A számításelmélet alapjai I.

11. előadás

előadó: Tichler Krisztián
ktichler@inf.elte.hu

KÖRNYEZETFÜGGŐ NYELVEK

Definíció

Egy $G = \langle N, T, P, S \rangle$ grammatikát **hossz-nemcsökkentőnek** mondunk, ha minden szabályának alakjára az alábbiak valamelyike teljesül

- $S \to \varepsilon$, de ha van ilyen szabály, akkor más szabály nem tartalmazhatja jobboldalán S-et.
- $u \to v$, ahol $u, v \in (T \cup N)^+$ és $|u| \leq |v|$.

A környezetfüggő grammatikák nyilván hossz-nemcsökkentőek.

Tétel

Minden hossz-nemcsökkentő grammatika környezetfüggő nyelvet generál.

Bizonyítás: (vázlat) Minden hossz-nemcsökkentő $G = \langle N, T, P, S \rangle$ grammatikához megadható egy vele ekvivalens $G' = \langle N', T, P', S \rangle$ környezetfüggő grammatika.

Hossznemcsökkentő grammatika

1. lépés: Álterminálisok bevezetése

A szokásos módon feltehető, terminálisok csak $A \to a$ $(A \in N, a \in T)$ alakú szabályok jobboldalán fordulnak elő.

2. lépés: Környezetfüggő szabályokkal való helyettesítés

Legyen $X_1 X_2 \cdots X_n \to Y_1 Y_2 \cdots Y_m \ \ (m \geq n)$ egy hossznemcsökkető szabály. Ezt az alábbi csupa 1-es típusú szabályokkal szimulálhatjuk:

$$
\begin{gathered}
X_1 X_2 \cdots X_n \to Z_1 X_2 \cdots X_n, \\
Z_1 X_2 \cdots X_n \to Z_1 Z_2 X_3 \cdots X_n, \\
\vdots \\
Z_1 Z_2 \cdots Z_{n-1} X_n \to Z_1 Z_2 \cdots Z_n Y_{n+1} \cdots Y_m \quad (n \leq m), \\
Z_1 Z_2 \cdots Z_n Y_{n+1} \cdots Y_m \to Y_1 Z_2 \ldots Z_n Y_{n+1} \cdots Y_m, \\
\vdots \\
Y_1 \cdots Y_{n-1} Z_n Y_{n+1} \cdots Y_m \to Y_1 Y_2 \cdots Y_m,
\end{gathered}
$$

ahol $Z_1, Z_2, \ldots, Z_n$ új nemterminálisok.

Hossz-nemcsökkentő grammatika

Meggondolható, hogy a $Z_1, \ldots, Z_n$ új volta miatt a szabályokat csak ebben a sorrendben lehet és kell végrehajtani, ezért az új grammatika is ugyanazt a nyelvet generálja. Csináljuk meg ezt a szabálytranszformációt az összes „rossz” szabályra. Az így kapott G' grammatika már 1-típusú és $L(G)$-t generálja.

1. Példa (csak egyetlen hosszú szabály):
Az $ABC \to DEFGH$ szabály a következő környezetfüggő szabályokkal helyettesíthető:

$ABC \to XBC$

$XBC \to XYC$

$XYC \to XYZGH$

$XYZGH \to DYZGH$

$DYZGH \to DEZGH$

$DEZGH \to DEFGH$

$(X, Y, Z$ új nemterminálisok)

Hossz-nemcsökkentő grammatika

2. Példa: $G = \langle\{S, B\}, \{a, b, c\}, P, S\rangle$

$P = \{S \to abc, S \to aSBc, cB \to Bc, bB \to abb\}$

Egy a G-vel ekvivalens 1-es típusú grammatika szabályai:

$S \to DEF$
$S \to DSBF$
$FB \to Z_1B$
$Z_1B \to Z_1F$
$Z_1F \to BF$

$EB \to Z_2B$
$Z_2B \to Z_2EE$
$Z_2EE \to DEE$
$D \to a$
$E \to b$
$F \to c$

(D, E, F, Z_1, Z_2 új nemterminálisok)

Nem környezetfüggetlen környezetfüggő nyelv

Következmény

$\mathcal{L}_2 \subset \mathcal{L}_1$

Bizonyítás: (vázlat) $\mathcal{L}_2 \subseteq \mathcal{L}_1$, ugyanis minden 2-es típusú grammatikához van vele ekvivalens Chomsky normálformájú grammatika. A Chomsky normálformájú grammatikák azonban környezetfüggőek.

Láttuk, hogy $L = \{a^n b^n c^n \mid n \geq 1\} \notin \mathcal{L}_2$ (Bar-Hillel lemmával).

Az alábbi hossz-nemcsökkentő $G = \langle \{S, B\}, \{a, b, c\}, P, S \rangle$ grammatika viszont L-et generálja, ahol

$P = \{S \to abc, S \to aSBc, cB \to Bc, bB \to bb\}$

Valóban, $S \Rightarrow^* a^{n-1}abc(Bc)^{n-1} \Rightarrow^* a^n bB^{n-1}c^n \Rightarrow^* a^n b^n c^n$.

Másrészt teljes indukcióval belátható, hogy minden w mondatformában $|w|_a = |w|_b + |w|_B = |w|_c$ és b, B, c nem állhat a előtt. ($|w|_t$ a w szóban előforduló t betűk száma.) Mivel minden B b mellé kell kerüljön ezért a generált szavak L-beliek.

Kuroda normálforma

Definíció

Egy $G = \langle N, T, P, S \rangle$ grammatikát **Kuroda normálformájúnak** mondunk, ha minden szabályának alakjára az alábbiak valamelyike teljesül

- $S \rightarrow \varepsilon$, de ha van ilyen szabály, akkor más szabály nem tartalmazhatja jobboldalán S-et.
- $A \rightarrow a$, ahol $A \in N, a \in T$,
- $A \rightarrow BC$, ahol $A, B, C \in N$,
- $AB \rightarrow AC$, ahol $A, B, C \in N$,
- $BA \rightarrow CA$, ahol $A, B, C \in N$.

A Kuroda normálformájú grammatikák nyilván környezetfüggőek.

Tétel

Minden környezetfüggő grammatika G grammatikához van vele ekvivalens Kuroda normálformájú G' grammatika.

Kuroda normálforma

Bizonyítás: (vázlat)

1. lépés: Álterminálisok bevezetése

A szokásos módon feltehető, terminálisok csak $A \to a$ $(A \in N, a \in T)$ alakú szabályok jobboldalán fordulnak elő.

2. lépés: Környezetfüggetlen szabályok hosszredukciója

Szintén a Chomsky normálformánál látott módon.

3. lépés: Környezetfüggő láncmentesítés

Az A-ból láncszabályokkal elérhető nemterminálisok $H(A) = \{B \in N \mid A \Rightarrow^* B\}$ halmazának meghatározása a Chomsky normálformánál látott módon. A szabályrendszer módosítása:

$$P' := \{A_1 \cdots A_n \to w \mid w \notin N \wedge \exists B_1 \cdots B_n \to w \in P : \\ B_i \in H(A_i)\ (\forall 1 \leq i \leq n)\}.$$

4. lépés: Környezetfüggő szabályok hosszredukciója

Az $X_1 \cdots X_m \to Y_1 \cdots Y_n$ alakú szabályok szimulációja, ahol $n \geq m \geq 2$.

Kuroda normálforma

Ha $n = m = 2$, akkor a következő lépésre ugorhatunk. Különben a szabály szimulációja a $Z_1, Z_2, \ldots, Z_{n-2}$ új nemterminálisok bevezetésével:

$$\begin{aligned} X_1X_2 &\to Y_1Z_1, \\ Z_1X_3 &\to Y_2Z_2, \\ &\vdots \\ Z_{m-3}X_{m-1} &\to Y_{m-2}Z_{m-2}, \end{aligned}$$

Továbbá ha $n = m$, akkor

$$Z_{m-2}X_m \to Y_{m-1}Y_m,$$

egyébként $(n > m)$ esetén:

$$\begin{aligned} Z_{m-2}X_m &\to Y_{m-1}Z_{m-1}, \\ Z_{m-1} &\to Y_mZ_m, \\ &\vdots \\ Z_{n-3} &\to Y_{n-2}Z_{n-2}, \\ Z_{n-2} &\to Y_{n-1}Y_n. \end{aligned}$$

Kuroda normálforma

5. lépés: Az $AB \to CD,\ A \neq C, B \neq D$ szabályok eliminációja

Végül a nem Kuroda-normálformájú szabályok sémája ekkor $AB \to CD$ $(A, B, C, D \in N)$. Átalakításukhoz szabályonként egyedi W új nemterminálisokat vezetünk be és a fenti szabályt az alábbi szabályokkal szimuláljuk:

$$AB \to AW,$$
$$AW \to CW,$$
$$CW \to CD.$$

A kapott G' grammatika ekvivalens G-vel. Ugyanis az átalakított grammatikában ezen 3 szabály bármelyikének alkalmazása implikálja a másik 2 alkalmazását ebben a sorrendben. Meggondolható, hogy az $AB \to AW$ szabályalkalmazás hátratolható közvetlenül az $AW \to CW$ szabályalkalmazás elé, míg a $CW \to CD$ szabályalkalmazás előrehozható közvetlenül az $AW \to CW$ szabályalkalmazás utánra.

Kuroda normálforma – példa

Példa:

$S \rightarrow C \mid AABC$

$A \rightarrow ABC \mid a$

$B \rightarrow b$

$C \rightarrow B \mid bA$

$ABC \rightarrow ABaC$

1-2. lépés után:

$S \rightarrow C \mid AD$

$A \rightarrow AF \mid a$

$B \rightarrow b$

$C \rightarrow B \mid YA$

$ABC \rightarrow ABXC$

$D \rightarrow AE$

$E \rightarrow BC$

$F \rightarrow BC$

$X \rightarrow a$

$Y \rightarrow b$

Kuroda normálforma – példa

3. lépés: $H(S) = \{S, C, B\}$, $H(C) = \{C, B\}$, minden más Z nemterminálisra $H(Z) = \{Z\}$.

$S \rightarrow AD \mid YA \mid b$
$A \rightarrow AF \mid a$
$B \rightarrow b$
$C \rightarrow YA \mid b$
$ABC \rightarrow ABXC$
$ACC \rightarrow ABXC$
$ASC \rightarrow ABXC$
$ABS \rightarrow ABXC$
$ACS \rightarrow ABXC$
$ASS \rightarrow ABXC$
$D \rightarrow AE$
$E \rightarrow BC$
$F \rightarrow BC$
$X \rightarrow a$
$Y \rightarrow b$

Kuroda normálforma – példa

4. lépés:

$S \to AD \mid YA \mid b$

$A \to AF \mid a$

$B \to b$

$C \to YA \mid b$

$AB \to AZ_1 \quad Z_1 C \to BZ_2 \quad Z_2 \to XC$

$AC \to AZ_3 \quad Z_3 C \to BZ_4 \quad Z_4 \to XC$

$AS \to AZ_5 \quad Z_5 C \to BZ_6 \quad Z_6 \to XC$

$AB \to AZ_7 \quad Z_7 S \to BZ_8 \quad Z_8 \to XC$

$AC \to AZ_9 \quad Z_9 S \to BZ_{10} \quad Z_{10} \to XC$

$AS \to AZ_{11} \quad Z_{11} S \to BZ_{12} \quad Z_{12} \to XC$

$D \to AE$

$E \to BC$

$F \to BC$

$X \to a$

$Y \to b$

Kuroda normálforma – példa

5. lépés:

$S \to AD \mid YA \mid b$

$A \to AF \mid a$

$B \to b$

$C \to YA \mid b$

$AB \to AZ_1 \quad Z_1C \to Z_1W_1 \quad Z_1W_1 \to BW_1 \quad BW_1 \to BZ_2 \quad Z_2 \to XC$

$AC \to AZ_3 \quad Z_3C \to Z_3W_2 \quad Z_3W_2 \to BW_2 \quad BW_2 \to BZ_4 \quad Z_4 \to XC$

$AS \to AZ_5 \quad Z_5C \to Z_5W_3 \quad Z_5W_3 \to BW_3 \quad BW_3 \to BZ_6 \quad Z_6 \to XC$

$AB \to AZ_7 \quad Z_7S \to Z_7W_4 \quad Z_7W_4 \to BW_4 \quad BW_4 \to BZ_8 \quad Z_8 \to XC$

$AC \to AZ_9 \quad Z_9S \to Z_9W_5 \quad Z_9W_1 \to BW_5 \quad BW_5 \to BZ_{10}$

$AS \to AZ_{11} \quad Z_{11}S \to Z_{11}W_6 \quad Z_{11}W_6 \to BW_6 \quad BW_6 \to BZ_{12}$

$D \to AE \qquad Z_{10} \to XC$

$E \to BC \qquad Z_{12} \to XC$

$F \to BC$

$X \to a$

$Y \to b$

A környezetfüggő nyelvek szóproblémája

Állítás: Eldönthető, egy $G = \langle N, T, P, S \rangle$ hossz-nemcsökkentő grammatika és $u \in T^*$ szó esetén $u \overset{?}{\in} L(G)$.

Bizonyítás: Ha $u = \varepsilon$, akkor $u \in L(G) \Leftrightarrow S \to \varepsilon \in P$.

$n = |u| \geq 1$ esetén legyen $r = \sum_{i=1}^{n} |T \cup N|^i$. Ekkor r a $T \cup N$ halmaz legfeljebb n hosszú, nemüres szavainak száma.

Mivel G hossz-nemcsökkentő, ezért u levezetései nem tartalmaznak n-nél hosszabb mondatformát,így u minden r-nél hosszabb levezetése tartalmaz ismétlődő mondatformát.

Ebből következően ha $S \Rightarrow_G^* u$, akkor u-nak létezik legfeljebb r hosszú levezetése is, hiszen egy levezetésben az ismétlődő mondatformák közötti levezetést kihagyva ugyanannak a szónak egy rövidebb levezetését kapjuk.

Tehát $u \in L(G)$, akkor és csak akkor, ha G legfeljebb r hosszú levezetéssel generálható. Utóbbiak viszont algoritmikusan előállíthatók.

Egy 0-típusú normálforma

Tétel

Bármely $G = \langle N, T, P, S \rangle$ 0-típusú grammatikához van vele ekvivalens G' grammatika, ahol G' minden szabályának alakjára az alábbiak valamelyike teljesül

- $S \to \varepsilon$, de ha van ilyen szabály, akkor más szabály nem tartalmazhatja jobboldalán S-et.
- $A \to a$, ahol $A \in N, a \in T$,
- $A \to B$, ahol $A, B \in N$,
- $A \to BC$, ahol $A, B, C \in N$,
- $AB \to B$, ahol $A, B \in N$,
- $AB \to AC$, ahol $A, B, C \in N$,
- $BA \to CA$, ahol $A, B, C \in N$.

Megjegyzés: a 0-típusú ε-mentesítést láttuk a zártsági tétel bizonyításában.

Egy 0-típusú normálforma

Bizonyítás: (vázlat)

Legyen $G = \langle N, T, P, S \rangle$ egy tetszőleges 0-típusú grammatika. Ekvivalens átalakításokkal a fenti alakra hozzuk.

1. lépés: 0. típusú ε-mentesítés

- Minden $u \to \varepsilon$ alakú szabályt, ahol $u \in (N \cup T)^+$, helyettesítsük az $uX \to X$ és $Xu \to X$ alakú szabályokkal minden egyes $X \in (N \cup T)$-re.
- A kapott G' grammatikára $L(G') = L(G) \setminus \{\varepsilon\}$. Ha $\varepsilon \notin L(G)$, akkor G' ekvivalens G-vel. Ha $\varepsilon \in L(G)$, akkor adjuk hozzá G'-höz az $S' \to S \,|\, \varepsilon$ szabályokat, ahol S' új nemterminális. Ez esetben S' legyen az új kezdőszimbólum.

2. lépés: Álterminálisok bevezetése

A szokásos módon.

3. lépés: Hossznemcsökkentő szabályok hosszredukciója

A Kuroda NF-nál látott módon.

Egy 0-típusú normálforma

4. lépés: Hosszcsökkentő szabályok hosszredukciója

Legyen $X_1 \cdots X_m \to Y_1 \cdots Y_n$ egy hosszcsökkentő szabály ($X_i, Y_j \in N$), azaz $m > n \geq 1$. Ezt a szabályt helyettesíthetjük az alábbi szabályhalmazzal, ahol $U_1, \dots U_m$ és $Z_{n+1}, \dots Z_m$ új, a szabályhoz bevezetett nemterminálisok:

$$\begin{array}{rcl} X_{m-1}X_m & \to & Z_m U_m, \\ X_{m-2}U_m & \to & Z_{m-1}U_{m-1}, \\ & \vdots & \\ X_n U_{n+2} & \to & Z_{n+1}U_{n+1}, \\ X_{n-1}U_n & \to & U_{n-1}Y_{n-1} \\ & \vdots & \\ X_1 U_2 & \to & U_1 Y_1 \\ U_1 Y_1 & \to & Y_1. \end{array} \qquad \begin{array}{rcl} Z_m U_m & \to & U_m, \\ Z_{m-1}U_{m-1} & \to & U_{m-1}, \\ & \vdots & \\ Z_{n+1}U_{n+1} & \to & U_n Y_n, \end{array}$$

5. lépés: Az $AB \to CD,\ A \neq C, B \neq D$ szabályok eliminációja

A Kuroda NF-nál látott módon.

Egy 0-típusú normálforma

Példa:

$S \rightarrow AB \mid BAB$

$AB \rightarrow \varepsilon$

$BAb \rightarrow bA$

$A \rightarrow a \mid SS$

1. lépés:

$S' \rightarrow S \mid \varepsilon$

$S \rightarrow AB \mid BAB$

$SAB \rightarrow S$ $\quad ABS \rightarrow S$

$AAB \rightarrow A$ $\quad ABA \rightarrow A$

$BAB \rightarrow B$ $\quad ABB \rightarrow B$

$aAB \rightarrow a$ $\quad ABa \rightarrow a$

$bAB \rightarrow b$ $\quad ABb \rightarrow b$

$BAb \rightarrow bA$

$A \rightarrow a \mid SS$

Egy 0-típusú normálforma

2-3. lépés:

$S' \rightarrow S \mid \varepsilon$

$S \rightarrow AB \mid BC$

$SAB \rightarrow S$ $\quad ABS \rightarrow S$

$AAB \rightarrow A$ $\quad ABA \rightarrow A$

$BAB \rightarrow B$ $\quad ABB \rightarrow B$

$XAB \rightarrow X$ $\quad ABX \rightarrow X$

$YAB \rightarrow Y$ $\quad ABY \rightarrow Y$

$BAY \rightarrow YA$

$A \rightarrow a \mid SS$

$C \rightarrow AB$

$X \rightarrow a$

$Y \rightarrow b$

Egy 0-típusú normálforma

4. lépés:

$S' \to S \,|\, \varepsilon$

$S \to AB \,|\, BC$

$AB \to Z_3U_3 \;\; Z_3U_3 \to U_3 \;\; SU_3 \to Z_2U_2 \;\; Z_2U_2 \to U_1S \;\; U_1S \to S$

$ABS \to S$-t hasonlóan...

$\vdots$

$ABb \to Y$-t hasonlóan...

$AY \to Z_{33}U_{33} \;\; Z_{33}U_{33} \to U_{32}A \;\; BU_{32} \to U_{31}Y \;\; U_{31}Y \to Y$

$A \to a \,|\, SS$

$C \to AB$

$X \to a$

$Y \to b$

Egy 0-típusú normálforma

5. lépés:

$S' \to S \,|\, \varepsilon$

$S \to AB \,|\, BC$

$AB \to AW_1 \quad AW_1 \to Z_3W_1 \quad Z_3W_1 \to Z_3U_3$

$Z_3U_3 \to U_3$

$SU_3 \to SW_2 \quad SW_2 \to Z_2W_2 \quad Z_2W_2 \to Z_2U_2$

$\vdots$

$U_{31}Y \to Y$

$A \to a \,|\, SS$

$C \to AB$

$X \to a$

$Y \to b$

0-típusú nyelvek algoritmikus problémái

Kevés pozitív eredmény mondható 0-típusú nyelvek algoritmikus kérdései kapcsán.

Bizonyítás nélkül megemlítjük, hogy a 0-típusú grammatikák szóproblémája algoritmikusan eldönthetetlen.

Egy parciálisan eldöntő (helyes választ adó, azonban nem mindig termináló) algoritmust azonban készíthetünk.

Legyen $G = \langle N, T, P, S \rangle$ 0-típusú grammatika és $u \in T^*$. Készíthetünk egy végtelen gráfot, melynek csúcsai $(N \cup T)^*$ elemeivel címkézettek és $x \in (N \cup T)^*$-ból akkor és csak akkor van irányított él $y \in (N \cup T)^*$-ba, ha $x \Rightarrow_G y$. A gráf minden csúcsa véges kifokú, hiszen P véges és egy szabályt legfeljebb annyi kezdőpozícióban alkalmazhatunk, amennyi a szó hossza.

Tehát $u \in L(G)$ akkor és csak akkor, ha ebben a gráfban egy S-ből indított szélességi bejárás megtalálja u-t. (Ha $u \notin L(G)$, akkor tipikusan nem terminál az algoritmus.)

További érdekesebb normálformák

Tétel

Bármely $G = \langle N, T, P, S\rangle$ 3-típusú grammatikához van vele ekvivalens G' grammatika, ahol G' minden szabályának alakjára az alábbiak valamelyike teljesül

- $S \to \varepsilon$, de ha van ilyen szabály, akkor más szabály nem tartalmazhatja jobboldalán S-et.
- $A \to a$, ahol $A \in N, a \in T$,
- $A \to aB$, ahol $A, B \in N, a \in T$,

Bizonyítás: Tudjuk, hogy van G-vel ekvivalens $G'' = \langle N'', T, P'', S''\rangle$ 3-as normálformájú grammatika. Minden $A \to aB, B \to \varepsilon \in P''$ esetén adjuk hozzá az $A \to a$ szabályt a szabályrendszerhez és hagyjuk el az ε-szabályokat. Továbbá ha volt $S'' \to \varepsilon \in P''$ szabály, akkor legyen S_0 az új kezdőszimbólum és adjuk hozzá a szabályrendszerhez az $S_0 \to S'' \mid \varepsilon$ szabályokat, majd az $S_0 \to S''$ láncszabályt a szokásos módon elimináljuk. $\square$

További érdekesebb normálformák

Definíció

Egy $G = \langle N, T, P, S \rangle$ környezetfüggetlen grammatikát **Greibach normálformájúnak** mondunk, ha minden szabálya $A \to a\alpha$ alakú, ahol $A \in N$, $a \in T$ és $\alpha \in N^*$.

Tétel

Minden ε-mentes G környezetfüggetlen grammatikához megkonstruálható vele ekvivalens Greibach normálformájú G' környezetfüggetlen grammatika.

(Nem bizonyítjuk.)

Megjegyzés: Amennyiben G nem ε-mentes, akkor ezen felül szokásos módon a korlátozott ε-szabályt is meg kell engedni.

További érdekesebb normálformák

Tétel

Bármely $G = \langle N, T, P, S \rangle$ 0-típusú grammatikához van vele ekvivalens G' grammatika, ahol G' minden szabályának alakjára az alábbbiak valamelyike teljesül

- $A \rightarrow \varepsilon$, ahol $A \in N$,
- $A \rightarrow a$, ahol $A \in N, a \in T$,
- $A \rightarrow BC$, ahol $A, B, C \in N$,
- $AB \rightarrow CD$, ahol $A, B, C, D \in N$.

(Nem bizonyítjuk.)