Számítási modellek

2: Alapok, Grammatikák, Nyelvek

Alapok, terminológia

- **Ábécé**: szimbólumok/betűk egy véges, nemüres halmaza.
- **Szavak** vagy **sztringek** egy V ábécé felett: Egy V ábécé elemeiből képzett véges sorozatok.
- Egy $u = t_1 \dots t_n$ **szó hossza**: u-ban lévő betűk száma n.
  Jelölés: $|u| = n$
- Az **üres szó** ε: a 0 hosszú szó ($|\varepsilon| = 0$).
- V^* : szavak halmaza V felett, amely tartalmazza az üres szót (ε) is.
- $V^+ = V^* \setminus \{\varepsilon\}$: nemüres szavak halmaza V felett.

- Példa:
  Legyen $V = \{a,b\}$, ekkor *ab* és *baaabb* szavak V felett.

Alapok, terminológia

- Legyen V egy ábécé és legyenek u és v V feletti szavak. Ekkor az uv szót az u és v szavak **konkatenált**jának nevezzük. (Az u szó betűi után írjuk a v szó betűit.)
- $|uv| = |u| + |v|$.

- Példa:
  Legyen $V = \{a,b\}$, $u = abb$ és $v = aaab$.
  Ekkor $uv = abbaaab$, $vu = aaababb$.

Alapok, terminológia

Tulajdonságok:

- A konkatenáció asszociatív, de általában nem kommutatív.
  - ha $u,v \in V^*$, $u \neq v$, akkor uv különbözik vu szótól, kivéve ha V egyetlen betűt tartalmaz (nem kommutatív).
  - ha $u,v,w \in V^*$, akkor $u(vw) = (uv)w$ (asszociatív).
- V^* **zárt** a konkatenáció műveletre (azaz bármely $u,v \in V^*$ esetén teljesül $uv \in V^*$).
- A konkatenáció egységelemes művelet, az **egységelem** az ε (azaz minden $u \in V^*$ esetén teljesül $u = u\varepsilon = \varepsilon u$).

Alapok, terminológia

- Legyen i nemnegatív egész szám és legyen u egy V ábécé feletti szó ($u \in V^*$). Az u szó ***i*-edik hatványa** u^i az u szó i darab példányának konkatenáltja.
- Konvenció: $u^0 = \varepsilon$.

- Példa:
  Legyen $V = \{a,b\}$ és $u = abb$ egy szó V felett.
  Ekkor $u^0 = \varepsilon$, $u^1 = abb$, $u^2 = abbabb$, $u^3 = abbabbabb$, ...

Alapok, terminológia

- Legyenek u és v szavak V felett. A szavak u és v **azonosak**, ha mint szimbólumsorozatok elemről-elemre megegyeznek, azaz, $|u|=|v|$ ha minden $i = 1,\ldots,|u|$, az u szó i-edik betűje megegyezik a v szó i-edik betűjével.
- Legyen V egy ábécé és legyenek u és v szavak V felett.
  Az u **részszava** v-nek,
  ha $v = xuy$ teljesül valamely $x,y \in V^*$ szavakra.
- Az u szó **valódi részszava** v-nek, ha x és y közül legalább az egyik nem üres, vagyis ha $xy \neq \varepsilon$.
- Ha $x = \varepsilon$, akkor az u szó a v szó **prefixe**.
- Ha $y = \varepsilon$, akkor az u szó a v szó **szuffixe**.

Alapok, terminológia

- Példa:
  Legyen $V = \{a,b\}$ és $u = abb$.
  - Az u részszavai: $\varepsilon, a, b, ab, bb, abb$.
  - Az u valódi részszavai: $\varepsilon, a, b, ab, bb$.
  - Az u prefixei: ε, a, ab, abb.
  - Az u szuffixei: ε, b, bb, abb.

Alapok, terminológia

- Legyen u egy szó V felett. Az u szó **tükörképe** vagy **fordítottja** az a u^{-1} szó, amelyet úgy kapunk, hogy u szimbólumait fordított sorrendben írjuk le.
- Legyen $u = a_1 \ldots a_n$, $a_i \in V$, $1 \leq i \leq n$. Ekkor $u^{-1} = a_n \ldots a_1$.
- $(u^{-1})^{-1} = u.$
- $(u^{-1})^i = (u^i)^{-1}$ teljesül, $i = 1, 2, \ldots$

- Példa:
  Legyen $V = \{a,b\}$, $u = abba$ és $v = aabbba$
  Ekkor $u^{-1} = abba$ (palindrom) és $v^{-1} = abbbaa$.

Alapok, terminológia

- Legyen V egy ábécé és L egy tetszőleges részhalmaza V^*-nek.
  Ekkor L-t egy V feletti nyelvnek nevezzük.
- Az **üres nyelv** (egy olyan nyelv, amely nem tartalmaz egyetlen szót sem) jelölése $\emptyset$.
- Egy V ábécé feletti L nyelv **véges nyelv**, ha véges számú szót tartalmaz, egyébként **végtelen**.

- Példa:
  Legyen $V = \{a,b\}$ egy ábécé.
  - $L_1 = \{a, b, \varepsilon\}$.
  - $L_2 = \{a^i b^i \mid i \geq 0\}$.
  - $L_3 = \{uu^{-1} \mid u \in V^*\}$.
  - $L_4 = \{(a^n)^2 \mid n \geq 1\}$.
  - $L_5 = \{u \mid u \in \{a, b\}^+ , N_a(u) = N_b(u)\}$, ahol $N_a(u)$ és $N_b(u)$ az a és a b szimbólumok u-beli előfordulásainak számát jelöli.

  L_1 egy véges nyelv, a többi végtelen.

Alapok, terminológia

- Egy **generatív grammatika** G egy (N, T, P, S) 4-es, ahol
  - N és T diszjunkt véges ábécék (azaz $N \cap T = \emptyset$).
  - N elemeit **nemterminális** szimbólumoknak nevezzük
  - T elemeit **terminális** szimbólumoknak nevezzük.
  - $S \in N$ a **kezdő szimbólum** (axióma).
  - P rendezett (x,y) párok véges halmaza, ahol $x,y \in (N \cup T)^*$ és x legalább egy nemterminális szimbólumot tartalmaz.
  - P elemeit **átírási szabályoknak** (**produkciós szabályoknak**) vagy röviden **szabályoknak** nevezik. $x \to y$ használható (x,y) helyett, ahol $\to \notin (N \cup T)$.

Alapok, terminológia

- Példa:
  - $G_1 = (\{S, A, B\}, \{a, b, c\}, \{S \to c, S \to AB, A \to aA, B \to \varepsilon, abb \to aSb\}, S)$ nem generatív grammatika.

  - $G_2 = (\{S, A, B, C\}, \{a, b, c\}, \{S \to a, S \to AB, A \to Ab, B \to \varepsilon, aCA \to aSc\}, S)$ generatív grammatika.

Alapok, terminológia

- Legyen $G = (N,T,P,S)$ egy generatív grammatika és $u,v \in (N \cup T)^*$. A v szó **közvetlenül** vagy **egy lépésben levezethető** az u szóból G-ben, ha $u = u_1xu_2$ és $v = u_1yu_2$, ahol $u_1,u_2 \in (N \cup T)^*$ és $x \to y \in P$. Jelölés: $u \Rightarrow_G v$.

- Legyen $G = (N,T,P,S)$ egy generatív grammatika és $u\ v \in (N \cup T)^*$. A v szó **levezethető** az u szóból G-ben, jelölés $u \Rightarrow^*_G v$,
  - ha $u = v$, vagy
  - van olyan szó $z \in (N \cup T)^*$, amelyre $u \Rightarrow^*_G z$ és $z \Rightarrow_G v$.
  - $\Rightarrow^*$ a $\Rightarrow$ reflexív és tranzitív lezártja.
  - $\Rightarrow^+$ a $\Rightarrow$ tranzitív lezártja.

Alapok, terminológia

- Legyen $G = (N,T,P,S)$ egy generatív grammatika és $u,v \in (N \cup T)^*$.
  A v szó ***k* lépésben levezethető** u szóból G-ben, $k \geq 1$, ha van olyan szavakból álló sorozat $u_1, \ldots, u_{k+1} \in (N \cup T)^*$, amelyre $u=u_1$, $v=u_{k+1}$, és $u_i \Rightarrow_G u_{i+1}$, $1 \leq i \leq k$.
- A v szó **levezethető** u szóból G-ben, ha
  - vagy $u = v$,
  - vagy van olyan $k \geq 1$ szám, amelyre v levezethető u-ból k lépésben.

Alapok, terminológia

- Legyen $G = (N,T,P,S)$ egy generatív grammatika.
  A G által **generált nyelv** $L(G)$:
  $L(G) = \{w \mid S \Rightarrow^*_G w, w \in T^*\}$

- Azaz, $L(G)$ olyan szavakból áll, amelyek levezethetők S-ből és T^*-beliek.

Alapok, terminológia

- Példa:
  Legyen $G = (N,T,P,S)$ egy generatív grammatika, ahol
  $N = \{S,A,B\}$, $T = \{a,b\}$,
  $P = \{S \to aSb,\ S \to ab,\ S \to ba\}$.
  Ekkor $L(G) = \{a^n abb^n\ ,\ a^n bab^n \mid n \geq 0\}$.

- Példa:
  Legyen $G = (N,T,P,S)$ egy generatív grammatika, ahol
  $N = \{S,X,Y\}$, $T = \{a,b,c\}$,
  $P = \{S \to abc,\ S \to aXbc,\ Xb \to bX\ ,\ Xc \to Ybcc,\ bY \to Yb,$
  $aY \to aaX\ ,\ aY \to aa\}$.
  Ekkor $L(G) = \{a^n b^n c^n \mid n \geq 1\}$.

Alapok, terminológia

- Minden grammatika generál egy nyelvet, de ugyanaz a nyelv több különböző grammatika által is generálható.

- Két grammatika **ekvivalens**, ha ugyanazt a nyelvet generálják.
- Két nyelv **gyengén ekvivalens**, ha legfeljebb az üres szóban különböznek.

Chomsky hierarchia

- Legyen $G = (N,T,P,S)$ egy generatív grammatika.
  A G grammatika i-típusú, i = 0, 1, 2, 3, ha P szabályhalmazra a következők teljesülnek:
  - i = 0: nincs korlátozás.
  - i = 1: P minden szabálya $u_1Au_2 \rightarrow u_1vu_2$ alakú, ahol $u_1, u_2, v \in (N \cup T)^*$, $A \in N$, és $v \neq \varepsilon$, kivéve a $S \rightarrow \varepsilon$ szabályt, ha létezik ilyen szabály P-ben.
    Ha P tartalmazza az $S \rightarrow \varepsilon$ szabályt, akkor S nem fordul elő egyetlen szabály jobb oldalán sem.
  - i = 2: P minden szabálya $A \rightarrow v$ alakú, ahol $A \in N$ és $v \in (N \cup T)^*$.
  - i = 3: P minden szabálya vagy $A \rightarrow uB$ vagy $A \rightarrow u$ alakú, ahol $A, B \in N$ és $u \in T^*$.

Chomsky hierarchia

- Egy L **nyelv *i*-típusú**, $i = 0, 1, 2, 3$, ha i-típusú grammatikával generálható.
- $\mathcal{L}_i$, $i = 0, 1, 2, 3$, jelöli az i-típusú nyelvek osztályát (családját).

Chomsky hierarchia

- 0-típusú grammatikákat **mondatszerkezetű** (**phrase-structured**) grammatikáknak is nevezzük.
- 1-típusú grammatikák a **környezetfüggő** (**context-sensitive**) grammatikák. Egy A nemterminális valamely előfordulása v szóval csak u_1 és u_2 kontextus jelenlétében helyettesíthető.
- 2-típusú grammatikákat **környezetfüggetlen** (**context-free**) grammatikáknak nevezzük. Egy A nemterminális v-vel való helyettesítése bármely kontextusban megengedett.
- 3-típusú grammatikákat **reguláris** (**regular**) vagy **véges állapotú** (**finite state**) grammatikáknak hívjuk, a véges állapotú automatákkal való kapcsolatuk miatt.

- A 0,1,2,3-típusú nyelvek osztályait rendre **rekurzívan felsorolható**, **környezetfüggő**, **környezetfüggetlen**, valamint **reguláris nyelvosztály**nak nevezzük.

Chomsky hierarchia

Nyelvészeti háttér

”The cunning fox hastily ate the leaping frog.”

- $S \rightarrow A + B$ (S: sentence, A: noun phrase, B: verb phrase)
- $A \rightarrow C + D + E$ (C : article, D: adjective, E : noun)
- $B \rightarrow G + B$ (G : adverb)
- $B \rightarrow F + A$ (F : verb)
- $C \rightarrow$ the
- $D \rightarrow$ cunning
- $E \rightarrow$ fox
- $G \rightarrow$ hastily
- $F \rightarrow$ ate
- $D \rightarrow$ leaping
- $E \rightarrow$ frog

Chomsky hierarchia

Nyelvészeti háttér

- \+ (space)
- cunning – leaping , fox – frog (felcserélhető, de más lesz a jelentés)
- szintaktikailag helyes mondat

Chomsky hierarchia

- Nyilvánvaló, hogy $\mathcal{L}_3 \subseteq \mathcal{L}_2 \subseteq \mathcal{L}_0$ and $\mathcal{L}_1 \subseteq \mathcal{L}_0$.
- Megmutatható, hogy a következő is fennáll (Chomsky hierarchia): $\mathcal{L}_3 \subset \mathcal{L}_2 \subset \mathcal{L}_1 \subset \mathcal{L}_0$.

- Az $\mathcal{L}_2$ és $\mathcal{L}_1$ nyelvosztály közötti tartalmazási reláció nem látható azonnal a grammatikák definíciójából.
- Szintén $\mathcal{L}_1$-et generálják az u.n. hossznemcsökkentő grammatikák. Ezek $p \rightarrow q$ szabályaira $|p| \leq |q|$ teljesül, kivéve az $S \rightarrow \varepsilon$ alakú szabályt, feltéve, hogy P-ben létezik ilyen szabály.
  Ha $S \rightarrow \varepsilon \in P$, akkor S nem fordul elő P egyetlen szabályának jobb oldalán sem.

Nyelvekre vonatkozó műveletek

- Legyen V egy ábécé és legyenek L_1, L_2 nyelvek V felett (azaz, $L_1 \subseteq V^*$, $L_2 \subseteq V^*$)
  - **unió:** $L_1 \cup L_2 = \{u \mid u \in L_1$ vagy $u \in L_2\}$.
  - **metszet:** $L_1 \cap L_2 = \{u \mid u \in L_1$ és $u \in L_2\}$.
  - **különbség:** $L_1 - L_2 = \{u \mid u \in L_1$ és $u \notin L_2\}$.

- Példa:

  Legyen $V = \{a,b\}$ egy ábécé és $L_1 = \{a,b\}$ és $L_2 = \{\varepsilon,a,bbb\}$ nyelvek V felett. Ekkor

  $L_1 \cup L_2 = \{\varepsilon, a, b, bbb\}$

  $L_1 \cap L_2 = \{a\}$

  $L_1 - L_2 = \{b\}$

Nyelvekre vonatkozó műveletek

- Az $L \subseteq V^*$ nyelv **komplementere** a V-re az $\overline{L} = V^* - L$ nyelv.

- Példa:
  Legyen $V = \{a\}$ egy ábécé és legyen $L = \{a^{4n} \mid n \geq 0\}$.
  Ekkor $\overline{L} = V^* - \{a^{4n} \mid n \geq 0\}$.

Nyelvekre vonatkozó műveletek

- Legyen V egy ábécé és L_1, L_2 nyelvek V felett (azaz, $L_1 \subseteq V^*$, $L_2 \subseteq V^*$). L_1 és L_2 **konkatenációja** $L_1L_2 = \{u_1u_2 \mid u_1 \in L_1 , u_2 \in L_2 \}$.

- Megjegyzés:
  A következő egyenlőségek minden L nyelvre érvényesek:
  - $\varnothing L = L\varnothing = \varnothing$ és
  - $\{\varepsilon\}L = L\{\varepsilon\} = L$.

Nyelvekre vonatkozó műveletek

- L^i jelöli az L nyelv ***i*-edik iterációját** (a konkatenáció műveletre nézve), ahol $i \geq 1$. Konvenció: $L^0 = \{\varepsilon\}$.
- Az L nyelv **iteratív lezártja** (vagy **Kleene lezártja**):
  $L^* = \bigcup_{i \geq 0} L^i$.
- L **pozitív lezártja**: $L^+ = \bigcup_{i \geq 1} L^i$.

- Megjegyzés:
  Ha $\varepsilon \in L$, akkor $L^+ = L^*$. Máskülönben, $L^+ = L^* - \{\varepsilon\}$.

Nyelvekre vonatkozó műveletek

- Példa (konkatenáció):
  Legyen $V = \{a, b\}$ és $L_1 = \{a, b\}$, $L_2 = \{\varepsilon, a, bbb\}$,
  $L_3 = \{a^{4n}b^{4n} \mid n \geq 0\}$ és $L_4 = \{a^{7n}b^{7n} \mid n \geq 0\}$. Ekkor
  - $L_1L_2 = \{a, b, aa, ba, abbb, bbbb\}$,
  - $L_3L_4 = \{a^{4n}b^{4n}a^{7m}b^{7m} \mid n \geq 0, m \geq 0\}$.

Nyelvekre vonatkozó műveletek

- Legyen V egy ábécé és $L \subseteq V^*$. Ekkor az L nyelv **tükörképe** (vagy **fordítottja**) $L^{-1} = \{u^{-1} \mid u \in L\}$.

- Megjegyzések:
  - $(L^{-1})^{-1} = L$,
  - $(L_1 L_2 \ldots L_n)^{-1} = L_n^{-1} \ldots L_2^{-1} L_1^{-1}$,
  - $(L^i)^{-1} = (L^{-1})^i$, ahol $i \geq 0$,
  - $(L^*)^{-1} = (L^{-1})^*$.

Nyelvekre vonatkozó műveletek

- Példa (tükör, fordított):
  Legyen $V = \{a,b\}$ és $L = \{\varepsilon, a, abb\}$ egy nyelv V felett. Ekkor
  $L^{-1} = \{\varepsilon, a, bba\}$.

Nyelvekre vonatkozó műveletek

- Egy $L \subseteq V^*$ nyelv **prefix nyelve**
  $PRE(L) = \{ u \mid u \in V^*, uv \in L$ valamely $v \in V^*$-re $\}$.

- Megjegyzés:
  Definíció szerint, $L \subseteq PRE(L)$ minden $L \in V^*$ nyelvre.

- Egy $L \subseteq V^*$ nyelv **szuffix nyelve**
  $SUF(L) = \{ u \mid u \in V^*, vu \in L$ valamely $v \in V^*$-re $\}$.

Nyelvekre vonatkozó műveletek

- Legyen V_1 és V_2 két ábécé. A $h : V_1^* \to V_2^*$ leképezést **homomorfizmusnak** nevezzük, ha a következő feltételek teljesülnek:
  - minden $u \in V_1^*$ szóra pontosan egy $v \in V_2^*$ szó létezik, amelyre $h(u) = v$.
  - $h(uv) = h(u)h(v)$, minden $u, v \in V_1^*$-ra.

- Megjegyzések:
  - A fenti feltételekből következik, hogy $h(\varepsilon) = \varepsilon$.
    Valamint, minden $u \in V_1^*$-ra $h(u) = h(\varepsilon u) = h(u\varepsilon)$.
  - Minden $u = a_1a_2 \ldots a_n$, $a_i \in V_1$, $1 \leq i \leq n$, szóra teljesül, hogy $h(u) = h(a_1)h(a_2) \ldots h(a_n)$.
    Ez azt jelenti, hogy elegendő a h leképezést V_1 elemeire megadni, ez automatikusan kiterjesztődik V_1^*-ra.

Nyelvekre vonatkozó műveletek

- Egy homomorfizmus $h : V_1^* \rightarrow V_2^*$ **ε-mentes** ha minden $u \in V_1^+$-ra $h(u) \neq \varepsilon$.

- Legyen $h : V_1^* \rightarrow V_2^*$ egy homomorfizmus.
  Egy $L \subseteq V_1^*$ nyelv ***h*-homomorf képe** a következő nyelv:
  $h(L) = \{w \in V_2^* \mid w = h(u), u \in L\}$

- Példa (homomorfizmus):
  Legyen $V_1 = V_2 = \{a,b\}$ két ábécé. Legyen $h : V_1^* \rightarrow V_2^*$ egy homomorfizmus, ahol $h(a) = bbb$, $h(b) = ab$ és $L = \{a, abba\}$.
  Ekkor $h(L) = \{bbb, bbbababbbb\}$.

Nyelvekre vonatkozó műveletek

- Egy h homomorfizmus egy **izomorfizmus**, ha teljesül:
  $\forall\, u,v \in V_1^*$: ha $h(u) = h(v)$, akkor $u = v$.

- Példa (izomorfizmus, decimális számok bináris reprezentációja):
  $V_1 = \{0, 1, 2, \ldots, 9\}$, $V_2 = \{0, 1\}$,
  $h(0) = 0000$, $h(1) = 0001$, $\ldots$, $h(9) = 1001$

Irodalom

- Handbook of Formal Languages, G. Rozenberg, A. Salomaa, (eds.), Springer–Verlag, Berlin–Heidelberg, 1997.
- Gy. E. Révész, Introduction to Formal Languages, Dover Publications, Inc., New York, 2012.