Formális Nyelvek - 1.

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

A kurzus célja, hogy megismerkedjünk a **formális nyelvek és automaták elméletének**, a **számítástudomány egyik tradícionális ágának** alapjaival.

Irodalom:

1. Csima Judit és Friedl Katalin: Nyelvek és automaták, BMGE jegyzet, 2013. (weben elérhető)

2. Révész György, Bevezetés a formális nyelvek elméletébe, Tankönyvkiadó, 1977.

3. Fülöp Zoltán, Formális nyelvek és szintaktikus elemzésük, Polygon, Szeged, 2004.

4. Bach Iván: Formális nyelvek, Typotex, 2001.

Irodalom:

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

További irodalom:

2. A. Salomaa, Formal Languages, Acedemic Press, 1973.

3. K. Krithivasan, Rama, R., Introduction to Formal Languages, Automata Theory and Computation, Pearson, 2009.

4. J. E. Hopcroft, Rajeev Motwani, J.D. Ullman, Introduction to Automata Theory, Languages, and Computation. Second Edition. Addison-Wesley, 2001.

5. M. Sipser: Introduction to the Theory of Computation, 2nd edition, Thomson Course Technology, 2006.

Tudnivalók

Az előadások alapjául az irodalomjegyzék magyar- és angol nyelvű elemei szolgálnak. Az táblaképek (a diák) segítik a felkészülést és kizárólag tanulási célokra használhatók.

A diák szerzői jogvédelem alatt állnak, mind részben, mind egészben kizárólag az előadó (Csuhaj Varjú Erzsébet) honlapján tehetők publikussá!!!

A kurzus vizsgával zárul, **a vizsga során az előadásokon elhangzott anyagot kérem számon**. Vizsgázni csak legalább elégséges gyakorlati jegy birtokában lehet. A vizsga előtt három héttel részletes információt adok a számonkérés alapjául szolgáló anyagról és a számonkérés módjáról.

Az előadások látogatása nem kötelező, de erősen ajánlott. A gyakorlatok látogatása kötelező, a lehetséges hiányzások számát a gyakorlatvezetők ismertetni fogják a gyakorlatokon. A gyakorlati jegy megszerzésének feltételeit is a gyakorlatvezetők fogják meghatározni és ismertetni.

Az előadásvázlatokat (a diákat) .pdf file formájában az előadások után felteszem a Neptun-meet-street-be.

Minden kedden 10-12 óra között fogadóórám van a Déli tömb 2.511-es hivatali szobámban, ahol az érdeklődőket szeretettel várom.

Ha bármilyen kérdésük van a tantárggyal kapcsolatban, keressenek meg emailben (csuhaj@inf.elte.hu) vagy a fogadóórán.

Jó tanulást kívánok!

Budapest, 2016. február

Csuhaj Varjú Erzsébet

tanszékvezető egyetemi tanár

A kurzus tartalmának rövid leírása

1. **Bevezetés, a formális nyelv fogalma:** alapvető fogalmak és jelölések,szavak, nyelvek, grammatikák, a grammatikák Chomsky-féle hierarchiája.

2. **Műveletek nyelveken:** definíciók, Chomsky-féle nyelvosztályok bizonyos zártsági tulajdonságai.

3. **Környezetfüggetlen grammatikák és nyelvek:** redukált grammatikák, normálformák, levezetési fa. Reguláris grammatikák, reguláris nyelvek, reguláris kifejezések. A generált nyelvek és nyelvosztályok bizonyos fontos tulajdonságai.

4. **Környezetfüggő- és mondatszerkezetű grammatikák:** hossz-nemcsökkentő grammatikák, normálformák. A generált nyelvek és nyelvosztályok tulajdonságai. Nyelvosztályok Chomsky-féle hierarchiája.

A kurzus tartalmának rövid leírása - folytatás

1. **Automaták és nyelvek:** véges automaták, veremautomaták. Az automaták tulajdonságai, a felismert nyelvosztályok, az automaták és a grammatikák kapcsolatai.

2. Lineárisan korlátozott automata, Turing gép.

3. **Szintaktikai elemzés:** kapcsolat szintaxis és szemantika között; $LL(k)$ és $LR(k)$ grammatikák.

A formális nyelvek és automaták elmélete - a gyökerek

A nyelv **grammatikájának** fogalma már kb. időszámításunk előtt az IV. században felmerült Indiában (Panini).

Fontosabb lépések:

- Axel Thue, Emil Post, matematika, a XX. század eleje.
- W. Mc Culloch, W. Pitts, 1943, az idegrendszer modellje - a véges állapotú gép;

  S.C. Kleene, 1956, neurális háló - a véges automata.
- Noam Chomsky, 1959, matematikai model, az angol nyelv grammatikájának matematikai modellje.
- Programnyelvek, ALGOL 60, 1960

Mivel foglalkozik a formális nyelvek és automaták elmélete?

A formális nyelvek elmélete szimbólumsorozatok halmazaival foglalkozik.

Célja - többek között - véges, tömör leírását adni az ilyen halmazoknak.

Az elmélet módszereket ad formális nyelvek definiálására, a formális elemek nyelvhez való tartozásának eldöntésére, a nyelvi elemek struktúrájának felismerésére és leírására.

A szimbólum fogalmát alapfogalomnak tekintjük, ezért nem definiáljuk.

Milyen tudományágakhoz kapcsolódnak a formális nyelvek és automaták?

- A természetes nyelvek gépi feldolgozása, matematikai modellezése, matematikai nyelvészet,
- programozási nyelvek, fordítóprogramok elmélete,
- kódelmélet,
- képfeldolgozás,

- mintafelismerés,

- fejlődő rendszerek modellezése (Lindenmayer rendszerek),

- molekuláris számítástudomány (DNS számítás),

- multi-ágens rendszerek formális leírásai,

- stb.

Alapfogalmak és jelölések - I

Szimbólumok véges nemüres halmazát **ábécének** nevezzük.

Példa: $V = \{a, b, c\}$

Egy V ábécé elemeiből képzett véges sorozatokat V feletti **szavaknak** vagy **sztringeknek**- más szóval füzéreknek- mondunk. A 0 hosszúságú sorozatot **üres szónak** nevezzük és ε-nal jelöljük.

Példa: Legyen az ábécé $V = \{a, b, c\}$ és akkor $aaabbbccc$ egy szó.

A V ábécé feletti szavak halmazát (beleértve az üres szót is) V^*-gal, a nemüres szavak halmazát V^+**-szal jelöljük.**

Alapfogalmak és jelölések - II

Legyen V egy ábécé és legyenek u, v V feletti szavak (azaz, legyen $u, v \in V^*$). Az uv szót az u és v szavak **konkatenáltjának** vagy más szóval összefűzésének nevezzük.

Példa: Legyen az ábécé $V = \{a, b, c\}$, legyenek $u = abb$ és $v = cbb$ szavak. Akkor $uv = abbcbb$ az u és v konkatenáltja.

Megjegyzés: A konkatenáció mint művelet asszociatív, de általában nem kommutatív.

Példa: Legyen $u = ab$, $v = ba$, akkor $uv = abba$ és $vu = baab$.

Alapfogalmak és jelölések - II - folytatás

Legyen V egy ábécé. Megállapíthatjuk, hogy V^* **zárt a konkatenáció műveletére nézve** (azaz, bármely $u, v \in V^*$ esetén $uv \in V^*$ teljesül), továbbá a **konkatenáció egységelemes művelet**, ahol az egységelem ε (azaz, bármely $u \in V^*$ esetén $u\varepsilon \in V^*$ és $\varepsilon u \in V^*$).

Alapfogalmak és jelölések - III

Legyen i nemnegatív egész szám és legyen w a V ábécé feletti szó ($w \in V^*$).

A w **szó i-edik hatványa** alatt a w szó i példányának konkatenálját értjük és w^i-vel jelöljük.

Példa: Legyen az ábécé $V = \{a, b, c\}$, és legyen $w = abc$. Akkor $w^3 = abcabcabc$.

Konvenció alapján minden $w \in V^*$ szóra $w^0 = \varepsilon$.

Alapfogalmak és jelölések - IV

Legyen V egy ábécé és legyen w egy V feletti szó (azaz, legyen $w \in V^*$).

A w **szó hosszán** a w szót alkotó szimbólumok számát értjük (azaz, w mint sorozat hosszát) és $|w|$-vel jelöljük.

Példa: Legyen az ábécé $V = \{a, b, c\}$ és legyen $w = abcccc$. Akkor w hossza 6.

Az üres szó hossza - nyilvánvalóan - 0, azaz $|\varepsilon| = 0$.

Alapfogalmak és jelölések - V

Egy V ábécé feletti két u és v szót azonosnak nevezünk, ha mint szimbólumsorozatok egyenlőek (azaz, mint sorozatok elemről-elemre megegyeznek.)

Legyen V ábécé és legyenek u és v szavak V felett. Az u szót a v szó **részszavának** nevezzük, ha $v = xuy$ teljesül valamely x és y V feletti szavakra.

Az u szót a v szó **valódi részszavának** mondjuk, ha $u \neq v$ és $u \neq \varepsilon$.

Példa: Legyen $V = \{a, b, c\}$ ábécé és legyen $v = aabbbcc$ szó. Az $u = abbbc$ szó valódi részszava v-nek.

Alapfogalmak és jelölések - V - folytatás

Legyen V ábécé és legyenek u, v, x, y szavak V felett, továbbá legyen $v = xuy$. Ha $x = \varepsilon$, akkor u-t a v szó **prefixének** vagy kezdőszeletének, ha $y = \varepsilon$, akkor u-t a v szó **szufixének** vagy utótagjának hívjuk.

Legyen $v = aabbbcc$ szó. Az $u = aabbb$ szó prefixe, a $bbbcc$ szó szufixe v-nek.

Alapfogalmak és jelölések - VI

Legyen u egy V ábécé feletti szó. Az u szó **tükörképe** vagy **fordítottja** alatt azt a szót értjük, amelyet úgy kapunk, hogy u szimbólumait megfordított sorrendben írjuk. Az u szó tükörképét u^{-1} -gyel jelöljük.

Legyen $u = a_1 \ldots a_n$, $a_i \in V$, $1 \leq i \leq n$. Ekkor $u^{-1} = a_n \ldots a_1$.

Alapfogalmak és jelölések - VII

Legyen V ábécé és legyen L tetszőleges részhalmaza V^*-nak. Akkor L-et egy V feletti **nyelvnek** nevezzük.

Az **üres nyelv** - amely egyetlen szót sem tartalmaz - jelölése $\emptyset$.

Egy V ábécé feletti nyelvet **véges nyelvnek** mondunk, ha véges számú szót tartalmaz, ellenkező esetben **végtelen nyelvről** beszélünk.

Példák nyelvekre

Legyen $V = \{a, b\}$ ábécé.

Akkor $L_1 = \{a, b, \varepsilon\}$ véges nyelv, $L_2 = \{a^i b^i \mid 0 \leq i\}$ végtelen nyelv.

Példa L_2-beli szavakra: $ab, aabb, aaabbb, \ldots$

Legyen $L_3 = \{uu^{-1} | u \in V^*\}$.

Példa L_3-beli szavakra: $u = ababb$, $u^{-1} = bbaba$ és $uu^{-1} = ababbbbaba$.

Nyelvekre vonatkozó műveletek - I

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

- $L_1 \cup L_2 = \{u \mid u \in L_1$ vagy $u \in L_2\}$ - az L_1 és az L_2 nyelv **uniója**;
- $L_1 \cap L_2 = \{u \mid u \in L_1$ és $u \in L_2\}$ - az L_1 és az L_2 nyelv **metszete**;
- $L_1 - L_2 = \{u \mid u \in L_1$ 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 ábécé és legyenek L_1, L_2 nyelvek V felett ($L_1 \subseteq V^*$ és $L_2 \subseteq V^*$).

- $L_1L_2 = \{u_1u_2 \mid u_1 \in L_1, u_2 \in L_2\}$ - az L_1 és az L_2 nyelv **konkatenációja**;
- Minden L nyelvre fennállnak a következő egyenlőségek:

  $\emptyset L = L\emptyset = \emptyset$ és $\{\varepsilon\}L = L\{\varepsilon\} = L$.
- L^i jelöli az L nyelv i-**edik hatványát**:
  $L^0 = \{\varepsilon\}$, $L^i = L^{i-1}L$, $i \geq 1$.
- Az L nyelv **lezártja** (Kleene-lezártja) alatt az

  $L^* = \bigcup_{i \geq 0} L^i$ nyelvet értjük.

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

Nyilvánvalóan, ha $\varepsilon \in L$, akkor $L^+ = L^*$.

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:

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

Nyelvekre vonatkozó leképezések:

Legyen V_1, V_2 ábécé.

A $h : V_1^* \to V_2^*$ leképezést **homomorfizmusnak** nevezzük, ha $h(uv) = h(u)h(v)$ minden $u, v \in V_1^*$ esetén.

A fenti tulajdonság alapján $h(\varepsilon) = \varepsilon$. (Minden $u \in V_1^*$-ra $h(u) = h(\varepsilon u) = h(u\varepsilon)$.)

Nyilvánvaló, hogy minden $u = a_1 a_2 \dots a_n$ szóra, ahol $a_i \in V_1, 1 \leq i \leq n$, fennáll, hogy $h(u) = h(a_1)h(a_2)\dots h(a_n)$.

Nyelvekre vonatkozó leképezések:

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

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

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

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

nyelvet értjük.

Nyelvekre vonatkozó leképezések:

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:

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

Formális Nyelvek - 2.

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

Nyelvek sokféle módon előállíthatók, például **logikai formulával, strukturális rekurzióval, algoritmussal, matematikai gépekkel, vagy produkciós rendszerekkel**.

A produkciós rendszerekkel való előállítás egyik módja a **nyelvek generálása grammatikával**.

Generatív grammatika - Definíció

Egy G **generatív grammatikán** (grammatikán vagy (generatív) nyelvtanon) egy (N, T, P, S) négyest értünk, ahol

- N és T diszjunkt ábécék, a **nemterminális** és a **terminális** szimbólumok ábécéi;
- $S \in N$ a **kezdőszimbólum** (axióma),
- P **véges halmaza** (x, y) alakú rendezett pároknak, ahol $x, y \in (N \cup T)^*$ és x legalább egy nemterminális szimbólumot tartalmaz.

  A P halmaz elemeit **átírási szabályoknak** (röviden szabályoknak) vagy **produkcióknak** nevezzük.

  Az (x, y) jelölés helyett használhatjuk az $x \to y$ jelölést is, ahol a $\to$ szimbólum nem eleme az $(N \cup T)$ halmaznak.

Példa Generativ Grammatikára

Legyen $G = (N, T, P, S)$ egy generatív grammatika, ahol

$N = \{S\}$ a nemterminálisok ábécéje,

$T = \{a, b\}$ a terminálisok ábécéje, és

$$P = \{S \to aSb, \quad S \to ab, \\ S \to ba\}$$

a szabályok halmaza.

Közvetlen levezetési lépés - Definíció

Legyen $G = (N, T, P, S)$ egy generatív grammatika és legyen $u, v \in (N \cup T)^*$.

Azt mondjuk, hogy a v szó **közvetlenül** vagy **egy lépésben levezethető** az u szóból G-ben és ezt

$$u \Longrightarrow_G v$$

módon jelöljük, ha $u = u_1 x u_2$, $v = u_1 y u_2$, $u_1, u_2 \in (N \cup T)^*$ és $x \to y \in P$.

Példa közvetlen levezetésre

Legyen $G = (N, T, P, S)$ egy generatív grammatika, ahol $N = \{S\}$ a nemterminálisok ábécéje, $T = \{a, b\}$ a terminálisok ábécéje és $P = \{S \to aSb, S \to ab, S \to ba\}$ a szabályok halmaza.

Legyen $u = aaaSbbb$.

Akkor $v = aaaaSbbbb$ közvetlenül (egy lépésben) levezethető u-ból, azaz

$$u \Longrightarrow_G v,$$

ugyanis $u_1 = aaa$, $u_2 = bbb$, $x = S$, $y = aSb$ és $S \to aSb \in P$.

Levezetés - Definíció

Legyen $G = (N, T, P, S)$ egy generatív grammatika és legyen $u, v \in (N \cup T)^*$.

Azt mondjuk, hogy a v szó k **lépésben levezethető** az u szóból G-ben, $k \geq 1$, ha létezik olyan $u_1, \ldots, u_{k+1} \in (N \cup T)^*$ szavakból álló sorozat, amelyre $u = u_1$, $v = u_{k+1}$, valamint $u_i \Longrightarrow_G u_{i+1}$, $1 \leq i \leq k$ teljesül.

A v szó **levezethető** az u szóból G-ben, ha vagy $u = v$, vagy létezik olyan $k \geq 1$ szám, hogy a v szó az u szóból k lépésben levezethető.

Levezetés

Legyen $G = (N, T, P, S)$ egy tetszőleges generatív grammatika és legyen $u, v \in (N \cup T)^*$.

Azt mondjuk, hogy a v szó levezethető az u szóból G-ben és ezt

$$u \Longrightarrow_G^* v$$

módon jelöljük, ha vagy $u = v$ vagy valamely $z \in (N \cup T)^*$ szóra fennáll, hogy $u \Longrightarrow_G^* z$ és $z \Longrightarrow_G v$ teljesül.

$\Longrightarrow^*$ a $\Longrightarrow$ reláció reflexív tranzitív lezártját jelöli.

A $\Longrightarrow$ reláció tranzítív lezártját $\Longrightarrow^+$-val jelöljük.

A kezdőszimbólumból levezethető sztringeket **mondatformának** nevezzük.

A generált nyelv - Definíció

Legyen $G = (N, T, P, S)$ egy tetszőleges generatív grammatika. A G **grammatika által generált** $L(G)$ **nyelv** alatt az

$$L(G) = \{w | S \Longrightarrow_G^* w, w \in T^*\}$$

szavakból álló halmazt értjük.

Azaz, a G grammatika által generált nyelv a T^* halmaz azon elemeiből áll, amelyek levezethetők a G grammatika S kezdőszimbólumából.

Példa

Legyen $G = (N, T, P, S)$ egy generatív grammatika, ahol $N = \{S\}$, $T = \{a, b\}$ és $P = \{S \to aSb, S \to ab, S \to ba\}$.

Akkor $L(G) = \{a^n abb^n, a^n bab^n | n \geq 0\}$.

Példa egy levezetésre:

$S \Longrightarrow_G aSb \Longrightarrow_G aaSbb \Longrightarrow_G aababb.$

Példa

Legyen $G = (N, T, P, S)$ egy generatív grammatika, ahol $N = \{S, X, Y\}$, $T = \{a, b, c\}$. Legyen

$$\begin{aligned} P = \{S \to abc, \quad & S \to aXbc, \\ Xb \to bX, \quad & Xc \to Ybcc, \\ bY \to Yb, \quad & aY \to aaX, \quad aY \to aa\}. \end{aligned}$$

Akkor $L(G) = \{a^n b^n c^n | n \geq 1\}$.

Példa egy levezetésre:

$$S \Longrightarrow_G aXbc \Longrightarrow_G abXc \Longrightarrow_G abYbcc \Longrightarrow_G aYbbcc \Longrightarrow_G aabbcc.$$

Ekvivalens Grammatikák és Nyelvek

Két generatív grammatikát (gyengén) **ekvivalensnek** nevezünk, ha ugyanazt a nyelvet generálják.

Két nyelvet **gyengén ekvivalensnek** mondunk, ha legfeljebb az üres szóban különböznek.

A Chomsky-féle hierarchia

A $G = (N, T, P, S)$ **generatív grammatikát** i**-típusúnak mondjuk**, $i = 0, 1, 2, 3$, ha P szabályhalmazára teljesülnek a következők:

- $i = 0$: Nincs korlátozás.
- $i = 1$: P minden szabálya $u_1Au_2 \to u_1vu_2$ alakú, ahol $u_1, u_2, v \in (N \cup T)^*$, $A \in N$, és $v \neq \varepsilon$, kivéve az $S \to \varepsilon$ alakú szabályt, feltéve, hogy P-ben ilyen szabály létezik. Ha P tartalmazza az $S \to \varepsilon$ szabályt, akkor S nem fordul elő P egyetlen szabályának jobboldalán sem.
- $i = 2$: P minden szabálya $A \to v$ alakú, ahol $A \in N$ és $v \in (N \cup T)^*$.
- $i = 3$: P minden szabálya vagy $A \to uB$ vagy $A \to u$, alakú, ahol $A, B \in N$ és $u \in T^*$.

Példa

Legyen $G = (N, T, P, S)$ egy generatív grammatika, ahol $N = \{S\}$, $T = \{a, b\}$ és $P = \{S \to aSb, S \to ab, S \to ba\}$.

Ez a grammatika 2-típusú (környezetfüggetlen).

A Chomsky-féle hierarchia - folytatás

Legyen $i = 0, 1, 2, 3$. Egy L nyelvet i-**típusúnak mondunk**, ha i-típusú grammatikával generálható.

Az i-típusú nyelvek osztályát $\mathcal{L}_i$-vel jelöljük.

A 0-típusú grammatikát **mondatszerkezetű grammatikának**, az 1-típusú grammatikát **környezetfüggő grammatikának**, a 2-típusú grammatikát **környezetfüggetlen grammatikának** is nevezzük. A 3-típusú grammatikát **reguláris** vagy **véges állapotú** grammatikának is mondjuk.

A 0,1,2,3-típusú nyelvek osztályait rendre **rekurzíven felsorolható**, **környezetfüggő**, **környezetfüggetlen**, valamint **reguláris nyelvosztálynak** is mondjuk.

A Chomsky-féle hierarchia - folytatás

Nyilvánvaló, hogy $\mathcal{L}_3 \subseteq \mathcal{L}_2 \subseteq \mathcal{L}_0$ és $\mathcal{L}_1 \subseteq \mathcal{L}_0$.

A későbbiekben megmutatjuk, hogy

$$\mathcal{L}_3 \subset \mathcal{L}_2 \subset \mathcal{L}_1 \subset \mathcal{L}_0.$$

Megjegyzés: Könnyen észrevehetjük, hogy a $\mathcal{L}_2$ és a $\mathcal{L}_1$ nyelvosztályok közötti, a tartalmazásra vonatkozó reláció nem azonnal látható a megfelelő grammatikák definíciójából.

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 $\bar{a}_i \rightarrow a_i$ 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 $\bar{a} \rightarrow a$ 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 a lezárás (a $*$) 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 használhatunk 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 megkonstruálunk egy G_u grammatikát, amely az $(L_1 \cup L_2)$ nyelvet generálja. Majd bevezetünk egy új S_1 kezdőszimbólumot 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

$$u_j = v_j v_j'$$

alakú valamely v_j és v_j'-re, ahol $S \Longrightarrow_G^* v_j$ 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 $\varepsilon \notin L$ és $\varepsilon \notin L'$ 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 többi 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 $L(G_c) \subseteq L(G)L(G')$. A G_c grammatikában minden 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

A lezárás:

(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.

A lezárás - folytatás

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

Legyen

$G_* = (N \cup \{S_0, S_1\}, T, P \cup \{S_0 \to \varepsilon, S_0 \to S, S_0 \to S_1S\} \cup \{S_1a \to S_1Sa \mid a \in T\} \cup \{S_1a \to Sa \mid a \in T\}, S_0)$ grammatika.

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

A lezárás - 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, $L(G_*) \subseteq L^*$.

A lezárás - 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 elhagyjuk az $S \to \varepsilon$ szabályt.

(b) $i = 0$ esetben a következőképpen járunk el: Jelölje P_ε azon $u \to \varepsilon$ alakú szabályok halmazát, amelyek elemei P-nek.

Akkor legyen $P_1 = (P - P_\varepsilon) \bigcup_u P_u$, ahol $P_u = \{uX \to X, Xu \to X \mid X \in (N \cup T), u \to \varepsilon \in P\}$.

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

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

Formális nyelvek - 3.

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

Környezetfüggetlen grammatikák normálformái

Grammatikai transzformációkkal nyert grammatikák,

- melyek bizonyos szintaktikai feltételeknek vagy **tulajdonságoknak** tesznek eleget,
- **általában valamilyen szempontból egyszerűbbek, mint az eredeti grammatikák**,
- de **ugyanazon típusba** tartoznak,
- és **ugyanazt a nyelvet generálják**.

Tétel (ε-mentesítés)

Minden $G = (N, T, P, S)$ **környezetfüggetlen grammatikához** meg tudunk konstruálni egy **vele ekvivalens** $G' = (N', T, P', S')$ **környezetfüggetlen grammatikát** úgy, hogy

- G' minden szabályának jobboldala **nemüres szó**,
- **kivéve** azt az esetet, ha az **üres szó benne van a** G **által generált nyelvben**,
- mely esetben $S' \to \varepsilon$ az **egyetlen** olyan szabály, melynek jobboldala az üres szó és ekkor S' **nem fordul elő** a G' egyetlen szabályának jobboldalán sem.

Bizonyításvázlat:

Tekintsük a $G = (N, T, P, S)$ környezetfüggetlen grammatikát. Ha P nem tartalmaz $X \to \varepsilon$ alakú szabályt, akkor $G' = G$.

Tegyük fel, hogy P-ben van $X \to \varepsilon$ alakú szabály. Definiáljuk az $U_i \subseteq N$ halmazokat a következőképpen:

$$U_1 = \{X \mid X \to \varepsilon \in P\},$$

$$U_{i+1} = U_i \cup \{X \mid X \to u \in P \text{ és } u \in U_i^*\},\ i \geq 1.$$

Nyilvánvaló, hogy az U_i sorozat, $i = 1, 2, \ldots$, a tartalmazásra nézve hierarchiát alkot és van olyan k index, hogy $U_k = U_{k+1}$ és így $U_k = U_j$ minden $j \geq k$-ra.

Legyen $U = U_k$.

Ekkor azonnal látható, hogy $X \Longrightarrow_G^* \varepsilon$ akkor és csak akkor, ha $X \in U$.

(Vagyis, $\varepsilon \in L(G)$ akkor és csak akkor, ha $S \in U$.)

Bizonyításvázlat - folytatás:

Ezután megkonstruáljuk a P_1 szabályhalmazt a következőképpen:

Minden olyan $X \to u$ szabály benne van P_1-ben, amelyre $u \neq \varepsilon$ és van olyan $v \in (N \cup T)^*$ sztring, hogy $X \to v \in P$ és u-t v-ből úgy kapjuk meg, hogy U-beli nemterminálisok valahány, azaz nulla vagy több előfordulását elhagyjuk v-ből. P_1-ben nincs más szabály.

(**Példa**: Legyen $A, B \in U$ és $C \notin U$, akkor az $S \to ACAB$ szabályból a következő szabályokat képezzük: $S \to ACAB$, $S \to CAB$, $S \to ACB$, $S \to ACA$, $S \to CB$, $S \to CA$, $S \to AC$, $S \to C$.)

Legyen $G_1 = (N, T, P_1, S)$. Ekkor látható, hogy $L(G_1) \subseteq L(G) - \{\varepsilon\}$, hiszen minden $X \to u$ szabály alkalmazása megfelel az $X \to v$ szabály alkalmazásának, amelyet valahány $Z \Longrightarrow_G^* \varepsilon$ levezetés alkalmazásával kombinálunk, ahol $Z \in U$ és Z előfordul v-ben.

Megfordítva, ha $S \Longrightarrow_G^* u$ és $u \neq \varepsilon$, akkor $S \Longrightarrow_{G_1}^* u$, hiszen az $X \to \varepsilon$ típusú szabályok alkalmazása elkerülhető P_1 megfelelő szabályának alkalmazásával.

Bizonyításvázlat - folytatás:

A fentiek alapján $L(G_1) = L(G) - \{\varepsilon\}$. Ha $\varepsilon \notin L(G)$, akkor $G' = G_1$.

Ha $\varepsilon \in L(G)$, akkor vesszük a $G' = (N \cup \{S'\}, T, P_1 \cup \{S' \to \varepsilon, S' \to S\}, S')$ grammatikát, amely az $L(G)$ nyelvet generálja.

ε-mentes grammatika

Definíció

A G grammatikát ε-**mentesnek** nevezzük, ha egyetlen szabályának jobboldala sem az üres szó.

Tétel

Minden környezetfüggetlen G grammatikához meg tudunk konstruálni egy G' ε-mentes környezetfüggetlen grammatikát, amelyre $L(G') = L(G) - \{\varepsilon\}$ teljesül.

Az állítás közvetlen következménye a megelőző állításnak, az előző bizonyításban konstruált G_1 grammatika pontosan ilyen.

Környezetfüggetlen grammatikák Chomsky-normálformája

Definíció

A $G = (N, T, P, S)$ környezetfüggetlen grammatikát **Chomsky-normálformájúnak** mondjuk, ha minden egyes szabálya vagy

1. $X \to a$, ahol $X \in N$, $a \in T$, vagy

2. $X \to YZ$, ahol $X, Y, Z \in N$ alakú.

Chomsky-normálforma - folytatás:

Tétel:

Minden ε-mentes $G = (N, T, P, S)$ környezetfüggetlen grammatikához meg tudunk konstruálni egy vele **ekvivalens** $G' = (N', T, P', S)$ **Chomsky-normálformájú** környezetfüggetlen grammatikát.

Bizonyításvázlat:

Ha minden szabály $X \to a$ vagy $X \to YZ$ alakú, ahol $a \in T$, $X, Y, Z \in N$, akkor $G' = G$. Ha nem, akkor a következőképpen járunk el.

Az általánosság megszorítása nélkül feltehetjük, hogy a P szabályhalmaz elemei terminális szimbólumokat csak az $X \to a$, $a \in T$ alakú szabályokban tartalmaznak (egy megelőző normálforma tételhez hasonlóan járunk el). Ebben az esetben minden további szabály $X \to u$ alakú, ahol $u \in N^+$.

3) Ekkor minden

$$X \to Y_1 Y_2 \ldots Y_k,\ k \geq 3$$

alakú szabályt helyettesítünk egy

$$\begin{aligned} X &\to Y_1 Z_1, \\ Z_1 &\to Y_2 Z_2, \\ &\ldots, \\ Z_{k-2} &\to Y_{k-1} Y_k, \end{aligned}$$

szabályhalmazzal, ahol $Z_1, \ldots, Z_{k-2}$ új, a szabályhoz bevezetett nemterminálisok.

Bizonyításvázlat - folytatás:

Így egy $G_1 = (N', T, P_1, S)$ grammatikát kapunk, ahol P_1 olyan szabályokból áll, amelyek alakja az alábbi három típus valamelyike:

1. $X \to a$, $X \in N'$, $a \in T$,

2. $X \to Y$, $X, Y \in N'$,

3. $X \to YZ$, $X, Y, Z \in N'$.

Az N' halmaz N elemeiből, valamint azokból az új nemterminálisokból áll, amelyeket külön-külön bevezettünk P azon szabályaihoz, amelyek jobboldalának hosszát csökkentettük.

Bizonyításvázlat - folytatás:

Könnyű belátni, hogy a $L(G_1) = L(G)$.

Legyen $S = u_1 \Longrightarrow u_2 \ldots \Longrightarrow^* u_m = u$ levezetés G-ben, ahol $u \in T^*$ és $u_j \in (N \cup T)^*$, $1 \leq j \leq m-1$.

Akkor az $u_i \Longrightarrow u_{i+1}$, $1 \leq i \leq m-1$, közvetlen levezetési lépésben P valamely szabályát alkalmazzuk.

Ha a szabály $X \to a$, vagy $X \to Y$, vagy $X \to YZ$ alakú, akkor a közvetlen levezetési lépés megfelel egy közvetlen levezetési lépésnek G_1-ben.

Ha valamely $X \to Y_1Y_2 \ldots Y_k,\ k \geq 3$ alakú szabályt alkalmaztuk, akkor a megelőzőek alapján létezik G_1-ben egy $u_i \Longrightarrow^* u_{i+1}$ levezetés, amely megfelel a szabály alkalmazásának. Így $L(G) \subseteq L(G_1)$.

$L(G_1) \subseteq L(G)$ is fennáll, mivel minden $X \to Y_1Y_2 \ldots Y_k,\ k \geq 3$ alakú szabályhoz új, szabályonként különböző nemterminálisokat vezettünk be, vagyis nem fordulhat elő olyan eset, hogy a szabályok alkalmazása közben olyan mondatforma is megjelenik, amely nem vezet $L(G)$-beli szóra.

Vagyis, $L(G_1) = L(G)$.

Bizonyításvázlat - folytatás:

A továbbiakban az $X \to Y$ **alakú** szabályokat, ahol X és Y nemterminálisok, és amelyeket **láncszabályoknak nevezünk**, elimináljuk a szabályhalmazból.

Ezen célból minden egyes $X \in N'$ nemterminálisra definiáljuk az $U_i(X)$ halmazokat a következőképpen:

$$U_1(X) = \{X\},$$

$$U_{i+1}(X) = U_i(X) \cup \{Y \mid Y \to Z \in P_1,\ Z \in U_i(X)\},\ i = 1, 2, \ldots.$$

Nyilvánvaló, hogy van olyan k természetes szám, hogy $U_k(X) = U_{k+1}(X)$, és így $U_k(X) = U_l(X)$ teljesül minden l-re, ahol $l \geq k$.

Legyen $U_k(X) = U(X)$.

Látható, hogy $Y \Longrightarrow^* X$ akkor és csak akkor, ha $Y \in U(X)$.

Bizonyításvázlat - folytatás:

Definiáljuk P'-t a következőképpen:

1. $X \rightarrow a \in P'$ akkor és csak akkor, ha van olyan $A \in N'$, ahol $X \in U(A)$ és $A \rightarrow a \in P_1$,

2. $X \rightarrow YZ \in P'$ akkor és csak akkor, ha van olyan $A \in N'$, ahol $X \in U(A)$ és $A \rightarrow YZ \in P_1$.

További szabály nincs P'-ben.

Látható, hogy $X \rightarrow a \in P'$ akkor és csak akkor, ha $X \Longrightarrow^*_{G_1} a$ és $X \rightarrow YZ \in P'$ akkor és csak akkor, ha $X \Longrightarrow^*_{G_1} A \Longrightarrow^*_{G_1} YZ$ teljesül valamely A-ra.

Ezek alapján megmutatható, hogyha egy terminális szó generálható a G_1 grammatikával, akkor generálható a G' grammatikával is, és a fordított állítás is fennáll. Vagyis, $L(G) = L(G')$.

Következmények:

Tétel: (szóprobléma)

Minden $G = (N, T, P, S)$ környezetfüggetlen grammatika esetében eldönthető, hogy egy tetszőleges $u \in T^*$ szó benne van-e G grammatika által generált nyelvben vagy sem.

Bizonyításvázlat:

Mivel az előzőek alapján az, hogy az üres szó benne van-e a G grammatika által generált nyelvben eldönthető, elég az $u \neq \varepsilon$ esetre szorítkozni. Az általánosság megszorítása nélkül feltehetjük, hogy G Chomsky-normálformájú. Ekkor az u szó $k = 2|u| - 1$ lépésben levezethető G-ben ($|u|$ az u szó hosszát jelöli). (Lásd a G szabályainak alakját). Minthogy a G grammatikában a legfeljebb k lépésben levezethető szavak véges halmazt alkotnak, ezért el tudjuk dönteni, hogy u előfordul-e ebben a halmazban vagy sem.

Következmények:

Korollárium:

Minden $G = (N, T, P, S)$ környezetfüggetlen grammatika és minden $L \subseteq T^*$ véges nyelv esetében eldönthető, hogy igazak-e a következő állítások: $L \subseteq L(G)$, valamint $L \cap L(G) = \emptyset$.

Formális nyelvek - 4.

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

Aktív, elérhető nemterminálisok

Definíció

A környezetfüggetlen grammatika egy nemterminálisát **aktívnak** nevezzük, ha levezethető belőle terminális szó; egyébként **inaktívnak** vagy **nem aktívnak** mondjuk.

A környezetfüggetlen grammatika egy nemterminálisát **elérhetőnek** nevezzük, ha legalább egy olyan mondatformában előfordul, amely a kezdőszimbólumból levezethető; egyébként **nem elérhetőnek** mondjuk.

Hasznos/nem hasznos nemterminálisok

Definíció

Egy nemterminálist **hasznosnak** nevezünk, ha aktív és elérhető, egyébként **nem hasznos**.

(Egy nemterminális nem hasznos, ha vagy inaktív, vagy nem elérhető, vagy mindkét tulajdonság teljesül rá.)

Redukált grammatika

Definíció

Egy környezetfüggetlen grammatika redukált, ha minden nemterminálisa aktív és elérhető.

Redukált grammatika - folytatás

Tétel

Minden környezetfüggetlen grammatikához meg tudunk konstruálni egy vele ekvivalens redukált környezetfüggetlen grammatikát.

Megjegyzés:

A nem elérhető és a nem aktív nemterminálisok, valamint azok a szabályok, amelyekben előfordulnak, meghatározhatók és eliminálhatók anélkül, hogy a generált nyelv megváltozna.

Bizonyításvázlat:

Legyen $G = (N, T, P, S)$ egy környezetfüggetlen grammatika. Tekintsük az alábbi halmazokat:

$$A_1 = \{X \mid X \to u \in P, u \in T^*\},$$
$$A_{i+1} = A_i \cup \{X \mid X \to w \in P, w \in (T \cup A_i)^*\},\ i = 1, 2, \ldots$$

Nyilvánvaló, hogy az A_i, $i = 1, 2, \ldots,$ halmazok nemcsökkenő hierarchiát alkotnak a tartalmazásra nézve. Így létezik olyan k szám, hogy $A_k = A_l$ teljesül minden $l \geq k$-ra. Ekkor A_k a G grammatika aktív nemterminálisainak halmaza.

Bizonyításvázlat - folytatás:

Ezután tekintsük az

$$R_1 = \{S\},$$

$$R_{i+1} = R_i \cup \{Y \mid X \to uYw \in P,\ X \in R_i, u, w \in (N \cup T)^*\},\ i = 1, 2, \ldots$$

halmazokat.

Az R_i, $i = 1, 2, \ldots,$ halmazok a tartalmazásra nézve nemcsökkenő hierarchiát alkotnak. Így létezik olyan m szám, hogy $R_m = R_l$ minden $l \geq m$ esetben. Könnyen látható, hogy az R_m halmaz G elérhető nemterminálisainak halmaza.

Az A_k és az R_m halmazok kiszámolása után eliminálunk minden olyan nemterminálist, amely nem eleme az $A_k \cap R_m$ halmaznak együtt azokkal a szabályokkal amelyekben előfordulnak. A fenti procedúrát megismételjük mindaddig, amíg egy redukált grammatikát nem kapunk.

Környezetfüggetlen grammatikák - levezetési fa

A **környezetfüggetlen grammatikák** levezetéseit ún. **levezetési fákkal** is jellemezhetjük.

A levezetési fa a szó előállításának lehetőségeiről ad információkat.

A **levezetési fa** egy **irányított gráf**, amely speciális tulajdonságoknak tesz eleget.

Levezetési fa - folytatás

Legyen V véges nemüres halmaz, amelynek elemeit **csúcsoknak** nevezzük.

Az **élek** E halmaza csúcsok rendezett párjaiból álló halmaz, azaz, $E \subseteq V \times V$.

Minden $e = (n_1, n_2)$ élre $s(e) = n_1$ az él kiindulási csúcsa és $t(e) = n_2$ a végcsúcsa.

Élek egy $e_0, e_1, \ldots, e_k$ sorozatát az $s(e_0)$-ból kiinduló $t(e_k)$-ig vezető k hosszúságú **irányított útnak** nevezzük, ha $s(e_{i+1}) = t(e_i)$, ahol $i = 0, 1, \ldots, k-1$.

Levezetési fa - folytatás

A (V, E) rendezett párt **irányított fának** nevezzük, ha van olyan $r \in V$ **csúcs**, amelyre teljesülnek a következők:

1. Az E halmaz egyetlen élének végcsúcsa sem azonos r-rel.
2. Minden r-től különböző csúcshoz V-ben létezik egy r-ből kiinduló irányított út.

Az r **csúcsot a fa gyökerének** nevezzük. Minden fának egyetlen gyökere van.

A fa minden csúcsa gyökere a fa valamely **részfájának**.

Az n **csúcs leszármazottjainak azokat az** n' **csúcsokat nevezzük**, amelyekre $(n, n') \in E$ (azaz, (n, n') él).

Azokat a csúcsokat, amelyeknek nincs leszármazottjuk, **levélnek** mondjuk.

Levezetési fa

A környezetfüggetlen grammatikák levezetéseit fákkal is leírhatjuk.

A $G = (N, T, P, S)$ grammatika feletti egy **levezetési fának** nevezünk egy fát, ha teljesülnek a következők:

- A levezetési fa **gyökerének címkéje** S.
- **Minden további csúcs címkéje** $(N \cup T \cup \{\varepsilon\})$ **valamely eleme**.
- Ha egy **csúcs címkéje** X és **leszármazottjainak címkéi balról jobbra olvasva rendre** $X_1, \ldots X_m$, $m \geq 1$, **akkor** $X \to X_1 \ldots X_m \in P$**.**
- Minden **levél címkéje** a $T \cup \{\varepsilon\}$ halmaz valamely eleme, és ha egy csúcsnak leszármazottja ε, akkor ennek a csúcsnak ez az **egyetlen leszármazottja** van.

A levezetési fa levelei címkéinek sorozata a **levezetési fa** határa.

Levezetési fa - folytatás

Minden, a G grammatikában történő **levezetéshez** hozzá tudunk rendelni egy **levezetési fát**.

A levezetési fa **nem minden esetben** adja meg a levezetés során alkalmazott **szabályok sorrendjét**.

Két levezetés lényegében azonos, ha csak a **szabályok alkalmazásának sorrendjében** különbözik, azaz, **ugyanahhoz a levezetési fához tartozik**.

Legbaloldalibb levezetés

Egy környezetfüggetlen grammatika feletti **levezetési fa** egy **egyetlen** legbaloldalibb levezetést határoz meg.

A **legbaloldalibb levezetés** során minden levezetési lépésben arra a nemterminálisra kell szabályt alkalmazni, amely a levezetési lépéshez tartozó mondatformában balról a legelső.

(Példa: Ha $u_1A_1u_2A_2\dots A_nu_{n+1}$, $u_i \in T^*$, $1 \leq i \leq n+1$, $A_j \in N$, $1 \leq j \leq n$, mondatforma a $G = (N, T, P, S)$ grammatikában, akkor legbaloldalibb levezetés esetén az A_1 nemterminálist kell helyettesíteni.)

A környezetfüggetlen grammatika által generált nyelv üres volta

Tétel

Minden környezetfüggetlen grammatikáról eldönthető, hogy az általa generált nyelv az üres nyelv-e vagy sem.

Bizonyításvázlat:

Legyen $G = (N, T, P, S)$ környezetfüggetlen grammatika. Az általánosság megszorítása nélkül feltehetjük, hogy G ε-mentes.

Legyen n a G nemterminális szimbólumainak száma.

Tegyük fel, hogy létezik egy $S \Longrightarrow_G^* u$ levezetés G-ben, ahol $u \in T^*$.

Tekintsük az ezen levezetéshez tartozó levezetési fát.

Ha a leghosszabb út hossza ebben a levezetési fában nagyobb, mint n, akkor van olyan v szó $L(G)$-ben, amely levezetési fájában a leghosszabb út hossza nem nagyobb, mint n.

Ez nyilvánvaló, hiszen ha az út hossza nagyobb, mint n, akkor legalább egy nemterminális szimbólum legalább kétszer fordul elő ezen az úton.

Bizonyításvázlat - folytatás:

Tekintsünk két azonos címkéjű csúcsot az úton és helyettesítsük a fa gyökeréhez közelebb eső csúcshoz tartozó részfát a másik csúcshoz tartozó részfával. Akkor továbbra is terminális szót kapunk.

Megismételve ezt az eljárást annyiszor, ahányszor szükséges, addig csökkenthetjük az út hosszát, amíg legfeljebb n hosszúságú utat kapunk.

Ebből következően, ha $L(G)$ nem üres, akkor léteznie kell egy szónak a nyelvben, amelyhez tartozó levezetési fában a leghosszabb út nem hosszabb, mint n.

Minthogy azokat a levezetési fákat, amelyekre az igaz, meg tudjuk határozni, a kérdést el tudjuk dönteni.

Korollárium

Az, hogy egy $G = (N, T, P, S)$ környezetfüggetlen grammatika A nemterminálisa inaktív-e, eldönthető.

A probléma ekvivalens azzal a problémával, hogy a $G_A = (N, T, P, A)$ grammatika által generált nyelv üres-e. Ha $L(G_A)$ üres, akkor A nem aktív.

Korollárium

Az, hogy egy $G = (N, T, P, S)$ környezetfüggetlen grammatika A nemterminálisa elérhető-e, eldönthető.

Bizonyításvázlat:

Távolítsunk el a P szabályhalmazból minden szabályt, amelynek A van a baloldalán. Jelöljük az így kapott szabályhalmazt P_1-gyel. Tekintsük a

$$G_A^\varepsilon = ((N - \{A\} \cup T, \{A\}, P_1 \cup \{X \to \varepsilon \mid X \in (N - \{A\} \cup T)\}, S)$$

grammatikát.

Ha A nem elérhető, akkor $L(G_A^\varepsilon) = \{\varepsilon\}$, máskülönben tartalmaznia kell egy olyan szót, amelyben A legalább egyszer előfordul. Megkonstruáljuk a G' környezetfüggetlen grammatikát, amelyre $L(G') = L(G_A^\varepsilon) - \{\varepsilon\}$. Minthogy el tudjuk dönteni, hogy $L(G')$ üres-e vagy sem, el tudjuk dönteni A elérhetőségét is G-ben.

Bar-Hillel vagy pumpáló lemma

Tétel

Minden L környezetfüggetlen nyelvhez meg tudunk adni két p és q természetes számot úgy, hogy minden olyan szó L-ben, amely hosszabb, mint p

$$uxwyz$$

alakú, ahol $|xwy| \leq q$, $xy \neq \varepsilon$, továbbá minden

$$ux^iwy^iz$$

alakú szó is benne van az L nyelvben minden $i \geq 0$ egész számra $(u, x, w, y, z \in T^*)$.

Bizonyításvázlat:

Legyen L ε-mentes nyelv, amelyet a $G = (N, T, P, S)$ Chomsky-normálformájú grammatika generál.

Tegyük fel, hogy G nemterminálisainak száma n és legyen $p = 2^n$, valamint legyen $q = 2^{n+1}$.

Ha $|\beta| > p$ valamely β szóra L-ben, akkor a β levezetési fájában a leghosszabb út hossza nagyobb, mint n. (A fa bináris fa, azaz minden csúcsából legfeljebb két csúcs származik és, ha a bináris fában a leghosszabb út hossza k, akkor a leveleinek száma legfeljebb 2^k.)

Tekintsük az utolsó $n + 1$ élét a leghosszabb útnak. Akkor lennie kell egy A nemterminálisnak, amely ezen az úton legalább kétszer előfordul.

Bizonyításvázlat - folytatás:

Feleljen meg az a részfa, amely az A első előfordulásához tartozik ezen az úton (abból indul ki) az $S \Longrightarrow^* \alpha$, $\alpha \in T^*$ levezetésnek, az A második előfordulásához tartozó részfa pedig feleljen meg az $A \Longrightarrow^* w$, $w \in T^*$ levezetésnek.

Mivel A kétszer fordult elő az úton, van két olyan $x, y \in T^*$ szó, hogy $\alpha = xwy$ és $A \Longrightarrow^* xAy$.

Ezen kívül

$$S \Longrightarrow^* uAz \Longrightarrow^* uxAyz \Longrightarrow^* uxwyz = u\alpha z = \beta.$$

Az A nemterminális adott előfordulásainak pozíciójából $|\alpha| \leq 2^{n+1}$ következik. Továbbá, az $A \Longrightarrow^* xAy$ levezetéshez során legalább egy $B \to CD$ alakú szabály alkalmazása szükséges, így $|xy| \neq \varepsilon$.

Láthatjuk, hogy $S \Longrightarrow^* uwz$ és $S \Longrightarrow^* ux^iwy^iz$, $i \geq 1$, levezetések G-ben.

Következmény

Léteznek nem környezetfüggetlen mondatszerkezetű nyelvek.

Ilyen például az $L = \{a^n b^n c^n | n \geq 1\}$.

Következmény - folytatás

Megmutatjuk, hogy L nem környezetfüggetlen. Tegyük fel az ellenkezőjét. Legyen $w = a^q b^q c^q \in L$, ahol p, q a Bar-Hillel lemmának megfelelő konstansok. Nyilvánvaló, hogy $|w| > q > p$. (Lásd a lemma bizonyítását.) Ekkor bármely $u, v, x, y, z \in \{a, b, c\}^*$-ra, amelyre $w = uxvyz$, $|xvy| \leq q$, $|xy| > 0$, a lemma alapján fennáll, hogy $uvz \in L$. Viszont, mivel xy az $\{a, b, c\}$-ből legalább egy betűt nem tartalmaz, így uvz nem lehet L eleme, és így a nyelv nem teljesíti a Bar-Hillel lemma feltételeit, vagyis a nyelv nem környezetfüggetlen.

Következmény

Tétel

Eldönthető, hogy egy környezetfüggetlen grammatika végtelen nyelvet generál-e vagy sem.

Bizonyításvázlat:

Csak ε-mentes $G = (N, T, P, S)$ grammatikával foglalkozunk. Megmutatjuk, hogy $L(G)$ akkor és csak akkor végtelen, ha tartalmaz olyan β szót, hogy $p < |\beta| \leq p + q$ teljesül.

1) Ha G rendelkezik ezzel a tulajdonsággal, akkor a Bar-Hillel lemma alapján az állítás fennáll.

2) Megmutatjuk, hogy a fordított állítás is teljesül. Ha $L(G)$ végtelen, akkor tartalmaznia kell egy β szót, amelyre $p < |\beta|$. Megmutatjuk, hogy ekkor tartalmaz egy szót, amelyre $p < |\beta| \leq p + q$ fennáll.

Bizonyításvázlat - folytatás:

Tegyük fel az ellenkezőjét, azaz, azt, hogy minden β szóra, ahol $p < |\beta|$, az teljesül, hogy $p + q < |\beta|$.

De ha $p < |\beta|$, akkor β alakja $uxwyz$, ahol $uwz \in L(G)$ és $|uwz| < |\beta|$, mivel $xy \neq \varepsilon$.

Ha $p < |uwz|$, akkor a fenti érvelés megismételhető mindaddig, amíg egy $\beta' = u'x'w'y'z'$ szót kapunk, amelyre $p < |\beta'|$ és $|u'w'z'| \leq q$ teljesül.

Ekkor, mivel $|x'w'y'| \leq q$, a Bar-Hillel lemma alapján a $p < |u'x'w'y'z'| \leq p + q$ egyenlőtlenséget kapjuk, amely ellentmond feltételezésünknek.

(A $p + q$ felső korlát alapján a megfelelő levezetési fa megkereshető.)

Formális nyelvek - 5.

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

Lineáris grammatikák és reguláris nyelvek

Definíció

Egy $G = (N, T, P, S)$ környezetfüggetlen grammatikát lineárisnak nevezünk, ha minden szabálya vagy

1. $A \to u$, $A \in N$, $u \in T^*$ vagy

2. $A \to u_1 B u_2$ alakú, ahol $A, B \in N$ és $u_1, u_2 \in T^*$.

Továbbá G-t bal-lineárisnak, illetve jobb-lineárisnak mondjuk, ha $u_1 = \varepsilon$, illetve $u_2 = \varepsilon$ minden 2. alakú szabályra.

Megjegyzés: A jobb-lineáris grammatikák azonosak a reguláris grammatikákkal (3-típusúak).

Lineáris grammatikák és nyelvek

Definíció

Egy L nyelvet lineárisnak, bal-lineárisnak, illetve jobb-lineárisnak mondunk, ha van olyan G lineáris, bal-lineáris, illetve jobb-lineáris grammatika, amelyre $L = L(G)$ teljesül.

Tétel

Minden bal-lineáris grammatika reguláris (3-típusú) nyelvet generál.

Bizonyításvázlat

Legyen $G = (N, T, P, S)$ bal-lineáris grammatika és legyen $N = \{S, A_1, \ldots, A_n\}$. Az általánosság megszorítása nélkül feltehetjük, hogy S nem fordul elő egyetlen szabály jobboldalán sem.

Megkonstruálunk egy $G' = (N, T, P', S)$ jobb-lineáris grammatikát, amelyre $L(G) = L(G')$ teljesül.

Legyen

1. $S \to u \in P'$ akkor és csak akkor, ha $S \to u \in P$, $u \in T^*$,
2. $S \to uA_k \in P'$ akkor és csak akkor, ha $A_k \to u \in P$, $u \in T^*$,
3. $A_j \to uA_k \in P'$ akkor és csak akkor, ha $A_k \to A_j u \in P$, $u \in T^*$,
4. $A_j \to u \in P'$ akkor és csak akkor, ha $S \to A_j u \in P$, $u \in T^*$.

Bizonyításvázlat - folytatás

Megmutatjuk, hogy $L(G) = L(G')$.

Legyen $w \in L(G)$. Ha $S \to w \in P$, akkor $S \to w \in P'$, így $w \in L(G')$.

Egyébként w-hez van G-ben egy

$$S \Longrightarrow A_{i_1} w_1 \Longrightarrow \ldots \Longrightarrow A_{i_{m-1}} w_{m-1} \ldots w_1 \Longrightarrow w_m \ldots w_1 = w$$

levezetés. Azonban ekkor G'-ben létezik egy

$$S \Longrightarrow w_m A_{i_{m-1}} \Longrightarrow w_m w_{m-1} A_{i_{m-2}} \Longrightarrow \ldots \Longrightarrow w_m \ldots w_2 A_{i_1} \Longrightarrow w_m \ldots w_1 = w$$

levezetés, azaz $w \in L(G')$. Így $L(G) \subseteq L(G')$. A fordított állítás igaz volta a szimmetria következménye.

Következmény

Korollárium

$\mathcal{L}_3$ zárt a tükrözés (megfordítás) műveletére nézve és minden reguláris nyelv generálható bal-lineáris grammatikával.

A bizonyítás azonnal adódik. Minden jobb-lineáris G grammatikához meg tudunk konstruálni egy G' bal-lineáris grammatikát, amely $L(G)$ tükörképét generálja. Ha $L \in \mathcal{L}_3$, akkor L^{-1} generálható a bal-lineáris G grammatikával és a megelőző tétel alapján L^{-1} reguláris. Akkor $(L^{-1})^{-1}$ szintén reguláris.

Példa

Legyen $L_1 = \{a^n b^n c^k \mid n \geq 1, k \geq 1\}$ és $L_2 = \{a^k b^n c^n \mid n \geq 1, k \geq 1\}$.

L_1 és L_2 lineáris nyelvek és

$$L_1 \cap L_2 = \{a^n b^n c^n \mid n \geq 1\}$$

környezetfüggő, de nem környezetfüggetlen nyelv.

L_1 generálható a $G_1 = (\{S, A\}, \{a, b, c\}, P_1, S)$ grammatikával, ahol
$P_1 = \{S \to Sc, S \to Ac, A \to ab, A \to aAb\}$ és
L_2 generálható a $G_2 = (\{S, B\}, \{a, b, c\}, P_2, S)$ grammatikával, ahol
$P_2 = \{S \to aS, S \to aB, B \to bc, B \to bBc\}$.

3-típusú grammatikák normálformája

Tétel

Minden 3-típusú, azaz reguláris nyelv generálható egy olyan grammatikával, amelynek szabályai vagy

1. $X \to aY$, ahol $X, Y \in N$ és $a \in T$ vagy

2. $X \to \varepsilon$ alakúak, ahol $X \in N$.

Bizonyításvázlat

Legyen $G = (N, T, P, S)$ 3-típusú grammatika. Ismeretes, hogy G szabályai vagy:

1. $A \to uB$ vagy

2. $A \to u$ alakúak, ahol $A, B \in N$ és $u \in T^*$.

Bizonyításvázlat - folytatás

Tegyük fel, hogy $|u| > 1$.

Legyen $u = a_1 \ldots a_n$, $n \geq 2$. Helyettesítsünk minden $A \to a_1 \ldots a_n B$ szabályt az $\{A \to a_1 Z_1, Z_1 \to a_2 Z_2, \ldots, Z_{n-1} \to a_n B\}$ szabályhalmazzal, ahol $Z_1, \ldots, Z_{n-1}$ új, a szabályhoz bevezetett nemterminálisok.

Hasonlóan, az $A \to a_1 \ldots a_m$, $m \geq 1$ alakú szabályokat helyettesítsük a $\{A \to a_1 Y_1, Y_1 \to a_2 Y_2, \ldots, Y_{m-1} \to a_m Y_m, Y_m \to \varepsilon\}$ szabályhalmazokkal, ahol $Y_1, \ldots, Y_m$ a szabályhoz bevezetett új nemterminálisok.

Az így kapott új P_1 szabályhalmaz elemei $X \to aY$, $X \to Y$, $X \to \varepsilon$ alakúak, ahol $X, Y \in N$ és $a \in T$. Ezután elimináljuk a láncszabályokat.

Bizonyításvázlat - folytatás

Legyen N' a P_1 szabályhalmazban előforduló nemterminálisok halmaza (az új nemterminálisokat is beszámítva). Legyen bármely $X \in N'$ nemterminálisra $U(X) = \{Y \mid Y \Longrightarrow^* X\}$.

Definiáljuk P'-t a következőképpen:

1. $X \to aY$, akkor és csak akkor, ha létezik olyan $Z \in N'$, amelyre $X \in U(Z)$ és $Z \to aY \in P_1$,

2. $X \to \varepsilon$, akkor és csak akkor, ha létezik olyan $Z \in N'$, amelyre $X \in U(Z)$ és $Z \to \varepsilon \in P_1$.

Könnyen látható, hogy $L(G) = L(G')$, ahol $G' = (N', T, P', S)$.

Korollárium

Minden ε-mentes reguláris (3-típusú) grammatikához konstruálható egy vele ekvivalens reguláris grammatika, amelynek szabályai

1. $X \to aY$ vagy

2. $X \to a$ alakúak, ahol $X, Y \in N$ és $a \in T$.

Környezetfüggetlen nyelvek egy zártsági tulajdonsága

Tétel

A környezetfüggetlen nyelvek osztálya, azaz $\mathcal{L}_2$ zárt a reguláris (3-típusú) nyelvekkel való metszetre nézve.

Bizonyításvázlat

Legyen L egy tetszőleges környezetfüggetlen nyelv ($L \in \mathcal{L}_2$) és legyen let L' ($L' \in \mathcal{L}_3$) reguláris nyelv. Megmutatjuk, hogy $L \cap L'$ környezetfüggetlen.

Először is, tegyük fel, hogy $\varepsilon \notin L$ és az L nyelvet a $G = (N, T, P, S)$ Chomsky-normálformájú grammatika, az L' nyelvet pedig a $G' = (N', T', P', S')$, a megelőzőekben ismertetett normálformájú grammatika generálja.

Legyen $\{X_1, \ldots, X_k\}$ a G' azon nemterminálisainak halmaza, amelyekre $X_i \to \varepsilon$, $1 \leq i \leq k$ teljesül.

Bizonyításvázlat - folytatás

Definiáljuk a következő grammatikákat:

$$G_i = (N' \times (N \cup T) \times N', T \cup T', P'', (S', S, X_i)),$$

$i \in \{1, \ldots, k\}$ ahol

1. $(X, A, Y) \rightarrow (X, B, Z)(Z, C, Y) \in P''$ minden $X, Y, Z \in N'$ nemterminálisra akkor és csak akkor, ha $A \rightarrow BC \in P$,

2. $(X, A, Y) \rightarrow (X, a, Y) \in P''$ minden $X, Y \in N'$ nemterminálisra, akkor és csak akkor, ha $A \rightarrow a \in P$,

3. $(X, a, Y) \rightarrow a \in P''$, akkor és csak akkor, ha $X \rightarrow aY \in P'$.

Megmutatjuk, hogy $L \cap L' = \bigcup_{i=1}^{k} L(G_i)$.

Bizonyításvázlat - folytatás

Legyen $w = a_1 \ldots a_n,\ a_i \in T,\ 1 \leq i \leq n$.

1. w akkor és csak akkor vezethető le G-ben, ha minden i-re és N'-beli nemterminálisok minden $Z_1, \ldots, Z_{n-1}$ sorozatára létezik

$(S', S, X_i) \Longrightarrow_{G_i}{}^{*} (S', a_1, Z_1)(Z_1, a_2, Z_2) \ldots (Z_{n-1}, a_n, X_i)$ alakú levezetés G_i-ben. Továbbá,

2. w akkor és csak akkor vezethető le G'-ben, ha létezik nemterminálisok $Z_1, \ldots, Z_{n-1}$ sorozata és $X_i \in N'$, ahol $X_i \to \varepsilon \in P'$ úgy, hogy

$$S' \Longrightarrow_{G'} a_1 Z_1 \Longrightarrow_{G'} a_1 a_2 Z_2 \Longrightarrow_{G'} \ldots \Longrightarrow_{G'} a_1 \ldots a_n X_i \Longrightarrow_{G'} a_1 a_2 \ldots a_n.$$

Bizonyításvázlat - folytatás

Ez azt jelenti, hogy w akkor és csak akkor vezethető le G'-ben, ha vannak olyan $Z_1, \ldots, Z_{n-1}$ nemterminálisok N'-ben és van olyan levezetés G_i-ben, ahol

$(S', a_1, Z_1)(Z_1, a_2, Z_2) \ldots (Z_{n-1}, a_n, X_i) \Longrightarrow_{G_i}^{*} a_1 a_2 \ldots a_n$ fennáll.

Ebből az következik, hogy $w \in L \cap L'$ akkor és csak akkor, ha $w \in L(G_i)$ valamely i-re.

Ha $\varepsilon \in L$, akkor megkonstruálunk egy $(L - \{\varepsilon\}) \cap L'$-t generáló környezetfüggetlen grammatikát és hozzáadjuk az $S_0 \to \varepsilon, S_0 \to S$ szabályokat.

Formális nyelvek - 6.

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

Hossz-nemcsökkentő grammatikák

Definíció

Egy $G = (N, T, P, S)$ 0-típusú grammatikát hossz-nemcsökkentőnek mondunk, ha bármely $u \to v \in P$ szabályra $|u| \leq |v|$ teljesül.

Tétel

Minden hossz-nemcsökkentő grammatika környezetfüggő nyelvet generál.

Bizonyításvázlat:

Legyen $G = (N, T, P, S)$ egy tetszőleges hossz-nemcsökkentő grammatika. Az általánosság megszorítása nélkül feltehetjük, hogy terminális szimbólum csak $A \to a$ alakú szabályban szerepel a P szabályhalmazban, ahol A nemterminális és a terminális.

Legyen $u \to v$ olyan szabály P-ben, amelyre $|u| \geq 2$ és $u \to v$ alakja legyen

$$X_1X_2 \dots X_m \to Y_1Y_2 \dots Y_n,$$

$n \geq m \geq 2$. Akkor az előbbi megjegyzés alapján $X_i, Y_j \in N$, $1 \leq i \leq m$, $1 \leq j \leq n$.

Legyen $n > m$. Helyettesítsünk minden

$$X_1X_2 \dots X_m \to Y_1Y_2 \dots Y_n$$

alakú szabályt a következő szabályhalmazzal, ahol $Z_i \notin N$, $1 \leq i \leq m$, új, a szabályhoz bevezetett nemterminálisok. (Különböző szabályokhoz páronként diszjunkt új nemterminálishalmazokat vezetünk be.)

Bizonyításvázlat - folytatás:

$$X_1X_2\dots X_m \to Y_1Y_2\dots Y_n$$

Helyettesítve:

$$\begin{aligned}
X_1X_2\dots X_m &\to Z_1X_2\dots X_m\\
Z_1X_2\dots X_m &\to Z_1Z_2\dots X_m\\
&\vdots\\
Z_1\dots Z_{m-1}X_m &\to Z_1\dots Z_{m-1}Z_mY_{m+1}\dots Y_n\\
Z_1\dots Z_{m-1}Z_mY_{m+1}\dots Y_n &\to Y_1\dots Z_{m-1}Z_mY_{m+1}\dots Y_n\\
&\vdots\\
Y_1\dots Y_{m-1}Z_mY_{m+1}\dots Y_n &\to Y_1\dots Y_{m-1}Y_mY_{m+1}\dots Y_n\\
&.
\end{aligned}$$

Az $n = m$ eset értelemszerű módosítással adódik.

Könnyen látható, hogy így minden egyes hossz-nemcsökkentő szabályt helyettesíthetünk környezetfüggő szabályok halmazával úgy, hogy a generált nyelv nem változik.

Bizonyításvázlat - szemléltetés:

$$X_1X_2\ldots X_m \Longrightarrow Z_1X_2\ldots X_m \Longrightarrow Z_1Z_2\ldots X_m$$

$$\Longrightarrow^* Z_1Z_2\ldots Z_{m-1}X_m \Longrightarrow Z_1Z_2\ldots Z_{m-1}Z_mY_{m+1}\ldots Y_n$$

$$\Longrightarrow^* Y_1Y_2\ldots Y_{m-1}Z_mY_{m+1}\ldots Y_n \Longrightarrow Y_1Y_2\ldots Y_{m-1}Y_mY_{m+1}\ldots Y_n.$$

$$\begin{array}{r}
X_1X_2\ldots X_m \to Z_1X_2\ldots X_m \\
Z_1X_2\ldots X_m \to Z_1Z_2\ldots X_m \\
\vdots \\
Z_1\ldots Z_{m-1}X_m \to Z_1\ldots Z_{m-1}Z_mY_{m+1}\ldots Y_n \\
Z_1\ldots Z_{m-1}Z_mY_{m+1}\ldots Y_n \to Y_1\ldots Z_{m-1}Z_mY_{m+1}\ldots Y_n \\
\vdots \\
Y_1\ldots Y_{m-1}Z_mY_{m+1}\ldots Y_n \to Y_1\ldots Y_{m-1}Y_mY_{m+1}\ldots Y_n \\
.
\end{array}$$

Korollárium

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

Bizonyítás: Az $L = \{a^n b^n c^n \mid n \geq 1\}$ nyelv hossz-nemcsökkentő grammatikával generálható, így környezetfüggő, viszont nem generálható környezetfüggetlen grammatikával.

Megjegyzés: A hossz-nemcsökkentő tulajdonság íly módon ekvivalens a környezetfüggőséggel, azzal a kivétellel, hogy környezetfüggő grammatikák esetében az $S \to \varepsilon$ szabály megléte megengedett.

Megjegyzések a korolláriumhoz

Legyen $G = (\{S, A, B\}, \{a, b, c\}, P, S)$, ahol

$$\begin{aligned} \{S \to abc, \quad & S \to aAbc, \quad Ab \to bA, \\ Ac \to Bbcc, \quad & bB \to Bb, \quad aB \to aaA, \\ aB \to aa\} \quad & \end{aligned}$$

Akkor $L = \{a^n b^n c^n \mid n \geq 1\}$ és G hossz-nemcsökkentő. Legyen $S \Longrightarrow^* a^i A b^i c^i$. Akkor alkalmazzuk a 3. szabályt, majd a 4. szabályt és az $a^i b^i B b c^{i+1}$ szót kapjuk. Ezután csak az 5. szabályt alkalmazhatjuk és $a^i B b^{i+1} c^{i+1}$-hez jutunk. Ebből a szóból vagy az $a^{i+1} A b^{i+1} c^{i+1}$ szót vagy az $a^{i+1} b^{i+1} c^{i+1}$ szót nyerhetjük.

Megjegyzések a korolláriumhoz

Megmutatjuk, hogy L nem környezetfüggetlen. Tegyük fel az ellenkezőjét. Legyen $w = a^q b^q c^q \in L$, ahol p, q a Bar-Hillel lemmának megfelelő konstansok. Nyilvánvaló, hogy $|w| > q > p$. (Lásd a lemma bizonyítását.) Ekkor bármely $u, v, x, y, z \in \{a, b, c\}^*$-ra, amelyre $w = uxvyz$, $|xvy| \leq q$, $|xy| > 0$, a lemma alapján fennáll, hogy $uvz \in L$. Viszont, mivel xy az $\{a, b, c\}$-ből legalább egy betűt nem tartalmaz, így uvz nem lehet L eleme, és így a nyelv nem teljesíti a Bar-Hillel lemma feltételeit, vagyis a nyelv nem környezetfüggetlen.

Bar-Hillel lemma: Minden L környezetfüggetlen nyelvhez meg tudunk adni két p és q természetes számot úgy, hogy minden olyan w szó L-ben, amely hosszabb, mint p $w = uxvyz$ alakú, ahol $|xvy| \leq q$, $xy \neq \varepsilon$, és minden $ux^i vy^i z$ szó is benne van az L nyelvben minden $i \geq 0$ egész számra $(u, x, v, y, z \in T^*)$.

Tartalmazás (szóprobléma)

Tétel:

Minden környezetfüggő $G = (N, T, P, S)$ grammatika és minden $u \in T^*$ szó esetén eldönthető, hogy G-ben u levezethető-e vagy sem.

Bizonyításvázlat:

Ha $u = \varepsilon$, akkor a válasz triviális. Ha $u \in T^+$, akkor tekintsük az összes olyan véges

$$S = u_0, u_1, u_2, \ldots, u_n = u$$

sorozatot, ahol $u_i \in (N \cup T)^+$ és $|u_i| \leq |u_{i+1}|$ teljesül $0 \leq i \leq n-1$-re , valamint $u_i \neq u_j$, hacsak $i \neq j$.

Azon $v \in (N \cup T)^+$ szavak száma, amelyekre $|v| \leq |u|$ fennáll, véges, ezért a fenti sorozatok száma is véges.

Ezért szisztematikusan elő tudjuk az összes ilyen sorozatot állítani és le tudjuk ellenőrizni, hogy $u_i \Longrightarrow u_{i+1}$, $0 \leq i \leq n-1$ teljesül-e vagy sem.

Ha van ilyen sorozat, akkor $u \in L(G)$. (A legrövidebb levezetésben nincs szóismétlés.)

Kuroda normálforma

Definíció:

Egy hossz-nemcsökkentő $G = (N, T, P, S)$ grammatikát Kuroda normálformájúnak mondunk, ha minden egyes szabálya

1. vagy $A \to a$,
2. vagy $A \to B$,
3. vagy $A \to BC$
4. vagy $AB \to CD$

alakú, ahol $a \in T$ és $A, B, C, D \in N$.

Kuroda normálforma - folytatás

Tétel:

Minden hossz-nemcsökkentő grammatikához meg tudunk konstruálni egy vele ekvivalens Kuroda normálformájú grammatikát.

Bizonyításvázlat:

Legyen $G = (N, T, P, S)$ egy hossz-nemcsökkentő grammatika, és tegyük fel, hogy terminális szimbólumok csak $A \to a$ alakú szabályokban fordulnak elő, ahol $A \in N, a \in T$.

Ha $u \to v \in P$, valamint $|u| = 1$ és $|v| > 2$, akkor a $u \to v$-t 3. alakú (azaz, $A \to BC$ alakú) szabályokkal tudjuk helyettesíteni a Chomsky normálformára hozás algoritmusának megfelelően.

Ha $|u| = |v| = 2$, akkor az $u \to v$ szabály 4. alakú (azaz, $AB \to CD$ alakú).

Vagyis elegendő azzal az esettel foglalkoznunk, ha $|u| \geq 2$ és $|v| > 2$.

Bizonyításvázlat - folytatás

Legyen $u = X_1X_2\ldots X_m$ és $v = Y_1Y_2\ldots Y_n$, $2 \leq m$, $2 < n$, $X_i \in N, 1 \leq i \leq m$ és $Y_j \in N, 1 \leq j \leq n$. Tegyük fel, hogy $n > m$, az $n = m$ eset értelemszerű módosítással adódik. Akkor az

$$X_1X_2\ldots X_m \to Y_1Y_2\ldots Y_n$$

szabályt helyettesítjük a

$$\begin{array}{rr} X_1X_2 \to Y_1Z_2 & Z_m \to Y_mZ_{m+1} \\ Z_2X_3 \to Y_2Z_3 & Z_{m+1} \to Y_{m+1}Z_{m+2} \\ \vdots & \\ Z_{m-1}X_m \to Y_{m-1}Z_m & Z_{n-1} \to Y_{n-1}Y_n \end{array}$$

szabályokkal, ahol $Z_2, \ldots, Z_{n-1}$ új, a szabályhoz bevezetett nemterminálisok. (Minden fenti alakú szabályhoz páronként diszjunkt új nemterminálishalmazt vezetünk be.)

Bizonyításvázlat - szemléltetés

A $X_1X_2\dots X_m \to Y_1Y_2\dots Y_n$ szabály végrehajtásának szimulálása:

$$X_1X_2X_3\dots X_{m-1}X_m \Longrightarrow Y_1Z_2X_3\dots X_{m-1}X_m \Longrightarrow Y_1Y_2Z_3\dots X_{m-1}X_m$$

$$\Longrightarrow Y_1Y_2Y_3\dots Z_{m-1}X_m \Longrightarrow Y_1Y_2Y_3\dots Y_{m-1}Z_m \Longrightarrow Y_1Y_2Y_3\dots Y_{m-1}Y_mZ_{m+1}$$

$$\Longrightarrow Y_1Y_2Y_3\dots Y_mY_{m+1}Z_{m+2} \Longrightarrow^* Y_1Y_2Y_3\dots Y_{m-1}Y_{m+1}Y_{m+2}\dots Y_{n-1}Y_n$$

Az alkalmazott szabályok:

$$\begin{array}{rr}
X_1X_2 \to Y_1Z_2 & Z_m \to Y_mZ_{m+1} \\
Z_2X_3 \to Y_2Z_3 & Z_{m+1} \to Y_{m+1}Z_{m+2} \\
\vdots & \\
Z_{m-1}X_m \to Y_{m-1}Z_m & Z_{n-1} \to Y_{n-1}Y_n
\end{array}$$

Bizonyításvázlat - folytatás

Legyen N' az új és az eredeti nemterminálisok halmaza, P' pedig a P halmazból megmaradó és a fenti átalakítással kapott szabályok halmaza. Akkor könnyen látható, hogy a $G' = (N', T, P', S)$ grammatikára $L(G) \subseteq L(G')$ teljesül. A fordított irányú tartalmazás is igazolható.

Kuroda normálforma - folytatás

Korollárium

Minden ε-mentes környezetfüggő nyelv generálható Kuroda normálformájú grammatikával.

0-**típusú grammatikák egy normálformája**

Tétel

Minden $G = (N, T, P, S)$ 0-típusú grammatikához létezik egy vele ekvivalens G' generatív grammatika, amelynek minden egyes szabályának alakja egyike az alábbiaknak:

$$\begin{array}{rl} S \to \varepsilon, & A \to a, \\ A \to B, & A \to BC, \\ AB \to AC, & AB \to CB, \\ AB \to B, & \end{array}$$

ahol $a \in T$, $S, A, B, C, \in N$ és az S kezdőszimbólum csak a szabályok baloldalán fordul elő.

Bizonyításvázlat

Legyen $G = (N, T, P, S)$ egy tetszőleges 0-típusú grammatika. Tegyük fel, hogy P tartalmaz $u \to \varepsilon$ alakú szabályt, ahol $u \in (N \cup T)^+$. Ekkor minden $u \to \varepsilon$ alakú szabályt, ahol $u \in (N \cup T)^+$, helyettesítsünk $uX \to X$ és $Xu \to X$ alakú szabályokkal minden egyes $X \in (N \cup T)$-re. Jelöljük az újonnan kapott szabályhalmazt P'-vel. Könnyen látható, hogy a $G' = (N, T, P', S)$ grammatika az $L(G') = L(G) - \{\varepsilon\}$ nyelvet generálja.

Ha $\varepsilon \in L(G)$, akkor adjuk hozzá P-hez az $S' \to \varepsilon$ szabályt, ahol $S' \notin N$.

Ezután P'-t átalakítjuk úgy, hogy terminális szimbólumok csak 2. alakú, azaz $A \to a$, $(A \in N, a \in T)$ alakú szabályokban fordulhassanak elő. Ekkor minden további szabály $u \to v$ alakú lesz, ahol $u, v \in N^+$.

A hossz-nemcsökkentő szabályok (azaz, ahol $|u| \leq |v|$) esetében a Kuroda normálformára hozás során alkalmazott módszert alkalmazzuk.

Bizonyításvázlat - folytatás

Ezután a hossz-csökkentő szabályokkal foglalkozunk. Minden $X_1 \dots X_m \to Y_1 \dots Y_n$, ahol $m > n \geq 1$, és $X_i, Y_j \in N$, alakú szabályt helyettesítünk egy szabályhalmazzal, ahol $U_1, \dots U_m$, és $Z_1, \dots Z_n$ új, a szabályhoz bevezetett nemterminálisok:

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

Itt $AB \to CD$ típusú szabályokat váltunk ki, négy 5. vagy 6. alakú szabállyal ($AB \to AC$, $AB \to CB$). Így egy olyan grammatikát kapunk, amely ugyanazt a nyelvet generálja, mint az eredeti G grammatika.

Bizonyításvázlat - szemléltetés

A szabály alkalmazása:

$$X_1X_2\dots X_{m-2}X_{m-1}X_m \Longrightarrow X_1X_2\dots X_{m-2}Z_mU_m \Longrightarrow X_1X_2\dots X_{m-2}U_m \Longrightarrow X_1X_2\dots Z_{m-1}U_{m-1}$$
$$\Longrightarrow X_1X_2\dots U_{m-1} \Longrightarrow^* X_1X_2\dots X_nU_{n+2} \Longrightarrow X_1X_2\dots Z_{n+1}U_{n+1} \Longrightarrow X_1X_2\dots U_nY_n$$
$$\Longrightarrow X_1X_2\dots U_nY_n \Longrightarrow X_1U_2Y_2\dots Y_n \Longrightarrow^* U_1Y_1Y_2\dots Y_n \Longrightarrow Y_1Y_2\dots Y_n$$

$$\begin{array}{rl}
X_{m-1}X_m \to Z_mU_m & Z_mU_m \to U_m \\
X_{m-2}U_m \to Z_{m-1}U_{m-1} & Z_{m-1}U_{m-1} \to U_{m-1} \\
\vdots & \\
X_nU_{n+2} \to Z_{n+1}U_{n+1} & Z_{n+1}U_{n+1} \to U_nY_n \\
X_{n-1}U_n \to U_{n-1}Y_{n-1} & \\
\vdots & \\
X_1U_2 \to U_1Y_1 & \\
U_1Y_1 \to Y_1, &
\end{array}$$

Formális nyelvek - 7. 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

Reguláris kifejezések

Motiváció

Ismeretes, hogy **minden véges nyelv reguláris**. Tudjuk továbbá, hogy az $\mathcal{L}_3$ **nyelvosztály** (a reguláris nyelvek osztálya) **zárt az unió, a konkatenáció és az iteráció lezártja** műveletekre nézve.

Következésképpen, kiindulva véges számú véges nyelvből és az előzőekben felsorolt, ún. reguláris műveleteket véges sokszor alkalmazva reguláris nyelvet kapunk.

Kérdés az, hogy vajon ezzel az **eljárással minden reguláris nyelvet elő tudunk-e állítani**, azaz, ez a módszer elégséges-e az $\mathcal{L}_3$ nyelvosztály leírására?

Reguláris kifejezések

Definíció

Legyenek V és $V' = \{\varepsilon, \cdot, +, *, (,)\}$ diszjunkt ábécék. A $V \cup V'$ ábécé feletti reguláris kifejezéseket rekurzív módon a következőképpen definiáljuk:

1. ε reguláris kifejezés V felett.

2. Minden $a \in V$ reguláris kifejezés V felett.

3. Ha R reguláris kifejezés V felett, akkor $(R)^*$ is reguláris kifejezés V felett.

4. Ha Q és R reguláris kifejezések V felett, akkor $(Q) \cdot (R)$ és $(Q) + (R)$ is reguláris kifejezések V felett.

A $*$, $\cdot$ és $+$ szimbólumok rendre az iteráció lezártjára, a konkatenációra és az unióra utalnak, azt jelölik.

A reguláris kifejezések által jelölt nyelv

Minden reguláris kifejezés jelöl (meghatároz) valamely reguláris nyelvet.

A $V \cup V'$ ábécé felett megadott R reguláris kifejezés által jelölt nyelvet L_R-el jelöljük és a következőképpen definiáljuk:

- $L_\varepsilon = \{\varepsilon\}$,
- $L_a = \{a\}$, minden $a \in V$-re,
- Továbbá minden R, Q reguláris kifejezésre $V \cup V'$ felett $L_{(R+Q)} = L_R \cup L_Q$, $L_{(R \cdot Q)} = L_R L_Q$, valamint $L_{(R)^*} = (L_R)^*$.

Például a az $\{a\}$ nyelvet, $a + b$ az $\{a\} \cup \{b\} = \{a, b\}$ nyelvet és $a \cdot b$ az $\{a\}\{b\} = \{ab\}$ nyelvet jelöli.

Példák

Legyen $V = \{a, b\}$. Az alábbi reguláris kifejezések mellett az általuk jelölt nyelv található.

Megjegyzés:

A zárójelek egyrésze elhagyható, ha a műveleteken precedenciát definiálunk. A szokásos sorrend $*, \cdot, +$.

- a^* ugyanaz, mint $(a)^*$ és az $\{a\}^*$ nyelvet jelöli.
- $(a+b)^*$ ugyanaz, mint $((a)+(b))^*$ és az $\{a,b\}^*$ nyelvet jelöli.
- $a^* \cdot b$ ugyanaz, mint $((a)^*) \cdot (b)$ és az $\{a\}^* b$ nyelvet jelöli.
- $b + a \cdot b^*$ ugyanaz, mint $(b) + ((a) \cdot (b)^*)$ és a $\{b\} \cup \{a\}\{b\}^*$ nyelvet jelöli.
- $(a+b) \cdot a^*$ ugyanaz, mint $((a)+(b)) \cdot ((a)^*)$ és az $\{a,b\}\{a\}^*$ nyelvet jelöli.

Egyenlőségek reguláris kifejezésekre

Könnyen látható, hogy $\{a, b\}\{a\}^* = \{a\}\{a\}^* \cup \{b\}\{a\}^*$. Így

$$(a + b) \cdot a^* = a \cdot a^* + b \cdot a^*,$$

azaz, a két reguláris kifejezés ugyanazt a nyelvet jelöli.

Legyenek P, Q, R reguláris kifejezések. Akkor P, Q és R helyébe reguláris kifejezéseket írva fennállnak az alábbi egyenlőségek.

$$P + (Q + R) = (P + Q) + R \qquad P \cdot (Q \cdot R) = (P \cdot Q) \cdot R$$

$$P + Q = Q + P \qquad P \cdot (Q + R) = P \cdot Q + P \cdot R$$

$$(P + Q) \cdot R = P \cdot R + Q \cdot R \qquad P^* = \varepsilon + P \cdot P^*$$

$$\varepsilon \cdot P = P \cdot \varepsilon = P \qquad P^* = (\varepsilon + P)^*$$

Egyenlőségek reguláris kifejezésekre - folytatás

Ha a fenti egyenlőségekben a P, Q, R reguláris kifejezéseket reguláris kifejezésekkel helyettesítjük, reguláris kifejezéseket kapuk.

Azonban sem az előbbi egyenlőségekből, sem egyenlőségek más véges halmazából nem kaphatjuk meg az összes reguláris kifejezést kizárólag helyettesítés segítségével.

Még egy további szabályra van szükségünk, nevezetesen, ha

$P = R + P \cdot Q$ és $\varepsilon \notin Q$, akkor $P = R \cdot Q^*$.

Egyenlőségek reguláris kifejezésekre - folytatás

A teljesség biztosítása céljából még hozzáadjuk az $\emptyset$ szimbólumot a reguláris kifejezések halmazához, amely az üres nyelvet jelöli.

Ebben az esetben nincs szükségünk a ε szimbólumra, mivel $\emptyset^* = \{\varepsilon\}$.

Így, a definícióban helyettesíthetjük a ε szimbólumot az $\emptyset$ szimbólummal.

Ekkor helyettesítjük ε-t a megelőző axióma rendszerben $(\emptyset)^*$-gal és még egy további axiómát tekintünk:

$$\emptyset \cdot P = P \cdot \emptyset = \emptyset.$$

Az egyenlőségek, valamint a helyettesítés és a fenti feltételes egyenlőség elégséges ahhoz, hogy levezessünk minden érvényes egyenlőséget reguláris kifejezések között.

Reguláris kifejezések versus reguláris nyelvek

Tétel

Minden reguláris kifejezés egy reguláris (3-típusú) nyelvet jelöl, és megfordítva, minden reguláris nyelvhez megadható egy, ezen nyelvet jelölő reguláris kifejezés.

Bizonyításvázlat

1) Az állítás első fele a megelőző diszkusszióból következik.

2) Megmutatjuk, hogy minden L reguláris nyelvhez, amelyet a $G = (N, T, P, S)$ normálformában adott reguláris grammatika generál, meg tudunk konstruálni egy reguláris kifejezést, amely az L nyelvet jelöli.

Legyen $N = \{A_1, \ldots, A_n\}$, $n \geq 1$, $S = A_1$. (G minden szabálya vagy $A_i \to aA_j$ vagy $A_i \to \varepsilon$ alakú, ahol $a \in T$, $1 \leq i, j \leq n$.)

Azt mondjuk, hogy az $A_i \Longrightarrow^* uA_j$ ($u \in T^*$) levezetés **érinti** az A_m nemterminálist, ha A_m előfordul valamely közbülső mondatformában A_i és uA_j között a levezetésben.

Az $A_i \Longrightarrow^* uA_j$ levezetést k-**megszorítottnak** nevezzük, ha $0 \leq m \leq k$ teljesül minden A_m nemterminálisra, amely a levezetésben előfordul.

Bizonyításvázlat - folytatás

Definiáljuk a következő halmazokat:

$E_{i,j}^k = \{u \in T^* \mid$ létezik $A_i \Longrightarrow^* uA_j$ k-megszorított levezetés $\}$.

k**-szerinti indukcióval** bizonyítjuk, hogy az $E_{i,j}^k$ nyelvhez létezik őt jelölő reguláris kifejezés i,j,k-ra, ahol $0 \leq i,j,k \leq n$.

Bázis: $i \neq j$ esetén az $E_{i,j}^0$ halmaz vagy üres vagy T-beli betűkből áll. ($a \in E_{i,j}^0$, akkor és csak akkor, ha $A_i \to aA_j \in F$.) Ha $i = j$, akkor $E_{i,i}^0$ tartalmazza ε-t és nulla vagy több elemét T-nek, így $E_{i,j}^0$ reguláris kifejezéssel jelölhető.

Indukciós lépés: tegyük fel, hogy rögzített k-ra, $0 < k \leq n$, az $E_{i,j}^{k-1}$ nyelvek mindegyike jelölhető reguláris kifejezéssel. Akkor minden i,j,k-ra fennáll, hogy

$$E_{i,j}^k = E_{i,j}^{k-1} + E_{i,k}^{k-1} \cdot (E_{k,k}^{k-1})^* \cdot E_{k,j}^{k-1}.$$

Ekkor az indukciós feltevés alapján $E_{i,j}^k$ szintén jelölhető reguláris kifejezéssel.

Legyen I_ε azon i indexek halmaza, amelyekre $A_i \to \varepsilon$. Akkor $L(G) = \cup_{i \in I_\varepsilon} E_{1,i}^n$, azaz, L reguláris kifejezéssel jelölhető.

Helyettesítés

Definíció

Legyen V egy ábécé, valamint legyen minden $a \in V$-re V_a ábécé és $s(a) \subseteq V_a^*$. Minden $u = a_1 a_2 \ldots a_n \in V^*$ szóra definiáljuk az u szó s helyettesítését a következőképpen:

$$s(u) = s(a_1)s(a_2)\ldots s(a_n).$$

Legyen továbbá $s(\varepsilon) = \varepsilon$. Az s helyettesítés kiterjeszthető bármely $L \subseteq V^*$ nyelvre a következő módon: $s(L) = \{w \mid w \in s(u), u \in L\}$.

Reguláris nyelvek zártsága a helyettesítésre nézve

Reguláris kifejezéseket használva könnyen látható, hogy az $\mathcal{L}_3$ nyelvosztály zárt a helyettesítésre nézve. A reguláris kifejezések halmaza nyilvánvalóan zárt a kifejezés minden betűjének valamely reguláris kifejezéssel való helyettesítésére nézve. (Lásd a megelőző diszkussziót).

Megjegyzés: A helyettesítés a homomorfizmus általánosítása.

Formális nyelvek - 8.

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

Automaták

Formális nyelvek megadása nemcsak generatív, hanem felismerő eszközökkel is lehetséges, azaz olyan számítási eszközök segítségével, amelyek szavak (sztringek) feldolgozására és azonosítására alkalmasak.

Ilyen eszköz például az automata, amely egy szó (sztring) mint input hatására kétféleképpen viselkedhet: vagy elfogadja, vagy elutasítja (igen vagy nem). Megjegyezzük, hogy olyan lehetőség is ésszerű, hogy az automata egyáltalán nem ad választ.

Az automaták analitikus eszközök, míg a grammatikák szintetizáló megközelítést alkalmaznak.

Véges automata

Definíció

A véges automata egy rendezett ötös,

$$A = (Q, T, \delta, q_0, F),$$

- ahol Q állapotok véges nemüres halmaza,
- T input szimbólumok ábécéje,
- $\delta : Q \times T \to Q$, az ún. állapot-átmeneti (röviden átmeneti) függvény,
- $q_0 \in Q$ az ún. kezdeti állapot (kezdőállapot),
- $F \subseteq Q$ az ún. elfogadó állapotok halmaza.

Szokás az elfogadó állapot helyett végállapotot is mondani.

Véges automata - folytatás

A δ leképezés kiterjeszthető $Q \times T^*$ halmazból Q halmazba való $\widehat{\delta}$ leképezéssé a következő módon:

- $\widehat{\delta}(q, \varepsilon) = q$,

- $\widehat{\delta}(q, xa) = \delta(\widehat{\delta}(q, x), a)$ minden $x \in T^*$-ra és minden $a \in T$-re.

Véges automata - működési mód

A véges automata diszkrét időintervallumokban végrehajtott lépések sorozata által működik.

Minden egyes lépés során az automata elolvassa a következő input szimbólumot és átmegy egy olyan állapotba, amelyet az állapot-átmeneti függvény meghatároz az adott aktuális input szimbólumra és azon állapotra vonatkozóan, amelyben az automata az adott pillanatban van.

Véges automata - működési mód

Az input szó (input sztring) - vagy röviden az input - megadását elképzelhetjük oly módon, hogy a szó egy ún. input szalagon van, amely cellákból áll, és minden egyes cella tartalmaz egy szimbólumot.

Az input szalagon van egy ún. olvasófej, amelynek a mozgását egy vele összekötött, az állapotok változását kontrolláló ún. véges kontroll, valamint az a szimbólum határozza meg, amelyet az a cella tartalmaz, ahol az adott pillanatban az olvasófej áll.

Az olvasófej kezdetben az első cellán áll, a véges kontroll pedig a kezdőállapotot jelöli meg. Minden egyes pillanatban az olvasófej egy cellával jobbra mozdul, a véges kontroll pedig állapotát az automata aktuális állapotától és a cellában található betűtől függően változtatja meg.

Véges automata - működési mód

Kezdetben az A véges automata a q_0 kezdőállapotban van és az olvasófej az input szalagon levő $u \in T^*$ szó első betűjét dolgozza fel (olvassa el, a megfelelő cellán áll).

Ezután a véges automata lépések sorozatát végrehajtva elolvassa az input u szótát; betűről betűre haladva olvas és új állapotba kerül.

Véges automata - működési mód

Miután az u input szó utolsó betűjét is elolvasta a véges automata, vagy $q \in F$, azaz, elfogadó állapotba kerül, és akkor az u szót az automata elfogadja, vagy az új állapot nem lesz eleme F-nek, és ekkor az automata a szót nem fogadja el (visszautasítja).

Vegyük észre, hogy a fenti definíció alapján véges hosszúságú szót olvasva a véges automata sohasem fordul végtelen ciklusba, mivel minden egyes lépésben egy új szimbólumot kell elolvasnia. Ebből az adódik, hogy minden u szó esetén $|u|$ lépésben egyértelmű választ ad arra, hogy elfogadja-e a szót vagy visszautasítja.

Példa

Legyen

$$A = (Q, T, \delta, q_0, F)$$

véges automata, ahol

$$Q = \{q_0, q_1, q_2, q_3\}, \;\; T = \{a, b\}, \;\; F = \{q_0\}$$

és legyen

$$\begin{array}{ll} \delta(q_0, a) = q_2, & \delta(q_0, b) = q_1, \\ \delta(q_1, a) = q_3, & \delta(q_1, b) = q_0, \\ \delta(q_2, a) = q_0, & \delta(q_2, b) = q_3, \\ \delta(q_3, a) = q_1, & \delta(q_3, b) = q_2. \end{array}$$

Az A véges automata pontosan azokat a szavakat fogadja el, amelyek páros számú a betűt és páros számú b betűt tartalmaznak.

Példa - folytatás

$$\begin{array}{ll} \delta(q_0,a)=q_2, & \delta(q_0,b)=q_1, \\ \delta(q_1,a)=q_3, & \delta(q_1,b)=q_0, \\ \delta(q_2,a)=q_0, & \delta(q_2,b)=q_3, \\ \delta(q_3,a)=q_1, & \delta(q_3,b)=q_2. \end{array}$$

Tekintsük a $bbabab$ input szót.

Az állapotok sorozata akkor q_0, q_1, q_0, q_2, q_3, q_1, q_0.

Determinisztikus véges automata

A δ függvény egyértékű, ezért minden egyes (q, a) párra, ahol $(q, a) \in Q \times T$ egyetlen olyan p állapot létezik, amelyre $\delta(q, a) = p$ teljesül. Ezért ezt a véges automatát **determinisztikusnak** nevezzük.

Nemdeterminisztikus véges automata

Ha többértékű állapot-átmeneti függvényt is megengedünk, azaz δ a $Q \times T$ halmazból a 2^Q halmazba való leképezés, akkor **nemdeterminisztikus** véges automatáról beszélünk. Ebben az esetben aktuális állapotnak egy állapothalmaz valamely elemét, mintsem egyetlen állapotot tekinthetünk.

Ez azt jelenti, hogy a kezdeti állapot helyettesíthető egy $Q_0 \subseteq Q$ kezdőállapothalmazzal (kezdeti állapothalmazzal).

Az is előfordulhat, hogy egy a input szimbólum esetén $\delta(q, a)$ üres az aktuális állapotok mindegyikére.

Alternatív jelölés

Legyen $A = (Q, T, \delta, q_0, F)$ (determinisztikus) véges automata vagy nemdeterminsiztikus

Az A automata állapot-átmeneteit

$$qa \to p$$

alakú szabályok formájában is írhatjuk $p \in \delta(q, a)$ esetén.

Jelöljük M_δ-val az $A = (Q, T, \delta, Q_0, F)$ nemdeterminisztikus véges automata δ állapot-átmenet függvénye által az előbbi módon származó szabályok halmazát.

Ha minden egyes (q, a) párra egyetlen $qa \to p$ szabály van M_δ-ban, akkor a véges automata determinisztikus, egyébként nemdeterminisztikus.

Közvetlen redukció

Definíció

Legyen $A = (Q, T, \delta, q_0, F)$ egy véges automata és legyenek $u, v \in QT^*$ szavak (ahol QT^* a Q és T^* konkatenációját jelöli).

Azt mondjuk, hogy az A automata az u szót a v szóra redukálja egy lépésben vagy közvetlenül ($u \Longrightarrow_A v$), ha van olyan $qa \to p$ szabály M_δ-ban (azaz, $\delta(q, a) = p$) és van olyan $w \in T^*$ szó, amelyre $u = qaw$ és $v = pw$ teljesül.

Redukció

Definíció

Az $A = (Q, T, \delta, q_0, F)$ véges automata az $u \in QT^*$ szót a $v \in QT^*$ szóra redukálja ($u \Longrightarrow_A^* v$), ha vagy $u = v$, vagy van olyan $z \in QT^*$, amelyre $u \Longrightarrow_A^* z$ és $z \Longrightarrow_A v$ teljesül.

A $\Longrightarrow_A$ reláció, valamint tranzitív és reflexív lezártja, $\Longrightarrow_A^*$, lényegében a grammatikák elméletéből ismert levezetésnek felel meg.

Az automata által elfogadott nyelv

Az $A = (Q, T, \delta, q_0, F)$ véges automata által elfogadott (vagy másképpen felismert) nyelv alatt az

$$L(A) = \{u \in T^* | q_0 u \Longrightarrow_A^* p,\ q_0 \in Q_0 \text{ és } p \in F\}$$

szavak halmazát értjük.

Az üres szó, ε, akkor és csak akkor van benne az automata által elfogadott $L(A)$ nyelvben, ha $Q_0 \cap F \neq \emptyset$.

Minden $qa \to p \in M_\delta$ $(q, p \in Q, a \in T, \delta(q, a) = p)$ állapot-átmeneti szabály hossz-csökkentő, de könnyen belátható, hogy a redukció megfelel egy bal-lineáris grammatikabeli levezetés megfordítottjának.

Tétel

Minden A nemdeterminisztikus véges automatához meg tudunk adni egy 3-típusú G grammatikát úgy, hogy $L(G) = L(A)$ teljesül.

Bizonyításvázlat

Először megmutatjuk, hogy $L(A) \subseteq L(G)$.

Legyen $A = (Q, T, \delta, Q_0, F)$ egy nemdeterminisztikus véges automata. Definiáljuk a $G = (N, T, P, S)$ grammatikát úgy, hogy $N = Q \cup \{S\}$ és legyen

1. $p \to a \in P$ akkor és csak akkor, ha $q_0 a \to p \in M_\delta$ valamely $q_0 \in Q_0$-ra,
2. $p \to qa \in P$ akkor és csak akkor, ha $qa \to p \in M_\delta$,
3. $S \to p \in P$ akkor és csak akkor, ha $p \in F$,
4. $S \to \varepsilon \in P$ akkor és csak akkor, ha $Q_0 \cap F \neq \emptyset$.

Bizonyításvázlat - folytatás

1. Nyilvánvaló, hogy $\varepsilon \in L(G)$ akkor és csak akkor, ha $\varepsilon \in L(A)$.

2. Tegyük fel, hogy $u \neq \varepsilon \in L(A)$. Akkor van olyan $q_0 \in Q_0$ kezdőállapot és olyan $p \in F$ elfogadó állapot, hogy $q_0 u \Longrightarrow_A^* p$ teljesül. Tekintve ennek a redukciónak a megfordítását, a 2. csoportba tartozó szabályok használatával meg tudjuk konstruálni a $p \Longrightarrow_G^* q_0 u$ levezetést.

3. A G-beli levezetés utolsó lépése $p_1 \rightarrow q_0 a$ alakú szabály alkalmazása lenne, ahol $q_0 \in Q_0$. Ezért az 1. szabálycsoportban van $p_1 \rightarrow a$ alakú szabály. Ebből az adódik, hogy $p \Longrightarrow_G^* u$.

4. Végezetül, tekintsük az $S \rightarrow p$ szabályt a 3. szabálycsoportból és így az $S \Longrightarrow_G p \Longrightarrow_G^* u$ levezetéshez jutunk.

Vagyis, $L(A) \subseteq L(G)$.

Bizonyításvázlat - folytatás

Megmutatjuk, hogy $L(G) \subseteq L(A)$.

1. Legyen $u \in L(G)$ és $u \neq \varepsilon$, azaz legyen $S \Longrightarrow_G^* u$, ahol $u \in T^+$.

2. A G konstrukciója alapján akkor létezik a

$$S \Longrightarrow_G p \Longrightarrow_G^* p_1 v \Longrightarrow_G av = u$$

levezetés, ahol $p \in F$ és $p_1 \to a \in P$.

3. Ebből az következik, hogy $q_0 u \Longrightarrow_A^* u$, így $u \in L(A)$.

A G grammatika szabályai bal-lineárisak, de ismeretes, hogy minden bal-lineáris grammatikához létezik vele ekvivalens jobb-lineáris grammatika, így a tétel állítása fennáll.

Tétel

Minden 3-típusú G grammatikához meg tudunk adni egy A véges automatát úgy, hogy $L(A) = L(G)$ teljesül.

Bizonyításvázlat

1. Az általánosság megszorítása nélkül feltehetjük, hogy $G = (N, T, P, S)$ normálformában van (minden szabály vagy $X \to aY$, vagy $X \to \varepsilon$ alakú, ahol $X, Y \in N$ és $a \in T$).

2. Legyen $A = (Q, T, \delta, Q_0, F)$ véges automata úgy, hogy $Q = N$, $Q_0 = \{S\}$ és $F = \{Z \in N \mid Z \to \varepsilon \in P\}$. Legyen M_δ úgy definiálva, hogy

$$Xa \to Y \in M_\delta \text{ akkor és csak akkor, ha } X \to aY \in P.$$

3. Könnyen belátható, hogy az $S \Longrightarrow_G^* u$ G-beli levezetésre van A-ban olyan redukció, amelyre $Su \Longrightarrow_A^* Z$ teljesül, ahol $Z \in F$.

4. Megfordítva, minden A-beli előbbi alakú redukcióhoz tudunk egy megfelelő levezetést találni G-ben. Azaz, az állítás fennáll.

Formális nyelvek - 9.

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

Véges automata - folytatás

Ismeretes, hogy minden A nemdeterminisztikus véges automatához meg tudunk konstruálni egy G reguláris grammatikát úgy, hogy $L(G) = L(A)$ teljesül. Az is tudott, hogy minden G reguláris grammatikához meg tudunk konstruálni egy A véges automatát úgy, hogy $L(A) = L(G)$.

Kérdés: Létezik-e olyan reguláris nyelv, amely nemdeterminisztikus véges automatával felismerhető, de nem ismerhető fel determinisztikus véges automatával?

Válasz: Nem létezik.

Determinisztikus versus nemdeterminisztikus véges automata

Tétel

Minden $A = (Q, T, \delta, Q_0, F)$ nemdeterminisztikus véges automatához meg tudunk konstruálni egy $A' = (Q', T, \delta', q'_0, F')$ determinisztikus véges automatát úgy, hogy $L(A) = L(A')$ teljesül.

Bizonyításvázlat

Legyen Q' a Q halmaz összes részhalmazainak halmaza.

Definiáljuk a $\delta' : Q' \times T \to Q'$ függvényt a következőképpen:

$\delta'(q', a) = \bigcup_{q \in q'} \delta(q, a)$.

Továbbá legyen $q'_o = Q_0$ és $F' = \{q' \in Q' \mid q' \cap F \neq \emptyset\}$

Először megmutatjuk, hogy $L(A) \subseteq L(A')$. Az állítás bizonyításához megmutatjuk, hogy a következő lemma állítása teljesül.

Bizonyításvázlat - folytatás

Lemma 1

Minden $p, q \in Q$, $q' \in Q'$ és $u, v \in T^*$ esetén, ha

$$qu \Longrightarrow_A^* pv \text{ és } q \in q',$$

akkor van olyan $p' \in Q'$, hogy

$$q'u \Longrightarrow_{A'}^* p'v \text{ és } p \in p'.$$

Lemma 1 bizonyításának vázlata:

Az állítás a $qu \Longrightarrow_A^* pv$ redukcióban szereplő lépések száma szerinti indukcióval bizonyítható.

Nulla számú lépés esetében az állítás triviálisan fennáll. Tegyük fel, hogy teljesül valamely n lépésre, ahol $n \geq 0$.

Lemma 1 bizonyításának vázlata - folytatás

(**Lemma 1** : Minden $p, q \in Q$, $q' \in Q'$ és $u, v \in T^*$ esetén, ha

$$qu \Longrightarrow_A^* pv \text{ és } q \in q',$$

akkor van olyan $p' \in Q'$, hogy

$$q'u \Longrightarrow_{A'}^* p'v \text{ és } p \in p'.)$$

Álljon a $qu \Longrightarrow_A^* pv$ redukció $n+1$ lépésből. Akkor valamely $q_1 \in Q$ és $u_1 \in T^*$-re

$$qu \Longrightarrow_A q_1u_1 \Longrightarrow_A^* pv$$

teljesül. Így van olyan $a \in T$, amelyre $u = au_1$ és $q_1 \in \delta(q, a)$. De $\delta(q, a) \subseteq \delta'(q', a)$, hacsak $q \in q'$, így választhatjuk q_1'-et $q_1' = \delta'(q', a)$-nek, ahonnan

$$q'u \Longrightarrow_{A'} q_1'u_1$$

következik $q_1 \in q_1'$-re. Az indukciós feltevés alapján valamely $p' \in Q'$-re

$$q_1'u_1 \Longrightarrow_{A'}^* p'v \text{ és } p \in p'.$$

Azaz, Lemma 1 fennáll.

Bízonyításvázlat - folytatás

Legyen $u \in L(A)$, azaz teljesüljön $q_0 u \Longrightarrow_A^* p$ valamely $q_0 \in Q_0$-ra és legyen $p \in F$.

Ekkor Lemma 1 alapján valamely p'-re fennáll, hogy $q_0' u \Longrightarrow_{A'}^* p'$ és $p \in p'$.

Az F' definíciója alapján $p \in p'$-ből és $p \in F$-ből az következik, hogy $p' \in F'$, ahonnan $L(A) \subseteq L(A')$.

Bizonyításvázlat - folytatás

Ezután megmutatjuk, hogy $L(A') \subseteq L(A)$.

Az állítást Lemma 2 segítségével bizonyítjuk.

Lemma 2

Minden $p', q' \in Q'$, $p \in Q$ és $u, v \in T^*$ esetén, ha

$$q'u \Longrightarrow^*_{A'} p'v \text{ és } \; p \in p',$$

akkor van olyan $q \in Q$, hogy

$$qu \Longrightarrow^*_{A} pv \text{ és } \; q \in q'.$$

Bizonyításvázlat - folytatás

Lemma 2 bizonyításának vázlata: A lépések száma szerinti indukcióval történik. Nulla számú lépés esetén az állítás triviális.

Tegyük fel, hogy az állítás igaz n lépésre, ahol $n \geq 0$, és álljon a $q'u \Longrightarrow^*_{A'} p'v$ redukció $n+1$ lépésből.

Akkor

$$q'u \Longrightarrow^*_{A'} p'_1 v_1 \Longrightarrow_{A'} p'v,$$

ahol $v_1 = av$ valamely $a \in T$-re és $p'_1 \in Q'$.

Ekkor

$$p \in p' = \delta'(p'_1, a) = \bigcup_{p \in p'_1} \delta(p_1, a),$$

azaz, lennie kell olyan $p_1 \in p'_1$-nek, amelyre $p \in \delta(p_1, a)$.

Ekkor p_1-re teljesül, hogy $p_1 v_1 \Longrightarrow_A pv$, és az indukciós hipotézis alapján $q \in q'$, ahol $qu \Longrightarrow^*_A p_1 v_1$. Azaz, Lemma 2 állítása teljesül.

Bizonyításvázlat - folytatás

Végezetül legyen $q_0'u \Longrightarrow^*_{A'} p'$ és $p' \in F$. Az F' definíciója alapján van olyan $p \in p'$, ahol $p \in F$, és így Lemma 2 alapján valamely $q_0 \in q_0'$-ra teljesül $q_0u \Longrightarrow^*_A p$. Így $L(A') \subseteq L(A)$.

Ezzel a tétel bizonyítását befejeztük.

Példa

Legyen $A = (Q, T, \delta, Q_0, F)$ nemdeterminisztikus véges automata, ahol

$$Q = \{0, 1, 2\}, T = \{a, b\}, Q_0 = \{0\}, F = \{2\},$$
$$\delta(0, a) = \{0, 1\}, \quad \delta(0, b) = \{1\},$$
$$\delta(1, a) = \emptyset, \quad \delta(1, b) = \{2\},$$
$$\delta(2, a) = \{0, 1, 2\}, \quad \delta(2, b) = \{1\}.$$

Példa - folytatás

A vele ekvivalens determinisztikus véges automata állapotai a következők lesznek:

$$\emptyset, \{0\}, \{1\}, \{2\}, \{0,1\}, \{0,2\}, \{1,2\}\{0,1,2\}.$$

A további komponensek pedig $F = \{\{2\}, \{0,2\}, \{1,2\}, \{0,1,2\}\}$, valamint δ', ahol

$$\begin{aligned}
\delta'(\emptyset, a) &= \emptyset, & \delta'(\emptyset, b) &= \emptyset,\\
\delta'(\{0\}, a) &= \{0,1\}, & \delta'(\{0\}, b) &= \{1\},\\
\delta'(\{1\}, a) &= \emptyset, & \delta'(\{1\}, b) &= \{2\},\\
\delta'(\{2\}, a) &= \{0,1,2\}, & \delta'(\{2\}, b) &= \{1\},\\
\delta'(\{0,1\}, a) &= \{0,1\}, & \delta'(\{0,1\}, b) &= \{1,2\},\\
\delta'(\{0,2\}, a) &= \{0,1,2\}, & \delta'(\{0,2\}, b) &= \{1\},\\
\delta'(\{1,2\}, a) &= \{0,1,2\}, & \delta'(\{1,2\}, b) &= \{1,2\},\\
\delta'(\{0,1,2\}, a) &= \{0,1,2\}, & \delta'(\{0,1,2\}, b) &= \{1,2\}.
\end{aligned}$$

Következmények

Korollárium

A reguláris nyelvek osztálya ($\mathcal{L}_3$) zárt a komplemens műveletére nézve.

A bizonyítás vázlata :

Legyen L egy $A = (Q, T, \delta, q_0, F)$ véges automata által felismert nyelv. Akkor $\bar{L} = T^* - L$ felismerhető a $\bar{A} = (Q, T, \delta, q_0, Q - F)$ véges automatával.

Következmények

Korollárium

A reguláris nyelvek osztálya zárt a metszet műveletére nézve.

Bizonyításvázlat:

Ismeretes, hogy a reguláris nyelvek osztálya zárt az unió műveletére nézve. Mivel

$$L_1 \cap L_2 = \overline{\bar{L}_1 \cup \bar{L}_2},$$

az állítás következik.

Következmények

Korollárium

Eldönthető, hogy két reguláris grammatika azonos nyelvet generál-e vagy sem.

Bizonyításvázlat:

Legyenek G_1 és G_2 reguláris grammatikák, amelyek rendre az L_1 és L_2 nyelvet generálják. Az $L_3 = (L_1 \cap \bar{L}_2) \cup (\bar{L}_1 \cap L_2)$ nyelv szintén reguláris, így van olyan G_3 reguláris grammatika, amely L_3-at generálja. Ekkor azonban $L_1 = L_2$ akkor és csak akkor, ha $L_3 = \emptyset$, amely minden G_3 környezetfüggetlen grammatika esetében eldönthető.

A reguláris nyelvosztály egy algebrai karakterizációja

Legyen L egy T ábécé feletti nyelv.

Az L nyelv által indukált E_L reláció alatt egy olyan bináris relációt értünk a T feletti szavak halmazán, amelyre a következő teljesül: bármely $u, v \in T^*$-ra uE_Lv akkor és csak akkor, ha nincs olyan $w \in T^*$, hogy az uw és a vw szavak közül pontosan egy eleme L-nek.

Megjegyzés: E_L ekvivalencia reláció és jobb-invariáns. (Jobb-invariáns: ha uE_Lv, akkor uwE_Lvw is fennáll minden $w \in T^*$ szóra).

Az E_L reláció indexén ekvivalencia osztályainak számát értjük.

Tétel

$L \subseteq T^*$ akkor és csak akkor ismerhető fel determinisztikus véges automatával, ha E_L véges indexű.

Bizonyításvázlat

Tegyük fel, hogy $L \in T^*$ és E_L véges indexű. Legyenek $u_0 = \varepsilon$, $u_1, \ldots, u_n$ az egyes ekvivalencia osztályokat reprezentáló szavak és jelöljük $[u_i]$-vel, $0 \le i \le n$, az ekvivalencia osztályt.

Legyen $A = (Q, T, \delta, s_0, F)$ determinisztikus véges automata, ahol

$Q = \{[u_i] \mid 0 \le i \le n\}$, $[u_0]$ a kezdőállapot és $F = \{[u_j] \mid u_j \in L, 1 \le j \le n\}$.

Definiáljuk a δ állapot-átmenet függvényt a következő módon: $\delta([u_i], a) = [u_i a]$, $0 \le i \le n$.

Mivel E_L jobb-invariáns, ezért δ értékei nem függenek az $[u_i]$ osztályokat reprezentáló elemek választásától, továbbá, az elfogadó állapotok halmaza sem függ reprezentánsainak választásától.

A δ definíciójából közvetlenül adódik, hogy bármely $u \in T^*$ szóra $[q_0]u \Longrightarrow^* [u]$.

Ebből adódik, hogy L felismerhető egy determinisztikus véges automatával.

Bizonyításvázlat - folytatás

Tegyük most fel, hogy $L \subseteq T^*$ felismerhető egy $A = (Q, T, \delta, q_0, F)$ determinisztikus véges automatával.

Definiáljuk az E ekvivalencia relációt T^*-on a következőképpen: az u és v T feletti szavakra uEv akkor és csak akkor, ha van az A automatának olyan q állapota, hogy

$$q_0 u \Longrightarrow_A^* q \text{ és } q_0 v \Longrightarrow_A^* q$$

teljesül. (Nyilvánvaló, hogy E ekvivalencia reláció.)

Megmutatjuk, hogy uEv-ből uE_Lv következik. Tegyük fel az ellenkezőjét, vagyis, hogy valamely u', v', w' szavakra fennáll az, hogy $u'Ev'$, de csak egyike az $u'w'$ és a $v'w'$ szavaknak eleme az L nyelvnek.

Ekkor $q_0u'w' \Longrightarrow^* q_1$, ahol $q_1 \in F$ és $q_0v'w' \Longrightarrow^* q_2$, ahol $q_2 \notin F$. Mivel azonban $u'Ev'$ fennáll, ezért léteznie kell olyan $q \in Q$ állapotnak, hogy $q_0u \Longrightarrow_A^* q$ és $q_0u \Longrightarrow_A^* q$ teljesül.

Ez ellentmondás, így állításunk fennáll, ami azt jelenti, hogy E_L indexe nem nagyobb, mint E indexe, amely nem nagyobb az A automata állapotainak számánál, vagyis véges.

Példa a tétel használatára

Legyen $L = \{a^k b^k \mid 0 \leq k\}$. Megmutatjuk, hogy L nem reguláris nyelv. Tegyük fel az ellenkezőjét. Akkor létezik olyan determinisztikus véges automata, amely elfogadja L-et. Ebben az esetben E_L véges indexű. Ha E_L véges indexű, akkor $a^m E_L a^n$ fennáll valamely $m > n$ számpárra. De ez nem lehetséges, mivel csak az $a^m b^m$ és az $a^n b^m$ szavak egyike eleme L-nek. Ebből következően E_L nem lehet véges indexű, azaz, L nem reguláris nyelv.

Ez egy újabb megerősítése annak, hogy $\mathcal{L}_3 \subset \mathcal{L}_2$.

L generálható a $G = (\{S\}, \{a, b\}, \{S \rightarrow \varepsilon, S \rightarrow aSb\}, S)$ lineáris grammatikával.

Minimális számú állapottal rendelkező véges automata

Az A determinisztikus véges automata **minimális állapotszámú (minimális)**, ha nincs olyan A' determinisztikus véges automata, amely ugyanazt a nyelvet ismeri fel, mint A, de A' állapotainak száma kisebb, mint A állapotainak száma.

Tétel

Az L reguláris nyelvet felismerő minimális (állapotszámú) determinisztikus véges automata az izomorfizmus erejéig egyértelmű.

A bizonyítás az előző tétel alapján könnyen belátható.

Véges automata megkülönböztethető állapotai

Legyen $A = (Q, T, \delta, q_0, F)$ determinisztikus véges automata. Definiáljunk egy R relációt a Q állapothalmazon úgy, hogy pRq, ha minden egyes $x \in T^*$ input szóra fennáll, hogy $px \Longrightarrow_A^* r$ akkor és csak akkor, ha $qx \Longrightarrow_A^* r'$ valamely $r, r' \in F$ állapotokra. ($r = r'$ lehetséges).

Azt mondjuk, hogy p és q **megkülönböztethetők**, ha van olyan $x \in T^*$, amelyre vagy $px \Longrightarrow_A^* r$, $r \in F$, vagy $qx \Longrightarrow_A^* r'$, $r' \in F$, de mindkét redukció egyszerre nem áll fenn. Egyébként p és q nem megkülönböztethetők.

Ha p és q **nem megkülönböztethetők**, akkor $\delta(p, a) = s$ és $\delta(q, a) = t$ sem megkülönböztethetők egyetlen $a \in T$-re sem.

Összefüggő determinisztikus véges automata

Legyen $A = (Q, T, \delta, q_0, F)$ determinisztikus véges automata. A q állapotot a kezdőállapotból **elérhetőnek** mondjuk, ha létezik $q_0x \Longrightarrow^* q$ redukció, ahol x valamely T feletti szó.

Az $A = (Q, T, \delta, q_0, F)$ determinisztikus véges automatát **összefüggőnek** mondjuk, ha minden állapota elérhető a kezdőállapotból.

Összefüggő determinisztikus véges automata

Definiáljuk a H halmazt a következőképpen:

Legyen $H_0 = \{q_0\}$, $H_{i+1} = H_i \cup \{r \mid \delta(q,a) = r, q \in H_i, a \in T\}$, $i = 1, 2, \ldots$. Akkor létezik olyan $k \geq 0$, amelyre $H_k = H_l$, ahol $l \geq k$. Legyen $H = H_k$. Könnyen látható, hogy H azoknak az állapotoknak a halmaza, amelyek a kezdőállapotból elérhetők.

Ezután definiáljuk az $A' = (Q', T, \delta', q_0, F')$ determinisztikus véges automatát a következőképpen:

$Q' = H$, $F' = F \cap H$ és $\delta' : H \times T \to H$ úgy, hogy $\delta'(q,a) = \delta(q,a)$, hacsak $q \in H$.

Könnyen megmutatható, hogy A' összefüggő és ugyanazt a nyelvet fogadja el, amelyet A. Továbbá, A' az A legnagyobb összefüggő részautomatája.

Algoritmus véges determinisztikus automata minimalizálására

Tegyük fel, hogy az automata összefüggő. Eután partícionáljuk (bontsuk ekvivalencia osztályokra) a Q halmazt a következőképpen:

1.: Tekintsük a Q állapothalmazt. Osszuk két ekvivalenciaosztályra: F-re és $Q - F$-re. Az F-beli állapotok megkülönböztethetők a $Q - F$-beli állapotoktól. (Bármely elfogadó állapot megkülönböztethető egy nem elfogadó állapottól az üres szóval.) Ismételjük meg az ekvivalenciaosztályok további ekvivalenciaosztályokra való szétbontását addig, amíg a partícionálás változatlan marad a következőképpen:

2.: Ha adott egy partícionálás, $\mathcal{B} = B_1 \cup \ldots \cup B_s$ akkor minden egyes B_i ekvivalenciaosztály bármely két p és q állapota pontosan akkor marad együtt, ha minden $a \in T$-re a $\delta(p, a)$ és a $\delta(q, a)$ egyazon ekvivalenciaosztályba tartozik, egyébként a B_i ekvivalenciaosztályt szétbontjuk. Ezt az eljárást addig ismételjük, ameddig változás van, egyébként az eljárás leáll.

Algoritmus - folytatás

3.: Minden egyes B_i ekvivalenciaosztályt reprezentáljunk egy b_i szimbólummal. Konstruáljuk meg a determinisztikus véges automatát a következőképpen: legyen $A = (Q', T, \delta', q'_0, F')$, ahol

$$Q' = \{b_i \mid B_i \text{ a 2. eljárás során nyert ekvivalenciaosztály}\}.$$

$q'_0 = b_0$ a q_0-t tartalmazó ekvivalenciaosztály reprezentánsa. $\delta'(b_i, a) = b_j$, ha $\delta(q, a) = p$, $q \in B_i$, $p \in B_j$. $F' = \{b_f\}$ azon ekvivalenciaosztály reprezentánsa, amely F elemeit tartalmazza.

Formális nyelvek - 10.

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

Veremautomata

A veremautomata a véges automata általánosítása potenciálisan végtelen veremmel és véges kontrollal.

A verem esetében az új adat mindig a már meglevő veremtartalom tetejéhez adódik, kivétele fordított sorrendben történik.

A verem kezelési technikája ún. FILO (first-in-last-out).

Definíció

A veremautomata egy rendezett hetes

$$A = (Z, Q, T, \delta, z_0, q_0, F),$$

ahol

- Z a veremszimbólumok véges halmaza,
- Q az állapotok véges halmaza,
- T az inputszimbólumok véges halmaza,
- δ leképezése a $Z \times Q \times (T \cup \{\varepsilon\})$ halmaznak $Z^* \times Q$ véges részhalmazaiba, az ún. átmeneti függvény,
- $z_0 \in Z$ a kezdeti (kezdő) veremszimbólum,
- $q_0 \in Q$ a kezdeti (kezdő) állapot,
- $F \subseteq Q$ az elfogadó állapotok halmaza.

Definíció

Legyen $A = (Z, Q, T, \delta, z_0, q_0, F)$ veremautomata.

A veremautomata konfigurációja alatt egy uq alakú szót értünk, ahol $u \in Z^*$ a verem aktuális tartalma és $q \in Q$ az aktuális állapot.

A kezdeti konfiguráció $z_0 q_0$.

Megjegyzés: Ha $u = x_1 \ldots x_n$, ahol $x_i \in Z$, $1 \leq i \leq n$, akkor a verem tetején levő szimbólum x_n.

Közvetlen lépés a veremautomatában

Tegyük fel, hogy az A veremautomata olvasófeje az a inputszimbólumon áll, a veremautomata (a véges kontroll) q állapotban van, valamint a verem tetején levő szimbólum z.

Legyen $\delta(z, q, a) = \{(u_1, r_1), \ldots, (u_n, r_n)\}$, ahol $u_i \in Z^*$ és $r_i \in Q$, $1 \leq i \leq n$.

Akkor A következő állapota valamely r_i lesz és egyidejűleg z-t helyettesíti az u_i szóval, továbbá az olvasófej egy cellával jobbra lép az input szalagon.

Ha $\delta(z, q, \varepsilon)$ nem üres, akkor ún. ε-átmenet (ε-lépés) hajtható végre.

Szó elfogadása a veremautomatában

Ha az input szalag a $w \in T^*$ szót tartalmazza és a z_0q_0 kezdeti konfigurációból kiindulva és lépések sorozatát végrehajtva az A veremautomata egy up konfigurációba ér, ahol p elfogadó állapot, akkor azt mondjuk, hogy A elfogadta a w szót.

Definíció

Az A veremautomata az $\alpha \in Z^*QT^*$ szót a $\beta \in Z^*QT^*$ szóra redukálja egy lépésben, amelyet $\alpha \Longrightarrow_A \beta$-val jelölünk, ha létezik olyan $z \in Z$,$q, p \in Q$, $a \in T \cup \{\varepsilon\}$, $r, u \in Z^*$ és $w \in T^*$, hogy $(u, p) \in \delta(z, q, a)$ és $\alpha = rzqaw$ és $\beta = rupw$ teljesül.

Az A veremautomata az $\alpha \in Z^*QT^*$ szót a $\beta \in Z^*QT^*$ szóra redukálja, amelyet $\alpha \Longrightarrow_A^* \beta$-val jelölünk, ha vagy $\alpha = \beta$, vagy létezik olyan $\alpha_1, \ldots, \alpha_n \in Z^*QT^*$ szavakból álló véges sorozat, ahol $\alpha = \alpha_1$, $\beta = \alpha_n$, és $\alpha_i \Longrightarrow_A \alpha_{i+1}$, $1 \leq i \leq n - 1$.

Definíció

Az A veremautomata által (elfogadó állapottal) elfogadott nyelv

$$L(A) = \{w \in T^* \mid z_0 q_0 w \Longrightarrow_A^* up, \text{ ahol } u \in Z^*, p \in F\}.$$

Megjegyzés

A δ leképezést szabályok formájában is megadhatjuk. Az így nyert szabályhalmazt M_δ-val jelöljük. Tehát

1. $zqa \to up \in M_\delta$ hacsak $(u,p) \in \delta(z,q,a)$,

2. $zq \to up \in M_\delta$ hacsak $(u,p) \in \delta(z,q,\varepsilon)$

Példa

Legyen $L = \{a^n b^n \mid n \geq 2\}$. Az L nyelvet elfogadó veremautomata átmenetei a következők (csak azokat az átmeneteket tüntetjük fel, amelyek nemüresek):

$$
\begin{aligned}
(z_0 a, q_0,) &\in \delta(z_0, q_0, a), \\
(aa, q_0) &\in \delta(a, q_0, a), \\
(\varepsilon, q_1) &\in \delta(a, q_0, b), \\
(\varepsilon, q_1) &\in \delta(a, q_1, b), \\
(\varepsilon, q_2) &\in \delta(a, q_1, b), \\
(\varepsilon, q_2) &\in \delta(z_0, q_2, \varepsilon)
\end{aligned}
$$

Definíció

Az $A = (Z, Q, T, M, z_0, q_0, F)$ veremautomatát determinisztikusnak mondjuk, ha minden $(z, q) \in Z \times Q$ pár esetén

1. vagy $\delta(z, q, a)$ pontosan egy elemet tartalmaz minden $a \in T$ inputszimbólumra és $M(z, q, \varepsilon) = \emptyset$, vagy

2. $\delta(z, q, \varepsilon)$ pontosan egy elemet tartalmaz és $\delta(z, q, a) = \emptyset$ minden $a \in T$ inputszimbólumra.

Megjegyzés

A determinisztikus veremautomata elfogadó (felismerő) ereje kisebb, mint a (nemdeterminisztikus) veremautomatáé.

Példanyelvek:

$$L_1 = \{wcw^{-1} \mid w \in \{a,b\}^*\}$$

$$L_2 = \{ww^{-1} \mid w \in \{a,b\}^*\}$$

Míg az L_1 nyelvet el lehet fogadni determinisztikus veremautomatával, addig a másodikat nem, L_2-t csak nemdeterminisztikus veremautomatával lehet elfogadni.

Definíció

Az $N(A)$ nyelvet az A veremautomata üres veremmel fogadja el, ha

$$N(A) = \{w \in T^* \mid z_0 q_0 w \Longrightarrow_A^* p, \text{ ahol } p \in Q\}.$$

Vegyük észre, hogyha a verem üres, az automata működése blokkolódik, mivel nincs átmenet definiálva üres verem esetére. (Ezért van szükségünk a z_0 kezdeti veremszimbólumra.)

Lemma

Bármely A veremautomatához meg tudunk adni egy A' veremautomatát úgy, hogy $N(A') = L(A)$ teljesül.

Bizonyításvázlat

Legyen

$$A = (Z, Q, T, \delta, z_0, q_0, F)$$

veremautomata. Az

$$A' = (Z \cup \{z_0'\}, Q \cup \{q_0', q_h'\}, \delta', z_0', q_0', \emptyset)$$

veremautomatát a következőképpen definiáljuk: $z_0' \notin Z$, $q_0', q_h' \notin Q$, δ'-re pedig teljesülnek a következők:

$$\begin{aligned}
\delta'(z_0', q_0', \varepsilon) &= \{(z_0' z_0, q_0)\}, \\
\delta'(z, q, a) = \delta(z, q, a)&, z \in Z, q \in Q, a \in T, \\
\delta(z, q, \varepsilon) \subseteq \delta'(z, q, \varepsilon)&, z \in Z, q \in Q, \\
(\varepsilon, q_h') \in \delta'(z, q, \varepsilon)&, z \in Z \cup \{z_0'\}, q \in F \cup \{q_h'\}.
\end{aligned}$$

A' szimulálja A működését; a $z_0' \neq z_0$ veremszimbólum azért szükséges, hogy A' nem fogadhasson el olyan u szót, amely kiüríti az A automata vermét.

Bizonyításvázlat - folytatás

Legyen $w \in L(A)$. Akkor

$$z_0 q_0 w \Longrightarrow_A^* uq,$$

valamely $u \in Z^*$ veremszimbólum-sorozatra és $q \in F$ elfogadó állapotra. De akkor

$$z_0' q_0' w \Longrightarrow_{A'} z_0' z_0 q_0 w \Longrightarrow_{A'}^* z_0' uq \Longrightarrow_{A'}^* q_h'$$

is fennáll, ahonnan $w \in N(A')$ következik.

Bizonyításvázlat - folytatás

Legyen $w \in N(A')$. Akkor

$$z_0' q_0' w \Longrightarrow^*_{A'} q$$

valamely $q \in Q \cup \{q_0', q_h'\}$ állapotra.

A konstrukció alapján az első lépés $z_0' q_0' w \Longrightarrow_{A'} z_0' z_0 q_0$. Mivel a z_0' szimbólum csak valamely $(\varepsilon, q_h') \in \delta'(z_0', q, \varepsilon)$ átmenetet alkalmazó lépéssel törölhető a veremből, ezért lennie kell olyan $q \in F$ elfogadó állapotnak és $u \in Z^*$ szónak, amelyre

$$z_0' z_0 q_0 w \Longrightarrow^*_{A'} z_0' u q \Longrightarrow^*_{A'} q_h'$$

teljesül, ahol $z_0 q_0 w \Longrightarrow^*_A uq$. Azaz, $w \in L(A)$.

Lemma

Bármely A veremautomatához meg tudunk adni egy A' veremautomatát úgy, hogy $L(A') = N(A)$ teljesül.

Bizonyításvázlat

Legyen

$$A = (Z, Q, T, \delta, z_0, q_0, \emptyset)$$

veremautomata, amely üres veremmel az $N(A)$ nyelvet fogadja el. Megkonstruáljuk az A' veremautomatát, amely elfogadó állapottal az $L(A') = N(A)$ nyelvet fogadja el. Legyen

$$A' = (Z \cup \{z_0'\}, Q \cup \{q_0', q_f'\}, T, \delta', z_0', q_0', \{q_f'\}),$$

ahol $z_0' \notin Z$, $q_0', q_f' \notin Q$ és

$$\begin{aligned} \delta'(z_0', q_0', \varepsilon) &= \{(z_0' z_0, q_0)\}, \\ \delta'(z, q, a) &= \delta(z, q, a), z \in Z, q \in Q, a \in (T \cup \{\varepsilon\}), \\ \delta'(z_0', q, \varepsilon) &= \{(z_0', q_f')\}, q \in Q. \end{aligned}$$

Könnyen látható, hogy valahányszor A kiüríti a vermét, akkor A' elfogadó állapotba kerül, továbbá A' csak ebben az esetben kerül elfogadó állapotba. Így a $L(A') = N(A)$ teljesül.

Formális nyelvek - 11.

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

Tétel

Bármely G környezetfüggetlen grammatikához meg tudunk adni egy olyan A veremautomatát, amelyre $L(A) = L(G)$ teljesül.

Bizonyításvázlat

Legyen $G = (N, T, P, S)$ Chomsky normálformájú grammatika. Ha $\varepsilon \in L(G)$, akkor P tartalmazza az $S \to \varepsilon$ szabályt, de ebben az esetben S ne forduljon elő egyetlen szabály jobboldalán sem. Megkonstruálunk egy A veremautomatát, amelyre $L(G) = L(A)$ teljesül.

Bizonyításvázlat

Legyen $A = (N \cup \{z_0\}, Q, T, \delta, z_0, q_0, \{q_h\})$, ahol

$$Q = (\bigcup_{X \in N} q_X) \cup \{q_0, q_h\}$$

és a δ függvényt a következőképpen definiáljuk:

$z_0 q_0 \to z_0 q_S \in M_\delta$ akkor és csak akkor, ha $S \to \varepsilon \in P$,

$z_0 q_0 a \to z_0 q_X \in M_\delta$ akkor és csak akkor, ha $X \to a \in P$,

$Z q_Y a \to ZY q_X \in M_\delta$ minden $Z \in N \cup \{z_0\}, Y \in N, X \to a \in P$ esetén,

$Z q_Y \to q_X \in M_\delta$ akkor és csak akkor, ha $X \to ZY \in P$,

$z_0 q_S \to q_h \in M_\delta$.

Megjegyezzük, hogy ezek az átmenetek lényegében a G invertált szabályai.

Bizonyításvázlat - folytatás

Legyen $w \in L(A)$. Akkor $z_0 q_0 w \Longrightarrow_A^* u q_h$ valamely u veremszimbólumokból álló szóra. Az A konstrukciója alapján $u = \varepsilon$ kell, hogy teljesüljön, illetve, pontosabban

$$z_0 q_0 w \Longrightarrow_A^* z_0 q_S \Longrightarrow_A q_h,$$

kell, hogy fennálljon.

Ha $w = \varepsilon$, akkor az első lépés $z_0 q_0 \Longrightarrow_A z_0 q_S$ és így $S \to \varepsilon \in P$. Egyébként

$$z_0 q_0 w \Longrightarrow_A z_0 q_X v \Longrightarrow_A^* z_0 q_S \Longrightarrow_A q_h,$$

ahol $w = av$ valamely $a \in T$ inputszimbólumra, $v \in T^*$ és $X \to a \in P$. Könnyen megmutathatjuk, hogy a

$z_0 q_0 w \Longrightarrow_A^* z_0 q_S$ redukcióból az $S \Longrightarrow_G^* w$ levezetés következik.

Nevezetesen, minden 2,3 vagy 4 típusú átmenetnek van a P szabályrendszerben megfelelője, így meg tudjuk konstruálni a megfelelő levezetést G-ben. Azaz, $L(A) \subseteq L(G)$. A fordított irányú tartalmazást, $L(G) \subseteq L(A)$-t, hasonlóan könnyen igazolni tudjuk.

Tétel

Minden A veremautomatához meg tudunk adni egy környezetfüggetlen G grammatikát úgy, hogy $L(G) = N(A)$ teljesül.

Bizonyításvázlat:

Legyen $A = (Z, Q, T, \delta, z_0, q_0, \emptyset)$ veremautomata.

Definiáljuk a $G = (N, T, P, S)$ grammatikát úgy, hogy N elemei (q, x, p) alakú rendezett hármasok, ahol $q, p \in Q$ és $x \in Z$. Bevezetjük az S új szimbólumot és legyen $N = Q \times Z \times Q \cup \{S\}$.

Bizonyításvázlat - folytatás:

Definiáljuk a P szabályhalmazt a következőképpen:

1. Legyen $S \to (q_0, z_0, p) \in P$ minden $p \in Q$ állapotra.

2. Ha $xqa \to y_1 \dots y_m p_m \in M_\delta$, ahol $a \in (T \cup \{\varepsilon\})$, akkor minden $p_0, p_1, \dots, p_{m-1} \in Q$ állapotsorozatra legyen
$$(q, x, p_0) \to a(p_m, y_m, p_{m-1}) \dots (p_1, y_1, p_0)$$
szabály P-ben. Az $m = 0$, azaz, az $xqa \to p \in M_\delta$ esetben legyen $(q, x, p_0) \to a \in P$.

3. A P szabályhalmaz ne tartalmazzon további szabályt.

Bizonyításvázlat - folytatás:

Először az $L(G) \subseteq N(A)$ tartalmazást igazoljuk. Bizonyításához megmutatjuk, hogy minden x veremszimbólumra, q, p állapotpárra, valamint u inputszóra fennáll, hogyha $(q, x, p) \Longrightarrow^* u$ levezetés G-ben, akkor $xqu \Longrightarrow^* p$ redukció A-ban.

A bizonyítást a $(q, x, p) \Longrightarrow_G^* u$ levezetés lépéseinek száma szerinti indukcióval végezzük.

Egy lépés esetében nyilvánvalóan $u = a \in T \cup \{\varepsilon\}$ és $xqa \to p \in M_\delta$.

Bizonyításvázlat - folytatás:

Tegyük fel, hogy az állítás igaz minden legfeljebb n lépésből álló levezetésre és álljon a $(q, x, p) \Longrightarrow_G^* u$ levezetés $n+1$ lépésből.

Akkor a levezetés alakja

$$(q, x, p_0) \Longrightarrow a(p_m, y_m, p_{m-1}) \dots (p_1, y_1, p_0) \Longrightarrow_G^* u,$$

ahonnan az adódik, hogy léteznek olyan $u_m, u_{m-1}, \dots, u_1 \in T^*$ szavak, amelyekre $u = au_m u_{m-1} \dots u_1$ és $(p_i, y_i, p_{i-1}) \Longrightarrow_G^* u_i$ $1 \leq i \leq m$.

Az indukciós hipotézis alapján fennáll az

$$y_i p_i u_i \Longrightarrow_A^* p_{i-1}$$

redukció és G definíciója alapján

$$xqa \Longrightarrow_A y_1 \dots y_m p_m.$$

Így

$xqu = xqau_m \dots u_1 \Longrightarrow_A y_1 \dots y_m p_m u_m \dots u_1 \Longrightarrow_A^* y_1 \dots y_{m-1} p_{m-1} u_{m-1} \dots u_1 \Longrightarrow_A^* p_0 = p$.

Bizonyításvázlat - folytatás:

Így, ha $u \in L(G)$, akkor van olyan $p \in Q$, amelyre

$$S \Longrightarrow_G (q_0, z_0, p) \Longrightarrow_G^* u,$$

ahonnan

$$z_q q_0 u \Longrightarrow_A^* p,$$

azaz, $u \in N(A)$ adódik.

Bizonyításvázlat - folytatás:

A fordított irányú tartalmazás bizonyításához megmutatjuk, hogy a

$$xqu \Longrightarrow_A^* p$$

redukcióból a

$$(q, x, p) \Longrightarrow_G^* u$$

levezetés adódik.

A bizonyítást az automata lépéseinek (átmenetek) száma szerinti indukcióval végezzük.

Egyetlen lépés esetén $xqa \to p \in M_\delta$, ahonnan $(q, x, p) \to a \in P$.

Bizonyításvázlat - folytatás:

Egynél több lépés esetén az $xqu \Longrightarrow_A^* p$ redukció

$$xqu = xqav \Longrightarrow_A y_1 \ldots y_m p_m v \Longrightarrow_A^* p$$

alakú kell, hogy legyen.

Vizsgáljuk meg ezt a redukciót azután a lépés utánig, amikor az y_{m-1} szimbólum a verem tetején levő szimbólum.

Ekkor van olyan $u_m \in T^*$, amelyre $v = u_m v_1$ és $y_m p_m u_m \Longrightarrow_A^* p_{m-1}$ valamely p_{m-1} állapotra és így

$$xqu \Longrightarrow_A y_1 \ldots y_m p_m u_m v_1 \Longrightarrow_A^* y_1 \ldots y_{m-1} p_{m-1} v_1$$

.

Bizonyításvázlat - folytatás:

Folytatva a gondolatmenetet, azt kapjuk, hogy vannak olyan $u_{m-1}, \ldots, u_1$ szavak, hogy

$$u = au_m u_{m-1} \ldots u_1$$

és

$$y_i p_i u_i \Longrightarrow_A^* p_{i-1}, \text{ ha } i = 1, \ldots m.$$

Az indukciós hipotézis alapján ekkor

$$(p_i, y_i, p_{i-1}) \Longrightarrow_G^* u_i \text{ ha } i = 1, \ldots m$$

és a G grammatika definíciója alapján

$$(q, x, p) \Longrightarrow_G a(p_m, y_m, p_{m-1}) \ldots (p_1, y_1, p)$$

teljesül, ahonnan $(q, x, p) \Longrightarrow_G^* u$ következik.

Bizonyításvázlat - folytatás:

Így, ha $u \in N(A)$, akkor $z_0 q_0 u \Longrightarrow_A^* p$ valamely p állapotra, és így

$$S \Longrightarrow_G (q_0, z_0, p) \Longrightarrow_G^* u,$$

azaz, $N(A) \subseteq L(G)$.

Korollárium

Minden A veremautomatához meg tudunk adni egy A' veremautomatát úgy, hogy $L(A') = N(A)$.

Korollárium

Minden A veremautomatához meg tudunk adni egy G környezetfüggetlen grammatikát úgy, hogy $L(A) = L(G)$.