

A számításelmélet alapjai I.

1. előadás

előadó: Tichler Krisztián
ktichler@inf.elte.hu

Irodalom

- Révész György, Bevezetés a formális nyelvek elméletébe, Tankönyvkiadó, 1977.
- Bach Iván: Formális nyelvek, Typotex, 2001.
- Fülöp Zoltán, Formális nyelvek és szintaktikus elemzésük, Polygon, Szeged, 2004.
- Hunyadvári László, Manhertz Tamás, Automaták és formális nyelvek, Elektronikus előadásjegyzet, 2006.
- Csima Judit, Friedl Katalin: Nyelvek és automaták, BMGE jegyzet, 2013.
- A. Salomaa, Formal Languages, Academic Press, 1973.
- J. E. Hopcroft, Rajeev Motwani, J.D. Ullman, Introduction to Automata Theory, Languages, and Computation. Second Edition. Addison-Wesley, 2001.
- M. Sipser, Introduction to the Theory of Computation, 2nd ed., Thomson, 2006.

Miről fogunk tanulni?

- Alapfogalmak és jelölések, szavak, nyelvek, grammatikák, a grammatikák Chomsky-féle hierarchiája.
- Műveletek nyelveken: definíciók, Chomsky-féle nyelvosztályok bizonyos zártsági tulajdonságai.
- Reguláris grammatikák, reguláris nyelvek, reguláris kifejezések. Tulajdonságaik.
- Környezetfüggetlen grammatikák és nyelvek: redukált grammatikák, normálformák, levezetési fa.
- Környezetfüggő- és mondszerkezetű grammatikák: hossznemcsökkentő grammatikák, normálformák. A generált nyelvek és nyelvosztályok tulajdonságai.
- Automaták és nyelvek: véges automaták, veremautomaták. Az automaták tulajdonságai, a felismert nyelvosztályok, az automaták és a grammatikák kapcsolatai.
- Szintaktikai elemzés: kapcsolat szintaxis és szemantika között; környezetfüggetlen grammatikák és nyelvek egyértelműsége; $LL(k)$ és $LR(k)$ grammatikák.

Hol használják a formális nyelveket?

- A természetes nyelvek gépi feldolgozása, matematikai modellezése, matematikai nyelvészet,
- programozási nyelvek, fordítóprogramok elmélete,
- kódelmélet,
- képfeldolgozás,
- mintafelismerés,
- biológiai folyamatok modellezése (pl. Lindenmayer rendszerek),
- természet által inspirált új számítási modellek

Alapfogalmak, jelölések

ábécé, betű, szó, szó hossza

Ábécé: Egy véges, nemüres halmaz.
Az ábécé elemeit **betűk**nek nevezzük.

Egy V ábécé elemeiből képzett véges sorozatokat **V feletti szavaknak** vagy sztringeknek nevezzük. Egy $u = t_1 \cdot \dots \cdot t_n$ szóban lévő betűk számát (n) a szó **hosszának** nevezzük. Jelölés: $|u| = n$. A 0 hosszú sorozat jelölése ε , ezt **üres szónak** nevezzük ($|\varepsilon| = 0$).

Példa:

Legyen $V = \{a, b\}$, ekkor a és b a V ábécé két betűje. $abba$ és $aabba$ egy-egy V feletti szó. $|aabba| = 5$. $aabba$ és $baaba$ különböző szavak, bár mindkettő 5 hosszú valamint 3 a -t és 2 b -t tartalmaz. A betűk sorrendje számít!

Alapfogalmak, jelölések

az összes szavak halmaza

V^* jelöli a V **ábécé feletti szavak halmazát**, beleértve az üres szót is.

$V^+ = V^* \setminus \{\varepsilon\}$ a V **ábécé feletti, nemüres szavak halmazát** jelöli.

Példa: $V = \{a, b\}$, ekkor

$V^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, aab, \dots\}$.


V^* szavainak egy lehetséges felsorolása a következő, melyet **hosszlexikografikus (shortlex) rendezésnek** nevezünk.

- Feltesszük, hogy az ábécé rendezett. (A fenti példában, mondjuk a előbb van, mint b .)
- A rövidebb szó mindig megelőzi a hosszabbat.
- Az azonos hosszúságú szavak ábécésorrendben (lexikografikusan) követik egymást (balról jobbra az első eltérés alapján rakjuk sorba a két szót).

Műveletek szavakon

konkatenáció

Legyen V egy ábécé továbbá legyenek $u = s_1 \cdots s_n$ és $v = t_1 \cdots t_k$ V feletti szavak (azaz legyen $s_1, \dots, s_n, t_1, \dots, t_k \in V$). Ekkor az

 $uv := s_1 \cdots s_n t_1 \cdots t_k$ szót az u és v szavak **konkatenáltjának** nevezzük. (Az u szó betűi után írjuk a v szó betűit.)

Nyilván $|uv| = |u| + |v|$.

Példa:

Legyen $V = \{a, b\}$, valamint $u = abb$ és $v = aaaba$ egy-egy V feletti szó. Ekkor $uv = abbaaaba$ illetve $vu = aaabaabb$.

Műveletek szavakon

konkatenáció

A konkatenáció mint művelet asszociatív, de **általában nem kommutatív.**

- Ha $u, v \in V^*$, akkor uv általában eltér vu -tól
(Megjegyzés: unáris (egyetlen betűből álló) ábécé esetén persze fennáll a kommutativitás, de ez nagyon speciális eset.)
- Ha $u, v, w \in V^*$, akkor $u(vw) = (uv)w$ (asszociativitás).

V^* a konkatenáció műveletére nézve zárt (azaz bármely $u, v \in V^*$ esetén $uv \in V^*$ teljesül).

A konkatenáció egységelemes művelet, az egységelem ε (azaz bármely $u \in V^*$ esetén $u = u\varepsilon = \varepsilon u$).

Tehát adott ábécé feletti szavak a konkatenáció műveletére nézve egységelemes félcsoportot alkotnak, melynek ε az egységeleme.

Műveletek szavakon

i -edik hatvány

Legyen i nemnegatív egész szám és legyen u egy V ábécé feletti szó ($u \in V^*$). Az u szó i -edik hatványa alatt az u szó i darab példányának konkatenáltját értjük és u^i -vel jelöljük.

Formálisan $u^0 := \varepsilon$, $u^i := uu^{i-1}$ ($i \geq 1$).

Ekkor $u^{n+k} = u^n u^k$ teljesül ($k, n \in \mathbb{N}$)

Példa:

Legyen $V = \{a, b\}$ és legyen $u = abb$. Ekkor $u^0 = \varepsilon$, $u^1 = abb$, $u^2 = abbabb$ és $u^3 = abbabbabb$.

$(ab)^3 \neq a^3b^3$! ! ! !

$(ab)^3 = ababab$, $a^3b^3 = aaabbb$.

Műveletek szavakon

tükörkép

Legyen u egy V ábécé feletti szó. Az u szó u^{-1} -gyel jelölt **tükörképén** vagy fordítottján azt a szót értjük, amelyet úgy kapunk, hogy u betűit fordított sorrendben írjuk le.

Azaz ha $u = a_1 \cdots a_n$, $a_i \in V$, $1 \leq i \leq n$, akkor $u^{-1} = a_n \cdots a_1$.

Példa:

Legyen $V = \{a, b\}$, valamint $u = abba$ és $v = baabba$ egy-egy V ábécé feletti szó. Ekkor $u^{-1} = abba$ és $v^{-1} = abbaab$.

Ha $u = u^{-1}$, akkor u -t **palindrom** tulajdonságúnak, vagy palindrómának nevezzük.

Tehát $abba$ egy palindróma.

Az alábbi tulajdonságok könnyen meggondolhatóak:

- $\varepsilon^{-1} = \varepsilon$
- $(u^{-1})^{-1} = u$
- $(uv)^{-1} = v^{-1}u^{-1}$
- $(u^i)^{-1} = (u^{-1})^i$, minden $i \in \mathbb{N}$ -re.

Alapfogalmak, jelölések

részszó, prefix, suffix

Legyenek u és v V ábécé feletti szavak. u és v szót azonosnak mondjuk, ha mint szimbólumsorozatok elemről-elemre rendre megegyeznek. (Tehát $abb \neq bab$!!!)

Legyen V egy ábécé és legyenek u és v szavak V felett. Az u szó a v szó **részszava**, ha $v = xuy$ teljesül, valamely $x, y \in V^*$ szavakra. Az u szó a v szó **valódi részszava**, ha $u \neq v$ és $u \neq \varepsilon$.

Ha $x = \varepsilon$, akkor u -t a v szó **prefixének** (kezdőszeletének), ha pedig $y = \varepsilon$, akkor u -t a v szó **suffixének** (utótagjának, végszeletének) nevezzük.

Egy u szó **valódi prefixei/suffixei** az ε -tól és u -tól különböző prefixei/suffixei.

Alapfogalmak, jelölések

részszó, prefix, suffix

Példa:

Legyen $V = \{a, b\}$ és $u = abba$.

u részszavai: $\varepsilon, a, b, ab, bb, ba, abb, bba, abba$.

u valódi részszavai: $a, b, ab, bb, ba, abb, bba$.

u prefixei: $\varepsilon, a, ab, abb, abba$.

u suffixei: $\varepsilon, a, ba, bba, abba$.

u valódi prefixei: a, ab, abb .

u valódi suffixei: a, ba, bba .

abb nem suffixe u -nak!!! A suffix nem a tükörkép szó prefixe!!!

Alapfogalmak, jelölések

nyelv

Legyen V egy ábécé, V^* egy L részalmazát V feletti **nyelvnek** nevezzük.

Az üres nyelv (nyelv, amely egyetlen szót sem tartalmaz) jelölése \emptyset .

Egy V ábécé feletti nyelv véges nyelv, ha véges számú szót tartalmaz, ellenkező esetben végtelen.

Példák:

Legyen $V = \{a, b\}$ egy ábécé.

- ▶ $L_1 = \{a, ba, ababb, \varepsilon\}$.
- ▶ $L_2 = \{a^i b^i \mid i \geq 0\}$.
- ▶ $L_3 = \{uu^{-1} \mid u \in V^*\}$.
- ▶ $L_4 = \{a^n \mid n \geq 1\}$.
- ▶ $L_5 = \{u \mid u \in \{a, b\}^+, u\text{-ban ugyanannyi } a \text{ van, mint } b\}$.

L_1 véges nyelv, a többi végtelen.

Nyelvekre vonatkozó műveletek

halmazműveletek

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

$L_1 \cup L_2 = \{u \mid u \in L_1 \text{ vagy } u \in L_2\}$ (az L_1 és az L_2 nyelv **uniója**)

$L_1 \cap L_2 = \{u \mid u \in L_1 \text{ és } u \in L_2\}$ (az L_1 és L_2 nyelv **metszete**)

$L_1 - L_2 = \{u \mid u \in L_1 \text{ és } u \notin L_2\}$ (az L_1 és az L_2 nyelv **különbsége**)

Példák:

$L_1 = \{\varepsilon, ab, abb\}, L_2 = \{ab, bab\}$. Ekkor

$L_1 \cup L_2 = \{\varepsilon, ab, abb, bab\},$

$L_1 \cap L_2 = \{ab\},$

$L_1 - L_2 = \{\varepsilon, abb\}.$

Az $L \subseteq V^*$ nyelv **komplementere** a V ábécére nézve az $\bar{L} = V^* - L$ nyelv.

Példa:

$V = \{a\}, L = \{a^{2n+1} \mid n \in \mathbb{N}\}$, ekkor $\bar{L} = \{a^{2n} \mid n \in \mathbb{N}\}$

Nyelvekre vonatkozó műveletek

konkatenáció


Legyen V egy ábécé és $L_1, L_2 \subseteq V^*$. Az L_1 és az L_2 nyelvek **konkatenációján** az $L_1 L_2 = \{u_1 u_2 \mid u_1 \in L_1, u_2 \in L_2\}$ nyelvet értjük. Asszociatív, de nyilván **nem kommutatív** művelet.

Példák:

$$V = \{a, b\}, L_1 = \{ab, bb\}, L_2 = \{\varepsilon, a, bba\},$$

$$L_1 L_2 = \{ab\varepsilon, bb\varepsilon, aba, bba, abbba, bbbba\} \\ = \{ab, bb, aba, bba, abbba, bbbba\}. \quad (\text{mindenkit mindenkivel!})$$

$$L_3 = \{a, ab\}, L_4 = \{bc, c\}$$

 $L_3 L_4 = \{abc, ac, abbc\}$, $|L_3 L_4| < 2 \cdot 2 = 4$, mivel abc kétféleképpen is előállítható!!!


$$L_5 = \{a^{2n} b^{2n} \mid n \in \mathbb{N}\}, L_6 = \{a^{3n} b^{3n} \mid n \in \mathbb{N}\}$$

$$L_5 L_6 = \{a^{2n} b^{2n} a^{3k} b^{3k} \mid n, k \in \mathbb{N}\}.$$

Az $L_5 L_6$ konkatenációnak pl. az is eleme, ha L_5 46. elemét konkatenáljuk L_6 87. elemével!!! ($n = 46, k = 87$)

Nyelvekre vonatkozó műveletek

i -edik hatvány, Kleene lezárt

Észrevétel: $\{\varepsilon\}L = L\{\varepsilon\} = L$ minden L nyelvre. 

$$L_1(L_2L_3) = (L_1L_2)L_3 \quad (\text{asszociativitás})$$

Adott ábécé feletti nyelvek a konkatenáció műveletére nézve egységelemes félcsoportot alkotnak, melynek $\{\varepsilon\}$ az egységeleme.

Egy L nyelv **i -edik hatványát** $L^0 := \{\varepsilon\}$ és $L^i := LL^{i-1}$ ($i \geq 1$) definiálják.

Azaz L^i jelöli az L i -edik iterációját a konkatenáció műveletére nézve. Nyilván $L^{k+n} = L^k L^n$ ($k, n \in \mathbb{N}$).

Példa: $L = \{a, bb\}$.

$$L^0 = \{\varepsilon\},$$

$$L^1 = \{a, bb\},$$

$$L^2 = \{aa, abb, bba, bb bb\},$$

$$L^3 = \{aaa, aabb, abb a, abb bb, bbaa, bba bb, bb bb a, bb bb bb\},$$

Nyelvekre vonatkozó műveletek

Kleene lezárt

Az L nyelv **iteratív lezártja** (vagy röviden lezártja vagy Kleene-lezártja) alatt az $L^* = \bigcup_{i \geq 0} L^i$ nyelvet értjük.

Az L nyelv **pozitív lezártja** alatt az $L^+ = \bigcup_{i \geq 1} L^i$ nyelvet értjük.

Észrevétel:

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

Példa: $L = \{a, bb\}$.

$$L^0 = \{\varepsilon\},$$

$$L^1 = \{a, bb\},$$

$$L^2 = \{aa, abb, bba, bb bb\},$$

$$L^3 = \{aaa, aabb, abba, abbb, bbaa, bba bb, bbba, bbbb\},$$

$$L^4 = \{aaaa, aaaa bb, aaabba, \dots\}, \dots$$

Ezen szavak uniója együttesen alkotja L^* -t. L^+ ettől csak annyiban tér el, hogy nincs benne az ε szó.

Nyelvekre vonatkozó műveletek

tükörkép



Legyen V egy ábécé és legyen $L \subseteq V^*$. Ekkor $L^{-1} = \{u^{-1} \mid u \in L\}$ a **tükörképe** (megfordítása) az L nyelvnek.

Példa:

$V = \{a, b\}$, $L = \{ba^{2n+1} \mid n \in \mathbb{N}\}$, ekkor $L^{-1} = \{a^{2n+1}b \mid n \in \mathbb{N}\}$.

Néhány nyelvműveletek közötti összefüggés
(meggondolható/gyakorlaton igazolható):

- ▶ $(L^{-1})^{-1} = L$
- ▶ $(L_1 L_2 \cdots L_n)^{-1} = L_n^{-1} \cdots L_2^{-1} L_1^{-1}$
- ▶ $(L^i)^{-1} = (L^{-1})^i$, ahol $i \geq 0$
- ▶ $(L^*)^{-1} = (L^{-1})^*$,
- ▶ $L^* L^* = L^*$,
- ▶ $(L^*)^* = L^*$,
- ▶ $(L_1 \cup L_2)^* = (L_1^* L_2^*)^*$,
- ▶ $L_1(L_2 \cup L_3) = L_1 L_2 \cup L_1 L_3$, de $L_1(L_2 \cap L_3) \neq L_1 L_2 \cap L_1 L_3$.

Nyelvekre vonatkozó műveletek

prefixnyelv, suffixnyelv

Egy $L \subseteq V^*$ nyelv **prefixnyelvén** a



$\text{PRE}(L) = \{u \mid u \in V^*, uv \in L \text{ valamely } v \in V^* \text{-ra}\}$ nyelvet értjük.

Egy $L \subseteq V^*$ nyelv **suffixnyelvén** a

$\text{SUF}(L) = \{u \mid u \in V^*, vu \in L \text{ valamely } v \in V^* \text{-ra}\}$ nyelvet értjük.

Észrevétel: $L \subseteq \text{PRE}(L)$, $L \subseteq \text{SUF}(L)$.

Példa:


$$L = \{a^{2n+1}b \mid n \in \mathbb{N}\}.$$

$$\text{PRE}(L) = \{a^{2n+1}b \mid n \in \mathbb{N}\} \cup \{a^n \mid n \in \mathbb{N}\}.$$

$$\text{SUF}(L) = \{a^n b \mid n \in \mathbb{N}\} \cup \{\varepsilon\}.$$

Homomorfizmus

Legyen V_1 és V_2 két ábécé. A $h : V_1^* \rightarrow V_2^*$ leképezést **homomorfizmusnak** nevezzük, ha teljesülnek a következő feltételek:

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

Észrevételek:

- ▶ A fenti tulajdonságokból következik, hogy $h(\varepsilon) = \varepsilon$. Világos, hiszen $h(\varepsilon) = h(\varepsilon\varepsilon) = h(\varepsilon)h(\varepsilon)$, így csak $h(\varepsilon) = \varepsilon$ lehet.
- ▶ Minden $u = a_1 a_2 \cdots a_n$ szóra, ahol $a_i \in V_1$, $1 \leq i \leq n$, fennáll, hogy $h(u) = h(a_1)h(a_2) \cdots h(a_n)$. Ez azt jelenti, hogy elegendő a h leképezést V_1 elemein megadni, és ez automatikusan kiterjesztődik V_1^* -ra.

Nyelvekre vonatkozó műveletek

h-homomorf kép

Legyen $h : V_1^* \rightarrow V_2^*$ homomorfizmus. A h homomorfizmus **ε -mentes**, ha $h(u) \neq \varepsilon$ bármely $u \in V_1^+$ szóra.

Legyen $h : V_1^* \rightarrow V_2^*$ 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.

Példa:

$$V_1 = \{a, b\}, V_2 = \{b, c\}, L = \{a^n b a^n \mid n \in \mathbb{N}\}$$

$$h(a) = cc, h(b) = cbb.$$

$$\text{Ekkor } h(L) = \{c^{2n+1} b^2 c^{2n} \mid n \in \mathbb{N}\}.$$

Formális nyelvek megadása

- Véges nyelvek esetén felsorolással:

$$L_1 = \{ab, ba, abba, ca\}$$

- Formulával:

$$L_2 = \{a^n b^n \mid n \in \mathbb{N}\}$$

- Reguláris kifejezéssel:

$$L_3 = \{a^{2n+1}b \mid n \in \mathbb{N}\} \cup \{a^n \mid n \in \mathbb{N}\},$$

ez reguláris kifejezéssel így fog majd kinézni: $(aa)^*ab \mid a^*$.

- Végtelen nyelvek esetén felsoroló algoritmussal:

$L_4 = \{0, 1\}^*$, ez egy végtelen nyelv, a szavait felsoroló algoritmus pl. a 0 és 1 betűkből álló szavak hosszlexikografikus felsorolása.

Formális nyelvek megadása

- (Generatív) grammatikákkal

A grammatikák **szintetizáló** eszközök, egyetlen szimbólumból egy szabályrendszer segítségével szavakat lehet felépíteni. Azon szavak halmaza, melyeket fel lehet építeni egy nyelvet határoz meg, tehát a grammatika szabályrendszere meghatároz egy nyelvet.

- Matematikai gépek, automaták segítségével

Az automaták elemző, **analitikus** eszközök. Az automaták bemenete egy szó, kimenete egy bináris érték. Az automata szabályrendszere szerint feldolgozza a szavakat, csak bizonyos szavak esetén ad igen választ, ezen szavak egy nyelvet alkotnak, melyet az automata szabályrendszere határoz meg.

Figyelem! Nem feltétlen lehet minden eszközzel minden nyelvet megadni. Be fogjuk látni például, hogy reguláris kifejezéssel kevesebb nyelvet lehet leírni mint egy általános grammatikával, de az se elég az összes $\{0, 1\}$ ábécé feletti nyelv leírásához.

Nyelvcsaládok

Ha X egy halmaz, jelölje $\mathcal{P}(X)$ az X halmaz hatványhalmazát, azaz $\mathcal{P}(X) = \{A \mid A \subseteq X\}$.

Nyelvcsalád (vagy nyelvosztály) alatt nyelveknek egy halmazát értjük.

Tehát ha V egy ábécé:

- $a \in V$: betű
- $u \in V^*$: szó
- $L \subseteq V^*$ vagy $L \in \mathcal{P}(V^*)$: nyelv
- $\mathcal{L} \subseteq \mathcal{P}(V^*)$ vagy $\mathcal{L} \in \mathcal{P}(\mathcal{P}(V^*))$: nyelvcsalád

Példa: $V = \{a, b\}$

- véges nyelvek nyelvcsaládja
- csak b betűvel kezdődő szavakat tartalmazó nyelvek nyelvcsaládja
- reguláris kifejezésekkel leírható nyelvek nyelvcsaládja

Grammatikák

Definíció

Egy $G = \langle N, T, P, S \rangle$ rendezett négyest **grammatikának** (generatív grammatikának vagy nyelvtannak) nevezünk ha

- ▶ N és T diszjunkt véges ábécék (azaz $N \cap T = \emptyset$). N elemeit **nemterminális**, T elemeit pedig **terminális** szimbólumoknak nevezzük.
- ▶ $S \in N$ a grammatika **kezdőszimbóluma**.
- ▶ P rendezett (x, y) párok véges halmaza, ahol $x, y \in (N \cup T)^*$ és x legalább egy nemterminális szimbólumot tartalmaz.

A P halmaz elemeit **szabályoknak** (vagy átírási szabályoknak vagy produkcióknak) hívjuk. A gyakorlatban az (x, y) jelölés helyett szinte mindig az $x \rightarrow y$ jelölést használjuk amennyiben \rightarrow nem eleme $N \cup T$ -nek.

Grammatikák

Példák:

$G_1 = \langle \{S, A, B\}, \{a, b, c\}, \{S \rightarrow c, S \rightarrow AB, A \rightarrow aA, B \rightarrow \varepsilon, abb \rightarrow aSb\}, S \rangle$ **nem** grammatika, mivel minden szabály baloldalának tartalmaznia kell legalább egy N -beli szimbólumot.

$G_2 = \langle \{S, A, B, C\}, \{a, b, c\}, \{S \rightarrow a, S \rightarrow AB, A \rightarrow Ab, B \rightarrow \varepsilon, aCA \rightarrow aSc\}, S \rangle$ grammatika.

Generatív grammatikák

egylépéses levezetés

Definíció

Legyen $G = \langle N, T, P, S \rangle$ egy generatív grammatika és legyen $u, v \in (N \cup T)^*$. A v szó közvetlenül vagy **egy lépésben levezethető** az u szóból G -ben, jelölése $u \Rightarrow_G v$, ha $u = u_1 x u_2$ és $v = u_1 y u_2$, ahol $u_1, u_2 \in (N \cup T)^*$ és $x \rightarrow y \in P$.

Példa:

$G_2 = \langle \{S, A, B, C\}, \{a, b, c\}, \{S \rightarrow a, S \rightarrow AB, A \rightarrow Ab, B \rightarrow \varepsilon, aCA \rightarrow aSc\}, S \rangle$.

Ekkor $BB\mathbf{aCA}a \Rightarrow BB\mathbf{aSc}a$, $\mathbf{ABB} \Rightarrow \mathbf{Ab}BB$, $B\mathbf{B} \Rightarrow B$.

Grammatikák

többlépéses levezetés

Definíció

Legyen $G = \langle N, T, P, S \rangle$ egy grammatika és legyen $u, v \in (N \cup T)^*$. u -ból (több lépésben) **levezethető** v , ha $u = v$ vagy van olyan $n \geq 1$ és $w_0, \dots, w_n \in (N \cup T)^*$, hogy $w_{i-1} \Rightarrow_G w_i$ ($1 \leq i \leq n$) és $w_0 = u$ és $w_n = v$. Jelölés: $u \Rightarrow_G^* v$.

A kezdőszimbólumból levezethető szavakat **mondatformának** nevezzük.

Grammatikák

többlépéses levezetés

Példa:

$$G_2 = \langle \{S, A, B, C\}, \{a, b, c\}, \{S \rightarrow a, S \rightarrow AB, A \rightarrow Ab, B \rightarrow \varepsilon, aCA \rightarrow aSc\}, S \rangle.$$

Ekkor $BBaCAa \Rightarrow BBaSc a$ és $BBaSc a \Rightarrow BBaABca$, tehát $BBaCAa \Rightarrow^* BBaABca$.

$S \Rightarrow AB \Rightarrow AbB \Rightarrow AbbB$, tehát $AbbB$ egy mondatforma.

Megjegyzés: \Rightarrow_G bináris reláció az $(N \cup T)^*$ alaphalmazon. $A \Rightarrow_G^*$ reláció nem más, mint a \Rightarrow_G reláció reflexív, tranzitív lezártja. Ha egyértelmű melyik grammatikáról van szó, akkor \Rightarrow_G helyett röviden \Rightarrow -t írhatunk.

Reláció reflexív, tranzitív lezártja

- ▶ Legyenek X_1, \dots, X_n halmazok és $R \subseteq X_1 \times \dots \times X_n$, ekkor R -et **n -változós relációnak** nevezzük.
- ▶ $R \subseteq X \times Y$ és $S \subseteq Y \times Z$ bináris relációk **kompozíciója**
 $R \circ S = \{(x, z) \in X \times Z \mid \exists y \in Y : (x, y) \in R \wedge (y, z) \in S\}$
- ▶ $R^1 := R$, $R^i := R \circ R^{i-1}$ ($i \geq 2$) definiálja R **i -edik hatványát**
- ▶ Ha $R \subseteq X \times X$, akkor $S := R \cup \{(x, x) \mid x \in X\}$ az R reláció **reflexív lezártja**
- ▶ Ha $R \subseteq X \times X$, akkor $R^+ := \bigcup_{i=1,2,\dots} R^i$ az R reláció **tranzitív lezártja**
- ▶ Ha $R \subseteq X \times X$, akkor $R^* := R^+ \cup \{(x, x) \mid x \in X\}$ az R reláció **reflexív, tranzitív lezártja**

Reláció reflexív, tranzitív lezártja

Példa:

$X = \{1, 2, 3, 4, 5\}$, $R = \{(1, 2), (2, 3), (3, 4), (2, 5)\}$ ekkor

$S = \{(1, 2), (2, 3), (3, 4), (2, 5), (1, 1), (2, 2), (3, 3), (4, 4), (5, 5)\}$ az R reflexív lezártja

$$R^1 = R = \{(1, 2), (2, 3), (3, 4), (2, 5)\}$$

$$R^2 = \{(1, 3), (2, 4), (1, 5)\}$$

$$\text{☐ } R^3 = \{(1, 4)\}$$

$$R^4 = R^5 = \dots = \emptyset$$

$R^+ = \{(1, 2), (2, 3), (3, 4), (2, 5), (1, 3), (2, 4), (1, 5), (1, 4)\}$ az R tranzitív lezártja

$R^* = \{(1, 2), (2, 3), (3, 4), (2, 5), (1, 3), (2, 4), (1, 5), (1, 4), (1, 1), (2, 2), (3, 3), (4, 4), (5, 5)\}$ az R reflexív, tranzitív lezártja

Generált nyelv

Definíció

Legyen $G = \langle N, T, P, S \rangle$ egy tetszőleges generatív grammatika. A G **grammatika által generált nyelv** alatt az

$L(G) := \{w \mid S \Rightarrow_G^* w, w \in T^*\}$ szavakból álló halmazt értjük.

Vegyük észre, hogy a generált nyelv szavai mindig terminális szavak!

Generált nyelv – példa

1. Példa:

$G_1 = \langle N, T, P, S \rangle$, ahol $N = \{S, A, B\}$, $T = \{a, b\}$ és $P = \{S \rightarrow aSb, S \rightarrow ab, S \rightarrow ba\}$.

Ekkor $L(G_1) = \{a^n abb^n, a^n bab^n \mid n \geq 0\}$.

A G_1 -beli levezetések igen speciálisak, az egyetlen lehetőségünk, hogy n -szer alkalmazzuk az 1. szabályt, majd egyszer a 2.-at vagy a 3.-at.

Ugyanis teljes indukcióval könnyen látható, hogy a mondatformák legfeljebb 1 nemterminálist tartalmaznak. Így a 2. vagy 3. szabály alkalmazása véget vet a levezetésnek.

Szintén teljes indukcióval látható, hogy az 1. szabály n darab alkalmazása után $a^n Sb^n$ lesz a mondatforma, amiből az állításunk következik.

(Az indukciós lépés: $a^n Sb^n \Rightarrow a^n aSbb^n = a^{n+1} Sb^{n+1}$.)

Generált nyelv – példa

2. Példa:

$G_2 = \langle N, T, P, S \rangle$, ahol $N = \{S, A, B\}$, $T = \{a, b\}$ és
 $P = \{S \rightarrow ASB, S \rightarrow \varepsilon, AB \rightarrow BA, BA \rightarrow AB, A \rightarrow a, B \rightarrow b\}$.

Jelölés: Jelölje $|u|_t$ az u szóban szereplő t betűk számát.

Ekkor $L(G_2) = \{u \in \{a, b\}^* \mid |u|_a = |u|_b\}$, azaz G_2 pontosan azokat a szavakat generálja, amelyek ugyanannyi a -t és b -t tartalmaznak.

(\subseteq !)

Teljes indukcióval látható, hogy minden α mondatformára

$$|\alpha|_a + |\alpha|_A = |\alpha|_b + |\alpha|_B,$$

ugyanis kezdetben (S -re) fennáll az egyensúly, és ezt az egyensúlyt egyik szabály alkalmazása se tudja elrontani.

Mivel $L(G_2)$ szavai $\{a, b\}$ feletti, így minden generált u szóra $|u|_a = |u|_b$ teljesül.

Generált nyelv – példa

(\supseteq):

Ha egy u szó n darab a -t és b -t tartalmaz, akkor megadunk egy levezetési stratégiát.

- ▶ alkalmazzunk n 1. szabályt, így $A^n S B^n$ -t kapunk (ez teljes indukcióval az előzőekhez hasonlóan látható)
- ▶ majd egy 2.-at, így $A^n B^n$ -et kapjuk,
- ▶ a 3. és 4. szabályokkal a betűk tetszőleges sorrendje előállítható, állítsuk elő azt a sorrendet, ami u -ban szerepel (persze nagybetűkkel kis betűk helyett),
- ▶ végül az 5. és 6. szabályokkal váltsuk át a nagybetűket kicsikre.

Például $bbaaba$ levezetése:

$S \Rightarrow ASB \Rightarrow AASBB \Rightarrow AAASBBB \Rightarrow AAABBB \Rightarrow AABABB \Rightarrow$
 $ABAABB \Rightarrow BAAABB \Rightarrow BAABAB \Rightarrow BABAAB \Rightarrow BBAAAB \Rightarrow$
 $BBAABA \Rightarrow bBAABA \Rightarrow bbAABA \Rightarrow bbaABA \Rightarrow bbaaBA \Rightarrow$
 $bbaabA \Rightarrow bbaaba.$