IN2505-II Berekenbaarheidstheorie

College 6

Hans Tonino

Algoritmiekgroep
Faculteit EWI – TU Delft

6 mei 2009

Vorig college

- Hotel Hilbert
- Aftelbaarheid vs. Overaftelbaarheid
- Diagonalisering
- Overaftelbaarheid van $\mathbb{R}$

Intermezzo / kleine opfriscursus

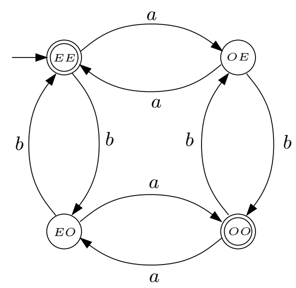
Waarin een klein gedeelte van de stof uit IN1305 II Automaten en Talen zal worden gepresenteerd (voor de nieuwkomers). Dit betreft de volgende paragrafen uit Sipser:

- H1, paragraaf 1.1 t/m pag. 43;
- H1, paragraaf 1.2 tot bewijs Theorem 1.39;
- H2, paragraaf 2.1 tot bewijs Theorem 2.9.

Het gaat met name om de begrippen en de belangrijkste eigenschappen.

Deterministische eindige automaten (DFA)

Informeel is een Deterministic Finite state Automaton (DFA) een Turingmachine zonder tape. Voorbeeld:



DFA formeel (Def. 1.5)

Een deterministische eindige automaat is een geordend vijftal $M = (Q, \Sigma, \delta, q_0, F)$, waarbij:

- Q een eindige verzameling van toestanden is;
- Σ een eindige verzameling is, het input alfabet;
- $\delta : Q \times \Sigma \to Q$ een totale functie is, de transitiefunctie;
- $q_0 \in Q$ de begintoestand is;
- $F \subseteq Q$ de verzameling eindtoestanden is.

Acceptatie van een woord door een DFA

Zij $M = (Q, \Sigma, \delta, q_0, F)$ een DFA en $a_1 a_2 \cdots a_n \in \Sigma^*$. M accepteert (herkent) $a_1 a_2 \cdots a_n$ d.e.s.d.a. er $r_0, r_1, \ldots, r_n \in Q$ bestaan zodanig dat:

1. $r_0 = q_0$;
2. $\delta(r_i, a_{i+1}) = r_{i+1} \quad (i = 0, 1, \ldots, n-1)$;
3. $r_n \in F$.

De taal herkend door M is gegeven door:

$$L(M) = \{w \in \Sigma^* \mid M \text{ accepteert } w\}.$$

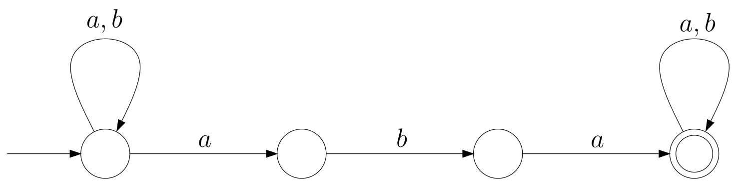
Reguliere talen (Def. 1.16)

Een taal L wordt een reguliere taal genoemd als er een DFA M bestaat zodanig dat $L(M) = L$.

Reguliere talen zijn ook bekend via reguliere expressies. Men kan bewijzen dat talen die beschreven kunnen worden m.b.v. reguliere expressies inderdaad regulier zijn in de zin van Def. 1.16 (zie Sipser par. 1.3).

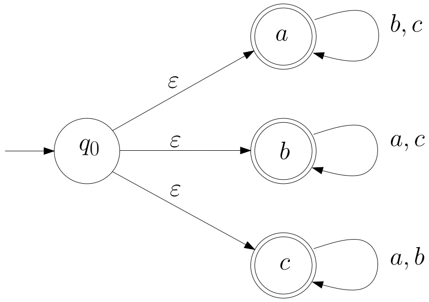
Niet-deterministische eindige automaten (NFA)

Net als bij TM's bestaat er een niet-deterministische variant van DFA's. Hierbij zijn ε-transities en "keuze"-transities mogelijk. Bijvoorbeeld:



Een NFA die ieder woord accepteert dat de substring aba bevat.

Voorbeeld NFA met ε-transities



NFA formeel (Def. 1.37)

Een niet-deterministische eindige automaat is een geordend vijftal $M = (Q, \Sigma, \delta, q_0, F)$, waarbij:

- Q, Σ, q_0 en F zijn gedefinieerd zoals bij een DFA;
- maar: $\boxed{\delta : Q \times (\Sigma \cup \{\varepsilon\}) \to \mathscr{P}(Q).}$

Acceptatie van woorden door NFA's

Zij $M = (Q, \Sigma, \delta, q_0, F)$ een NFA en $w \in \Sigma^*$.
M accepteert (herkent) W d.e.s.d.a. er zijn $a_1, a_2, \ldots, a_n \in \Sigma \cup \{\varepsilon\}$ met $w = a_1 a_2 \cdots a_n$ en er zijn $r_0, r_1, \ldots, r_n \in Q$ zodanig dat:

1. $r_0 = q_0$;
2. $r_{i+1} \in \delta(r_i, a_{i+1}), \quad (i = 0, 1, \ldots, n-1)$;
3. $r_n \in F$.

In transitiediagram: er bestaat een pad van q_0 naar r_n gelabeld met w.

Equivalentie DFA's en NFA's (Th. 1.39)

Voor iedere NFA bestaat een daaraan equivalente DFA.

Idee achter constructie NFA uit DFA: simuleer alle mogelijke paden die vanuit de starttoestand van de NFA kunnen worden bereikt; doe dit door alle mogelijke toestanden bij te houden waarin de NFA terecht kan komen. Als Q de verzameling toestanden is van de NFA, dan is $\mathscr{P}(Q)$ de verzameling toestanden van de bijbehorende DFA. Een toestand in $\mathscr{P}(Q)$ is nu een accepterende toestand als deze een accepterende toestand van de NFA bevat.

Afsluitingseigenschappen reguliere talen

De klasse van reguliere talen is gesloten onder het nemen van complement, vereniging en doorsnede. Met andere woorden, als L_1 en L_2 reguliere talen zijn, dan zijn ook $\overline{L_1}$, $L_1 \cup L_2$ en $L_1 \cap L_2$ regulier.

Niet-reguliere talen

Er bestaan niet-reguliere talen:

Zij $\Sigma = \{0, 1\}$. De taal $L \subseteq \Sigma^*$ gedefinieerd door

$$L = \{0^n 1^n \mid n \in \mathbb{N}\}$$

is niet regulier (Sipser, Ex. 1.73).

"Een DFA kan niet onthouden wat zij gelezen heeft."

Grammatica's

Een grammatica is een geordend viertal $G = (V, \Sigma, R, S)$, waarin:

- V is een eindige verzameling variabelen;
- Σ is een eindige verzameling eindsymbolen;
- S is de startvariabele $(S \in V)$;
- R is een eindige verzameling productieregels van de vorm

$$x \to y,$$

met $x \in (V \cup \Sigma)^+$ en $y \in (V \cup \Sigma)^*$.

Afleidingen in grammatica's

Als $x \to y$ een productieregel is, kan uit woord $w = uxv$ het woord $z = uyv$ worden afgeleid.
Notatie: $w \Rightarrow z$ als z (direct) afleidbaar is uit w.

Notatie: Als $w_1 \Rightarrow w_2 \Rightarrow \ldots \Rightarrow w_n$, dan schrijft men $w_1 \overset{*}{\Rightarrow} w_n$.

Let op: De $*$ geeft aan dat 0 of meer afleidingsstappen zijn gebruikt. Er geldt dus ook $w \overset{*}{\Rightarrow} w$.

Taal van een grammatica

Zij $G = (V, \Sigma, R, S)$ een grammatica. De taal van G, notatie $L(G)$ is gedefinieerd als

$$L(G) = \{w \in \Sigma^* \mid S \stackrel{*}{\Rightarrow} w\}.$$

Contextvrije grammatica's (Def. 2.2)

Een grammatica $G = (V, \Sigma, R, S)$ is contextvrij (CFG) d.e.s.d.a. alle productieregels in R de volgende vorm bezitten:

$$A \to x,$$

waarbij $A \in V$ en $x \in (V \cup \Sigma)^*$.

Voorbeeld CFG

Zij $G = (V, \Sigma, R, S)$ met $V = \{A, B, C, D\}$, $\Sigma = \{a, b, c, d\}$ en zij R gegeven door:

$$\begin{array}{lcl} S & \to & AB \mid DC \\ A & \to & \varepsilon \mid Aa \\ B & \to & \varepsilon \mid bBc \\ C & \to & \varepsilon \mid Cc \\ D & \to & \varepsilon \mid aDb \end{array}$$

Voorbeeld afleiding

$$\begin{array}{lcl} S & \to & AB \mid DC \\ A & \to & \varepsilon \mid Aa \\ B & \to & \varepsilon \mid bBc \\ C & \to & \varepsilon \mid Cc \\ D & \to & \varepsilon \mid aDb \end{array}$$

Dan:

$$S \Rightarrow D\underline{C} \Rightarrow \underline{D}Cc \Rightarrow a\underline{D}bCc \Rightarrow ab\underline{C}c \Rightarrow abc$$

en dus $S \stackrel{*}{\Rightarrow} abc$.

Contextvrije talen

Een taal L is contextvrij d.e.s.d.a. er een contextvrije grammatica G bestaat zodanig dat $L = L(G)$.

Chomsky normaalvorm (CNF)

Een CFG $G = (V, \Sigma, S, R)$ is in Chomsky normaalvorm d.e.s.d.a. iedere regel in R is van één van de volgende vormen:

1. $S \to \varepsilon$
2. $A \to a$
3. $A \to BC$

waarin $a \in \Sigma$ en $A, B, C \in V$ met B en C ongelijk aan S.

CFG's en CNF (Th. 2.8)

Voor iedere CFG G bestaat een CFG G' in Chomsky normaalvorm die dezelfde contextvrije taal voortbrengt, ofwel $L(G') = L(G)$.

Het bewijs van deze stelling wordt gegeven in de vorm van een constructie: een algoritme waarmee CFG's kunnen worden omgezet in equivalente CFG's in CNF.

Niet-contextvrije talen

Er bestaan niet-contextvrije talen:

Zij $\Sigma = \{a, b, c\}$. De taal $L \subseteq \Sigma^*$ gedefinieerd door

$$L = \{a^n b^n c^n \mid n \in \mathbb{N}\}$$

is niet contextvrij (Sipser, Ex. 2.36).

"Een PDA kan maar één keer poppen wat zij gepusht heeft."