Algebraická teorie regulárních jazyků

Ondřej Klíma

Ústav matematiky a statistiky
Masarykova Universita

MELA 2012

- Motivace: charakterizace důležitých přirozených tříd jazyků
- Příklady:
  - Star-free jazyky — jazyky dané regulárním výrazem, který nepoužívá hvězdičku, ale můze použít komplement.
  - Lineární temporální logika – jazyky popsané formulí LTL.
- Metoda: charakterizace pomocí vlastností algebraických struktur. (Primárně syntaktického monoidu).

Osnova

- Základy algebraické teorie jazyků – Eilenbergova korespondence
- Základy teorie konečných pologrup
- Pokroky z posledních let

Jak monoid rozpoznává jazyk?

Konečný monoid $(M, \cdot, 1)$ chápeme jako stroj, který určuje, zda slovo patří do jazyka.

Jazyky jsou podmnožiny A^* resp. A^+.

Tomu odpovídájí dvě souběžné teorie, kde se používají monoidy resp. pologrupy.

Úzké vazby. Někdy technického charakteru.

Dnes dáme přednost monoidům.

Jak monoid rozpoznává jazyk?

Konečný monoid $(M, \cdot, 1)$ chápeme jako stroj, který určuje zda slovo patří do jazyka.

Definice

Konečný monoid M rozpoznává jazyk $L \subseteq A^*$, jestliže existuje homomorfismus $\varphi : A^* \to M$ a podmnožina $F \subseteq M$ taková, že $\varphi^{-1}(F) = L$.

A^* volný monoid, tj. φ popsán obrazy písmen.

Máme zadán “stroj” (M, φ, F), kde M dán (např.) multiplikativní tabulkou, φ dáno obrazy písmen.

Vstup (slovo) $u = a_1 a_2 \dots a_n$ vyhodnotíme takto: určíme výsledek součinu $m = \varphi(u) = \varphi(a_1) \cdot \varphi(a_2) \cdot \dots \cdot \varphi(a_n)$ v M a máme $u \in L \iff m \in F$.

Od monoidu k DFA

Pokud máme (M, φ, F) můžeme uvažovat DFA $\mathcal{A} = (Q, A, \delta, i, T)$, kde

- $Q = M$,
- $i = 1$,
- pro $m \in M$ a $a \in A$ kladme $\delta(m, a) = m \cdot \varphi(a)$,
- $T = F$.

Tento automat rozpoznává stejný jazyk jako (M, φ, F).

Závěr: každý jazyk rozpoznávaný konečným monoidem je regulární.

Od DFA k monoidu

Pokud máme DFA $\mathcal{A} = (Q, A, \delta, i, T)$, pak každé písmeno $a \in A$ zadává transformaci τ_a množiny Q do sebe:

$$\tau_a : Q \to Q, \quad \tau_a(q) = \delta(q, a) \ .$$

Tyto transformace lze skládat, tj. obecně máme pro slovo $u \in A^*$

$$\tau_u : Q \to Q, \quad \tau_u(q) = \delta(q, u) \ .$$

Přičemž pro $u, v \in A^*$ platí $\tau_u \cdot \tau_v = \tau_{uv}$.
Zobrazení $\varphi : A^* \to \mathcal{T}(Q)$, kde $\varphi(u) = \tau_u$ je homomorfismus z A^* do monoidu transformací $(\mathcal{T}(Q), \cdot, \mathrm{id})$.
Slovo u je z jazyka L právě tehdy, když $\tau_u(i) \in T$.
Položme proto $F = \{\tau \in \mathcal{T}(Q) \mid \tau(i) \in T\}$.
Vidíme, že $(\mathcal{T}(Q), \varphi, F)$ rozpoznává stejný jazyk jako $\mathcal{A}$.

Závěr: Konečné monoidy rozpoznávají právě regulární jazyky.

“Minimální” monoid

Buď dán regulární jazyk L a uvažujme libovolný homomorfismus $\varphi : A^* \to M$, který ho rozpoznává.
Jednoduchá vlastnost jádra homomorfismu: pokud pro $u, v \in A^*$ platí $(u, v) \in \ker \varphi$, tj. $\varphi(u) = \varphi(v)$, pak u a v jsou “zaměnitelné”.
Přesněji: pokud $s, t \in A^*$ takové, že $sut \in L$, pak (vzhledem k $\varphi(sut) = \varphi(svt)$) platí $svt \in L$.
Dvojice (s, t) s vlastností $sut \in L$ se nazývá kontext slova u (v jazyce L). Pro $u \in A^*$ definujeme $C_L(u)$ jako množinu všech kontextů u v L, tj. $C_L(u) = \{(s, t) \in A^* \times A^* \mid sut \in L\}$.
Vlastnost: $\varphi(u) = \varphi(v)$ pak $C_L(u) = C_L(v)$.

"Minimální" monoid

Definice

Pro regulární jazyk $L \subseteq A^*$ definujeme na A^* relaci $\sim_L$ takto:

$$u \sim_L v \iff C_L(u) = C_L(v)\,.$$

- Vlastnost: $\ker \varphi \subseteq \sim_L$.
- Zřejmě: $\sim_L$ je relace ekvivalence.
- Lehce se ověří: $\sim_L$ je kongruence monoidu A^*.
  $(u \sim_L v \implies uw \sim_L vw, wu \sim_L wv)$
- Zřejmě: $u \in L \iff (\lambda, \lambda) \in C_L(u)$.

Věta

Pro regulární jazyk $L \subseteq A^*$ je $A^*/\sim_L$ monoid, který rozpoznává L a je mezi nimi minimální velikosti.

Syntaktický monoid

Terminologie: $\sim_L$ syntaktická kongruence,
$M_L = A^*/\sim_L$ syntaktický monoid.

Lemma

Syntaktický monoid M_L je izomorfní transformačnímu monoidu minimálního automatu jazyka L.

Důkaz: Chceme ukázat $\tau_u = \tau_v \iff u \sim_L v$.
"$\Longrightarrow$" Platí dle předchozího.
"$\Longleftarrow$" Pokud $\tau_u \neq \tau_v$ pak existuje stav p tak, že $\delta(p,u) \neq \delta(p,v)$. Z minimality automatu:
i) p je dosažitelný, tj. existuje $s \in A^*$ tak, že $\delta(i,s) = p$;
ii) stavy $q_1 = \delta(p,u) \neq \delta(p,v) = q_2$ lze rozlišit, tj. existuje $t \in A^*$ tak, že $\delta(q_1,t) \in T$ a $\delta(q_2,t) \notin T$ (nebo naopak).
Celkem $sut \in L$ a $svt \notin L$, tzn. $(s,t) \in C_L(u)$ a $(s,t) \notin C_L(v)$.
Tedy $u \not\sim_L v$. $\square$

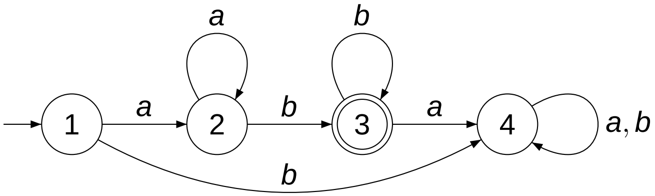
Příklad

Příklad

Nechť $A = \{a, b\}$ a $L = a^+b^+ = \{a^m b^n \mid m, n > 0\}$.

$C_L(ba) = \emptyset$ a máme $[ba] = A^*bA^*aA^*$;
$C_L(ab) = \{(a^i, b^j) \mid i, j \geq 0\} = a^* \times b^*$ a máme $[ab] = L$;
$C_L(a) = a^* \times b^+$ a máme $[a] = a^+ = \{a^i \mid i > 0\}$;
$C_L(b) = a^+ \times b^*$ a máme $[b] = b^+ = \{b^j \mid j > 0\}$;
$C_L(\lambda) = a^+ \times b^+ \cup a^+b^+ \times b^* \cup a^* \times a^+b^+$, tj. $[\lambda] = \{\lambda\}$.

$M_L = A^* / \sim_L = \{[\lambda], [a], [b], [ab], [ba]\}$, $F = \{[ab]\}$.



Piecewise Testable Languages

Definice

Regulární jazyk $L \subseteq A^*$ se nazývá piecewise testable, jestliže je konečnou Booleovskou kombinací jazyků tvaru

$$A^* a_1 A^* a_2 A^* \dots A^* a_\ell A^*, \quad \text{kde} \quad a_1, \dots, a_\ell \in A,\ \ell \geq 0\,.$$

Příklad

$L = a^+ b^+ = A^* a A^* b A^* \cap (A^* b A^* a A^*)^c$.

Základní otázka:
pro $L \subseteq A^*$ rozhodnout, zda L je “piecewise testable” jazyk.

Charakterizace “piecewise testable” jazyků — Simon

Theorem (Simon – 1972)

Regulární jazyk $L \subseteq A^*$ je “piecewise testable” právě tehdy, když jeho syntaktický monoid je $\mathcal{J}$-triviální.

Přímá implikace . . . jednoduchá.

Opačná implikace . . . náročnější . . . více důkazů:

- Simon (1972) . . . kombinatorika na slovech
- Stern (1985) . . . kombinatorika na (posloupnostech) slov
- Straubing and Thérien (1988) . . . uspořádané monoidy
- Almeida (1990) . . . profinitní metody
- Higgins (1997) . . . transformační monoidy
- Klíma (2009) . . . kombinatorika na slovech, elementární důkaz (4s) – Discrete Mathematics 2011
- Klíma, Polák (2011) . . . následující přednáška

$\mathcal{J}$-triviální monoidy

Greenovy relace $\mathcal{R}, \mathcal{L}, \mathcal{J}$ jsou základním nástrojem při studiu struktury pologrup. Zajímáme se o relaci “dělitelnosti”. Nejsme v komutativním monoidu:
$a \mid b \iff b = xa$ (resp. $b = ax$, resp. $b = xay$).

Definice

Na monoidu M uvažujeme relaci $\leq_{\mathcal{J}}$ definovanou takto:

$$a \leq_{\mathcal{J}} b \quad \text{právě tehdy, když} \quad (\exists x, y \in M)\ a = xby\ .$$

Potom $a \mathcal{J} b$ právě tehdy, když $a \leq_{\mathcal{J}} b$ a zároveň $b \leq_{\mathcal{J}} a$.
Monoid M se nazývá $\mathcal{J}$-triviální, jestliže z $a \mathcal{J} b$ plyne $a = b$.

Tedy M je $\mathcal{J}$-triviální právě tehdy, když $\leq_{\mathcal{J}}$ je uspořádání.

Charakterizace “piecewise testable” jazyků

Pro daný jazyk spočítáme syntaktický monoid a pro něj napočítáme relaci $\mathcal{J}$ a rozhodneme, zda je triviální.

Příklad

$L = a^+b^+$, $M_L = \{[\lambda], [a], [b], [ab], [ba]\}$,
$[ba] <_{\mathcal{J}} [ab] <_{\mathcal{J}} [a], [b] <_{\mathcal{J}} [\lambda]$, $[a]$ a $[b]$ nesrovnatelné.
Tzn. M_L je $\mathcal{J}$-triviální monoid.

Jde i efektivněji. Existují podmínky na minimální automat jazyka ekvivalentní s $\mathcal{J}$-trivialitou syntaktického monoidu. (Simon(1975), Stern(1985) - $O(n^5)$, Trahtman(2000) - $O(n^2)$)

Důležitost abecedy

Příklad

Je $L = A^*abA^*$ “piecewise testable” ?
Ano, pokud $A = \{a, b\}$. Potom $L = A^*aA^*bA^*$.
Pokud ovšem máme $c \in A$ další písmeno, pak

$$cac \sim_L c \quad \text{a} \quad ca \not\sim_L c \ ,$$

tzn. $[cac] \leq_{\mathcal{J}} [ca] \leq_{\mathcal{J}} [c] = [cac]$, tj. $[ca]\mathcal{J}[c]$.
V tomto případě není L “piecewise testable”.

Star-free jazyky

Příklad

Nechť $A = \{a, b\}$. Je jazyk $L = (ab)^+ = ab(ab)^*$ star-free?
Ano. $A^* = \emptyset^c$,
$L = (ab)^+ = aA^* \cap A^*b \cap (A^*aaA^*)^c \cap (A^*bbA^*)^c$.

Theorem (Schützenberger – 1966)

Regulární jazyk $L \subseteq A^*$ je “star-free” právě tehdy, když jeho syntaktický monoid je aperiodický.

Monoid M je aperiodický, jestliže obsahuje pouze triviální (tj. jednoprvkové) podgrupy.
Pozn: neutrální prvek nemusí být 1_M, ale libovolný idempotent.
Ekvivalentně: M je $\mathcal{H}$-triviální. (Zde $\mathcal{H} = \mathcal{R} \cap \mathcal{L}$.)

Star-free jazyky – poznámky

- Složitost $O(|M_L|^2)$.
- Pokud je vstup minimální automat, tak problém je *PSPACE*-úplný.
- Stejná vlastnost pro LTL. Jazyk je popsatelný formulí LTL právě tehdy, když jeho syntaktický monoid je aperiodický.

Třídy jazyků vs. třídy monoidů

Zkoumané třídy jazyků jsou často uzavřené na mnohé operace, např. Booleovské operace, homomorfní obrazy nebo vzory, kvocienty.

Uzávěrové operátory na třídách algeber (monoidů, pologrup) jsou založeny na základních algebraických konstrukcích: podalgebra, součin algeber, homomorfní obraz.

Tyto operace spolu úzce souvisí. Např. když máme dva automaty pro jazyky K a L, a chceme vyrobit automat pro jazyk $K \cap L$ nebo $K \cup L$, tak uvažujeme součin daných automatů. (Pro monoidy je to stejné.)

Variety jazyků

Definice

Varieta jazyků $\mathcal{V}$ přiřazuje každé konečné abecedě A množinu $\mathcal{V}(A)$ regulárních jazyků nad A tak, že

- $\mathcal{V}(A)$ je uzvařena na konečná sjednocení a průniky, a na komplementy ($\emptyset, A^* \in \mathcal{V}(A)$);
- $\mathcal{V}(A)$ je uzvařená na kvocienty (derivace), tj. $L \in \mathcal{V}(A)$, $u, v \in A^*$ implikuje $u^{-1}Lv^{-1} = \{ w \in A^* \mid uwv \in L \} \in \mathcal{V}(A)$;
- $\mathcal{V}$ je uzavřena na vzory homomorfismů, tj. $f : B^* \to A^*$, $L \in \mathcal{V}(A)$ implikuje $f^{-1}(L) = \{ v \in B^* \mid f(v) \in L \} \in \mathcal{V}(B)$.

Pseudovariety monoidů

Definice

Pseudovarieta monoidů je třída konečných monoidů uzavřená na podmonoidy, homomorfní obrazy a konečné součiny.

Eilenbergova korespondence: Variety jazyků jsou ve vzájemně jednoznačné korespondenci s pseudovarietami monoidů.

Eilenbergova korespondence

Theorem (Eilenberg – 1976)

Nechť pro každou pseudovarietu monoidů $\mathbf{V}$ *značí* $\alpha(\mathbf{V})$ *varietu regulárních jazyků definovanou předpisem*

$$\alpha(\mathbf{V})(A) = \{L \subseteq A^* \mid M_L \in \mathbf{V}\} .$$

Nechť pro každou varietu regulárních jazyků $\mathcal{L}$ *značí* $\beta(\mathcal{L})$ *pseudovarietu monoidů generovanou syntaktickými monoidy* M_L*, kde* L *je regulární jazyk nad nějakou konečnou abecedou* A *splňující* $L \in \mathcal{L}(A)$*.*
Potom α *a* β *jsou vzájemně inverzní izomorfismy mezi svazem všech pseudovariet konečných monoidů a svazem všech variet regulárních jazyků.*

Modifikace a zobecnění

Ne všechny studované třídy regulárních jazyků jsou uzavřené na všechny operace.

- (Pin 1995) Pozitivní variety regulárních jazyků – nepožadujeme komplement.
  Tyto variety korespondují s pseudovarietami uspořádaných monoidů.
  Syntaktický monoid je implicitně uspořádaný:
  $u \preceq_L v \iff C_L(v) \subseteq C_L(u)$.
  (Někteří davají přednost opačnému uspořádání.)
- (Polák 1999) Konjunktivní (disjunktivní) variety jazyků – nepožadujeme komplement ani sjednocení (průnik).
  Nová syntaktická struktura: idempotentní semiring.

Modifikace a zobecnění

- (Straubing 2002) $\mathcal{C}$-variety jazyků – vzory jen v některých homomorfismech (těch z (kategorie) $\mathcal{C}$), např. homomorfismy zachovávající délku slov, nebo injektivní homomorfismy, apod.
  Syntaktická struktura: homomorfismus na syntaktický monoid.
- Lze uvažovat $\mathcal{C}$-pozitivní variety, $\mathcal{D}$-konjunktivní variety.
- (Gehrke, Grigorieff, Pin 2008) Svazy regulárních jazyků – pouze sjednocení a průniky.

Jak charakterizovat pseudovariety

Namísto variet jazyků máme pseudovarietu monoidů **V** a chceme umět rozhodovat, zda zadaný monoid do ní patří.
Chceme nějaký efektivní popis (charakterizaci).
Z univerzální algebry víme, že variety algeber (navíc libovolné součiny) jsou charakterizovatelné množinami identit.

Příklad

Boolovské kombinace jazyků tvaru A^*aA^* odpovídají pseudovarietě všech polosvazů: $\mathbf{Sl} = \mathit{Mod}(xy \approx yx, xx \approx x)$.

Identity nestačí

Identity mohou posloužit, ale nestačí.

Příklad

Uvažujme třídu všech konečných aperiodických monoidů **A**.
Zde neplatí žádná identita.
(Pro danou identitu $u \approx v$ ($u \neq v$), lze vytvořit konečný nilpotentní monoid $\langle A \mid w = 0 \iff |w| \geq |uv| \rangle$, který identitu $u \approx v$ nesplňuje.)

Potřebujeme silnější nástroj než jsou identity.

Zobecnění identit

Co je vlastně identita?
Dvojice slov (obecně v jiných algebrách dvojice termů).
Při kontrole, zda $M \models u \approx v$, uvažujeme všechna možná dosazení do termů (slov) a porovnáváme výslednou hodnotu v M pro obě strany.
Tzn. pro slovo (term) u nad množinou promněných $X = \{x_1, \dots, x_n\}$ uvažujeme $u^M : M^n \to M$, $u^M(a_1, a_2, \dots, a_n) = \alpha(u)$, kde $\alpha : X^* \to M$ je dané tak, že $\alpha(x_i) = a_i$.

Přičemž $M \models u \approx v$ znamená $u^M = v^M$.

Zobecnění identit

Z univerzální algebry víme, že jsou to termové operace (tj. zobrazení která “komutují” s homomorfismy):
je-li $\gamma : M \to N$, pak platí

$$\gamma(u^M(a_1, \ldots, a_n)) = u^N(\gamma(a_1), \ldots, \gamma(a_n)) \, .$$

Implicitní operace

Máme i jiné operace než ty dané slovy (termy)? ANO

Definice

Implicitní n-arní operací π na třídě všech monoidů $\mathbf{M}$ je systém $(\pi^M)_{M \in \mathbf{M}}$ takový, že:
i) pro libovolný monoid M je $\pi^M : M^n \to M$;
ii) pro libovolný homomorfismus $\gamma : M \to N$ platí
$\gamma(\pi^M(a_1, \dots, a_n)) = \pi^N(\gamma(a_1), \dots \gamma(a_n))$.

Definice

Pseudoidentita je dvojice implicitních n-árních operací a platí $M \models \pi \approx \rho$ právě tehdy, když $\pi^M = \rho^M$.

Lze nahlédnout, že pokud $\gamma : M \to N$ je surjektivní homomorfismus a $M \models \pi \approx \rho$, pak $N \models \pi \approx \rho$.

Příklad implicitní operace

Příklad

Pro libovolný prvek a monoidu M, který má m prvků, lze dokázat, že podpologrupa $\{a^k \mid k > 0\}$ generovaná prvkem a obsahuje právě jeden idempotent.
Ten lze popsat jako $a^{m!}$.
Označíme-li $a^\omega = a^{m!}$ pak lze vidět, že x^ω je unární implicitní operace.

Příklad

$\mathbf{A} = Mod(x^\omega x \approx x^\omega)$.
$\mathbf{J} = Mod(x^\omega x \approx x^\omega, (xy)^\omega \approx (yx)^\omega)$.

Konstrukce implicitních operací – naivně

Lze ukázat, že pro libovolný konečný monoid M a implicitní operaci π existuje slovo u, tak, že $\pi^M = u^M$.
Pro každý monoid tak máme jakousi aproximaci $\overline{\pi_M} \in X^*$ implicitní operace π. Potom π je tedy možne zadat systémem $(\overline{\pi_M})_{M \in \mathbf{M}}$. Pokud místo všech monoidů vezmeme jen po jedné izomorfní kopii, můžeme dokonce uvažovat posloupnost slov.
Tato posloupnost slov bude konvergovat k π.
Tj. každá implicitní operace je takto dána nějakou posloupností.

Konstrukce implicitních operací – formálně

Na množině všech implicitních operací nad množinou proměnných X, kterou značíme $\overline{X}$, uvažujeme metriku, která vyjadřuje, jak velký konečný monoid potřebujeme, abychom odlišili danou dvojici implicitních operací.
Pro $\pi, \rho \in \overline{X}$ definujeme $r(\pi, \rho) = \min\{|M| \mid \pi^M \neq \rho^M\}$ a následně $d(\pi, \rho) = 2^{-r(\pi,\rho)}$.

Věta

- *d je ultrametrika na $\overline{X}$.*
- *$(\overline{X}, d)$ je úlpný kompaktní metrický prostor.*
- *Každá implicitní operace je limitou konvergentní posloupnosti termových operací. (X^* je hustá v $\overline{X}$.)*

Konstrukce implicitních operací – poznámky

Příklad

$x^\omega = \lim_{m\to\infty} x^{m!}$

Poznámky:

- Problém: $\overline{X}$ je nespočetná množina (již pro $|X| = 1$).
- $\overline{X}$ lze konstruovat z X^* stejně jako $\mathbb{R}$ z $\mathbb{Q}$. (Tzn. metrika d se bere jen na X^*.)
- Topologický přístup: konečné monoidy jsou implicitně vybaveny diskrétní topologií. Takže se uvažují monoidy s topologií (kompaktní, spojité násobení, reziduálně konečné).

Reitermanova věta

Theorem (Reiterman – 1982)

Třída konečných monoidů $\mathbf{V}$ *je pseudovarieta právě tehdy, když existuje množina pseudoidentit* Σ *taková, že* $\mathbf{V} = Mod(\Sigma)$.

Taková ekvacionální charakterizace je možná i v případě modifikací.

Konstrukce implicitních operací lze dělat pro libovolnou pseudovarietu $\mathbf{V} \subseteq \mathbf{M}$. Tedy implicitní operace v pseudovarietě $\mathbf{V}$ je $(\pi^M)_{M \in \mathbf{V}}$. Pak máme $\psi_{\mathbf{V}} : \overline{X} \to \overline{X}_{\mathbf{V}}$ a $\Sigma = \ker \psi_{\mathbf{V}}$.

Almeida pro pseudovarietu odpovídající varietě piecewise testable jazyků dokázal, že implicitní operace jsou tvaru $u_0 v_1^\omega u_1 \dots v_k^\omega u_k$, kde $u_0, v_1, \dots \in X^*$, takové, že ....

Polynomiální uzávěr

Brzozowski, Cohen (1971) definovali dot-depth hierarchii star-free jazyků (A^+). Monoidové verzi se říká Straubingova-Thérienova hierarchie star-free jazyků. Definují se pomocí polynomiálního uzávěru tříd jazyků.

Pro množinu jazyků $\mathcal{K}$ nad abecedou A označujeme $\mathrm{PPol}\mathcal{K}$ množinu všech pozitivních kombinací jazyků tvaru

$$L_0 a_1 L_1 a_2 \dots a_\ell L_\ell\,, \quad \text{kde} \quad a_i \in A,\, L_i \in \mathcal{K}\,.$$

Podobně $\mathrm{BPol}\mathcal{K}$ je množina všech Boolovských kombinací stejných jazyků.
(Pozitivní) polynomiální uzávěr $\mathrm{PPol}\mathcal{V}$, resp. $\mathrm{BPol}\mathcal{V}$, variety $\mathcal{V}$ se pak definuje takto: pro abecedu A klademe $\mathrm{PPol}\mathcal{V}(A) = \mathrm{PPol}(\mathcal{V}(A))$, resp. $\mathrm{BPol}\mathcal{V}(A) = \mathrm{BPol}(\mathcal{V}(A))$.

Straubingova-Thérienova hierarchie

Položme $\mathcal{V}_0 = \mathcal{T}$ triviální varietu: $\mathcal{T}(A) = \{\emptyset, A^*\}$.
Definujme nyní pro libovolné $n \in \mathbb{N}_0$ variety:

$$\mathcal{V}_{n+1/2} = \text{PPol}\mathcal{V}_n, \quad \mathcal{V}_{n+1} = \text{BPol}\mathcal{V}_n\,.$$

Za $\mathcal{V}_0$ lze vzít něco jiného ... dot-depth hierarchie, grupová hierarchie.
Mluvíme o “concatenation hierarchies”.

Jazyky z $\bigcup_{n\in\mathbb{N}} \mathcal{V}_n$ jsou právě star-free jazyky.
$\mathcal{V}_n \neq \mathcal{V}_{n+1/2} \neq \mathcal{V}_{n+1}$ (Brzozowski a Knast (1978) pro dot-depth).
Jazyky v $\mathcal{V}_1$ jsou právě piecewise testable jazyky.

Souvislost s logikou

Věta (McNaughton, Papert – 1971)

Star-free jazyky jsou přesně ty jazyky, které lze definovat v logice prvního řádu lineárního uspořádání. $(FO[<])$

Věta (Thomas – 1982)

Pro libovolné n *jsou jazyky z* $\mathcal{V}_n$ *právě ty, jež lze definovat formulemi, kde je nejvýše* n *alternací kvantifikátorů.*

Straubingova-Thérienova hierarchie - co víme

Pro odpovídající pseudovariety víme:

- $\mathbf{V}_{1/2} = Mod(x \preceq 1)$
- $\mathbf{V}_1 = \mathbf{J} = Mod(x^\omega x \approx x^\omega, (xy)^\omega \approx (yx)^\omega)$
- $\mathbf{V}_{3/2} = Mod(x^\omega y x^\omega \preceq x^\omega | \mathbf{c}(x) = \mathbf{c}(y))$

Lemma (Pin, Straubing – 1981)

Pro libovolnou abecedu A je $\mathcal{V}_{3/2}(A)$ množina všech jazyků tvaru

$$B_0^* a_1 B_1^* a_2 \dots a_\ell B_\ell^* \,, \qquad \textit{kde} \qquad a_i \in A,\, B_i \subseteq A\,.$$

Všechny podmínky lze efektivně kontrolovat.
(Jde je přetlumočit do řeči automatů. Pak polynomiální algoritmus vzhledem k velikosti vstupu (automatu). Pouze v případě $3/2$ hraje (exponenciální) roli velikost abecedy.)
$\mathbf{V}_2$ (a všechny ostatní) – otevřený problém (40 let).

Malcevův součin

Věta (Pin, Weil – 1995)

Pro libovolné n platí

$$\mathbf{V}_{n+1/2} = Mod(x^\omega y x^\omega \preceq x^\omega) \, ⓜ \, \mathbf{V}_n \ .$$

Platí obecně pro libovolnou pseudovarietu **V**.
Jde najít bázi “pseudoidentit”. Potom řešitelnost $\mathbf{V}_n$ dává řešitelnost $\mathbf{V}_{n+1/2}$.

Věta (Straubing – 1979)

Nejmenší pseudovarieta obsahující **V** *uzavřená na polynomiální uzávěr je tvaru* **A** ⓜ **V***.*

“Straubing-Thérien” versus “dot-depth”

Věta (Straubing – 1985)

Pro libovolné n platí $\mathbf{B}_n = \mathbf{V}_n * \mathbf{LI}$.

Důsledek:

Věta (Straubing – 1985)

Pro libovolné n je $\mathbf{B}_n$ řešitelná právě tehdy, když je rešitelná $\mathbf{V}_n$.

Level 2

Věta (Pin, Straubing – 1981)

$\mathbf{V}_2 = \mathbf{PJ}$.

Bohužel není efektivní.
Straubing (1986) uvažoval pseudovarietu **CJ** danou pseudoidentitami

$$(x^\omega p y^\omega q x^\omega)^\omega x^\omega p y^\omega s x^\omega (x^\omega r y^\omega s x^\omega)^\omega \approx (x^\omega p y^\omega q x^\omega)^\omega (x^\omega r y^\omega s x^\omega)^\omega ,$$

kde $x, y, p, q, r, s \in \overline{X}$ mají stejný obsah.

Speciální případ obecného případu (Pin, Weil 1995) pseudovariet $\mathbf{B}_1 ⓜ \mathbf{V}$, které mají stejnou bázi identit, kde $x, \ldots, s$ jsou stejné ve $\mathbf{V}$.
Tj. $\mathbf{CJ} = \mathbf{B}_1 ⓜ \mathbf{Sl}$. Pseudovarieta **CJ** je řešitelná (Straubing).

Level 2 – Straubingova hypotéza

Hypotéza (Pin, Straubing, Weil – 1986–2002?)

$\mathbf{V}_2 = \mathbf{CJ}$.

Věta (Straubing – 1986)

Pseudovariety $\mathbf{V}_2$ a $\mathbf{CJ}$ obsahují stejné monoidy generované dvojicí generátorů. Dále platí $\mathbf{V}_2 \subseteq \mathbf{CJ}$.

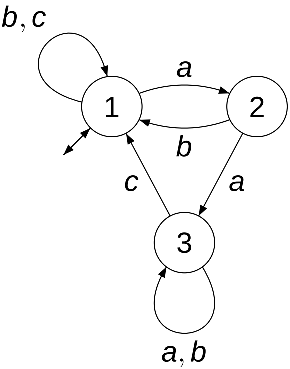
Tzn. pro jazyky nad dvoupísmenou abecedou umíme rozhodnout, zda jsou z variety $\mathcal{V}_2$.

Věta (Cowen – 1993)

Pseudovariety $\mathbf{V}_2$ a $\mathbf{CJ}$ obsahují stejné inverzní monoidy.

Straubingova hypotéza – protipříklad

Označme M transformační monoid (neúplného) automatu:



Věta (Almeida, Klíma – 2009)

$M \in \mathbf{CJ}$ a $M \notin \mathbf{V}_2$.

Straubingova hypotéza – protipříklad

^{*1}

a	$^*ab$
$^*ba$	$^*b$

$^*a^2$	$^*bca$	$^*bc$
$^*aca^2$	$^*aca$	ac
$^*baca^2$	$baca$	bac
$^*ca^2$	ca	$^*c$

^{*0}

Důkaz

Tvrzení $M \in \mathbf{CJ}$ se dokáže snadno z popisu $\mathbf{CJ} = \mathbf{B}_1 \, ⓜ \, \mathbf{Sl}$.

Pro důkaz $M \notin \mathbf{V}_2$ je klíčové najít nějakou novou pseudoidentitu, která platí v pseudovarietě $\mathbf{V}_2$ (a neplatí v M).

Lemma

Nechť π, ρ jsou implicitní operace takové, že $\mathbf{V}_{3/2} \models \pi \preceq \rho$. Potom $\mathbf{V}_2 \models \pi^\omega \approx \pi^\omega \rho \pi^\omega$.

Např. pro $\pi = z[(xy)^\omega x]^2 z$, $\rho = z(xy)^\omega xz$ platí.

Nový horní odhad - Almeida, Klíma – 2010

$\mathbf{F} = Mod(\{\pi^\omega \approx \pi^\omega \rho \pi^\omega \mid \mathbf{V}_{3/2} \models \pi \preceq \rho\})$.

- $\mathbf{V}_2 \subseteq \mathbf{F}$.
- Existuje monoid $N \in \mathbf{F}$ tak, že $N \notin \mathbf{CJ}$.
  Tudíž $\mathbf{V}_2 \neq \mathbf{F}$ a $\mathbf{F} \not\subseteq \mathbf{CJ}$
- $\mathbf{F}$ je řešitelná pseudovarieta (netriviální důkaz).
- Závěr: Nový horní odhad $\mathbf{CJ} \cap \mathbf{F}$ pro pseudovarietu $\mathbf{V}_2$, který je řešitelný.

Studované problémy

Některé studované otázky v teorii pseudovariet pologrup:

- Univerzální algebra – problém identit.
  Zde problém pseudoidentit, spíše problém ω-identit.
- Vlastnosti implicitních operací vzhledem k faktorizaci: např. pro slova $u, v, s, t \in X^*$ taková, že $uv = st$ existuje slovo w tak, že $u = sw$ a $t = wv$ nebo existuje slovo w tak, že $s = uw$ a $v = wt$. Platí to pro implicitní operace? (Pro danou pseudovarietu.)
- Jsou ω-termy uzavřené na faktory?
  (Opět závisí na pseudovarietě.)

Studované problémy

- Algebraické vlastnosti pseudovariet.
  (Margolis, Sapir, Weil – 1998, resp. Steinberg, Rhodes – 2004) Většina pseudovariet tvaru $\overline{\mathbf{H}}$ (třída všech konečných pologrup, jejichž podgrupy jsou z dané pseudovariety grup $\mathbf{H}$) jsou join- (resp. malcev-, resp. semidirect-) ireducibilní.

  Almeida, Klíma – 2011 (zasláno) – všechny.

Reference — základní studijnprameny

Survey o algebraické teorii jazyků (rozsáhlého seznamu referencí):

- J.-E. Pin: Syntactic semigroups, Chapter 10 in *Handbook of Formal Languages*, G. Rozenberg and A. Salomaa (eds.), Springer, 1997.

Knihy o algebraické teorii jazyků a teorii konečných pologrup:

- J.-E. Pin: *Varieties of Formal Languages*. North Oxford, London and Plenum, New-York, (1986).
- J. Almeida: *Finite Semigroups and Universal Algebra*, World Scientific, 1994.
- J. Rhodes and B. Steinberg: *The q-theory of finite semigroups*, Springer Monographs in Mathematics (2009)

Reference — Odkazy

Citované práce v přednášce publikované po roce 1997 (tj. ty co nejsou podchyceny v přehledovém článku J.E. Pina).

- J.-E. Pin, P. Weil: A conjecture on the concatenation product, Theor. Inform. Appl. 35 (2001), 597â618 (2002),
- A. N. Trahtman: Verifying piecewise testability of deterministic finite automaton. CIAA 2000.
- H. Straubing: On Logical Descriptions of Regular Languages. LATIN 2002: 528-538
- M. Kunc: Equational description of pseudovarieties of homomorphisms Theoretical Informatics and Applications 37, 243-254, 2003.
- L. Polák: On pseudovarieties of semiring homomorphisms, in Proc. Mathematical Foundations of Computer Science 2004, LNCS 3153, (2004), 635–647
- M. Gehrke, S. Grigorieff, J.-E. Pin: A topological approach to recognition, ICALP 2010, Part II, LNCS 6199, (2010), 151–162.
- S. Margolis, M. Sapir, and P. Weil, Irreducibility of certain pseudovarieties, Comm. Algebra 26 (1998), 779â792.
- J. Rhodes and B. Steinberg, Join irreducible pseudovarieties, group mapping, and KovÂ´acs-Newman semigroups, LATIN 2004: Theoretical informatics, LNCS 2976, (2004), 279–291.
  The q-theory of finite semigroups, Springer Monographs in Mathematics, Springer, 2009. 12. B. Steinberg, On pointlike sets and joins of pseudovarieties, Int. J. Algebra Comput.

Reference — Naše výsledky

- O. Klíma, Piecewise testable languages via combinatorics on words. *Discrete Mathematics* **311** (2011), 2124–2127.
- O. Klíma, L. Polák, On Biautomata. Non-Classical Models for Automata and Applications, NCMA 2011: 153-164, rozšířená verze přijata v Rairo-ITA
- J. Almeida and O. Klíma: A counterexample to a conjecture concerning concatenation hierarchies, Information Processing Letters, Vol 110 (2009) 4-7
- J.Almeida, O.Klíma: New decidable upper bound of the second level in the Straubing-ThÃ©rien concatenation hierarchy of star-free languages, Discrete Mathematics and Theoretical Computer Science, Vol 12 (2010), No 4 (Special issue - Authomatha 2009) 41-58
- J.Almeida, O.Klíma: On the Irreducibility of Pseudovarieties of Semigroups *zasláno*

Stranka: `www.math.muni.cz/∼klima/publications.html`