Postův korespondenční problém

- Postův sys. nad Σ je $S = \langle (\alpha_1, \beta_1), (\alpha_2, \beta_2) \ldots (\alpha_k, \beta_k) \rangle$, $k \geq 1$, $\forall\, 1 \leq i \leq k : \alpha_i, \beta_i \in \Sigma^+$.
- Řešení S je m-tice $(m \geq 1)$ $I = \langle i_1, i_2, \ldots, i_m \rangle$, kde $\forall\, 1 \leq j \leq m : 1 \leq i_j \leq k$, taková, že $\alpha_{i_1} \alpha_{i_2} \cdots \alpha_{i_m} = \beta_{i_1} \beta_{i_2} \cdots \beta_{i_m}$.
- PCP: Má daný PS řešení?

- Uvažte následující Postovy systémy. Naleznete řešení, nebo ukažte, že neexistuje.

a) $S_1 = \langle (ab, a), (aa, baaab), (aa, a) \rangle$. $\Sigma = \{a, b\}$

b) $S_2 = \langle (ab, a), (aaa, a), (ba, b) \rangle$ $\Sigma = \{a, b\}$

---

ad a) Má řešení, a to např. $I_1 = \langle 1, 2, 1, 3 \rangle$, konečně:

$$\alpha_1 \alpha_2 \alpha_1 \alpha_3 = ab \cdot aa \cdot ab \cdot aa$$

$$\beta_1 \beta_2 \beta_1 \beta_3 = a \cdot baaab \cdot a \cdot a$$

ad b) Nemá řešení, koneckně $\forall\, 1 \le i \le 3 : |\alpha_i| > |\beta_i|$, tedy libovolná konkatenace řetězců α_i bude delší než odpovídající konkatenace řetězců β_i.

- Dokažte, že problém víceznačnosti bezkontextových gramatik je nerozhodnutelný, a to redukcí založenou na PCP.

Důkaz redukcí z PCP:

- PCP lze charakterizovat jazykem PCP = $\{\langle S \rangle \mid S$ je PS, který má řešení$\}$. $\langle\cdot\rangle$ je vhodné kódování PS např. nad abecedou $\Sigma_1 = \{0, 1, \text{"}{,}\text{"}, \#\}$, kde pomocí "0" budeme unárně kódovat symboly ze Σ, "1" použijeme k oddělení symbolů v rámci α_i, β_i; "," použijeme k oddělení α_i a β_i a "#" použijeme k oddělení dvojic (α_i, β_i).
- Problém jednoznačnosti bezk. gr. lze charakterizovat jazykem VG = $\{\langle G \rangle \mid G$ je bezk. gr., která je

víceznačná. Zde $\langle\cdot\rangle$ je kódování bezk. gr.

např. nad $\Sigma_2 = \{0, 1, 2, \to, \#\}$, zde

- 0 : unární kódování terminálů
- 1 : unární kódování neterminálů ("1" = S)
- 2 : oddělovač terminálů a neterminálů
- $\to$: oddělovač levé a pravé strany pravidel
- # : oddělovač pravidel.

Ilustrace redukce na příkladu:

K PS $S_1 = \langle (\underset{\alpha_1}{ab}, \underset{\beta_1}{a}), (\underset{\alpha_2}{aa}, \underset{\beta_2}{baaab}), (\underset{\alpha_3}{aa}, \underset{\beta_3}{a}) \rangle$

přiřadí redukce G s pravidly

$$S \to S_1 \mid S_2$$

$$S_1 \to \alpha_1 S_1 1 \mid \alpha_2 S_1 2 \mid \alpha_3 S_1 3 \mid \#$$

$$= ab S_1 1 \mid aa S_1 2 \mid aa S_1 3 \mid \#$$

$$S_2 \to a S_2 1 \mid baaab S_2 2 \mid a S_2 3 \mid \#.$$

Pak např. pro $I = \langle 1, 2, 1, 3 \rangle$:

$$S \Rightarrow S_1 \Rightarrow ab S_1 1 \Rightarrow abaa S_1 21 \Rightarrow abaaab S_1 121 \Rightarrow$$

$\Rightarrow abaaabaaS_1 3121 \Rightarrow abaaabaa\#3121$

$S \Rightarrow S_2 \Rightarrow aS_2 1 \Rightarrow abaaabS_2 21 \Rightarrow abaaabaS_2 121 \Rightarrow$
$\Rightarrow abaaabaaS_2 3121 \Rightarrow abaaabaa\#3121$

↑ tato věta má 2 derivace $\Rightarrow$ gr. je víceznačná

b

- Navrhneme redukci $\sigma : \Sigma_1^* \rightarrow \Sigma_2^*$, která redukuje PCP na VG.
- TS M_σ implementující uvedenou redukci každému $x \in \Sigma_1^*$ přiřadí $\langle G_x \rangle$, kde G_x je gramatika, kterou M_σ sestaví následovně:

  1. M_σ ověří, zda x je platný kód PCP. Pokud ne $G_x = (\{S\}, \{a\}, \{S \rightarrow a\}, S)$.
  2. Pro x, který kóduje PS $S = \langle (\alpha_1, \beta_1), \ldots, (\alpha_k, \beta_k) \rangle$ pro nějaké $k \geq 1$, M_σ zkonstruuje gramatiku G_x následovně:
     a) $N = \{S, S_1, S_2\}$
     b) Σ bude stejná jako u PS rozšířená o dodání symbolů $\{\#, 1, \ldots, k\}$ (předp. že $\#$ nepatří do abecedy S) – M_σ pouze změní kód terminálů ze sekvencí 0 oddělených 1 na sekvence 1 oddělených 2.

c) Množina přepisovacích pravidel zahrne následující pravidla (a žádná jiná):

- $S \to S_1 \mid S_2$
- $\forall\, 1 \le i \le k: \; S_1 \to \alpha_i S_1 i \mid \#$
- $\forall\, 1 \le i \le k: \; S_2 \to \beta_i S_2 i \mid \#$

d) S je start. symbol.

- Snadno se nahlédne, že M_G lze opravdu realizovat jako úplný TS.
- Také se snadno nahlédne, že G_x je víceznačná tehdy a jen tehdy, když PS odpovídající řádu x má řešení.

Diagonalizace

- Ukažte, že množina $\mathbb{R}$ není spočetná.

(Důsledek: Nelze např. ke každému reálnému číslu přiřadit TS, který po spuštění běžel do nekonečna a vypisoval na pásku delší a delší rozvoj daného čísla.)..

důkaz diagonalizací:

- Postačí ukázat, že je nespočetná množina M reálných čísel z intervalu (0,1) mající právě jeden dekadický zápis (tj. neuvažujeme čísla jako např. $0{,}2000^{\omega} = 0{,}1999^{\omega}$)
- Důkaz vedeme sporem: tedy předpokládejme, že M je spočetná mn-a.
- Pak existuje bijekce $f: N \longleftrightarrow M$.
- S využitím f lze sestavit nekonečnou matici, jejíž jednotlivé řádky i budou odpovídat číslu f(i) z M. Sloupce budou pak odpovídat jednotlivým desetinným místům. Dostaneme tedy matici:

		1.	2.	3.	....
1. řádek	f(0)	d_{11}	d_{12}	d_{13}	...
2. řádek	f(1)	d_{21}	d_{22}	d_{23}	....
⋮	f(2)				
	⋮				

– Uvažme nyní číslo $0, e_1 e_2 e_3 \ldots$ takové, že

$\forall i \geq 1: \quad e_i = \begin{cases} 1 & \text{jestliže } d_{ii} \neq 1 \\ 2 & \text{jestliže } d_{ii} = 1 \end{cases}$

– Uvedené číslo evidentně patří do M a musí se nacházet na některém z řádků. To ale není možné, neboť u každého řádku i dojde k neshodě mezi e_i a d_{ii}. SPOR. $\square$

Uveďte základní ideu důkazu následujícího tvrzení:

$\forall L : L \in \mathcal{L}_1 \Rightarrow L \in REC$ (tedy je-li L kontextový jazyk, pak je také rekurzivní).

<u>Idea důkazu:</u>

- Protože $L \in \mathcal{L}_1$, existuje LOA M takový, že $L = L(M)$.
- LOA je zvláštní případ NTS, který neopustí tu část pásky, na níž je zapsán vstup.
- M lze převést na úplný LOA, protože délka vstupu je konečná

a tudíž M může provést pouze omezený počet kroků, aniž by cyklil.

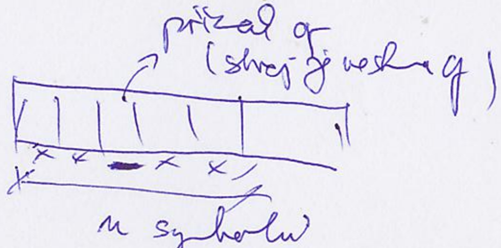
- Tento počet kroků lze shora omezit hodnotou:

$$(|\Gamma| \cdot (1+|Q|))^n,$$

kde n je délka vstupu.

Konečně na každém vstupním poli může být $|\Gamma|$ různých symbolů a u každého si můžeme pamatovat, zda na něm je hlava a pokud ano, jaký je stav řízení.

- Úplný LOA M' vytvořený z M může pracovat tak, že se na počátku vypočte hodnotu $(|\Gamma|(1+|Q|))^n$, zapíše si ji pomocí dalších příznaků na pásku a následně se chovat tak, že simuluje M, ale v každém kroce dekrementuje spočtené číslo. Pokud dojde k 0, odmítne.



$\in \Gamma$

$\rightarrow |\Gamma| \cdot |\Gamma| \cdots |\Gamma|$ různých obsahů $= |\Gamma|^n$

celkový počet konfig.:

$$(|\Gamma|(1+|Q|))^n$$

- Uvedené číslo $(|\Gamma|(1+|Q|))^n$ lze zapsat na n číslicích v $(|\Gamma|(1+|Q|))$-kové soustavě.

- Úplný LOA je úplným NTS. Ten lze algoritmicky převést na úplný DTS – postačí užít klasický alg. pro převod NTS na DTS.

$\square$

- Mějme gramatiku $G = (\{S\}, \{a,b\}, \{S \to aSb \mid \varepsilon\}, S)$ a $L = \{a^n b^n \mid n \geq 0\}$. Dokažte, že $L(G) = L$.

Důkaz:

1. Ukážeme, že $L(G) \subseteq L$. Indukcí nad délkou slova ukážeme, že $\forall w \in L(G): w \in L$.
   - Slova z $L(G)$ mají evidentně sudou délku (neboť každý přepis generuje právě 2 symboly, případně 0 symbolů), vystačí tudíž indukce přes sudá čísla.
   - pro délku $i = 0$: Jediné slovo délky 0 je ε. Toto slovo patří do $L(G)$, konečně $S \Rightarrow \varepsilon$. Ovšem $\varepsilon \in L$, protože $\varepsilon = a^0 b^0$.

- Předpokládejme, že tvrzení platí pro všechny $w \in L(G)$, kde $|w| \leq i$ pro nějaké sudé $i \geq 0$. Ukážeme, že platí i pro $w \in L(G)$, kde $|w| = i+2$.
- Protože $|w| \geq 2$, je zřejmé, že w bylo získáno derivací, která obsahuje alespoň 1 uplatnění pravidla $S \to aSb$. Tedy lze psát, že

  $$S \Rightarrow aSb \overset{*}{\Rightarrow} w = aw'b$$

  řetězec $w = aw'b$, protože žádné terminály v dané gr. není možno mazat.

  Tedy $S \overset{*}{\Rightarrow} w'$ a tedy $w' \in L(G)$, přičemž $|w'| = i$.
- Na w' lze uplatnit indukční předpoklad a tedy $w' \in L$, tedy $w' = a^m b^m$ pro $m \geq 0$.
- $w = aw'b = aa^m b^m b = a^{m+1} b^{m+1} \in L$.

2. Ukážeme, že $L \subseteq L(G)$. Indukcí nad délkou slova i ukážeme, že $\forall w \in L: w \in L(G)$.
   - Slova z L mají evidentně sudou délku

(konečné pro $w \in L$: $|w| = |a^n b^n| = 2n$ pro $n \geq 0$).
Postačí tedy opět indukce přes sudá čísla.

- Bázový případ – $i = 0$: $\varepsilon \in L$ ($a^0 b^0 = \varepsilon \in L$).
Ovšem $S \underset{G}{\Rightarrow} \varepsilon$ a tedy $\varepsilon \in L(G)$.

- Předpokládejme, že tvrzení platí pro $\forall w \in L$, kde $|w| \leq i$ pro nějaké sudé i. Ukážeme, že pak platí také $\forall w \in L$, kde $|w| = i+2$.

- Je-li $w \in L$ a má délku $i+2$, pak evidentně
$w = a^n b^n$ pro $n \geq 1$, neboli
$w = a\, a^{n-1} b^{n-1}\, b$.

- Ovšem $|a^{n-1} b^{n-1}| = i$ a současně $a^{n-1} b^{n-1} \in L$.
- z indukčního předpokladu platí, že $w \in L(G)$ a tedy $S \overset{*}{\underset{G}{\Rightarrow}} a^{n-1} b^{n-1}$.
- V G lze jistě derivovat $S \underset{G}{\Rightarrow} aSb$ pomocí $S \to aSb$.
- Tedy $S \underset{G}{\Rightarrow} aSb \overset{*}{\Rightarrow} a a^{n-1} b^{n-1} b = a^n b^n$, tedy

$a^n b^n \in L(G)$. $\square$

- Uveďte alespoň základní ideu důkazu uzavřenosti RE vůči substituci RE.

Idea důkazu:

- Mějme RE jazyk L nad Σ. Mějme dále RE jazyky L_a pro každé $a \in \Sigma$.
- Ukážeme, že substitucí L_a do L vznikne jazyk L', který je také RE.
- Vzhledem k výše uvedenému můžeme předpokládat, že existují TS M a M_a (pro každé $a \in \Sigma$) takové, že $L(M) = L$ a $L(M_a) = L_a$ (pro $\forall a \in \Sigma$).
- Sestavíme TS M' takový, že $L(M') = L'$.
- M' lze sestavit jako NTS vícepáskový TS, který pracuje následovně:

- M' nedeterministicky rozdělí svůj vstup w na sekvenci řetězců $w_1, w_2, \dots, w_n$ ($n \geq 0$). Tedy $w = w_1 \cdot w_2 \cdots w_n$.
- M' projde sekvenci $w_1, w_2, \dots, w_n$ a pro každé $w_i \in \Sigma^*$ ($1 \leq i \leq n$) zapíše na pomocnou pásku symbol $a_i \in \Sigma$ zvolený nedeterministicky.
- Následně M' projde sekvenci a_i pro $1 \leq i \leq n$. Pro každé i přepíše w_i na další pomocnou pásku. Na ní odsimuluje běh TS M_{a_i}. Pokud M_{a_i} odmítne (cyklí), M' odmítne (cyklí).
- M' ověří, zda $a_1 \dots a_n \in L$, a to tak, že na pásce $\triangle a_1 a_2 \dots a_n$ odsimuluje běh TS M. Pokud M přijme, M' přijme, jinak odmítne (cyklí).

$\square$