Formális Nyelvek - 2. Előadás

Csuhaj Varjú Erzsébet

Algoritmusok és Alkalmazásaik Tanszék
Informatikai Kar
Eötvös Loránd Tudományegyetem
H-1117 Budapest
Pázmány Péter sétány 1/c

E-mail: csuhaj@inf.elte.hu

Nyelvekre vonatkozó műveletek - I

Legyen V egy ábécé és legyenek L_1, L_2 nyelvek V felett ($L_1 \subseteq V^*$ és $L_2 \subseteq V^*$).

- $L_1 \cup L_2 = \{u \mid u \in L_1 \text{ vagy } u \in L_2\}$ - az L_1 és az L_2 nyelv **uniója**;
- $L_1 \cap L_2 = \{u \mid u \in L_1 \text{ és } u \in L_2\}$ - az L_1 és az L_2 nyelv **metszete**;
- $L_1 - L_2 = \{u \mid u \in L_1 \text{ és } u \notin L_2\}$ - az L_1 és az L_2 nyelv **különbsége**.
- Az $L \subseteq V^*$ nyelv **komplementere** a V ábécére vonatkozóan $\bar{L} = V^* - L$.

Nyelvekre vonatkozó műveletek - II

Legyen V egy ábécé és legyenek L_1, L_2 nyelvek V felett ($L_1 \subseteq V^*$ és $L_2 \subseteq V^*$).

- $L_1 L_2 = \{u_1 u_2 \mid u_1 \in L_1, u_2 \in L_2\}$ - az L_1 és az L_2 nyelv **konkatenációja**;
- L^i jelöli L i**-edik iterációját** (a konkatenáció műveletére nézve), ahol $i \geq 1$ és $L^0 = \{\varepsilon\}$.
- Minden L nyelvre fennállnak a következő egyenlőségek:

  $\emptyset L = L \emptyset = \emptyset$ és $\{\varepsilon\}L = L\{\varepsilon\} = L$.
- Az L nyelv az **iteratív lezártja** (lezártja vagy Kleene-lezártja) alatt az

  $L^* = \bigcup_{i \geq 0} L^i$ nyelvet értjük. A megfelelő műveletet az iteráció lezárásának mondjuk.
- Az L^+ nyelv alatt az $L^+ = \bigcup_{i \geq 1} L^i$ nyelvet értjük.

Nyilvánvalóan $L^+ = L^*$, ha $\varepsilon \in L$ és $L^+ = L^* - \{\varepsilon\}$, ha $\varepsilon \notin L$.

Nyelvekre vonatkozó műveletek - III

Legyen V egy ábécé és $L \subseteq V^*$.

$L^{-1} = \{u^{-1} | u \in L\}$ a **tükörképe** (megfordítása) az L nyelvnek.

Tulajdonság:

$(L^{-1})^{-1} = L$ and $(L^{-1})^i = (L^i)^{-1}$, $i \geq 0$.

Definíció:

$HEAD(L) = \{u | u \in V^*, uv \in L, v \in V^*\}$.

Nyilvánvaló, hogy $L \subseteq HEAD(L)$ bármely $L \in V^*$ nyelvre.

Nyelvekre vonatkozó műveletek - IV

Legyen V_1 és V_2 két ábécé.

A $h : V_1^* \to V_2^*$ leképezést **homomorfizmusnak** nevezzük, ha teljesülnek a következők:

- h egyértelmű, azaz, minden $u \in V_1^*$ szóra pontosan egy $v \in V_2^*$ szó létezik, amelyre $h(u) = v$ teljesül.
- $h(uv) = h(u)h(v)$, ha $u, v \in V_1^*$.

A két tulajdonság alapján $h(\varepsilon) = \varepsilon$. Nevezetesen, minden $u \in V_1^*$-ra $h(u) = h(\varepsilon u) = h(u\varepsilon)$.

Nyilvánvaló, hogy egy homomorfizmus teljesen definiált, ha V_1 minden egyes szimbólumára definiálva van. Ekkor minden $u = a_1 a_2 \ldots a_n$ szóra, ahol $a_i \in V_1, 1 \leq i \leq n$, fennáll, hogy $h(u) = h(a_1)h(a_2) \ldots h(a_n)$. Ez alapján elégséges a h leképezést V_1 elemeire definiálni, és ez automatikusan kiterjesztődik V_1^*-ra.

Nyelvekre vonatkozó műveletek - V

Legyen $h : V_1^* \to V_2^*$ homomorfizmus.

A h **homomorfizmus** ε**-mentes**, ha $h(u) \neq \varepsilon$ bármely $u \in V_1^+$ szóra.

Legyen $h : V_1^* \to V_2^*$ homomorfizmus. Az $L \in V_1^*$ nyelv h**-homomorf képén** a

$$h(L) = \{w \in V_2^* \mid w = h(u), u \in V_1^*\}$$

nyelvet értjük.

Nyelvekre vonatkozó műveletek - VI

A h homomorfizmust **izomorfizmusnak** nevezzük, ha bármely u és v V_1^*-beli szóra teljesül, hogyha $h(u) = h(v)$, akkor $u = v$.

Egy példa az izomorfizmusra a decimális számok bináris reprezentációja:

$$V_1 = \{0, 1, 2, \ldots, 9\}, V_2 = \{0, 1\},$$
$$h(0) = 0000, \;\; h(1) = 0001, \;\; \ldots, h(9) = 1001$$

Generatív grammatikák egy normálformája

Tétel:

Minden $G = (N, T, P, S)$ generatív grammatikához meg tudunk konstruálni egy vele ekvivalens és azonos típusú $G' = (N', T, P', S)$ generatív grammatikát úgy, hogy P' egyetlen szabályának baloldalán sem fordul elő terminális szimbólum.

A bizonyítás vázlata:

- 2- és 3-típusú grammatikák esetében az állítás azonnal adódik a definíciókból.

- Legyen $G = (N, T, P, S)$ 0-típusú vagy 1-típusú grammatika.

  Megkonstruáljuk a $G' = (N', T, P', S)$ grammatikát.

  Tekintsük az $N' = N \cup \bar{T}$ halmazt, ahol $\bar{T} = \{\bar{a} \mid a \in T\}$. Képezzük P'-t a P szabályhalmazból úgy, hogy minden $a \in T$ szimbólumot $\bar{a}$ szimbólumra cserélünk minden egyes olyan szabály mindkét oldalán P-ben, ahol a előfordul, továbbá az így kapott szabályhalmazhoz adjuk hozzá minden $a \in T$ szimbólumra a $\bar{a} \to a$ szabályt.

  Álljon P' az így kapott szabályokból.

Bizonyításvázlat - folytatás

(1) Megmutatjuk, hogy $L(G) \subseteq L(G')$.

Azonnal látható, hogyha $u = a_1 \dots a_n \in L(G)$, ahol $a_i \in T$, $1 \leq i \leq n$, akkor $v = \bar{a}_1 \dots \bar{a}_n$ levezethető G'-ben. Ekkor a $\bar{a}_i \to a_i$ szabályok alkalmazásával u is levezethető G'-ben.

Az üres szó esetében nyilvánvaló, hogyha $\varepsilon \in L(G)$, akkor $\varepsilon \in L(G')$ is teljesül.

Bizonyításvázlat - folytatás

(2) Megmutatjuk, hogy $L(G') \subseteq L(G)$.

Definiáljuk a h homomorfizmust úgy, hogy $h(\bar{a}) = a$ minden $\bar{a} \in \bar{T}$ szimbólumra és $h(x) = x$ minden $x \in (N \cup T)$ szimbólumra.

Ha $u \Longrightarrow_{G'} v$, akkor fennáll $h(u) \Longrightarrow_G^* h(v)$ is. Ha a v szó levezethető az u szóból valamely $\bar{a} \to a$ szabály alkalmazásával, akkor $h(u) = h(v)$. Egyébként az $u \Longrightarrow_{G'} v$ levezetés P valamely szabályának alkalmazását kívánja meg, és így $h(u) \Longrightarrow_G h(v)$ szintén fennáll. Vagyis, $u \Longrightarrow_{G'}^* v$ teljesülése maga után vonja $h(u) \Longrightarrow_G^* h(v)$ teljesülését. Azaz, ha $S \Longrightarrow_{G'}^* w$, ahol $w \in T^*$, akkor $S = h(S) \Longrightarrow_G^* h(w) = w$.

Nyelvosztályok zártsági tulajdonságai

Az unió, a konkatenáció, valamint az iteráció lezárása műveleteket együttesen **reguláris** műveleteknek nevezzük.

Tétel:

Az $\mathcal{L}_i$, $i = 0, 1, 2, 3$ nyelvosztályok mindegyike zárt a reguláris műveletekre nézve.

Tétel:

Az $\mathcal{L}_i$, $i = 0, 1, 2, 3$ nyelvosztályok mindegyike zárt a reguláris műveletekre nézve.

Bizonyításvázlat:

Legyen L és L' két i-típusú nyelv, ahol $i = 0, 1, 2, 3$. Tegyük fel, hogy L és L' rendre generálhatók az i-típusú $G = (N, T, P, S)$ és $G' = (N', T', P', S')$ grammatikákkal. Az általánosság megszorítása nélkül feltehetjük, hogy G és G' a korábbiakban ismertetett normálformában adott (a szabályok baloldalán nincs terminális szimbólum), valamint hogy $N \cap N' = \emptyset$.

Bizonyításvázlat - folytatás

Unió:

(1) $i = 0, 2, 3$ esetében legyen $S_0 \notin (N \cup N')$ és legyen

$$G_u = (N \cup N' \cup \{S_0\}, T \cup T', P \cup P' \cup \{S_0 \to S, S_0 \to S'\}, S_0).$$

(2) Nyilvánvaló, hogy G_u egyazon típusú, mint G és G'.

(3) Az is azonnal látható, hogy $L(G) \cup L(G') \subseteq L(G_u)$.

(4) $L(G_u) \subseteq L(G) \cup L(G')$ szintén fennáll, mivel N és N' diszjunktak és az $S_0 \to S$, $S_0 \to S'$ szabályok garantálják, hogy $L(G_u)$ egyetlen elemének levezetésekor sem használunk szabályt mind a P és mind a P' szabályhalmazból.

Bizonyításvázlat - folytatás

Az $i = 1$ és $\varepsilon \notin (L \cup L')$ esetben megkonstruálunk egy G_u grammatikát az előbbi módon.

Ha $i = 1$ és $\varepsilon \in (L \cup L')$, akkor először tekintjük az $L_1 = L - \{\varepsilon\}$ és az $L_2 = L' - \{\varepsilon\}$ nyelveket. Tegyük fel, hogy a G_1 és a G_2 grammatikák 1-típusúak, valamint rendre generálják az L_1 és az L_2 nyelveket. Ezután az előbbieknek megfelelően konstruálunk egy G_u grammatikát, amely az $(L_1 \cup L_2)$ nyelvet generálja. Majd bevezetünk egy új S_1 nemterminális szimbólumot és a G_u szabályhalmazához hozzáadjuk az $S_1 \to S_0$ és $S_1 \to \varepsilon$ szabályokat. Az így nyert grammatika az $(L \cup L')$ nyelvet generálja.

Bizonyításvázlat - folytatás

Konkatenáció:

Tekintsük először az $i = 0, 2$ eseteket. Legyen $S_0 \notin (N \cup N')$ és

$$G_c = (N \cup N' \cup \{S_0\}, T \cup T', P \cup P' \cup \{S_0 \to SS'\}, S_0).$$

(2) Nyilvánvaló, hogy G_c egyazon típusú, mint G és G'.

(3) Az is azonnal látható, hogy $L(G)L(G') \subseteq L(G_c)$.

Konkatenáció - folytatás

(4) Megmutatjuk, hogy $L(G_c) \subseteq L(G)L(G')$. Tekintsük az

$$S_0 \Longrightarrow u_1 \Longrightarrow u_2 \Longrightarrow \ldots \Longrightarrow u_m = u,\ m \geq 1$$

G_c-beli levezetést, ahol $u \in (T \cup T')^*$.

j-szerinti indukcióval megmutatjuk, hogy $u_j = v_j v'_j$ valamely v_j és v'_j-re úgy, hogy $S \Longrightarrow^*_G v_j$ és $S' \Longrightarrow^*_{G'} v'_j$ teljesül. A $j = 1$ esetben az állítás triviális, hiszen $u_1 = SS'$ kell, hogy legyen. Tegyük fel, hogy az állítás igaz u_j-re. Akkor viszont igaz u_{j+1}-re is, mivel N és N' diszjunktak, terminális szimbólum nem fordul elő a baloldalon, és ahhoz, hogy az u_{j+1} szót megkapjuk, vagy a v_j, vagy a v'_j mondatformát át kell írnunk. Ez alapján az $L(G_c)$ minden eleme egyben eleme az $L(G)L(G')$ nyelvnek is.

Konkatenáció - folytatás

(5) Legyen $i = 1$.

(a) Ha $\varepsilon \notin LL'$, akkor G_c-t az előzőeknek megfelelően konstruáljuk meg.

(b) Ha $\varepsilon \in LL'$, akkor először vegyük az $L_1 = L - \{\varepsilon\}$ és $L_2 = L' - \{\varepsilon\}$ nyelveket, és konstruáljuk meg G_c-t a fenti módon. Az LL' nyelv megegyezik a következő nyelvek valamelyikével:

$L_1L_2 \cup L_2$, $L_1L_2 \cup L_1$, $L_1L_2 \cup L_1 \cup L_2 \cup \{\varepsilon\}$,

attól függően, hogy $\varepsilon \in L$ és $\varepsilon \notin L'$, vagy fordítva, vagy ε mindkét nyelv eleme.

Mindegyik esetben $LL' \in \mathcal{L}_1$ következik abból, hogy $L_1L_2 \in \mathcal{L}_1$ és $\mathcal{L}_1$ zárt az unió műveletére nézve.

Konkatenáció - folytatás

Legyen $i = 3$.

A P szabályhalmazból megkonstruálunk egy P_1 szabályhalmazt úgy, hogy minden $A \to u$ alakú szabályt, ahol $A \in N$ és $u \in T^*$ felcserélünk egy $A \to uS'$ alakú szabályra ($S' \notin (N \cup T)$) és a többbi szabályt változatlanul hagyjuk. A

$$G_c = (N \cup N', T \cup T', P_1 \cup P', S)$$

grammatika nyilvánvalóan 3-típusú és generálja az $L(G)L(G')$ nyelvet.

Megmutatjuk, hogy $L(G_c) \subseteq L(G)L(G')$. Minden G_c-beli ww' terminális szóhoz vezető levezetés $S \Longrightarrow^*_{G_c} wS' \Longrightarrow^*_{G_c} ww'$ alakú, ahol ahhoz, hogy a w szót előállítsuk P-beli szabályokat, ahhoz, hogy a w' szó elemeit előállítsuk, P'-beli szabályokat kell használnunk. Azaz, $L(G_c) \subseteq L(G)L(G')$. A fordított irányú tartalmazás könnyen látható.

Bizonyításvázlat - folytatás

Az iteráció lezárása:

(1) Legyen $i = 2$ és legyen $S_0 \notin N$. Akkor

$$G_* = (N \cup \{S_0\}, T, P \cup \{S_0 \to \varepsilon, S_0 \to SS_0\}, S_0)$$

generálja az L^* nyelvet.

(2) Legyen $i = 3$. Definiáljuk a P_* szabályhalmazt úgy, hogy $A \to uS$ eleme P_*-nak minden $A \to u$ P-beli szabályra, ahol $u \in T^*$. Akkor

$$G_* = (N \cup \{S_0\}, T, P_* \cup P \cup \{S_0 \to \varepsilon, S_0 \to S\}, S_0)$$

grammatika generálja az L^* nyelvet.

Iteráció lezárása - folytatás

(3) Legyen $i = 0, 1$ és $\varepsilon \notin L$. Tegyük fel, hogy $S_0, S_1 \notin N$.

Legyen

$G_* = (N \cup \{S_0, S_1\}, T, P \cup \{S_0 \to \varepsilon, S_0 \to S, S_0 \to S_1 S\} \cup \{S_1 a \to S_1 S a \mid a \in T\} \cup \{S_1 a \to S a \mid a \in T\}, S_0)$ grammatika.

Legyen $L_* = L(G_*)$. Könnyen látható, hogy $L^* \subseteq L_*$. Megmutatjuk a fordított irányú tartalmazást.

Iteráció lezárása - folytatás

$G_* = (N \cup \{S_0, S_1\}, T, P \cup \{S_0 \to \varepsilon, S_0 \to S, S_0 \to S_1 S\} \cup \{S_1 a \to S_1 S a \mid a \in T\} \cup \{S_1 a \to S a \mid a \in T\}, S_0)$

Megmutatjuk, hogy $L(G_*) \subseteq L^*$.

Tekintsük az

$$S_0 \Longrightarrow_{G_*} u_1 \Longrightarrow_{G_*} u_2 \Longrightarrow_{G_*} \ldots \Longrightarrow_{G_*} u_m = u, \ m \geq 1$$

levezetést, ahol $u \in T^*$. Ha az első lépésben az $S_0 \to \varepsilon$ szabályt használjuk, akkor $m = 1$ és $u_m = \varepsilon \in L^*$. Ha $u_1 = S$, akkor $u \in L^*$, egyébként $u = S_1 S$ és minden j-re, $1 \leq j \leq m$ indukcióval j szerint megmutatható, hogy u_j alakja a következő két alak közül valamelyik:

(a) $S_1 v_1 \ldots v_k$, $k \geq 1$, ahol $S \Longrightarrow_G^* v_l$, $l = 1, \ldots, k$ és a $v_2, \ldots, v_k$ szavak mindegyike terminális szimbólummal kezdődik; vagy

(b) $v_1 \ldots v_k$, $k \geq 0$, ahol $S \Longrightarrow_G^* v_l$, $l = 0, \ldots, k$ és a $v_2, \ldots, v_k$ szavak mindegyike terminális szimbólummal kezdődik.

Az $u = S_1S$ eset az (a) esetnek felel meg. A G_* szabályait megvizsgálva láthatjuk, hogy u_{j+1} vagy (a), vagy (b) formájú, ha u_j rendelkezik ezzel a tulajdonsággal. azaz (a) vagy (b) formájú. Azaz, $L(G_*) \subseteq L^*$.

Iteráció lezárása - folytatás

Ha $i = 0, 1$ és $\varepsilon \in L$, akkor először veszünk egy G_1 grammatikát, ahol $L(G_1) = L - \{\varepsilon\}$.

(a) $i = 1$ esetében ez egyszerű, elhagyjuk az $S \to \varepsilon$ szabályt.

(b) $i = 0$ esetben minden $u \to \varepsilon$ szabályt az $uX \to X$ és az $Xu \to u$ szabályra cserélünk minden $X \in (N \cup T)$ szimbólum esetében. G_1 típusa ugyanaz marad, mint G típusa és $(L - \{\varepsilon\})^* = L^*$.

Korollárium

Ha az L nyelv i-típusú, $i = 0, 1, 2, 3$, akkor L^+ is az.

Néhány további tulajdonság

(1) A $\mathcal{L}_i$, ahol $i = 0, 1, 2$ is zárt a megfordítás (tükrözés) műveletére nézve.

(2) $\mathcal{L}_i$, ahol $i = 0, 1, 2$ zárt a homomorfizmus és $\mathcal{L}_1$ zárt a ε-mentes homomorfizmus műveletére nézve.

(3) Minden véges nyelv eleme $\mathcal{L}_3$-nak.

Irodalom:

György E. Révész, Introduction to Formal Languages, McGraw-Hill Book Company, 1983, Chapter 2.