Université Paris Cité
Automates et Analyse Lexicale

UFR Informatique
Année 2023–2024

Handout 11

Le théorème de Myhill-Nerode

1 Rappel: Relations d'équivalence et congruences droites

1.1 Relations d'équivalence

Une relation $\sim$ sur un ensemble U est une relation d'équivalence si

1. $\sim$ est *réflexive*: pour tout $x \in U$: $x \sim x$;
2. $\sim$ est *symétrique*: pour tous $x, y \in U$: si $x \sim y$ alors $y \sim x$;
3. $\sim$ est *transitive*: pour tous $x, y, z \in U$: si $x \sim y$ et $y \sim z$ alors $x \sim z$.

Quand $\sim$ est une relation d'équivalence sur U et $x \in U$ alors la *classe d'équivalence de x* est $[x]_\sim = \{y \in U \mid x \sim y\}$. Deux classes d'équivalence $[x]_\sim$ et $[y]_\sim$ sont soit égales (quand $x \sim y$), soit disjointes (quand $x \not\sim y$). Une classe d'équivalence ne peut pas être vide.

L'*indice* d'une relation d'équivalence $\sim$, noté *indice*$(\sim)$, est le nombre de ses classes d'équivalence, ce nombre peut être fini ou infini.

Une relation d'équivalence $\sim_1$ est un *raffinement* d'une relation d'équivalence $\sim_2$ si $x \sim_1 y$ implique que $x \sim_2 y$.

Si $\sim_1$ est un raffinement de $\sim_2$ et $\sim_1$ est d'indice fini, alors $\sim_2$ est également d'indice fini, et *indice*$(\sim_2) \leq$ *indice*$(\sim_1)$.

1.2 Congruences droites

Une relation d'équivalence $\sim$ sur Σ^* est une *congruence droite* si:

$$\forall x, y, z \in \Sigma^* : x \sim y \Rightarrow xz \sim yz$$

2 Congruence droite induite par un langage

Soit $L \subseteq \Sigma^*$ un langage. La relation $\sim_L$ sur Σ^* est définie par

$$x \sim_L y \text{ ssi } \forall w \in \Sigma^* : xw \in L \Leftrightarrow yw \in L$$

Propriétés de cette relation:

1. $\sim_L$ est une relation d'équivalence.
2. $\sim_L$ est une congruence droite.
3. $x \sim_L y$ si et seulement si $x^{-1}L = y^{-1}L$.
4. toute classe d'équivalence $[x]_{\sim_L}$ est soit incluse dans L, soit disjointe de L.

3 Congruence droite induite par un automate

Soit $A = (\Sigma, Q, q_0, F, \delta)$ un automate déterministe complet. La relation $\sim_A$ sur Σ^* est définie par

$$x \sim_A y \text{ ssi } \delta^*(q_0, x) = \delta^*(q_0, y)$$

Propriétés de cette relation:

1. $\sim_A$ est une relation d'équivalence
2. $\sim_A$ est une congruence droite
3. $|Q| \geq indice(\sim_A)$, donc $\sim_A$ est d'indice fini.
4. si L est le langage reconnu par A, alors $\sim_A$ est un *raffinement* de $\sim_L$, et on a donc que $indice(\sim_A) \geq indice(\sim_L)$.

4 Le théorème de Myhill-Nerode

Soit $L \subseteq \Sigma^*$ un langage.

1. L est rationnel si et seulement si $\sim_L$ est d'indice fini.
2. Si L est rationnel alors l'indice de $\sim_L$ est égal au nombre d'états du plus petit automate déterministe complet qui reconnaît L.