ELMÉLETI INFORMATIKA

I. rész

Formális nyelvek és automaták

Bevezetés, alapfogalmak

1. előadás

Bevezetés

A számítógépek programozására használatos **mesterséges nyelv**ek sokkal egyszerűbbek a természetes nyelveknél.

Egy ilyen mesterséges nyelv jellemzői:

- a fogalmi kör erősen leszűkített,
- a szókincs szegényes,
- a nyelvtani szabályok egyszerűek,
- a szabályhalmazban nincsenek kivételek.

Bevezetés

Programozási nyelvek esetén igen fontos feladat pontosan definiálni azokat a **forráskódok**at (**programok**at), amelyeket az adott programozási nyelvben megírva ***helyes***nek tekintünk.

Ez a probléma leginkább egy **fordítóprogram** (**compiler**) megtervezése során merül fel, amely egy magasabb szintű programozási nyelven megírt **forráskód**ból egy alacsonyabb szintű, ún. **tárgykód**ot állít elő. A fordítóprogram ugyanis első lépésben megvizsgálja, hogy a lefordítandó forráskód helyes-e, és csak azután fordítja le azt.

Bevezetés

Programozási nyelvek esetén igen fontos feladat pontosan definiálni azokat a **forráskódok**at (**programok**at), amelyeket az adott programozási nyelvben megírva ***helyes***nek tekintünk.

Kérdés: Mikor helyes egy forráskód?

Válasz: Ha teljesít bizonyos előre meghatározott szabályokat, feltételeket.

Például:

- megadott kulcsszóval (pl. **program**) kezdődik,
- előre megadott kulcsszavak használhatók benne előre megadott szabályok szerint,
- a program végén egy megadott jel (pl. **.**) szerepel,
- stb.

Szintaxis

Egy programozási nyelv esetén azon szabályok, feltételek összessége, amelyek segítségével definiálható, hogy az adott programozási nyelvben megírt forráskód mikor tekinthető ***helyes***nek.

Ha egy program teljesíti a szintaxisban megadott szabályokat, feltételeket, akkor a program **szintaktikailag helyes**.

Kérdés: Hogyan adható meg egy programozási nyelv szintaxisa?

Válasz: Legelterjedtebb a **generatív nyelvtan**nal történő szintaxis megadás.

1.1 példa: (egy aritmetikai kifejezés szintaxisának megadása)

1) Első lépésben meg kell adnunk, hogy milyen ***szimbólum***ok (*számok*, *betűk*, *műveleti jelek* stb.) szerepelhetnek a definiálandó aritmetikai kifejezésekben:

Változók: $A \quad B \quad C \quad D$

Konstansok: $0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9$

Műveleti jelek: $+ \quad * \quad (\;)$

$A * B + (D + 2 * A)$ **ez aritmetikai kifejezés**

$A) * BCC(** +3$ **ez nem aritmetikai kifejezés**

1.1 példa: (egy aritmetikai kifejezés szintaxisának megadása)

2) Meg kell adni, hogy a már megadott szimbólumokból milyen szabályok alapján épülnek fel az aritmetikai kifejezések:

$$\langle\text{KIFEJEZÉS}\rangle \longrightarrow \langle\text{TAG}\rangle$$
$$\langle\text{KIFEJEZÉS}\rangle \longrightarrow \langle\text{KIFEJEZÉS}\rangle + \langle\text{TAG}\rangle$$
$$\langle\text{TAG}\rangle \longrightarrow \langle\text{FAKTOR}\rangle$$
$$\langle\text{TAG}\rangle \longrightarrow \langle\text{TAG}\rangle * \langle\text{FAKTOR}\rangle$$
$$\langle\text{FAKTOR}\rangle \longrightarrow (\langle\text{KIFEJEZÉS}\rangle)$$
$$\langle\text{FAKTOR}\rangle \longrightarrow \langle\text{VÁLTOZÓ}\rangle$$
$$\langle\text{FAKTOR}\rangle \longrightarrow \langle\text{KONSTANS}\rangle$$
$$\langle\text{VÁLTOZÓ}\rangle \longrightarrow A$$
$$\langle\text{VÁLTOZÓ}\rangle \longrightarrow B$$
$$\langle\text{VÁLTOZÓ}\rangle \longrightarrow C$$
$$\langle\text{KONSTANS}\rangle \longrightarrow 0$$
$$\langle\text{KONSTANS}\rangle \longrightarrow 1$$

1.1 példa: (egy aritmetikai kifejezés szintaxisának megadása)

2) Meg kell adni, hogy a már megadott szimbólumokból milyen szabályok alapján épülnek fel az aritmetikai kifejezések:

$$\begin{array}{lcl}
\langle \text{KIFEJEZÉS} \rangle & \longrightarrow & \langle \text{TAG} \rangle \mid \langle \text{KIFEJEZÉS} \rangle + \langle \text{TAG} \rangle \\
\langle \text{TAG} \rangle & \longrightarrow & \langle \text{FAKTOR} \rangle \mid \langle \text{TAG} \rangle * \langle \text{FAKTOR} \rangle \\
\langle \text{FAKTOR} \rangle & \longrightarrow & (\langle \text{KIFEJEZÉS} \rangle) \mid \langle \text{VÁLTOZÓ} \rangle \mid \langle \text{KONSTANS} \rangle \\
\langle \text{VÁLTOZÓ} \rangle & \longrightarrow & A \mid B \mid C \mid D \\
\langle \text{KONSTANS} \rangle & \longrightarrow & 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9
\end{array}$$

Az **1.1 példá**ban tehát **aritmetikai kifejezés**nek azokat az A, B, C, D, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, $+$, $*$, $($, $)$ szimbólumokból álló jelsorozatokat (***szavakat*** vagy ***mondatokat***) fogjuk tekinteni, amelyek a fenti szabályok véges számú alkalmazásával felépíthetők.

Feladat: Vezessük le az $A * (B + 3)$ aritmetikai kifejezést!

$$\langle \text{KIFEJEZÉS} \rangle \Rightarrow \langle \text{TAG} \rangle \Rightarrow \langle \text{TAG} \rangle * \langle \text{FAKTOR} \rangle \Rightarrow \langle \text{TAG} \rangle * \big(\langle \text{KIFEJEZÉS} \rangle\big) \Rightarrow$$

$$\Rightarrow \langle \text{TAG} \rangle * \big(\langle \text{KIFEJEZÉS} \rangle + \langle \text{TAG} \rangle\big) \Rightarrow \langle \text{TAG} \rangle * (\langle \text{TAG} \rangle + \langle \text{TAG} \rangle) \Rightarrow$$

$$\Rightarrow \langle \text{FAKTOR} \rangle * (\langle \text{TAG} \rangle + \langle \text{TAG} \rangle) \Rightarrow \langle \text{FAKTOR} \rangle * (\langle \text{FAKTOR} \rangle + \langle \text{TAG} \rangle) \Rightarrow$$

$$\Rightarrow \langle \text{FAKTOR} \rangle * (\langle \text{FAKTOR} \rangle + \langle \text{FAKTOR} \rangle) \Rightarrow$$

$$\Rightarrow \langle \text{VÁLTOZÓ} \rangle * (\langle \text{FAKTOR} \rangle + \langle \text{FAKTOR} \rangle) \Rightarrow$$

$$\Rightarrow \langle \text{VÁLTOZÓ} \rangle * \big(\langle \text{VÁLTOZÓ} \rangle + \langle \text{FAKTOR} \rangle\big)$$

$$\begin{array}{lll}
\langle \text{KIFEJEZÉS} \rangle & \longrightarrow & \langle \text{TAG} \rangle \mid \langle \text{KIFEJEZÉS} \rangle + \langle \text{TAG} \rangle \\
\langle \text{TAG} \rangle & \longrightarrow & \langle \text{FAKTOR} \rangle \mid \langle \text{TAG} \rangle * \langle \text{FAKTOR} \rangle \\
\langle \text{FAKTOR} \rangle & \longrightarrow & (\langle \text{KIFEJEZÉS} \rangle) \mid \langle \text{VÁLTOZÓ} \rangle \mid \langle \text{KONSTANS} \rangle \\
\langle \text{VÁLTOZÓ} \rangle & \longrightarrow & A \mid B \mid C \mid D \\
\langle \text{KONSTANS} \rangle & \longrightarrow & 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9
\end{array}$$

Feladat: Vezessük le az $A * (B + 3)$ aritmetikai kifejezést!

$\langle\text{KIFEJEZÉS}\rangle \Rightarrow \langle\text{TAG}\rangle \Rightarrow \langle\text{TA}$

$\Rightarrow \langle\text{TAG}\rangle * (\langle\text{KIFEJEZÉS}\rangle +$

$\Rightarrow \langle\text{FAKTOR}\rangle * (\langle\text{TAG}\rangle + \langle\text{T}$

$\Rightarrow \langle\text{FAKTOR}\rangle * (\langle\text{FAKTOR}\rangle$

$$\langle\text{KIFEJEZÉS}\rangle \rightarrow \langle\text{TAG}\rangle \mid \langle\text{KIFEJEZÉS}\rangle + \langle\text{TAG}\rangle$$
$$\langle\text{TAG}\rangle \rightarrow \langle\text{FAKTOR}\rangle \mid \langle\text{TAG}\rangle * \langle\text{FAKTOR}\rangle$$
$$\langle\text{FAKTOR}\rangle \rightarrow (\langle\text{KIFEJEZÉS}\rangle) \mid \langle\text{VÁLTOZÓ}\rangle \mid \langle\text{KONSTANS}\rangle$$
$$\langle\text{VÁLTOZÓ}\rangle \rightarrow A \mid B \mid C \mid D$$
$$\langle\text{KONSTANS}\rangle \rightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$$

$\Rightarrow \langle\text{VÁLTOZÓ}\rangle * (\langle\text{FAKTOR}\rangle + \langle\text{FAKTOR}\rangle) \Rightarrow$

$\Rightarrow \langle\text{VÁLTOZÓ}\rangle * (\langle\text{VÁLTOZÓ}\rangle + \langle\text{FAKTOR}\rangle) \Rightarrow$

$\Rightarrow \langle\text{VÁLTOZÓ}\rangle * (\langle\text{VÁLTOZÓ}\rangle + \langle\text{KONSTANS}\rangle) \Rightarrow$

$\Rightarrow A * (\langle\text{VÁLTOZÓ}\rangle + \langle\text{KONSTANS}\rangle) \Rightarrow A * (B + \langle\text{KONSTANS}\rangle) \Rightarrow$

$\Rightarrow A * (B + 3)$

Feladat: Vezessük le az $A * (B+3)$ aritmetikai kifejezést!

$$\langle \text{KIFEJEZÉS} \rangle \Rightarrow \langle \text{TAG} \rangle \Rightarrow \langle \text{TAG} \rangle * \langle \text{FAKTOR} \rangle \Rightarrow \langle \text{TAG} \rangle * \big(\langle \text{KIFEJEZÉS} \rangle\big) \Rightarrow$$

$$\Rightarrow \langle \text{TAG} \rangle * \big(\langle \text{KIFEJEZÉS} \rangle + \langle \text{TAG} \rangle\big) \Rightarrow \langle \text{TAG} \rangle * (\langle \text{TAG} \rangle + \langle \text{TAG} \rangle) \Rightarrow$$

$$\Rightarrow \langle \text{FAKTOR} \rangle * (\langle \text{TAG} \rangle + \langle \text{TAG} \rangle) \Rightarrow \langle \text{FAKTOR} \rangle * (\langle \text{FAKTOR} \rangle + \langle \text{TAG} \rangle) \Rightarrow$$

$$\Rightarrow \langle \text{FAKTOR} \rangle * (\langle \text{FAKTOR} \rangle + \langle \text{FAKTOR} \rangle) \Rightarrow$$

$$\Rightarrow \langle \text{VÁLTOZÓ} \rangle * (\langle \text{FAKTOR} \rangle + \langle \text{FAKTOR} \rangle) \Rightarrow$$

$$\Rightarrow \langle \text{VÁLTOZÓ} \rangle * \big(\langle \text{VÁLTOZÓ} \rangle + \langle \text{FAKTOR} \rangle\big) \Rightarrow$$

$$\Rightarrow \langle \text{VÁLTOZÓ} \rangle * \big(\langle \text{VÁLTOZÓ} \rangle + \langle \text{KONSTANS} \rangle\big) \Rightarrow$$

$$\Rightarrow A * \big(\langle \text{VÁLTOZÓ} \rangle + \langle \text{KONSTANS} \rangle\big) \Rightarrow A * (B + \langle \text{KONSTANS} \rangle) \Rightarrow$$

$$\Rightarrow A * (B + 3)$$

Az **1.1 példá**ban az aritmetikai kifejezés szintaxisa tehát a következőképpen definiálható:

Egy szó akkor és csakis akkor aritmetikai kifejezés, ha az A, B, C, D, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, $+$, $*$, $($, $)$ szimbólumokból épül fel, és levezethető a $\langle \mathrm{KIFEJEZÉS} \rangle$-ből a megadott szintaktikai szabályok véges számú alkalmazásával.

A szintaxisnak ezt a megadási módját **generatív nyelvtan**nal való szintaxis megadásnak nevezzük.

Selye János Egyetem
GIK, Informatika Tanszék

Egy **generatív nyelvtan** definíciójának tehát több dolgot is kell tartalmaznia:

- meg kell adni azokat a szimbólumokat, melyekből a definiálandó nyelv szavai fognak állni – ezek a **terminális szimbólumok**,
- meg kell adni azokat a szimbólumokat, amelyek ugyan nem fognak szerepelni az nyelvtan által generált nyelv szavai között, de szükségesek a szintaktikai szabályok megadásához – ezek a **nemterminális szimbólumok**,
- meg kell adni a **szintaktikai** (levezetési) **szabályok**at,
- meg kell adni azt a nemterminális szimbólumot, amelyből a definiálandó nyelv valamennyi szava levezethető – ez a **kezdő nemterminális szimbólum**.

1.2 példa: (egy egyszerű programozási nyelv szintaxisa)

$$N = \{\, \langle\text{PROGRAM}\rangle, \langle\text{UTASÍTÁSLISTA}\rangle, \langle\text{UTASÍTÁS}\rangle, \langle\text{ÉRTÉKADÓ}\rangle, \langle\text{IF UTASÍTÁS}\rangle, \langle\text{WHILE UTASÍTÁS}\rangle, \langle\text{BLOKK}\rangle, \langle\text{RELÁCIÓ}\rangle, \langle\text{RELÁCIÓJEL}\rangle, \langle\text{KIFEJEZÉS}\rangle, \langle\text{TAG}\rangle, \langle\text{FAKTOR}\rangle, \langle\text{VÁLTOZÓ}\rangle, \langle\text{KONSTANS}\rangle \}$$

$$\Sigma = \{\, \textbf{.}, \textbf{;}, \textbf{:=}, \textbf{IF}, \textbf{THEN}, \textbf{ELSE}, \textbf{WHILE}, \textbf{DO}, \textbf{BEGIN}, \textbf{END}, <, >, =, \neq, +, *, (,), A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 \,\}$$

1.2 példa: (egy egyszerű programozási nyelv szintaxisa)

P:

$$\begin{aligned}
\langle \text{PROGRAM} \rangle &\longrightarrow \langle \text{UTASÍTÁSLISTA} \rangle\textbf{.} \\
\langle \text{UTASÍTÁSLISTA} \rangle &\longrightarrow \langle \text{UTASÍTÁS} \rangle \mid \langle \text{UTASÍTÁS} \rangle\textbf{;}\ \langle \text{UTASÍTÁSLISTA} \rangle \\
\langle \text{UTASÍTÁS} \rangle &\longrightarrow \langle \text{ÉRTÉKADÓ} \rangle \mid \langle \text{IF UTASÍTÁS} \rangle \mid \langle \text{WHILE UTASÍTÁS} \rangle \mid \langle \text{BLOKK} \rangle \\
\langle \text{ÉRTÉKADÓ} \rangle &\longrightarrow \langle \text{VÁLTOZÓ} \rangle := \langle \text{KIFEJEZÉS} \rangle \\
\langle \text{IF UTASÍTÁS} \rangle &\longrightarrow \textbf{IF}\ \langle \text{RELÁCIÓ} \rangle\ \textbf{THEN}\ \langle \text{UTASÍTÁS} \rangle\ \textbf{ELSE}\ \langle \text{UTASÍTÁS} \rangle \\
\langle \text{WHILE UTASÍTÁS} \rangle &\longrightarrow \textbf{WHILE}\ \langle \text{RELÁCIÓ} \rangle\ \textbf{DO}\ \langle \text{UTASÍTÁS} \rangle \\
\langle \text{BLOKK} \rangle &\longrightarrow \textbf{BEGIN}\ \langle \text{UTASÍTÁSLISTA} \rangle\ \textbf{END} \\
\langle \text{RELÁCIÓ} \rangle &\longrightarrow \langle \text{KIFEJEZÉS} \rangle \langle \text{RELÁCIÓJEL} \rangle \langle \text{KIFEJEZÉS} \rangle \\
\langle \text{RELÁCIÓJEL} \rangle &\longrightarrow < \mid > \mid = \mid \neq \\
\langle \text{KIFEJEZÉS} \rangle &\longrightarrow \langle \text{TAG} \rangle \mid \langle \text{KIFEJEZÉS} \rangle + \langle \text{TAG} \rangle \\
\langle \text{TAG} \rangle &\longrightarrow \langle \text{TAG} \rangle \mid \langle \text{TAG} \rangle * \langle \text{FAKTOR} \rangle \\
\langle \text{FAKTOR} \rangle &\longrightarrow (\langle \text{KIFEJEZÉS} \rangle) \mid \langle \text{VÁLTOZÓ} \rangle \mid \langle \text{KONSTANS} \rangle \\
\langle \text{VÁLTOZÓ} \rangle &\longrightarrow A \mid B \mid C \mid D \mid E \mid F \mid G \mid H \mid I \mid J \mid K \mid \ldots \mid X \mid Y \mid Z \\
\langle \text{KONSTANS} \rangle &\longrightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9
\end{aligned}$$

1.3 példa: (az ALGOL60 programozási nyelvben használatos számok szintaxisa)

$N = \{ \langle \text{szám} \rangle, \langle \text{előjel nélküli szám} \rangle, \langle \text{tizedes szám} \rangle, \langle \text{exponenciális rész} \rangle,$
$\langle \text{tizedes rész} \rangle, \langle \text{egész szám} \rangle, \langle \text{előjel nélküli egész szám} \rangle, \langle \text{számjegy} \rangle \}$

$\Sigma = \{ ., +, -, {}_{10}, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 \}$

1.3 példa: (az ALGOL60 programozási nyelvben használatos számok szintaxisa)

P:

$\langle \text{szám} \rangle \longrightarrow \langle \text{előjel nélküli szám} \rangle \mid +\langle \text{előjel nélküli szám} \rangle \mid -\langle \text{előjel nélküli szám} \rangle$

$\langle \text{előjel nélküli szám} \rangle \longrightarrow \langle \text{tizedes szám} \rangle \mid \langle \text{exponenciális rész} \rangle \mid \langle \text{tizedes szám} \rangle\langle \text{exponenciális rész} \rangle$

$\langle \text{tizedes szám} \rangle \longrightarrow \langle \text{előjel nélküli egész szám} \rangle \mid \langle \text{tizedes rész} \rangle \mid \langle \text{előjel nélküli egész szám} \rangle\langle \text{tizedes rész} \rangle$

$\langle \text{exponenciális rész} \rangle \longrightarrow {}_{10}\langle \text{egész szám} \rangle$

$\langle \text{tizedes rész} \rangle \longrightarrow .\langle \text{előjel nélküli egész szám} \rangle$

$\langle \text{egész szám} \rangle \longrightarrow \langle \text{előjel nélküli egész szám} \rangle \mid +\langle \text{előjel nélküli egész szám} \rangle \mid -\langle \text{előjel nélküli egész szám} \rangle$

$\langle \text{előjel nélküli egész szám} \rangle \longrightarrow \langle \text{számjegy} \rangle \mid \langle \text{számjegy} \rangle\langle \text{előjel nélküli egész szám} \rangle$

$\langle \text{számjegy} \rangle \longrightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

1.3 példa: (az ALGOL60 programozási nyelvben használatos számok szintaxisa)

A $-.98_{10}-06$ szám levezetése:

$$\langle \text{szám} \rangle \Rightarrow -\langle \text{előjel nélküli szám} \rangle \Rightarrow -\langle \text{tizedes szám} \rangle\langle \text{exponenciális rész} \rangle \Rightarrow$$
$$\Rightarrow -\langle \text{tizedes rész} \rangle\langle \text{exponenciális rész} \rangle \Rightarrow$$
$$\Rightarrow -.\langle \text{előjel nélküli egész szám} \rangle\langle \text{exponenciális rész} \rangle \Rightarrow$$
$$\Rightarrow -.\langle \text{számjegy} \rangle\langle \text{előjel nélküli egész szám} \rangle\langle \text{exponenciális rész} \rangle \Rightarrow$$
$$\Rightarrow -.9\langle \text{előjel nélküli egész szám} \rangle\langle \text{exponenciális rész} \rangle \Rightarrow$$
$$\Rightarrow -.9\langle \text{számjegy} \rangle\langle \text{exponenciális rész} \rangle \Rightarrow -.98\langle \text{exponenciális rész} \rangle \Rightarrow$$
$$\Rightarrow -.98_{10}\langle \text{egész szám} \rangle \Rightarrow -.98_{10}-\langle \text{előjel nélküli egész szám} \rangle \Rightarrow$$
$$\Rightarrow -.98_{10}-\langle \text{számjegy} \rangle\langle \text{előjel nélküli egész szám} \rangle \Rightarrow$$
$$\Rightarrow -.98_{10}-0\langle \text{előjel nélküli egész szám} \rangle \Rightarrow -.98_{10}-0\langle \text{számjegy} \rangle \Rightarrow$$
$$\Rightarrow -.98_{10}-06$$

Alapfogalmak

1.1 definíció: (ábécé)

Ábécének nevezzük szimbólumoknak egy tetszőleges véges, nem üres halmazát.

Megjegyzés: Az ábécét gyakran a görög Σ betűvel jelöljük (szigma).

1.2 definíció: (szó)

Legyen Σ egy ábécé. **Szó**nak nevezzük a Σ elemeiből képzett $a_1a_2\ldots a_k$ alakú sorozatot, ahol $k > 0$ és $a_1, a_2, \ldots a_k \in \Sigma$.

Megjegyzés: Azt a szót, amelyre $k = 0$ teljesül, és ezért egyetlen Σ halmazbeli elemet sem tartalmaz, üres szónak nevezzük és λ szimbólummal jelöljük.

1.4 példa: Σ = {A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z}

A Σ ábécé elemeiből képzett szavak: COMPUTER, NYELV, ZXBKW

1.5 példa: $\Sigma = \{0,1,2,3,4,5,6,7,8,9\}$

A Σ ábécé elemeiből képzett szavak: 0, 254987, 777

1.6 példa: $\Sigma = \{0,1\}$

A Σ ábécé elemeiből képzett szavak: 0111011010, 111, 11010001

1.7 példa: $\Sigma = \{a,b\}$

A Σ ábécé elemeiből képzett szavak: *aabaabba*, *aaab*, *babbba*

1.8 példa: Legyen Σ az összes ASCII karakter halmaza.

A Σ ábécé elemeiből képzett szavak:]Ae*w2+/wq, @xa+#0_d, M$%?

Σ^* – a Σ ábécé elemeiből képezhető összes szó halmaza

$$\Sigma^* = \left\{a_1 a_2 a_3 \dots a_k \;\middle|\; a_1, a_2, a_3, \dots a_k \in \Sigma \;\wedge\; k \geq 0\right\}$$

Σ^+ – a Σ ábécé elemeiből képezhető összes nem üres szó halmaza

$$\Sigma^+ = \left\{a_1 a_2 a_3 \dots a_k \;\middle|\; a_1, a_2, a_3, \dots a_k \in \Sigma \;\wedge\; k > 0\right\} = \Sigma^* \setminus \{\lambda\}$$

1.3 definíció: (szó hossza)

Egy w szó **hosszát** $|w|$-vel jelöljük, s ez egyenlő a szavat alkotó szimbólumok számával, azaz $|w| = k$, ha $w = a_1 a_2 \ldots a_k$.

Σ^n – a Σ ábécé elemeiből képezhető összes n hosszú szó halmaza

$$\Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \cdots \cup \Sigma^n \cup \cdots \qquad \Sigma^+ = \Sigma^1 \cup \Sigma^2 \cup \cdots \cup \Sigma^n \cup \cdots$$

1.4 definíció: (szavak egyenlősége)

A $w_1 = a_1 a_2 \ldots a_k$ és $w_2 = b_1 b_2 \ldots b_l$ szavak **egyenlők** $(w_1 = w_2)$, ha $k = l$ és $a_i = b_i$, minden $i = 1,2, \ldots, k$-ra.

1.5 definíció: (szavak konkatenációja)

Az $u, v \in \Sigma^*$ szavak **konkatenáció**ján a két szó összefűzésével, azaz egymás után történő leírásával kapott $uv \in \Sigma^*$ szót értjük.

Megjegyzés: Nyilvánvaló, hogy $|uv| = |u| + |v|$.

A konkatenáció művelet asszociatív, de nem kommutatív.

1.6 definíció: (szó k-adik hatványa)

A $w \in \Sigma^*$ szó k-**adik hatványán** k darab w szó konkatenációját értjük, és w^k-val jelöljük. Érvényes, hogy $w^0 = \lambda$.

1.7 definíció: (szó tükörképe)

A $w \in \Sigma^*$ szó visszafelé történő írását a w szó **tükörképé**nek nevezzük, és w^R-rel jelöljük.

Megjegyzés: Nyilvánvaló, hogy $(w^R)^R = w$ és $(uv)^R = v^R u^R$.

1.9 példa: Legyen $\Sigma = \{a, b\}$.

Az $abaab$ szó hossza: $|abaab| = 5$

Az $abaab$ és $bbab$ szavak konkatenációja: $abaabbbab$

A $bbab$ szó 3. hatványa: $bbab^3 = bbabbbabbbab$

Az $abaab$ szó tükörképe: $abaab^R = baaba$

A Σ^* a köv. elemeket tartalmazza: $\Sigma^* = \{\lambda, a, b, aa, ab, ba, bb, aaa, \ldots\}$

1.8 definíció: (formális nyelv)

A Σ^* halmaz valamely L részhalmazát **formális nyelv**nek nevezzük.

1.10 példa: A $\{00, 01001, 1101\}$ halmaz $\{0, 1\}$ ábécé feletti formális nyelv.

1.11 példa: A $\{\lambda, ab, aabb, aaabbb, \dots\}$ halmaz $\{a, b\}$ ábécé feletti formális nyelv.

Műveletek formális nyelvekkel

Mivel a nyelvek szavakat tartalmazó halmazok, ezért az ismert ***halmazműveletek*** elvégezhetők rajtuk:

- **egyesítés** – az $L_1 \cup L_2$ nyelv olyan szavakat tartalmaz, amelyek legalább az egyik nyelvbe beletartoznak,
- **metszet** – az $L_1 \cap L_2$ nyelv olyan szavakat tartalmaz, amelyek egyidejűleg mindkét nyelvbe beletartoznak,
- **különbség** – az $L_1 \setminus L_2$ nyelv olyan szavakat tartalmaz, amelyek az L_1 nyelvbe beletartoznak, de az L_2 nyelvbe nem,
- **komplementer** – az $\overline{L_1}$ nyelv olyan szavakat tartalmaz, amelyek nem tartoznak bele az L_1 nyelvbe.

Műveletek formális nyelvekkel

A nyelveken ezen kívül ún. ***reguláris műveletek*** is elvégezhetők:

- **konkatenáció** – az L_1 és L_2 nyelvek L_1L_2 konkatenációja olyan szavakat tartalmaz, amelyek L_1 nyelvbe tartozó szóval kezdődnek és L_2 nyelvbe tartozó szóval végződnek, azaz

$$L_1L_2 = \left\{uv \;\middle|\; u \in L_1 \;\wedge\; v \in L_2 \right\}$$

- **iteráció** – az L nyelv L^* iterátlja olyan szavakat tartalmaz, amelyeket az L nyelvbe tartozó tetszőleges számú szó konka-tenációjával kapunk, azaz

$$L^* = \{\lambda\} \cup L \cup LL \cup LLL \cup \cdots$$

1.12 példa: Legyen $L_1 = \{bb, a\}$ és $L_2 = \{a, ab, bbb\}$.

Ekkor $L_1L_2 = \{bba, bbab, bbbbb, aa, aab, abbb\}$

$$L_1L_1L_1 = \{bbbbbb, bbbba, bbabb, abbbb, bbaa, abba, aabb, aaa\}$$

$$L_1^* = \{\lambda, bb, a, bbbb, bba, abb, aa, bbbbbb, bbbbba, \ldots\}$$

Formális nyelvek megadása

Egy formális nyelv megadható

- szavainak felsorolásával

$$L_1 = \{ \lambda, 00, 1 \}$$

$$L_2 = \{ a, aa, aaa, ab, ba, aba \}$$

$$L_3 = \{ \lambda, ab, aabb, aaabbb, aaaabbbb, \ldots \}$$

- tulajdonság segítségével

$$L_4 = \left\{ a^n b^n \; ; \; n \geq 0 \right\}$$

$$L_5 = \left\{ ww^R \; ; \; w \in \Sigma^* \right\}$$

$$L_6 = \left\{ w \in \{a, b\}^* ; \; \#_a(w) = \#_b(w) \right\}$$

- generatív nyelvtan segítségével.

1.9 definíció: (generatív nyelvtan)

Generatív nyelvtannak nevezzük a $G = (N, \Sigma, P, S)$ rendezett elemnégyest, ahol:

N egy ábécé, a **nemterminális ábécé**,

Σ egy ábécé, a **terminális ábécé**, $N \cap \Sigma = \emptyset$,

P az $\alpha \longrightarrow \beta$ alakú **nyelvtani szabályok véges halmaza**, ahol $\alpha, \beta \in (N \cup \Sigma)^*$ és α tartalmaz legalább egy nemterminális szimbólumot,

S a **kezdő nemterminális szimbólum**, $S \in N$.

1.13 példa: A $G = (\{S, A\}, \{a, b\}, \{S \longrightarrow aAb, aA \longrightarrow aaAb, A \longrightarrow \lambda\}, S)$ rendezett elemnégyes egy generatív nyelvtan.

1.10 definíció: (mondatforma)

A $w \in (N \cup \Sigma)^*$ szimbólumsorozatot **mondatformá**nak nevezzük.

1.11 definíció: (mondat, szó)

A $w \in \Sigma^*$ szimbólumsorozatot **mondat**nak v. **szó**nak nevezzük.

Amennyiben egy $G = (N, \Sigma, P, S)$ nyelvtannak több olyan szabálya is van, melyeknek a bal oldala megegyezik, azaz pl. ha $\alpha \longrightarrow \beta_1$, $\alpha \longrightarrow \beta_2$, …, $\alpha \longrightarrow \beta_n$ az összes olyan P szabályhalmazbeli szabály, amelyeknek bal oldala α, akkor a továbbiakban ezt egyszerűbben így fogjuk írni:

$$\alpha \longrightarrow \beta_1 \mid \beta_2 \mid \ldots \mid \beta_n.$$

Ekkor a β_1, β_2, …, β_n mondatformákat az α ***alternatívái***nak nevezzük.

1.12 definíció: (közvetlen deriváció)

Az $(N \cup \Sigma)^*$ halmazon értelmezünk egy $\Longrightarrow_G$-vel jelölt relációt, amelyet **közvetlen derivációnak** (vagy *közvetlen levezetésnek*) nevezünk. Tetszőleges $\gamma, \delta \in (N \cup \Sigma)^*$-ra $\gamma \Longrightarrow_G \delta$ akkor és csakis akkor áll fenn, ha $\gamma = \alpha_1 \alpha \alpha_2$ és $\delta = \alpha_1 \beta \alpha_2$, és $\alpha \longrightarrow \beta \in P$.

Ilyenkor azt mondjuk, hogy a γ-ból az $\alpha \longrightarrow \beta$ nyelvtani szabály alkalmazásával **deriválható** (*levezethető*) a δ.

A $\Longrightarrow$ reláció reflexív és tranzitív lezártját a $\Longrightarrow^*$, míg tranzitív lezártját a $\Longrightarrow^+$ jelöli.

A $\Longrightarrow^*$ relációt *derivációnak* vagy *levezetésnek* nevezzük.

A reláció n-edik hatványának definíciója szerint az $\alpha \Longrightarrow^n \beta$ jelölés azt jelenti, hogy az α-ból kiindulva a P szabályhalmazbeli szabályok n-szeri alkalmazásával megkapható a β, azaz α-ból n lépésben deriválható (levezethető) β.

1.14 példa:

$$
\begin{aligned}
A &\rightarrow aBBb \mid AaA \\
B &\rightarrow \lambda \mid bCA \\
C &\rightarrow AB \mid a \mid b
\end{aligned}
$$

$$A \Rightarrow aBBb \Rightarrow abCABb \Rightarrow abCaBBbBb \Rightarrow abCaBbBb \Rightarrow abbaBbBb \Rightarrow abbaBbb \Rightarrow abbabb$$

1.13 definíció: (generált nyelv)

A $G = (N, \Sigma, P, S)$ nyelvtan által **generált nyelv** az alábbi módon adható meg:

$$L(G) = \{w \in \Sigma^* \mid S \Longrightarrow_G^* w\}.$$

1.14 definíció: (ekvivalens nyelvtanok)

A G_1 és G_2 generatív nyelvtanok, akkor és csakis akkor **ekvivalens**ek, ha $L(G_1) = L(G_2)$.