Transducteur minimal de la fonction $3n+1$

Soit $f(n) = \begin{cases} 3n+1 & \text{if } n \text{ est impair} \\ n/2 & \text{if } n \text{ est pair} \end{cases}$



Itération de la fonction $3n + 1$...

31, 94, 47, 142, 71, 214, 107, 322, 161, 484, 242, 121, 364, 182, 91, 274, 137, 412, 206, 103, 310, 155, 466, 233, 700, 350, 175, 526, 263, 790, 395, 1186, 593, 1780, 890, 445, 1336, 668, 334, 167, 502, 251, 754, 377, 1132, 566, 283, 850, 425, 1276, 638, 319, 958, 479, 1438, 719, 2158, 1079, 3238, 1619, 4858, 2429, 7288, 3644, 1822, 911, 2734, 1367, 4102, 2051, 6154, 3077, 9232, 4616, 2308, 1154, 577, 1732, 866, 433, 1300, 650, 325, 976, 488, 244, 122, 61, 184, 92, 46, 23, 70, 35, 106, 53, 160, 80, 40, 20, 10, 5, 16, 8, 4, 2, 1.

Un résultat utile

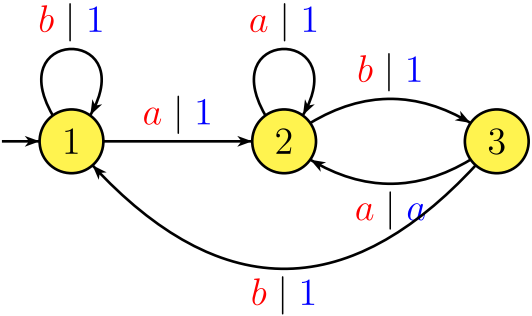
Soit $\varphi : A^* \to B^*$ une fonction séquentielle pure réalisée par $\mathcal{A} = (Q, A, B, q_0, \cdot, *)$. Soit L un langage régulier de B^* reconnu par $\mathcal{B} = (P, B, \cdot, p_0, F)$. Le produit en couronne de $\mathcal{B}$ par $\mathcal{A}$ est le transducteur séquentiel pur $\mathcal{B} \circ \mathcal{A} = (P \times Q, A, (p_0, q_0), \cdot)$ défini par $(p, q){\cdot}a = (p{\cdot}(q * a), q{\cdot}a)$.

Théorème

La langage $\varphi^{-1}(L)$ est reconnu par $\mathcal{B} \circ \mathcal{A}$.

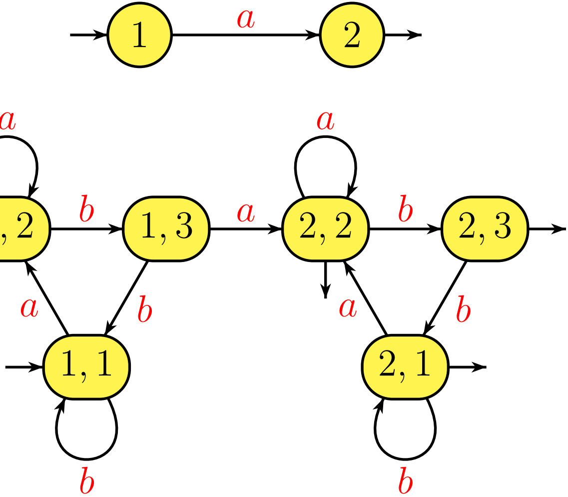
Nombre d'occurences de aba

Soit $\varphi(u) = a^n$, où n est le nombre d'occurrences de aba in u. Cette fonction est séquentielle pure :



Alors $\varphi^{-1}(a)$ est l'ensemble des mots contenant exactement une occurrence de aba.

Produit en couronne de deux automates



Produit en couronne de monoïdes

Idée : remplacer les automates par des monoïdes.

Le produit en couronne $M \circ N$ de deux monoïdes M et N est défini sur l'ensemble $M^N \times N$ par le produit

$(f_1, k_1)(f_2, k_2) = (f, k_1k_2)$ avec $f(k) = f_1(k)f_2(kk_1)$

Décompositions en produit en couronne

Théorème

(1) *Tout groupe résoluble divise un produit en couronne de groupes commutatifs,*

(2) *Tout monoïde $\mathcal{R}$-trivial divise un produit en couronne de copies de U_1,*

(3) *Tout monoïde apériodique divise un produit en couronne de copies de U_2,*

(4) *Tout monoïde divise un produit en couronne de groupes simples et de copies de U_2,*

Trace séquentielle d'un automate

Soit $\mathcal{A} = (Q, A, q_0, F, \cdot)$ un automate déterministe.
Soit $B = Q \times A$ et soit $\sigma \colon A^* \to B^*$ la fonction séquentielle pure définie par

$$\sigma(a_1 \cdots a_n) = (q_0, a_1)(q_0{\cdot}a_1, a_2) \cdots (q_0{\cdot}a_1 \cdots a_{n-1}, a_n)$$

$$q \xrightarrow{\ a \ \mid\ (q, a)\ } q{\cdot}a$$

Cette fonction est appelée trace séquentielle de $\mathcal{A}$.

Le principe du produit en couronne

Le principe du produit en couronne de Straubing donne une description des langages reconnus par la produit en couronne de deux monoïdes.

Proposition

Tout langage de A^* reconnu par $M \circ N$ est union finie de langages de la forme $U \cap \sigma^{-1}(V)$, où $\varphi : A^* \to N$ est un morphisme de monoïde reconnaissant U, σ est la trace séquentielle associée à φ et V est un langage de $(A \times N)^*$ reconnu par M.

Version automate

Théorème

Soit $L \subseteq A^*$ *un langage reconnu par un produit en couronne de la forme* $(P, Q \times A) \circ (Q, A)$. *Alors* L *est union finie de langages de la forme* $W \cap \sigma^{-1}(V)$, *où* $W \subseteq A^*$ *est reconnu by* (Q, A), σ *est la trace séquentielle associée à l'action* (Q, A) *et* $V \subseteq (Q \times A)^*$ *est reconnu par* $(P, Q \times A)$.

Applications du principe du produit en couronne

La classe des langages reconnus par un monoïde $\mathcal{R}$-trivial est la plus petite algèbre de Boole fermée pour l'opération $L \to LaA^*$.

La classe des langages reconnus par un groupe résoluble est la plus petite algèbre de Boole fermée pour les opérations $L \to (LaA^*)_{r,n}$.

La classe des langages reconnus par un monoïde apériodique est la plus petite algèbre de Boole fermée pour les opérations $L \to LaA^*$ et $L \to A^*aL$.

L'opération $L \to LaA^*$

Soit $\mathcal{A} = (Q, A, q_0, F, \cdot)$ un automate reconnaissant L. Soit $B = Q \times A$ et soit $\sigma\colon A^* \to B^*$ la trace séquentielle de $\mathcal{A}$

$$\sigma(a_1 \cdots a_n) = (q_0, a_1)(q_0{\cdot}a_1, a_2) \cdots (q_0{\cdot}a_1 \cdots a_{n-1}, a_n)$$

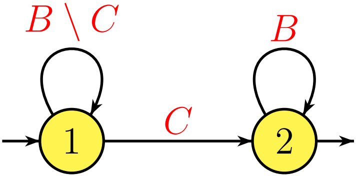
Soit $a \in A$ et soit $C = F \times \{a\} \subseteq B$. Alors

$$\sigma^{-1}(B^*CB^*) = LaA^*$$

En effet, $(q_0{\cdot}a_1 \cdots a_{k-1}, a_k) \in C$ signifie $a_k = a$ et $q_0{\cdot}a_1 \cdots a_{k-1} \in F$, i.e. $a_1 \cdots a_{k-1} \in L$.

Conséquence

Donc $\mathcal{B} \circ \mathcal{A}$ reconnaît LaA^*, où $\mathcal{B}$ est l'automate minimal de B^*CB^*.



En logique temporelle linéaire, l'opération duale $L \to A^*aL$ correspond à $\varphi \to F(\mathbf{a} \wedge X\varphi)$.

L'opération $L \to La$

Soit $\mathcal{A} = (Q, A, q_0, F, \cdot)$ un automate reconnaissant L. Soit $B = Q \times A$ et soit $\sigma\colon A^* \to B^*$ la trace séquentielle de $\mathcal{A}$

$$\sigma(a_1 \cdots a_n) = (q_0, a_1)(q_0 {\cdot} a_1, a_2) \cdots (q_0 {\cdot} a_1 \cdots a_{n-1}, a_n)$$

Soit $a \in A$ et soit $C = F \times \{a\} \subseteq B$. Alors

$$\sigma^{-1}(B^* C) = La$$

En effet, $(q_0 {\cdot} a_1 \cdots a_{n-1}, a_n) \in C$ signifie $a_n = a$ et $q_0 {\cdot} a_1 \cdots a_{n-1} \in F$, i.e. $a_1 \cdots a_{n-1} \in L$.

Produits à compteur

On note $(L_0aL_1)_{r,n}$ l'ensemble des mots u tels que le nombre de factorisations de u de la forme $u = u_0au_1$ avec $u_0 \in L_0$ et $u_1 \in L_1$, est congruent à r modulo n.

Alors $(LaA^*)_{r,n}$ est reconnu par $C_n \circ M$, où C_n est le groupe cyclique d'ordre n.

Langages sans-étoile

On note U_2 le monoïde $\{1, a_1, a_2\}$ defini par $a_1a_1 = a_2a_1 = a_1$ et $a_1a_2 = a_2a_2 = a_2$.

Il est facile de voir que tout langage de B^* reconnu par U_2 est une combinaison booléenne de langages de la forme B^*bC^*, où $b \in B$ et $C \subseteq B$.

Théorème

Soit T un monoïde. Tout langage de A^* reconnu par $U_2 \circ T$ est une combinaison booléenne de langages de la forme K ou $Ka(LbA^*)^c$ où $a, b \in A$ et K et L sont reconnus par T.

La formule clé

Soit L reconnu par $W = U_2 \circ T$. Soit $\eta : A^* \to W$, $\pi : W \to T$ et $\varphi = \pi \circ \eta : A^* \to T$. Soit $B = T \times A$ et soit $\sigma : A^* \to B^*$ la trace séquentielle définie par φ :

$$\sigma(a_1 a_2 \cdots a_n) = (1, a_1)(\varphi(a_1), a_2) \cdots (\varphi(a_1 \cdots a_{n-1}), a_n)$$

Alors, pour tout $C \subseteq B = T \times A$,

$$\sigma^{-1}(B^* b C^*) = \varphi^{-1}(m) a \Big(\bigcup_{(n,c) \notin C} \varphi^{-1}\Big(\big(m\varphi(a)\big)^{-1} n\Big) c A^* \Big)^c$$