

Automates en OCaml

On se donne le type `letter = A | B` et l'alias `word = letter list`.

1. Donnez une fonction `cmp_letter` dont la valeur de retour est nulle si ses deux arguments sont égaux, négative si la première lettre argument précède strictement la seconde et est positive sinon.
2. Donnez une fonction `cmp_word` qui implémente l'ordre lexicographique.
3. Proposez un type OCaml permettant de modéliser un dictionnaire dont les clefs sont des mots de Σ^* et les valeurs sont des éléments quelconques.

Le terme automate s'entendra dans le sens d'un automate fini déterministe complet ici, noté $(Q, q_0, \delta, \mathbb{1}_F)$. Dans la suite, on utilisera les types `state = word` et

```
automaton = {initial: state; transitions: state -> letter -> state; finals: state -> bool} .
```

4. Définir une fonction `delta_star` dont la valeur de renvoi est $\delta^*(q, w)$.
5. Définir une fonction `accepts` dont la valeur de renvoi est le booléen indiquant si le mot d'entrée est reconnu par l'automate d'entrée.

Deux mots u_1, u_2 sont séparés par v dans L si :

$$\mathbb{1}_L(u_1v) \neq \mathbb{1}_L(u_2v)$$

6. Définir une fonction qui vérifie si deux mots sont séparés par un troisième dans un langage reconnu par un automate donné.
7. Montrer que la relation d'inséparabilité par rapport à un langage L est une relation d'équivalence.
8. Montrer que $\delta_{\mathcal{A}}^*(q_0, u_1) = \delta_{\mathcal{A}}^*(q_0, u_2)$ implique l'inséparabilité par rapport au langage reconnu par l'automate $\mathcal{A}$.
9. En déduire que le nombre de classes d'équivalences de la relation d'inséparabilité par rapport à L est fini.

Ambiguïté et concision

Un automate sur Σ est un tuple $\mathcal{A} = (Q, T, I, F)$. Ici, I et F sont des ensembles et T est un ensemble de triplets état-lettre-état.

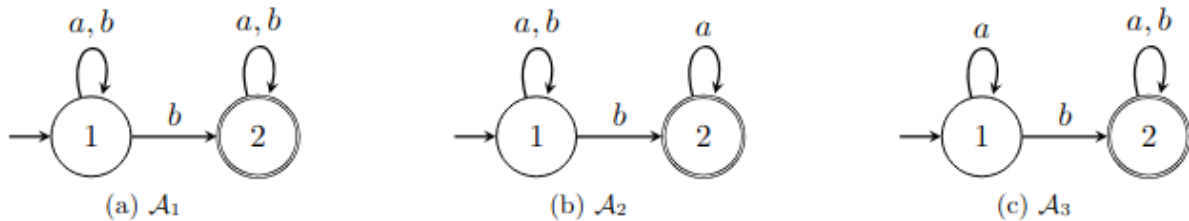


FIGURE 1 – Trois automates reconnaissant le même langage

Un calcul de $\mathcal{A}$ sur un mot $w = a_0 \dots a_{n-1}$ est une suite $q_0 \dots q_n \in Q^*$ finie d'états de Q telle que pour tout $i < n$, $(q_i, a_i, q_{i+1}) \in T$. Un calcul est acceptant si $q_n \in F$. Un mot w est ambigu sur $\mathcal{A}$ s'il existe au moins deux calculs acceptants différents de $\mathcal{A}$ sur w . On note $\text{Amb}(\mathcal{A})$ l'ensemble des mots ambigus sur $\mathcal{A}$. On dit que $\mathcal{A}$ est ambigu si $\text{Amb}(\mathcal{A})$ est non vide.

1. Les trois automates ci-dessus reconnaissent le même langage.
 - (i) Donnez (sans justifier) une expression régulière pour le langage ;
 - (ii) Pour chaque automate, justifiez s'il est déterministe ou non ;
 - (iii) Pour chaque automate, justifiez s'il est ambigu ou non.
2. Montrez qu'un automate déterministe n'est pas ambigu.
3. Justifiez rapidement qu'un automate co-déterministe (dont le miroir est déterministe) ne peut pas être ambigu.

Pour $n \in \mathbb{N}$ on pose Σ_n un alphabet à n lettres et K_n le langage des mots sur Σ_n^* dont au moins une lettre apparaît au moins deux fois.

4. Pour $\Sigma_3 = \{a, b, c\}$, donnez un automate acceptant K_3 et ayant au plus 5 états. Donnez un automate non ambigu acceptant K_3 et ayant au plus 9 états.
5. Donnez une expression régulière pour K_3 .
6. Montrez que pour tout $n \in \mathbb{N}$ il existe un automate avec au plus $n + 2$ états acceptant K_n .
7. Montrez que pour tout n il existe un automate non ambigu avec au plus $2^n + 1$ états acceptant K_n .
- 8.

Jusqu'à la fin de cette partie, on fixe un entier $n \geq 2$ et un automate non ambigu $\mathcal{A} = (Q, T, I, F)$ acceptant K_n . Soient u et v deux mots de Σ_n^* . Lorsque $u \cdot v \in K_n$, on note $q_{u,v}$ l'état de $\mathcal{A}$ atteint après avoir lu u lors de l'unique calcul acceptant de $\mathcal{A}$ sur $u \cdot v$.

9. Soient u, u' et v, v' quatre mots de Σ_n^* tels que $u \cdot v \in K_n$ et $u' \cdot v' \in K_n$. On suppose que $q_{u,v} = q_{u',v'}$.
 - (i) Montrez que $u \cdot v' \in K_n$ et $u' \cdot v \in K_n$.
 - (ii) Montrez que $q_{u,v} = q_{u,v'} = q_{u',v} = q_{u',v'}$.

Déterminisation et concision

Un automate sur Σ est un tuple $\mathcal{A} = (Q, T, I, F)$. Ici, I et F sont des ensembles et T est un ensemble de triplets état-lettre-état.

Un calcul de $\mathcal{A}$ sur un mot $w = a_0 \dots a_{n-1}$ est une suite $q_0 \dots q_n \in Q^*$ finie d'états de Q telle que pour tout $i < n$, $(q_i, a_i, q_{i+1}) \in T$. Un calcul est acceptant si $q_n \in F$. Un mot w est ambigu sur $\mathcal{A}$ s'il existe au moins deux calculs acceptants différents de $\mathcal{A}$ sur w . On note $\text{Amb}(\mathcal{A})$ l'ensemble des mots ambigus sur $\mathcal{A}$. On dit que $\mathcal{A}$ est ambigu si $\text{Amb}(\mathcal{A})$ est non vide.

On fixe l'alphabet $\Sigma = \{a, b\}$. Pour $n \geq 1$ un entier, on pose L_n le langage des mots dont la n -ième lettre en partant de la fin est un b .

1. Donnez un automate non ambigu acceptant L_3 , puis donnez un automate déterministe et complet acceptant L_3 .
2. Montrez que pour tout $n \geq 1$ entier, il existe un automate non ambigu $\mathcal{A}$ avec $n + 1$ états tel que $L(\mathcal{A}) = L_n$.
3. Montrez que pour tout $n \geq 1$ entier, il existe un automate déterministe et complet $\mathcal{B}$ avec 2^n états tel que $L(\mathcal{B}) = L_n$.

On veut à présent prouver que tout automate déterministe et complet acceptant L_n a au moins 2^n états.

4. Soit $\mathcal{B}$ un automate déterministe et complet. Soit $w \in \Sigma^*$. Montrez qu'il existe un unique calcul, non nécessairement acceptant, de $\mathcal{B}$ sur w .
5. Soit $\mathcal{B}$ un automate déterministe et complet et $w \in \Sigma^*$. On note q_w l'état atteint par l'unique calcul de $\mathcal{B}$ sur w , c'est-à-dire l'état q_m tel que $q_0 \dots q_m$ est un calcul de $\mathcal{B}$ sur w .
6. Soit $n \in \mathbb{N}$ et $\mathcal{B}$ un automate déterministe et complet reconnaissant L_n . Soient w et w' deux mots de Σ^* de longueur n . Montrez que $q_w = q_{w'}$ si et seulement si $w = w'$.
7. Soit $n \in \mathbb{N}$. Montrez que tout automate déterministe et complet reconnaissant L_n a au moins 2^n états.

Concision et ambiguïté - ENS Info Fondamentale 2023

Pour toutes les remarques ou corrections, vous pouvez m'envoyer un mail à galatee.hemery@gmail.com

1 Déterminisme et ambiguïté

Question 1.1. Il s'agit du langage des mots contenant au moins un b utilisant les lettres a et b .

Question 1.2. Les automates $\mathcal{A}_1$ et $\mathcal{A}_2$ sont non déterministes car $(1, b, 2) \in T$ et $(1, 2, 1) \in T$ pour ces deux automates.
L'automate $\mathcal{A}_3$ est déterministe car pour chaque couple état, lettre on a une unique transition.

Question 1.3. Soit $w \in L(\mathcal{A}_1)$.

Soit w s'écrit $a^i b a^j$ avec i, j des entiers naturels et le seul calcul acceptant est $\underbrace{1 \dots 1 2 \dots 2}_{i+1 \text{ fois } j+1 \text{ fois}}$.

Soit w s'écrit $w_1 b w_2 b w_3$ avec w_1, w_2, w_3 dans $\{a, b\}^*$ et on a au moins deux calculs acceptants : celui où l'on passe de l'état 1 à l'état 2 à l'aide du premier b et celui où l'on passe de l'état 1 à l'état 2 à l'aide du second.

Ainsi $\text{Amb}(\mathcal{A}_1) = \{a, b\}^* \{b\} \{a, b\}^* \{b\} \{a, b\}^*$.

Soit $w \in L(\mathcal{A}_2)$.

On peut écrire $w = w_1 b a^i$ en repérant le dernier b dans w .

L'unique calcul acceptant est celui qui utilise $(1, b, 2) \in T$ pour le dernier b . En effet, si la transition est utilisée avant, dans l'état 2, le calcul serait bloqué au b suivant.

Ainsi, $\text{Amb}(\mathcal{A}_2) = \emptyset$.

Dans $\mathcal{A}_3$, l'automate est déterministe et complet donc il existe un unique calcul pour un mot : pour chaque étape de calcul on suit l'unique transition correspondant au couple état courant, lettre suivante. Ainsi $\text{Amb}(\mathcal{A}_3) = \emptyset$.

Question 1.4. Soit $w = a_1 \dots a_n$ un mot de longueur n accepté par $\mathcal{A}$ où a_i désignent les lettres de w et $q_0 \dots q_n$ un calcul acceptant.

On sait donc $\forall i \in \{1, \dots, n\}, (q_{i-1}, a_i, q_i) \in T$.

On sait également que $q_0 \in I$ et $q_n \in F$.

On pose $\tilde{w} = a_n \dots a_0$. On a $q_n \in \tilde{I}$, $q_0 \in \tilde{F}$ et $\forall i \in \{1, \dots, n\}, (q_i, a_i, q_{i-1}) \in \tilde{T}$ donc $q_n \dots q_0$ est un calcul acceptant de $\tilde{\mathcal{A}}$ sur $\tilde{w}$.

Question 1.5. Soit $\mathcal{A}$ un automate ambigu, alors il existe w un mot de longueur n un entier tel que $q_0 \dots q_n$ et $q'_0 \dots q'_n$ sont des calculs acceptants distincts.

Ainsi, d'après la question précédente $q_n \dots q_0$ et $q'_n \dots q'_0$ sont des calculs acceptants distincts sur $\tilde{\mathcal{A}}$ pour $\tilde{w}$ donc $\tilde{\mathcal{A}}$ est ambigu.

On obtient la réciproque en remarquant que $\tilde{\tilde{\mathcal{A}}} = \mathcal{A}$.

Question 1.6. Si $\mathcal{A}$ est déterministe, il existe au plus un calcul de $\mathcal{A}$ pour un mot donné car pour chaque étape de calcul on suit l'unique transition correspondant au couple état courant, lettre suivante.

Ainsi, il ne peut exister deux calculs acceptants d'un même mot et $\mathcal{A}$ n'est pas ambigu.