1 ABECEDA, SLOVO, JAZYK

CÍL KAPITOLY

Po úspěšném a aktivním absolvování této kapitoly

Budete umět:

- Rozumět pojmům slovo, jazyk, abeceda v rámci teoretické informatiky
- Vytvořit příklady slov, jazyků, abeced
- Použít operace s jazyky

Budete umět

ČAS POTŘEBNÝ KE STUDIU



180 minut

KLÍČOVÁ SLOVA KAPITOLY ABECEDA, SLOVO, JAZYK

Abeceda, slovo, jazyk, množiny, operace s jazyky

Klíčová slova

1.1 Abeceda, slovo, jazyk

Abeceda

Základním stavebním kamenem mluveného slova jsou hlásky, psaného textu pak písmena, stavebnice se skládá z jednotlivých kostiček, dům z cihel, oken, trubek atd. V teoretické informatice těmto základním prvkům říkáme **symboly**. Obvykle jsou to písmena latinky, mohou to však být různé značky, geometrické útvary… Pokud sestavíme množinu všech typů symbolů, které se v daném textu, skládance,… vyskytují, dostáváme abecedu.

DEFINICE 1-1



Abeceda je konečná množina. Její prvky nazýváme symboly, znaky nebo písmena.

ŘEŠENÝ PŘÍKLAD 1-1



Uveďte příklady různých abeced.

Řešení příkladu

A = { a, b, c, d, e, f, …, y, z } – abeceda tvořená malými písmeny latinky, bez diakritiky

B = { A, Á, E, É, I, Í, O, Ó, U, Ú, Ů, Y, Ý } – abeceda tvořená samohláskami i s diakritikou

C = { α, β, …, ω } – abeceda tvořená malými písmeny řecké abecedy

D = { a, b, c } – tříprvková abeceda

E = { a } – jednoprvková abeceda

F = { ■, ●, ◆, ⬟, ⬣, ▲ } – abeceda tvořená grafickými symboly

G = { ·, – } – abeceda pro morseovku

*

Ze symbolů abecedy můžeme skládat slova.

DEFINICE 1-2



Slovo (řetězec) nad abecedou V je posloupnost symbolů $a_1\, a_2 \ldots a_n$, kde $a_i \in V$, $1 \le i \le n$.

Prázdné slovo je slovo nulové délky a značíme jej **ε**.

ŘEŠENÝ PŘÍKLAD 1-2



Uveďte příklady slov ze symbolů z abeced uvedených výše.

Řešení příkladu

s_A = sobota je slovo vytvořené ze symbolů abecedy A

s_C = αβχδ je slovo ze symbol abecedy C

s_F = ●▲■ je slovo tvořené symboly z abecedy F

*

DEFINICE 1-3



Délka slova ω = $a_1a_2\ldots a_n$, je přirozené číslo n, tedy počet symbolů slova. Značíme ji |ω|. Píšeme |ω| = n

ŘEŠENÝ PŘÍKLAD 1-3



Uveďte délky slov z příkladu 1-2.

Řešení příkladu

1. $|s_A| = 7$
2. $|s_C| = 4$
3. $|s_F| = 3$

*

DEFINICE 1-4

Počet výskytů symbolu a ve slově w značíme $|w|_a$.

ŘEŠENÝ PŘÍKLAD 1-4

Uveďte $|w|_a$ pro

a) $w = xyzzyxxzx$, $a = x$

b) $w = \pi\delta\beta\alpha\varphi\alpha\beta\pi\varphi\delta\varphi\alpha\beta\mu$, $a = \delta$

c) $w =$ ☺☹😐☺☹😐☹☺😐☺, $a =$ ☺

Řešení příkladu

a) počet symbolů x ve slově $xyzzyxxzx$ je 3 a tedy $|xyzzyxxzx|_x = 3$;

b) počet symbolů δ ve slově $\pi\delta\beta\alpha\varphi\alpha\beta\pi\varphi\delta\varphi\alpha\beta\mu$ je 2 a tedy $|\pi\delta\beta\alpha\varphi\alpha\beta\pi\varphi\delta\varphi\alpha\beta\mu|_\delta = 2$;

c) počet symbolů ☺ ve slově ☺☹😐☺☹😐☹☺😐☺ je 4 a tedy |☺☹😐☺☹😐☹☺😐☺|$_{☺}$ = 4;

*

DEFINICE 1-5

Jazyk je množina konečných slov (tj. slov konečné délky) nad danou abecedou.

Jazyk tedy nemusí být nutně konečná množina slov. Pro definici toho, jaká slova patří do jazyka, používáme dva způsoby zápisu.

Jazyk zapisujeme:

1. Výčtem prvků

   {leden, únor, březen, duben}

   {dveře, okno, stůl, postel}

   {ABC, BAC, ACB, BCA, CAB, CBA}

   {a, aa, aaaa, aaaaaa, a^8, a^{10}, ...}

2. Vyjádřením závislosti

   {x | x splňuje podmínku}

ŘEŠENÝ PŘÍKLAD 1-5



Zadejte dané jazyky výčtem prvků:

1. $\{a^i \mid 1 \leq i \leq 4\}$
2. $\{a^i \mid i > 5\}$
3. $\{a^i b^i \mid i \leq 2\}$
4. $\{a^i b^j \mid i \leq 2; j = 3,4\}$
5. $\{a^i b^j c^k \mid i = j; i,j,k > 0\}$
6. $\{a^i b^j \mid i \neq j; i,j>0\}$

Řešení příkladu

1. $\{a^1, a^2, a^3, a^4\}$
2. $\{a^6, a^7, a^8, ...\}$
3. {ab, aabb}
4. $\{a^3b^3, a^3b^4, a^2b^3, a^2b^4\}$
5. $\{abc, abc^2, a^2b^2c, a^2b^2c, a^2b^2c^3, a^3b^3c, ...\}$
6. $\{ab^2, a^2b, ab^3, a^3b\}$

*

SAMOSTATNÝ ÚKOL 1



Srovnejte jazyky: $\{ a^ib^ic^i \}$? $\{ a^ib^jc^k \mid i=j=k \}$

$\{ a^ib^ic^i \}$? $\{ (abc)^i \}$, pro $i, j, k \in N$

DEFINICE 1-6



V^* je jazyk, který obsahuje všechna slova vytvořená ze symbolů abecedy V, včetně prázdného slova.

$V^+ = V^* - \{\varepsilon\}$ $\quad \varnothing^* = \varepsilon$ $\quad \varnothing^+ = \varnothing$

ŘEŠENÝ PŘÍKLAD 1-6



Pro abecedu V = {0,1} uveďte, jak vypadá V^* a V^+.

Řešení příkladu

$V^* = \{\varepsilon, 0, 1, 00, 01, 10, 11, 000, 100, 010, ...\}$

$V^+ = \{0,1, ...\}$

*

1.2 Operace s jazyky a slovy

Jazyky jsou množiny, provádíme s nimi stejné operace jako s množinami, jako jsou sjednocení, průnik, doplněk a množinový rozdíl.

DEFINICE 1-7



Sjednocení jazyků L_1, L_2 značíme $L_1 \cup L_2$. Výsledkem je jazyk, který obsahuje slova jazyků L_1 i L_2.

Zápis: $L_1 \cup L_2 = \{w \mid w \in L_1 \vee w \in L_2\}$

ŘEŠENÝ PŘÍKLAD 1-7



Určete $L_1 \cup L_2$:

$L_1 = \{a, aa, aaa\}$

$L_2 = \{b, bb, bbb\}$

Řešení příkladu

$L_1 \cup L_2 = \{a, b, aa, bb, aaa, bbb\}$

*

DEFINICE 1-8



Průnik jazyků L_1, L_2 značíme $L_1 \cap L_2$. Výsledkem je jazyk, který obsahuje slova z obou jazyků. L_1 a L_2 zároveň.

Zápis: $L_1 \cap L_2 = \{w \mid w \in L_1 \wedge w \in L_2\}$

ŘEŠENÝ PŘÍKLAD 1-8



Určete $L_1 \cap L_2$:

$L_1 = \{a, aa, aaa\}$

$L_2 = \{a^2, a^4, a^6\}$

Řešení příkladu

$L_1 \cap L_2 = \{a^2\}$

*

ŘEŠENÝ PŘÍKLAD 1-9



Určete $L_1 \cap L_2$:

$L_1 = \{a^i \mid i = 2k\}$

$L_2 = \{a^i \mid i = 2k+1\}; k > 0$

Řešení příkladu

$L_1 \cap L_2 = \varnothing$

*

ŘEŠENÝ PŘÍKLAD 1-10



Určete $L_1 \cap L_2$:

$L_1 = \{a^i b^j c^k \mid i,j,k \geq 0; i = j\}$

$L_2 = \{ a^i b^j c^k \mid i,j,k \geq 0; j = k\}$

Řešení příkladu

$L_1 \cap L_2 = \{ a^i b^j c^k \mid i,j,k \geq 0; i = j = k\}$

*

DEFINICE 1-9



Zřetězením *slov* $v = a_1a_2...a_n$ a $u = b_1b_2...b_m$ rozumíme slovo $v.u = a_1a_2...a_nb_1b_2...b_m$.

Tato operace je asociativní, to znamená, že platí $(u \,.\, v) \,.\, w = u \,.\, (v \,.\, w)$

ŘEŠENÝ PŘÍKLAD 1-11



Určete, jak vypadá zřetězení slov u.v a v.u

u = abba

v = bba

Řešení příkladu

u . v = abba . bba = abbabba

v . u = bba . abba = bbaabba

*

ŘEŠENÝ PŘÍKLAD 1-12



Ověřte asociativitu zřetězení:

u = abb, v = aaa, w = bbb

Řešení příkladu

(u . v) . w = (abb . aaa) . bbb = abbaaa . bbb = abbaaabbb

u . (v . w) = abb . (aaa . bbb) = abb . aaabbb = abbaaabbb

*

DEFINICE 1-10



Slovo u je **podslovo** slova v jestliže existují slova x a y taková, že platí: u = x . u . y

Pokud x = ε, pak u = prefix (předpona)

Pokud y = ε, pak u = sufix (přípona)

Podslova jsou všechna taková slova, která jsou obsažena ve slově původním. Nejdelším podslovem je slovo samotné, nejkratším pak slovo prázdné.

ŘEŠENÝ PŘÍKLAD 1-13



Určete nějaké podslovo a předponu slova aabab.

Řešení příkladu

předpona: a, aa, aab, aaba

podslovo: ab, ba, ... (např. bb není podslovem)

*

ŘEŠENÝ PŘÍKLAD 1-14



Uveďte všechna podslova slova 0011. Najděte taková slova délky 2 a 3, která nejsou podslovem slova 0011.

Řešení příkladu

Podslova: ε, 0, 1, 00, 01, 11, 001, 011, 0011

Nejsou podslovem: 10, 000, 010, 100, 101, 110, 111.

*

DEFINICE 1-11

n-tou mocninu slova u pro přirozené n značíme u^n a definujeme následovně:

- $u^0 = \varepsilon$,
- $u^n = u^{n-1} . u$

Platí tedy:

$u^o = \varepsilon$,
$u^1 = u$,
$u^2 = u . u$,
$u^3 = u . u . u$

DEFINICE 1-12

Zřetězením jazyků L_1, L_2 nazýváme jazyk $L_1 . L_2 = \{u . v \mid u \in L_1, v \in L_2\}$

Zřetězením jazyků tedy myslíme takový jazyk, jehož slova jsou tvořena ze dvou podslov. První podslovo je slovem jazyka L_1 a druhé podslovo je slovo, které náleží do jazyk L_2.

ŘEŠENÝ PŘÍKLAD 1-15

Uveďte všechna možná zřetězení jazyků:

$L_1 = \{0^n \mid n \geq 0\}$

$L_2 = \{1\}$

$L_3 = \{01, 11\}$

Řešení příkladu

$L_1 . L_2 = \{10^n \mid n \geq 0\}$

$L_2 . L_1 = \{0^n1 \mid n \geq 0\}$

$L_1 . L_3 = \{0^{n+1}1, 0^n11 \mid n \geq 0\}$

$L_3 . L_1 = \{010^n, 110^n \mid n \geq 0\}$

$L_2 . L_3 = \{101, 111\}$

$L_3 . L_2 = \{011, 111\}$

*

ŘEŠENÝ PŘÍKLAD 1-16



Uveďte všechna možná zřetězení jazyků:

$L_1 = \{1^n \mid n > 0\}$

$L_2 = \{0, 00\}$

Řešení příkladu

$L_1 . L_2 = \{1^n0 \mid n > 0\} \cup \{1^n00 \mid n > 0\}$

$L_2 . L_1 = \{01^n \mid n > 0\} \cup \{001^n \mid n > 0\}$

*

DEFINICE 1-13



Je-li L jazyk nad abecedou Σ, pak L^n značí **n-tou mocninu jazyka** L a platí: $L^0 = \{\varepsilon\}$, $L^{n+1} = L^n.L$;

Platí vzorec: $L^i = L . L^{i-1}$

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

$L^1 = L$

$L^2 = L . L$

$L^3 = L . L^2$

L^+ je **pozitivní iterace** jazyka L a $L^+ = L \cup L^2 \cup L^3 \ldots = L^+ = \bigcup_{n=1}^{\infty} L^n$

L* značí **iteraci jazyka** L a platí: $L^* = L^0 \cup L^1 \cup L^2 \ldots = L^* = \bigcup_{n=0}^{\infty} L^n$

ŘEŠENÝ PŘÍKLAD 1-17



Vytvořte L^2 a L^3 z jazyka L = {a, ab}.

Řešení příkladu

$L^2 = \{a^2, aab, aba, abab\}$

$L^3 = \{aaa, aaab, aaba, aaba, abaa, aabab, ababa\}$

*

DEFINICE 1-14



Zrcadlový obraz (reverze) slova $w = a_1a_2...a_n$ se nazývá slovo $w^R = a_na_{n-1}...a_2a_1$.

$\varepsilon^R = \varepsilon$

ŘEŠENÝ PŘÍKLAD 1-18



Určete zrcadlový obraz slova w = aab.

Řešení příkladu

w^R = bba

*

DEFINICE 1-15



Zrcadlový obraz (reverze) jazyka L se nazývá jazyk $L^R = \{w^R \mid w \in L\}$.

ŘEŠENÝ PŘÍKLAD 1-19



Určete zrcadlový obraz jazyka $L = \{a^i b^i c^i \mid i > 0\}$.

Řešení příkladu

$L^R = \{c^i b^i a^i \mid i > 0\}$

*

DEFINICE 1-16



Substituce (nahrazení) je zobrazení $f: \Sigma \to 2^{\Delta^*}$ abecedy Σ do podmnožin množiny Δ^*.

Pro každé $a \in \Sigma$ definujeme $f(a) \subseteq \Delta^*$

Substituce se nazývá **nevypouštějící**, jestliže $\forall a \in \Sigma$ platí,že $\varepsilon \notin f(a)$

Substituce slova vzniká rozšířením substituce $f: \Sigma \to 2^{\Delta^*}$ na $f: \Sigma^* \to 2^{\Delta^*}$ a definujeme ji takto:

- $f(\varepsilon) = \varepsilon$
- $f(xa) = f(x)\, f(a)$

Čili substituce prázdného slova je opět prázdné slovo a substituce řetězce odpovídá zřetězení substitucí.

Substituci jazyků definujeme tak, že pro každé $L \in \Sigma^*$ je substituce sjednocením substitucí slov jazyka, čili:

$$f(L) = \bigcup_{x \in L} f(x).$$

Říkáme, že jazyk f(L) vznikl z L substitucí f.

ŘEŠENÝ PŘÍKLAD 1-20



Určete f(ab):

f(a) = {1, 2}; f(b) = {3, 4}.

Řešení příkladu

f(ab) = f(a) . f(b) = {1, 2} . {3, 4} = {13, 14, 23, 24}

*

ŘEŠENÝ PŘÍKLAD 1-21



Určete f(ab):

f(a) = {aa, bb}; f(b) = {ab, ba}.

Řešení příkladu

f(ab) = f(a) . f(b) = {aaab, aaba, bbab, bbba}

*

DEFINICE 1-17



Homomorfismus je jednoprvková substituce.

Jestliže každé f(a) obsahuje právě jedno slovo, pak se f nazývá homomorfismus.

SAMOSTATNÝ ÚKOL 2



Pro abecedy $\Sigma = \{a, b, c\}$ a $\Delta = \{1, 2\}$ navrhněte tři různé substituce tak, aby jedna byla nevypouštějící, druhá vypouštějící homomorfismus a třetí nevypouštějící homomorfismus.

SAMOSTATNÝ ÚKOL 3



Lze pro abecedy z předchozího příkladu vytvořit substituci takovou, aby ke každému slovu jazyka $f(L) \subseteq \Delta^*$ vzniklého substitucí jazyka $L \subseteq \Sigma^*$ existovalo právě jedno slovo z jazyka L?

1.3 Konečná reprezentace jazyků

V dnešní době počítacích strojů vzniká potřeba zpracovávat jazyky mechanicky. Například kontrolovat správnost jejich syntaxe, provádět příkazy, které jsou zakódované v jednotlivých slovech,….Vzniká otázka, jak takový jazyk, který může být nekonečný (ale spočetný), můžeme reprezentovat právě pro potřeby počítačů. Uvedeme si jednoduchý příklad: Mějme jazyk, jehož slova se skládají pouze s jedniček a nul. Podmínkami pro to, aby slovo patřilo do jazyka jsou: 1) stejný počet jedniček a nul, 2) slovo začíná posloupností 001 a 3) délka slova je sudá. Asi nikomu z nás nedá velkou práci představit si, jak vypadají slova tohoto jazyka a pro slova rozumné délky i ověřit, jestli náleží nebo nenáleží do jazyka. Jak si s takovým úkolem poradí počítač? Obecně je nemožné udělat pro něj slovník slov, která patří do jazyka a nechat ho jen porovnávat, jestli takové slovo zná. Máme dvě možnosti. První vychází právě s možnosti vytvoření slovníku.

1. Najdeme takový postup (**proceduru**), který je schopný vygenerovat všechna slova. (následně může zkontrolovat, jestli původní slovo vygeneroval či nikoliv)
2. Najdeme takový **algoritmus**, který nám dá odpověď, jestli dané slovo patří do jazyka, či nikoliv. (rozpozná, jestli dané slovo náleží do jazyka)

Algoritmus je výpočetní proces, který se zastaví pro každý vstup. **Procedura** se pro některé vstupy zastavit nemusí. Z toho vyplývá, že každý algoritmus je i procedura, ne každá procedura je však algoritmus.

SHRNUTÍ KAPITOLY



Shrnutí modulu

Tato kapitola seznámila studenty s pojmy: abeceda, slovo, jazyk, operace s (nad) jazyky.