

ZÁKLADNÍ POJMY Z TEORIE JAZYKŮ

ABECEDA - KONEČNÁ NEPRÁZDNÁ MNOŽINA.

PRVKY ABECEDY SE NAZÝVAJÍ PÍSMENA.

ŘETĚZEC (SLVO) - LIBOVOLNÁ, KONEČNÁ
POSLoupNOST^a (TJ. USPOřáDANÁ M-VICE)
PÍSMEN ABECEDY.

PŘ: ABECEDA $\Sigma = \{a, b, c\}$

PŘÍKLADY
ŘETĚZCŮ

ab
aabaaba
aba

ϵ [EMPTY = PRÁZDNÝ
ŘETĚZEC]

UZÁVĚR ABECEDY Σ^* - MNOŽINA

VŠECH NEPRÁZDNÝCH ŘETĚZCŮ VYTVOŘE-
NÝCH Z PÍSMEN ABECEDY Σ .

ZNAČENÍ: Σ^+

ITERACE ABECEDY Σ - KROUŽENÍ

VŠECH ŘETĚZÍ VYTVOŘENÝCH
Z PÍSMEN ABECEDY Σ .

ZNAČENÍ: Σ^*

ZJEDNĚNÉ: $\Sigma^* = \Sigma^+ \cup \{e\}$

PŘ: $A = \{0, 1\}$
 $2 = n$

$A^+ = \{ \underbrace{0, 1}_{2 = n}, \underbrace{00, 01, 10, 11}_{4 = n^2},$

$000, 001, 010, 011,$

$100, 101, 110, 111, \dots \}$ $8 = n^3$

$A^* = \{e, 0, 1, 00, 01, 10, 11, 000, \dots\}$

OPERACE NA ŘETĚZCI

MECHT $u = a_1 a_2 \dots a_n \in \Sigma^*$
 $v = b_1 b_2 \dots b_m \in \Sigma^*$

[u A v JSOU ŘETĚZCE NAH ABECEDOU
 Σ , OBECNĚ RŮZNÉ DĚLKY]

ŘĚTĚZENÍ : $\Sigma^* \times \Sigma^* \rightarrow \Sigma^*$

$$u \cdot v = a_1 a_2 \dots a_n b_1 b_2 \dots b_m$$

[ŘĚTĚZENÍ ŘETĚZCŮ u A v]

OBECNĚ NEPLATÍ $u \cdot v = v \cdot u$

[OPERACE NEMÍ KOMUTATIVNÍ]

ŘETĚZENÍ S PRAZDNÝM ŘETĚZCEM :

$$e \cdot u = u \cdot e = u \quad \forall u \in \Sigma^*$$

MOCMINA ŘETĚZCE : $\Sigma^* \times N_0 \rightarrow \Sigma^*$

$$M^0 = \epsilon$$

$$M^1 = M$$

$$M^2 = M \cdot M$$

\vdots

$$M^n = M \cdot M^{n-1} = M^{n-1} \cdot M$$

REVERZE (OBRÁŤENÍ) ŘETĚZCE $\Sigma^* \rightarrow \Sigma^*$

$$M = a_1 a_2 \dots a_{n-1} a_n$$

$$M^R = a_n a_{n-1} \dots a_2 a_1$$

[REVERZE ŘETĚZCE M]

DĚLKA ŘETĚZCE : $\Sigma^* \rightarrow N_0$

$$M = M_1 M_2 \dots M_n$$

$$|M| = n$$

$$|\epsilon| = 0$$

DEFINICE JAZYKA NAD PRÁZDNOU ABECEDOU

MECHT Σ JE KONEČNÁ NEPRÁZDNÁ ABECEDA. JAZYKEM L NAZVEME LIBOVOLNOU MNOŽINU ŘETĚZCŮ NAD ABECEDOU Σ .
 $L \subseteq \Sigma^*$.

TATO DEFINICE POHLÍŽÍ NA JAZYK JAKO NA MNOŽINU ŘETĚZCŮ. TATO DEFINICE NEMÁ VE SPORU S CHÁPÁNÍM PŘILOŽENÝM ANI PROGRAMOVACÍM JAZYKŮ.

OPERACE NAD JAZYKY

JAZYKY CHÁPEME JAKO MNOŽINY \Rightarrow
LŽE NAD NIMI DEFINOVAT STEJNÉ OPERACE
JAKO NAD MNOŽINAMI.

NECHť $L_1 \subseteq \Sigma^*$ A $L_2 \subseteq \Sigma^*$ JSOU
JAZYKY NAD Σ .

SJEDNOCENÍ JAZYKŮ $L = L_1 \cup L_2$

$$L = \{w \mid w \in \Sigma^* \wedge (w \in L_1 \vee w \in L_2)\}$$

PRŮNIK JAZYKŮ $L = L_1 \cap L_2$

$$L = \{w \mid w \in \Sigma^* \wedge (w \in L_1 \wedge w \in L_2)\}$$

DOPLNĚK JAZYKA $L = \overline{L_1}$

$$L = \{w \mid w \in \Sigma^* \wedge w \notin L_1\}$$

ROZDÍL JAZYKŮ $L = L_1 / L_2$

$$L = \{w \mid w \in \Sigma^* \wedge (w \in L_1 \wedge w \notin L_2)\}$$

ZŘETĚŽENÍ JAZYKŮ $L = L_1 L_2$

$$L = \{w \mid w \in \Sigma^* \wedge (w = uv \wedge u \in L_1 \wedge v \in L_2)\}$$

ZÁKLADNÍ ÚLOHA TEORIE JAZYKŮ

ŘETĚZEC

JAK ZJISTIT, ŽDA ~~JAZYK~~ PATŘÍ A NEPATŘÍ
DO DANÉHO JAZYKA?

[ÚLOHA SYNTAKTICKÉ ANALÝZY]

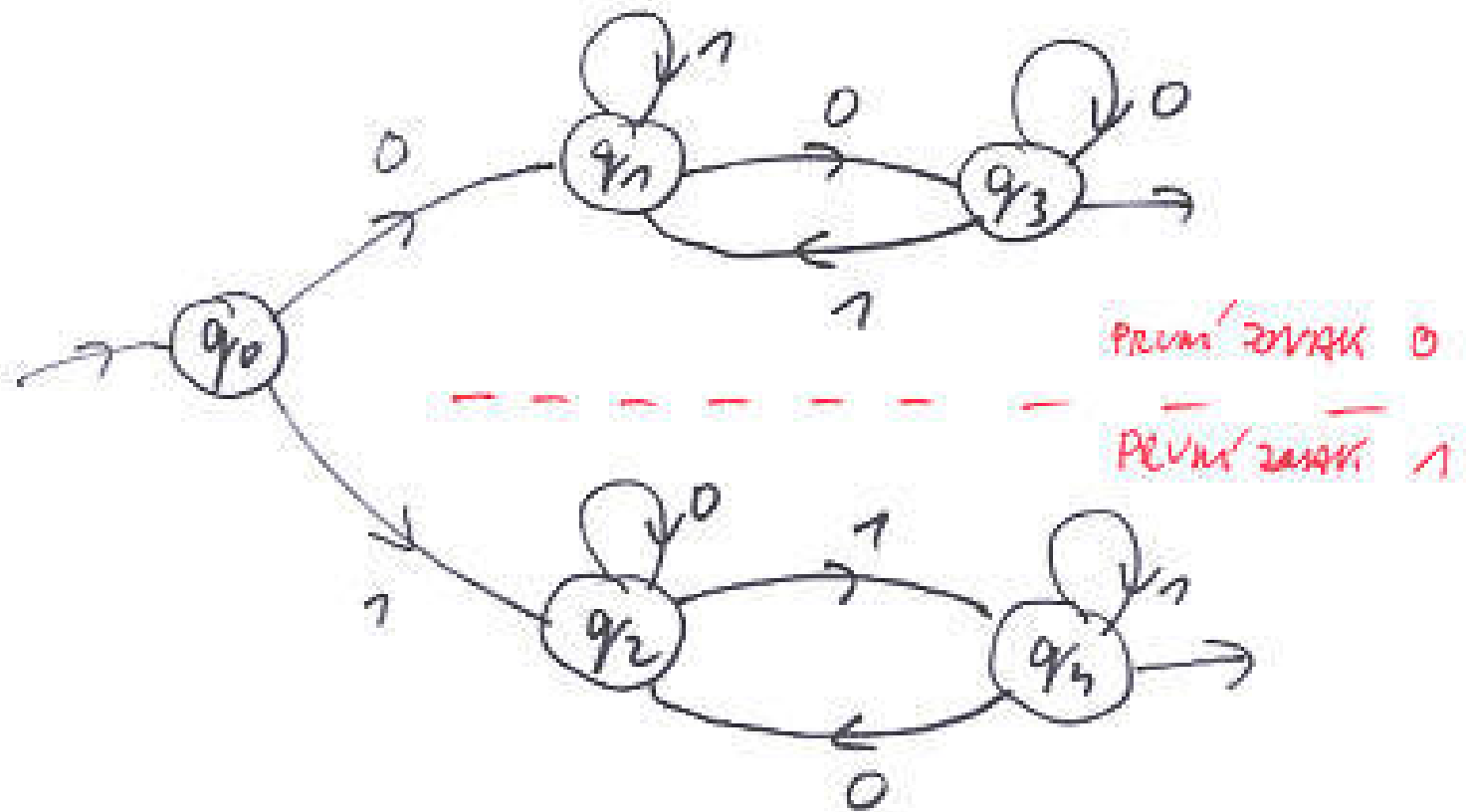
PŘI ŘEŠENÍ TÉTO ÚLOHY SE DÁTO LÉČÍ 1. PŘEKLAD
ŘETĚZCE" (Z PROGRAMOVACÍHO JAZYKA DO
KÓDU POČÍTAČE NEBO Z JEDNOHO PŘÍROZUMNÉHO
JAZYKA DO DRUHÉHO)

[SEMANTICKÁ ANALÝZA, GENEROVÁNÍ]

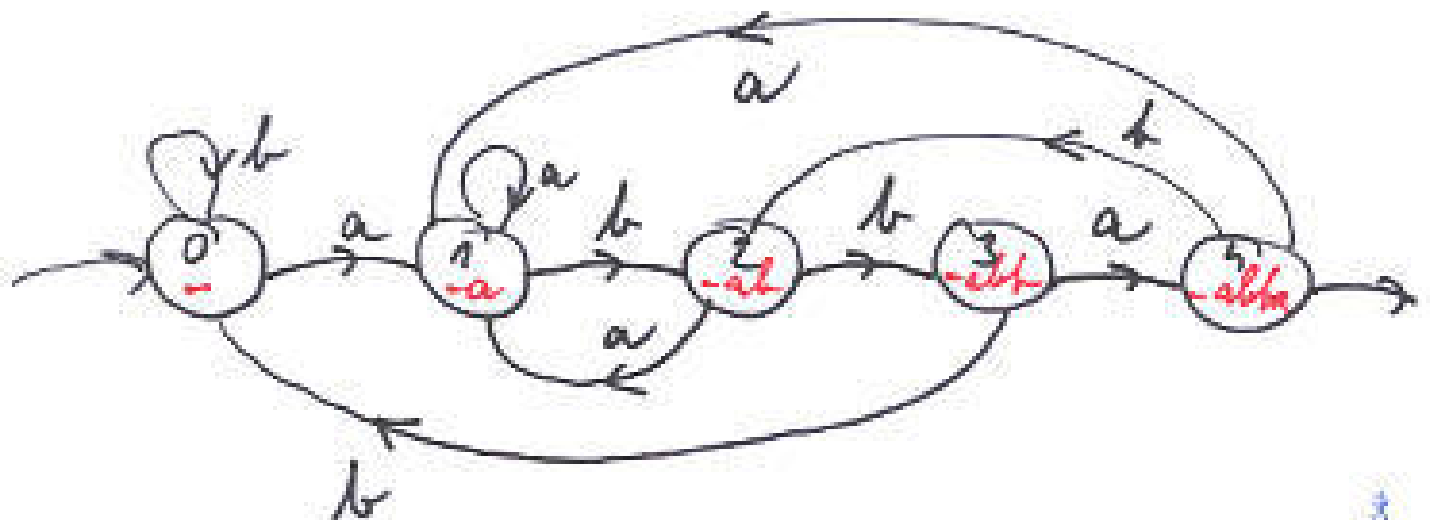
U PŘÍROZUMNÝCH JAZYKŮ - SLOŽITÉ, ALGO-
RITICKY TĚŽKO ŘEŠITELNÉ.

U FORMALNÍCH JAZYKŮ - ZA VLŠTÝCH
PŘEDPOKLADŮ LZE SYNTAKTICKOU ANALÝZU
ŘEŠIT ROZPOZNAVACÍH KONEČNÝM AUTOMATEM.

Př: $L_1 = \{w \mid w \in \{0,1\}^* \wedge w \text{ začíná a končí stejným symbolem} \wedge |w| \geq 2\}$



Př: $L_2 = \{w \mid w \in \{a,b\}^* \wedge w \text{ končí řetězcem abba}\}$



Ř: $L_3 = \{ w \mid w = 0^m 1^n \wedge m \geq 1 \}$

(POPIS SPOLEČNOU VLASTNOSTÍ)

$L_3 = \{ 01, 0011, 000111, 00001111, \dots \}$

(VÝČET)

AUTOMAT SI MUSÍ „PROSTŘEDNICTVÍM SVYCH STAVŮ“ PAMATOVAT POČET 0 NA ZAČÁTKU ŘETĚZCE.

KOLIK MUSÍ MÍT AUTOMAT STAVŮ?

PŘÍMEJENŤÍM TOLIK, KOLIK NUL MŮŽE BÝT NA ZAČÁTKU ŘETĚZCE (TJ m).

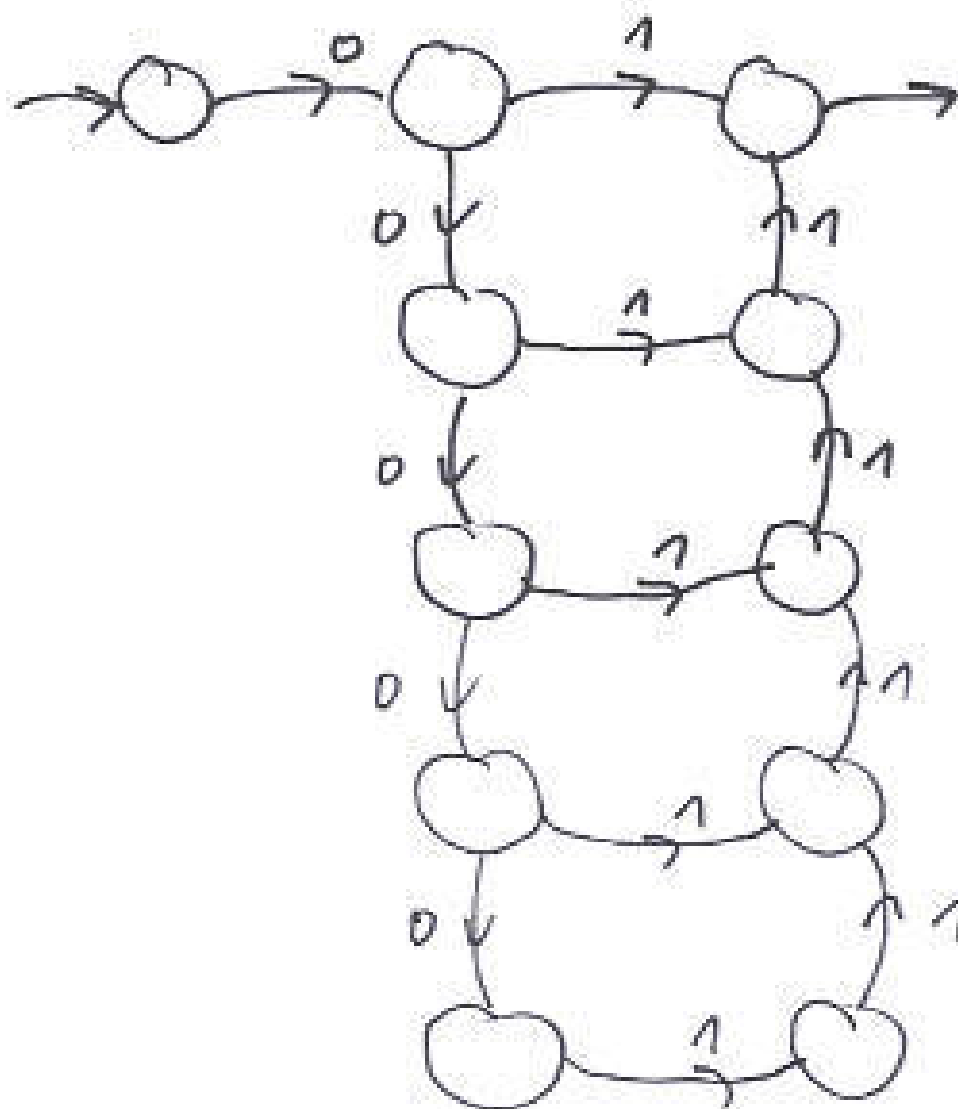
m NEMÍ ŠKORA OMEZENÁ \Rightarrow

\Rightarrow POČET STAVŮ NEMÍ OMEZEN \Rightarrow

KONZPOR, NA ŘEŠENÍ PROBLÉMU NESTRČIT

KONEČNÝ AUTOMAT.

KONEČNÝH AUTOMATŮ LZE ÚLOHU ŘEŠIT
POUZE ZA PŘEDPOKLADU, ŽE BUDE m SMĚRA
ODĚŘENO, NAPŘ. $1 \leq m \leq 5$:



ZÁVĚR : JSOU JAZYKY, KTERÉ LZE ROZ-
POZNAT KA (TJ. PO ZPRACOVÁNÍ PŘETĚŽE
AUTOMAT SKONČÍ V KONEČNÉM STAVU),
JSOU I JAZYKY nerozpoznatelné KA.
CO O TOM ROZHODUJEME ?

žatím zde používá uvedeně slovní
popis jazyka. Je nutné specifikovat
jazyk přesněji (specifikace větmi nebo
společnou vlastností množin).

PRECIZNĚJŠÍ PŘÍSTUP:

POPIS JAZYKA — AKCEPTAČNÍ (AUTOMAT,
KTERÝ JAZYK ROZPO-
ZNÁVA. **NEMUSÍ BYT**
KONEČNÝ).

— GENERATIVNÍ (GRAMATIKA,
TJ. PRAVIDLY PRO
VYTVAŘENÍ ŘETĚZCŮ.)

K AKCEPTAČNÍMU POPISU:

KAŽDÝ ROZPOZNÁVAČÍ KONEČNÝ AUTOMAT
JE DVOZNAČNĚ DEFINUJE JAZYK (MNOŽINA
VĚTÍ ŘETĚZCŮ, KTERÉ PŘEVEDOU AUTOMAT Z
POČÁTEČNÍHO STAVU DO NĚKTERÉHO Z KONEČNÝCH)

KE GENERATIVNÍMU POPISU:

ZÁKLADNÍ MYŠLENKA: PODOBŮ FORMALNÍCH
PRAVIDEL POPISAT „SPRAVNÉ ŘETĚZCE“

(ANALOGIE GRAMATICKÝCH PRAVIDEL V PŘI-
LOŽENÝCH JAZYCÍCH)

PŘ: GRAMATIKA PRO GENEROVÁNÍ JEDNODU-
HÝCH ANGLICKÝCH VĚT A DĚŘE-
NOU MNOŽINOU ANGLICKÝCH SLOV.

PŘÍKLADY ŘETĚZŮ, KTERÉ LZE ODVODIT

JOHN DRINKS TEA

MARY DRINKS GIN

OUR JOHN BUYS GIN

CELKEM LZE VYTVOŘIT 3.2.2.3 ŘETĚZŮ.

JAK SE ŘETĚZVE ODVODZÍ?

VYMAŽÍ SE OD POČÁTEČNÍHO PODOL-
NÉHO SYMBOLU $\langle \text{VĚTA} \rangle$. POSTUPNĚ
MAHAZUJEME PODOLNÉ SYMBOLY JEJICH
ROZVOJI DEFINOVANÝMI ODVODZOVACÍMI
PRAVIDLY.

ODVODZOVÁNÍ KONČÍ TEHDY, KDYŽ UŽ
V ODVODZENÉ ŘETĚZI JSOU POUZE
ANGLICKÁ SLOVA A NEMÍTAN ŽÁDNÝ
PODOLNÝ SYMBOL – ŮJEKY MÁŽEV VĚTNÉ-
HO ŮLEMU.

PRAVIDLA PRO ODVOZOVÁNÍ ŘETĚZCŮ:

<VĚTA> → <PODH. ČÁST> <PŘÍS. ČÁST> ①

<PODH. ČÁST> → <PODDĚT> ②

<PODH. ČÁST> → <PŘÍMASTEK> <PODDĚT> ③

<PŘÍS. ČÁST> → <PŘÍSUDK> ④

<PŘÍS. ČÁST> → <PŘÍSUDK> <PŘEDĚT> ⑤

<PODDĚT> → JOHN ⑥

<PODDĚT> → MARY ⑦

<PŘÍMASTEK> → OUR ⑧

<PŘÍMASTEK> → YOUR ⑨

<PŘÍSUDK> → DRINKS ⑩

<PŘÍSUDK> → BUYS ⑪

<PŘEDĚT> → TEA ⑫

<PŘEDĚT> → GIN ⑬

PŘÍKLADY ODVOZOVÁNÍ:

$\langle \text{VĚTA} \rangle \xRightarrow{(1)} \langle \text{PODM. ČÁST} \rangle \langle \text{PŘÍS. ČÁST} \rangle \xRightarrow{(5)}$
 $\Rightarrow \langle \text{PŘÍVĚSTEK} \rangle \langle \text{PODMĚT} \rangle \langle \text{PŘÍS. ČÁST} \rangle \xRightarrow{(8)} \text{OUR} \langle \text{PODMĚT} \rangle \langle \text{PŘÍS. ČÁST} \rangle \xRightarrow{(6)}$
 $\Rightarrow \text{OUR JOHN} \langle \text{PŘÍS. ČÁST} \rangle \xRightarrow{(5)} \text{OUR JOHN} \langle \text{PŘÍVĚSTEK} \rangle \langle \text{PODMĚT} \rangle \xRightarrow{(10)}$
 $\Rightarrow \text{OUR JOHN DRINKS} \xRightarrow{(13)} \text{OUR JOHN DRINKS GIN}$
 $\quad \quad \quad \underbrace{\hspace{10em}}_{\langle \text{PODMĚT} \rangle}$

SHRNUTÍ: POMOCNÉ SYMBOLY (ČESKÉ NÁZVY VĚTVYCH ČLENŮ)
POČÍTELMÍ SYMBOLY ($\langle \text{VĚTA} \rangle$)
CÍLOVÉ SYMBOLY (ANGlická slova)
ODVOZOVACÍ PRAVIDLA

DEFINICE GRAMMATIKY

GRAMMATIKA G JE UVEDENÁ JAKO

$$G = (N, T, S, P)$$

N ... MNOŽINA NETERMINÁLNÍCH SYMBOŮ

T ... MNOŽINA TERMINÁLNÍCH SYMBOŮ

$S \in N$... POČÁTEČNÍ SYMBOLE

P ... MNOŽINA PŘEPISOVACÍCH PRAVIDEL

PŘ: PROGRAMOVÝ JAZYK PASCAL JE
POPSÁN GRAMMATIKOU VE FORMĚ SYNTAKTICKÝCH
DIAGRAMŮ (WIRTH) [TO JSOU PŘEPISOVACÍ-
PRAVIDLA]. POČÁTEČNÍ SYMBOLE - SLOVO
PROGRAM