A szóprobléma eldöntése

Formális nyelvek, 8. gyakorlat

Célja: A szóprobléma megoldásának bemutatása a különböző Chomsky-osztályok esetén

Fogalmak: i. típusú nyelvtanok, Chomsky-normálforma, szóprobléma, elemzés, összes levezetések gráfja, szélességi bejárás, mélységi bejárás, gyakorlati megoldhatóság, CYK algoritmus, véges determinisztikus automata (VDA), az automaták megadásának lehetőségei.

Feladatok jellege: Konkrét nyelvtanra az összes levezetések gráfjának felírása, az utak és a levezetések közötti összefüggés bemutatása, a szélességi bejárás algoritmusának elkészítése struktogramként, a CYK algoritmus egy példán keresztül, utalva a műveleti igényre, automatakészítés néhány egyszerű nyelv esetében.

2008/09 I. félév

A szóprobléma

Szóprobléma

Adott egy u terminális szó és egy G nyelvtan. Benne van-e az u szó a G nyelvtan által generált $L(G)$ nyelvben, azaz levezethető-e u a nyelvtan szabályaival a kezdőszimbólumból? Adjunk minél hatékonyabb algoritmust a kérdés eldöntésére!

0. típus: Csak parciálisan rekurzív algoritmus ismeretes.
1. típus: Nem ismert hatékony (polinomiális) algoritmus. Csak exponenciálisat tudunk.
A fenti két algoritmus az összes levezetések gráfjának szélességi bejárásán alapul.
2. és 3. típus: Létezik hatékony (polinomiális) algoritmus.

CYK algoritmus

Polinomiális algoritmus a 2. típusú szóproblémára

Cocke-Younger-Kasami (CYK) algoritmus: $\mathcal{O}(n^3)$

Adott egy környezetfüggetlen $G = \langle T, N, \mathcal{P}, S\rangle$ nyelvtan Chomsky-normálformában adva.
Az algoritmus adott $u \in T^*$ esetén eldönti, hogy "$u \in L(G)$"-e.

Legyen $u = t_1 \dots t_n$, $t_i \in T$. Legyen A_i a $P_i \in \mathcal{P}$ szabály bal-, β_i pedig a jobboldala. ($A_i \in N$, $\beta_i \in T \cup N^2$.)

A CYK algoritmus rekurzíven definiál $H_{i,j}$, $1 \le i \le j \le n$ halmazokat $(j-i)$ szerint növekvő sorrendben.

$$H_{i,i} := \{A_j \mid \beta_j = t_i\}$$

$$H_{i,j} := \{A_k \mid \beta_k \in \bigcup_{h=i}^{j-1} H_{i,h}H_{h+1,j}\} \quad (i<j)$$

Ha $S \in H_{1,n}$, akkor $u \in L(G)$, különben $u \notin L(G)$.

CYK algoritmus

1.feladat:

Elemezzük CYK algoritmussal az *aabbcc* szót az alábbi G nyelvtan esetén:

$S \to AB \mid BC$
$A \to XA \mid a$
$X \to a$
$C \to YC \mid c$
$Y \to c$
$B \to UV \mid VW$
$U \to XX$
$W \to YY$
$V \to ZZ$
$Z \to b$

CYK algoritmus
CYK tábla

Megoldás:

					$\{S\}$					
				$\{S\}$		$\{S\}$				
			$\{B\}$		$\emptyset$		$\{B\}$			
		$\emptyset$		$\emptyset$		$\emptyset$		$\emptyset$		
	$\{A,U\}$		$\emptyset$		$\{V\}$		$\emptyset$		$\{C,W\}$	
$\{A,X\}$		$\{A,X\}$		$\{Z\}$		$\{Z\}$		$\{Y,C\}$		$\{Y,C\}$
a		a		b		b		c		c

Tehát $aabbcc \in L(G)$.

Lineáris algoritmus a 3. típusú szóproblémára

Legyen $G = \langle T, N, \mathcal{P}, S \rangle$ egy 3. típusú nyelvtan normálformában adva. Továbbá legyen a levezetendő szó $u = t_1 \cdots t_n$. A következő algoritmus lineáris időben eldönti 3. típusú nyelvtanok szóproblémáját.

Az algoritmus rekurzívan kiszámol egy a nyelvtani jelek halmazának részhalmazaiból álló sorozatot.

$H_0 = \{S\}$
$H_{i+1} = \{A \in N \mid \exists B \in H_i \wedge B \to t_{i+1}A \in \mathcal{P}\}$.

Legyen továbbá $F = \{A \in N \mid A \to \varepsilon \in \mathcal{P}\}$.

$u \in L(G) \Leftrightarrow H_n \cap F \neq \emptyset$.

Lineáris algoritmus a 3. típusú szóproblémára
Példa

2.feladat:

$S \to aA \mid bS \mid \varepsilon$
$A \to aA \mid aS$

Levezethetőek-e az S kezdőszimbólumból az alábbi szavak?

$ba, aaba, aababa, bbaabaabaaba, b^7a(aba)^{2007}ab^7$?

Nem, nem, nem, nem, igen.

Rögzített nyelvtan esetén a szóprobléma eldöntése különböző szavakra gépiesen végezhető, ha rendelkezünk egy táblázattal, mely a lehetséges H_i halmaz és betű párokra rendre megadja a H_{i+1} eredményhalmazt.

Véges determinisztikus automaták

Véges determinisztikus automata (VDA) alatt a következő 5-öst értjük: $\mathcal{A} = \langle Q, T, \delta, q_0, F \rangle$, ahol

Q	az állapotok (véges) halmaza
T	egy ábécé, a bemenő ábécé
$\delta : Q \times T \to Q$	az állapotátmeneti függvény
$q_0 \in Q$	kezdőállapot
$F \subseteq Q$	a végállapotok halmaza.

A VDA egy ütemben kiolvassa a központi egység állapotát, az input szó aktuális szimbólumát, ennek függvényében új állapotba kerül és az input szó következő betűjére áll az olvasófej (azaz, jobbra lép).

Véges determinisztikus automaták

Általánosított állapotátmeneti függvény: $\delta : Q \times T^* \to Q$
(az állapotátmeneti függvény egyértelműen meghatározza, annak kiterjesztése; rekurzív definíció)
0. Ha $\ell(u) = 0$ (azaz $u = \varepsilon$), akkor $\delta(q, u) = q$.
1. Ha $\ell(u) = 1$, akkor $\delta(q, u)$ mint az eddigiekben.
2. Ha $u = tv$, $t \in T$, $v \in T^+$, akkor legyen $\delta(q, u) = \delta(\delta(q, t), v)$.

$\mathcal{A}$ elfogad (vagy felismer) egy u szót ha $\delta(q_0, u) \in F$, azaz ha az inputszó teljes elolvasása után az automata állapota elfogadó.

Az elfogadott szavak $L(\mathcal{A})$ halmazát az automata által elfogadott (vagy felismert) nyelvnek nevezzük.

Véges determinisztikus automaták

VDA-k megadása

3. Feladat: $T = \{a, b, c\}$. Adjunk VDA-t mely a legfeljebb 5 hosszú szavakat fogadja el!

Megoldás: *I.*

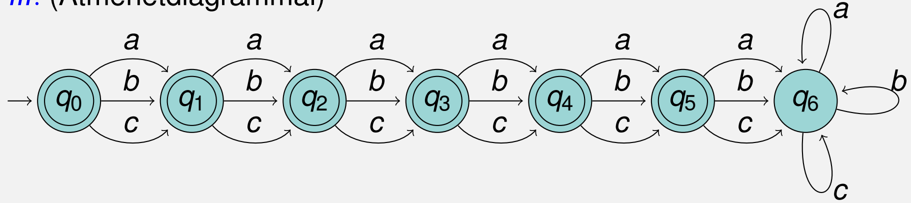
$\langle\{q_0, \dots, q_6\}, \{a, b, c\}, \delta, q_0, \{q_0, \dots, q_5\}\rangle$

$\delta(q_i, t) = q_{i+1},\ (i{=}0, \dots, 5,\ t \in \{a, b, c\})$
$\delta(q_6, t) = q_6.\ (t \in \{a, b, c\})$

II.

	a	b	c
$\leftrightarrows q_0$	q_1	q_1	q_1
$\leftarrow q_1$	q_2	q_2	q_2
$\leftarrow q_2$	q_3	q_3	q_3
$\leftarrow q_3$	q_4	q_4	q_4
$\leftarrow q_4$	q_5	q_5	q_5
$\leftarrow q_5$	q_6	q_6	q_6
q_6	q_6	q_6	q_6

III. (Átmenetdiagrammal)



VDA használata a 3. típusú szóprobléma eldöntésére

4.feladat:
$S \to aA \,|\, bS \,|\, \varepsilon$
$A \to aA \,|\, aS$

Levezethetőek-e az S kezdőszimbólumból az alábbi szavak?

$ba, aaba, aababa, bbaabaabaaba, b^7a(aba)^{2007}ab^7$?

	a	b
$\leftrightarrows \{S\}$	$\{A\}$	$\{S\}$
$\{A\}$	$\{S, A\}$	$\{\}$
$\leftarrow \{S, A\}$	$\{S, A\}$	$\{S\}$
$\{\}$	$\{\}$	$\{\}$

Házi feladat

1. Elemezzük CYK algoritmussal az 1. feladatban adott nyelvtan esetén a következő 2 szót: *aaabcc* és *abbccc*.
2. Adjunk VDA-t mely a legalább 4 hosszú szavakat fogadja el!
3. Adjunk VDA-t mely a 7-tel osztható számokat fogadja el!