Langages rationnels

option informatique

Introduction

Classification des langages formels

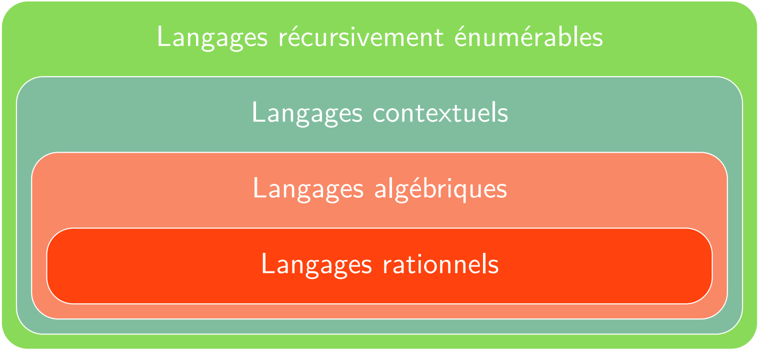
- Au concept de **langage formel**, il convient d'ajouter celui de **grammaire formelle**. Une grammaire définit un ensemble de règles syntaxiques visant à construire un langage formel.
- S'appuyant sur ce concept, en 1956, Noam Chomsky définit une classification des langages formels à quatre niveaux appelée **hiérarchie de Chomsky**.

Classification des langages formels

Du langage plus restrictif au langage plus général, la **hiérarchie de Chomsky** est la suivante.

- Les **langages rationnels** (type 3) peuvent être définis à partir d'**expressions rationnelles**. Nous verrons qu'ils sont reconnus par des **automates finis**. (Au programme)
- Les **langages algébriques** (type 2) sont reconnaissables par des automates à pile non déterministes. (HP)
- Les **langages contextuels** (type 1) sont reconnaissables par des machines de Turing non déterministes à ruban de longueur bornée. (HP)
- Les **langages récursivement énumérables** (type 0) sont reconnaissables par des machines de Turing. (HP)

Hiérachie de Chomsky



Expressions rationnelles

Expressions rationnelles

Soit Σ un alphabet complété par les symboles $\{\varepsilon, \varnothing, +, \cdot, *, (,)\}$ non éléments de Σ.

Définition 1

Une **expression rationnelle** est une formule construite de manière inductive sur le langage étendu $\Sigma \cup \{\varepsilon, \varnothing, +, \cdot, *, (,)\}$.

- $\varnothing$ est une expression rationnelle qui désigne le langage vide.
- ε est une expression rationnelle qui désigne le langage $\{\varepsilon\}$.
- Tout symbole a de Σ est une expression rationnelle qui désigne le langage $\{a\}$.
- Si r et s sont deux expressions rationnelles qui désignent les langages L_r et L_s, alors $r + s$, $r \cdot s$, r^* sont des expressions rationnelles qui désignent respectivement les langages $L_r \cup L_s$, $L_r \cdot L_s$, L^*.

Régles d'écriture

- Pour simplifier l'écriture des expressions rationnelles, on s'accorde sur les règles de priorité des opérateurs. Par priorité décroissante, on a : $*$, $\cdot$, $\cup$. Ces règles améliorent la lisibilité des formules.
- Par exemple, sur l'alphabet $\Sigma = \{a, b, c\}$, on écrit :

$$\begin{array}{rcl}
\{a\} \cdot (\{b\} \cdot \{c\})^* & \rightarrow & a(bc)^* \\
(\{a\}^* \cdot \{b\}) \cup \{c\} & \rightarrow & a^* b + c \\
\{\varepsilon\} \cup (\{a\} \cdot \{b\}^* \cdot \{a\}^*) & \rightarrow & \varepsilon + ab^* a^*
\end{array}$$

Langage dénoté par une expression rationnelle

Définition 2

À toute expression rationnelle r, on associe le **langage $\mathcal{L}(r)$** par :

- Si $r = \varnothing$, $\mathcal{L}(r) = \varnothing$.
- Si $r = \varepsilon$, $\mathcal{L}(r) = \{\varepsilon\}$.
- Si $r = a \in A$, $\mathcal{L}(r) = \{a\}$.
- Si $r = (r_1 + r_2)$, $\mathcal{L}(r) = \mathcal{L}(r_1) \cup \mathcal{L}(r_2)$.
- Si $r = (r_1 \cdot r_2)$, $\mathcal{L}(r) = \mathcal{L}(r_1) \cdot \mathcal{L}(r_2)$.
- Si $r = (r_1)^*$, $\mathcal{L}(r) = \mathcal{L}(r_1)^*$.

On dit que $\mathcal{L}(r)$ est un **langage dénoté** par r.

Exemples

- Soit $\Sigma = \{a, b\}$. $r = aA^*$ est une expression rationnelle. Le langage $\mathcal{L}(r)$ associé est formé des mots commençant par a.
- Si $\Sigma = \{a, \ldots, z\}$, le langage formé des mots ayant pour facteur le mot acteur ou le mot actrice peut être décrit par les expressions rationnelles :
  - $\Sigma^*(\text{acteur} + \text{actrice})\Sigma^*$
  - $\Sigma^*\text{act}(\text{eur} + \text{rice})\Sigma^*$
  - $\Sigma^*\text{a}(\text{cteur} + \text{ctrice})\Sigma^*$

☛ **Pour s'entraîner :** exercices 1, 2 (td langages rationnels)

Expressions équivalentes

- Une expression rationnelle dénote un unique langage.
- Mais un langage donné peut être dénoté par plusieurs expressions rationnelles différentes.
- Déterminer si deux **expressions rationnelles** sont **équivalentes** ou trouver l'expression rationnelle la plus courte dénotant un langage donné sont des problèmes difficiles.

Définition 3

Deux expressions rationnelles r_1 et r_2 sur un alphabet Σ sont **équivalentes** si elles décrivent le même langage : $\mathcal{L}(r_1) = \mathcal{L}(r_2)$.

Par abus de notation, on note parfois $r_1 \equiv r_2$.

☛ **Pour s'entraîner :** exercice 3 (td langages rationnels)

  - arrtes@gmail.com

Langages rationnels

Langages rationnels

L'importance des **langages rationnels** est double.

- D'un **point de vue théorique**, ils forment une classe de langages suffisamment simple pour pouvoir être étudiée, suffisamment riche pour contenir toutes sortes de langages intéressants, suffisamment naturelle pour être définie de manières différentes, et suffisamment flexible pour être stable par un certain nombre d'opérations.
- D'un **point de vue pratique et informatique**, ils permettent un traitement algorithmique efficace de problèmes tels que la **recherche de motifs** dans un texte.

Langage rationnel

Définition 4

Soit Σ un alphabet. On dit qu'un **langage** $L \subset \Sigma^*$ est **rationnel** s'il existe une expression rationnelle r sur Σ telle $L = \mathcal{L}(r)$.
On note $\mathbf{Rat}(\Sigma)$ l'ensemble des langages rationnels sur Σ.

- L'ensemble $\mathrm{Rat}(\Sigma)$ sur A est la plus petite partie de A^* qui contient $\varnothing$, $\{\varepsilon\}$, $\{a\}$ pour tout $a \in \Sigma$ et qui est stable par réunion, concaténation et itération.
- Étant donnée une expression rationnelle r et un mot $u \in \Sigma^*$, un problème fondamental est de savoir si $u \in \mathcal{L}(r)$. La réponse à cette question est l'objet du **cours sur les automates**.

☛ **Pour s'entraîner :** exercices 4, 5, 6, 7 (td langages rationnels)