Uniwersytet Warszawski
Wydział Matematyki, Informatyki i Mechaniki

Wojciech Moczydłowski, Jr
Nr albumu: 174746

Jednoregułowe systemy przepisywania słów

Praca magisterska
na kierunku INFORMATYKA

Praca wykonana pod kierunkiem
dr hab. Pawła Urzyczyna

Marzec 2002

Pracę przedkładam do oceny

Data Podpis autora pracy:

Praca jest gotowa do oceny przez recenzenta

Data Podpis kierującego pracą:

Streszczenie

W pracy pogłębiono wynik McNaughtona z [MacNau95] tyczący się terminacji jednoregułowych systemów przepisywania słów — dowiedziona jest rozstrzygalność terminacji dla pewnej klasy systemów. Ponadto wprowadzono nowe pojęcia przydatne w badaniach — jednowątkowość oraz automat ze stosem odpowiadający systemowi.

Słowa kluczowe

przepisywanie słów, systemy pół-Thue'owskie

Klasyfikacja tematyczna

F. Theory of computation
F.4 Mathemathical logic and formal languages
F.4.2 Grammars and other rewriting systems

Spis treści

Wprowadzenie

> The further removed from usefulness or practical application, the more important.
>
> *Axel Thue*

Historia systemów przepisywania słów sięga początku XX wieku. W roku 1910 Axel Thue w pracy [Thue10] sformułował definicję tego, co obecnie nazywamy systemem Thue'owskim i podjął próbę ataku na problem słów — który, jak okazało się w roku 1947, był nierozstrzygalny. Z systemami przepisywania słów związane są w naturalny sposób ciekawe problemy decyzyjne — wspomniany już problem słów, terminacji, silnej normalizacji, osiągalności, posiadania własności Churcha-Rossera, pętlenia się i innych. W ogólnym przypadku (dla systemu z dowolną ilością reguł) odpowiedzi są znane — na ogół każdy interesujący problem jest nierozstrzygalny. Nierozstrzygalność silnej normalizacji pokazano np. w [Her71]. Translację maszyn Turinga do pewnego systemu pół-Thue'owskiego, a zarazem dowód nierozstrzygalności problemu terminacji można znaleźć w [Davis58]. Większość istotnych wyników znajduje się w doskonałej monografii [BoOt93].

Prędzej czy później musiało nasunąć się pytanie o znaczenie ilości reguł w systemie — podobnie jak w przypadku maszyn Turinga, gdzie można pytać o najmniejszą ilość stanów i symboli alfabetu koniecznych do np. skonstruowania maszyny uniwersalnej. Na ten temat powstało kilka prac, np. [MatSen96].

Ponad 10 lat temu powstała pierwsza istotna praca na temat jednoregułowych systemów przepisywania słów — [Kur90]. W swojej pracy doktorskiej Kurth podaje kilka podstawowych kryteriów terminacji i opisuje oparte na nich "sito Kurtha" służące do odsiania "łatwych" systemów, które umożliwiło mu kompletną klasyfikację terminacji systemów, w których długość prawej strony jest mniejsza od 7. Prace powstałe później można podzielić z grubsza na dwie klasy. Prace z pierwszej klasy charakteryzują terminację systemu bezpośrednio na podstawie struktury jego jedynej reguły — tu można wymienić [ZanGes95], [Ges97] i [Ges98]. Druga klasa zaś obejmuje prace bardziej ogólne, zajmujące się ogólnymi własnościami — np. [Wra92], gdzie podana jest rozstrzygalność własności Churcha-Rossera, [MacNau95] czy [STKK97]. Podział ten jest oczywiście arbitralny i nieostry, lecz swą pracę zaliczyłbym do drugiej klasy. Jako referencję polecam pracę habilitacyjną [Ges01], w której zebrane są wszystkie istotne wyniki, acz podane czasami w mało strawnej formie.

W niniejszej pracy pogłębione nieco są wyniki z [MacNau95]. W sekcjach od 1. do 7. przedstawione są konieczne definicje, w sekcji 8. przedstawione jest nowe pojęcie służące do badania jednoregułowych systemów — jednowątkowość, które to pojęcie charakteryzujemy w sekcji 9. W następnych sekcjach podana jest konstrukcja automatu ze stosem dla systemu oraz sformułowane i udowodnione jest główne twierdzenie.

1. Abstrakcyjne systemy redukcyjne

Symbol ξ zawsze oznaczać będzie liczbę porządkową taką, że $1 \leq \xi \leq \omega$. Symbole n, m, l, i, j zawsze będą oznaczać liczby naturalne. Przez f^n oznaczamy n-krotną iterację funkcji f. Minimum zbioru pustego to $-\infty$. Podciąg $\{a_{n_k}\}$ ciągu $\{a_n\}$ jest *spójny*, jeśli ciąg $\{n_k\}$ składa się z kolejnych liczb naturalnych.

Definicja 1.1 *System redukcyjny* $\mathcal{A}$ jest to para $\langle A, \to \rangle$, gdzie $\to \subseteq A \times A$.

Stosujemy notację infiksową: $a \to b$ oznacza $\to (a, b)$. Ciąg $\{w_n\}_{n<\xi}$ taki, że:

$$\forall n \ (n > 0 \land n < \xi \Rightarrow w_{n-1} \to w_n)$$

nazywamy *derywacją*. Derywację symbolicznie zapisujemy tak:

$$w_0 \to w_1 \to \dots$$

O redukcji $w_{i-1} \to w_i$ mówimy, że jest to i-ta redukcja (krok redukcyjny) w derywacji $w_0 \to w_1 \to \dots$

Derywacja jest skończona, gdy $\xi < \omega$, w przeciwnym przypadku jest nieskończona. Piszemy $x \to^n y$ gdy istnieje derywacja długości n od x do y.

Definicja 1.2 $\to^*$ jest to domknięcie przechodnio-zwrotne relacji $\to$.

Definicja 1.3 System $\mathcal{A}$ ma własność Churcha-Rossera (jest konfluentny), jeśli:

$$\forall a, b, c \in A \ (a \to^* b \land a \to^* c \Rightarrow \exists d \in A \ (b \to^* d \land c \to^* d))$$

Definicja 1.4 System $\mathcal{A}$ ma własność silnej normalizacji (terminuje), jeśli nie istnieje homomorfizm $f : \langle \mathbf{N}, < \rangle \to \langle A, \to \rangle$.

Oczywiście system terminuje wtedy i tylko wtedy, gdy nie ma nieskończonej derywacji.

2. Słowa

Niech Σ będzie ustalonym skończonym alfabetem. Przez Σ^* oznaczamy zbiór wszystkich skończonych słów nad Σ. Przez ε oznaczamy słowo puste. Litery a, b i c z ewentualnymi indeksami zawsze oznaczać będą litery z Σ. Symbole x, y, z, ewentualnie z indeksami, zawsze oznaczać będą słowa z Σ^*. Przez $|x|$ oznaczamy długość słowa x. Piszemy $x \sqsubseteq y$, jeśli słowo x jest podsłowem słowa y, czyli gdy $y = x_1 x x_2$ dla pewnych słów x_1, x_2.

Definicja 2.1 $OVL(x, y) \equiv \{z \mid x = x_1 z \land y = z y_1\}$

3. Systemy przepisywania słów

Niech $S = \{(s_1, t_1), (s_2, t_2), \dots, (s_n, t_n) \mid (s_i, t_i) \in \Sigma^* \times \Sigma^*\}$. Zbiór S w naturalny sposób wyznacza relację na Σ^* zdefiniowaną następująco:

$$x \to_S y \equiv \exists i \ (x = x_1 s_i x_2 \land y = x_1 t_i x_2)$$

System redukcyjny $\langle S, \rightarrow_S \rangle$ zapisujemy tak:

$$\begin{array}{ccc} s_1 & \rightarrow & t_1 \\ s_2 & \rightarrow & t_2 \\ & \vdots & \\ s_n & \rightarrow & t_n \end{array}$$

Będziemy zajmować się szczególnym przypadkiem, gdy $|S| = 1$. Od tej pory symbole S, s i t są zarezerwowane dla jedynej reguły systemu $S = s \rightarrow t$. Gdy $s \sqsubseteq t$, to system ma trywialną nieskończoną derywację. Gdy $|t| \leq |s|$, to system terminuje. **Wykluczamy zatem takie systemy z rozważań**.

Definicja 3.1 System jest sof (ma własność sof[1]), jeśli $OVL(s, s) = \{\varepsilon\}$.

Definicja 3.2 Jeśli istnieją niepuste słowa y, z takie, że $s = yx$ i $t = xz$ ($s = xy$ i $t = zx$), to można powiedzieć, że słowo s nakłada się na słowo t. Wtedy słowo x nazywamy lewą (prawą) nakładką systemu[2]. Słowo y nazywamy lewą (prawą) s-frazą x, zaś słowo z lewą (prawą) t-frazą x. Inaczej mówiąc, lewa nakładka to po prostu element $OVL(s, t)$, jej lewa s-fraza to reszta słowa s, a jej lewa t-fraza to reszta słowa t.

Twierdzenie 1 *Jeśli system jest sof, to ma własność Churcha-Rossera.*

Dowód [Wra92]

Od tej pory rozważamy wyłącznie systemy o własności sof, poza przypadkami gdzie explicite będzie stwierdzone inaczej.

Definicja 3.3 System ma pętlę długości n, gdy $\exists x, y \ (x \rightarrow^n y \wedge x \sqsubseteq y)$.

Czasami stosować będziemy następującą konwencję notacyjną dla derywacji — kolejne słowa zapisujemy pod sobą, redukowany redeks s pogrubiamy, a powstałe słowo t pochylamy.

Przykład 1 Weźmy system $s = ab \rightarrow bbaa = t$. Oto przykładowa derywacja (z pętlą długości 2):

a**ab**
a*b*_baa_
_bbaa_baa

Definicja 3.4 Niech $w_0 \rightarrow w_1 \rightarrow \ldots$ będzie derywacją. Wtedy definiujemy

$$R(i) \equiv |x_1|$$

gdzie x_1 jest takie, że:

$$w_i = x_1 s x_2 \wedge w_{i+1} = x_1 t x_2$$

[1] Z angielskiego *self-overlap free*.
[2] Angielski odpowiednik to *overlap*.

4. Pozycje

Definicja 4.1 Niech słowo x ma n liter. Pozycje w słowie x to liczby naturalne z przedziału $[0, n]$. Słowo y jest w x na pozycji i, gdy:

$$x = a_1 a_2 \ldots a_i y x_1$$

W szczególności x jest w x na pozycji 0, a $R(i)$ jest pozycją redeksu w słowie w_i.

Jeśli słowo x jest podsłowem w y, to występuje w nim na jakiejś pozycji n. Piszemy $\overline{x}$ na oznaczenie pary (x, n), jeśli wiadomo o jakie słowo y chodzi. W takiej sytuacji będziemy pisali również $\overline{x} \sqsubseteq y$. Pozycje $\overline{x} = (x, n)$ to wszystkie pozycje od n do $n + |x|$.

Definicja 4.2 Jeśli $w_i = xzy$, to mówimy, że redukcja $w_i \to w_{i+1}$ nie dotyka $\overline{z}$, jeśli redukowane s znajduje się w x bądź w y.

Definicja 4.3 Gdy $w_{i-1} = xsy \to xty = w_i$, to mówimy że pozycje w słowie w_{i-1} wewnątrz $\overline{s}$ (czyli od $|x| + 1$ do $|x| + |s| - 1$) są niszczone. O pozycjach od $|x|$ do $|x| + |t|$ ze słowa w_i mówimy, że są tworzone. Zbiór pozycji utworzonych w trakcie redukcji $xsy \to xty$ zapisujemy jako $C(xsy \to xty)$.

Popatrzmy na słowo w_{i-1} w pewnej derywacji $w_0 \to w_1 \to \ldots$ i na pozycje w tym słowie, które nie są niszczone w redukcji $w_{i-1} \to w_i$. Każdej takiej pozycji p możemy w naturalny sposób przyporządkować pozycję p' w w_i: gdy p jest na lewo od redeksu, to $p' = p$, w przeciwnym przypadku $p' = p + |t| - |s|$. Oznaczmy to przyporządkowanie literą T_i i rozszerzmy je na wszystkie pozycje w redukowanym słowie przyporządkowując niszczonym pozycjom -1. Ściślej, niech $w_{i-1} = xsy \to xty = w_i$. Wtedy:

$$T_i(n) = \begin{cases} n & \text{jeśli } n \leq |x| \\ -1 & \text{jeśli } |x| < n < |x| + |s| \\ n + |t| - |s| & \text{w przeciwnym przypadku} \end{cases}$$

Definicja 4.4 Dla derywacji $w_0 \to w_1 \to \ldots$ oznaczamy $C_i \equiv C(w_{i-1} \to w_i)$ (pozycje powstałe w i-tym kroku derywacji).

Stwierdzenie 1 *Dla derywacji $w_0 \to w_1 \to \ldots$ zachodzi:*

$$C_i = \Big((\{0, 1, \ldots, |w_i|\}) \setminus (T_i^{\rightarrow}(\{0, 1, \ldots, |w_{i-1}|\})) \Big) \cup \{R(i), R(i) + |t|\}$$

Definicja 4.5 Dla derywacji $w_0 \to w_1 \to \ldots$ funkcję $Y_i : (P(\mathbf{N}), \mathbf{N}) \to P(\mathbf{N})$ definiujemy następująco:

$$\begin{aligned} Y_i(A, 0) &= A \\ Y_i(A, n) &= T_{i+n-1}^{\rightarrow}(Y_i(A, n-1)) \setminus \{-1\} \end{aligned}$$

Intuicyjnie, $Y_i(A, n)$ to zbiór pozycji ze zbioru A, które pozostają obecne w słowie w_{i+n} (oczywiście po przenumerowaniu). Zdefiniujmy jeszcze często używaną funkcję pomocniczą: $Z_i : \mathbf{N} \to \mathbf{N}$, $Z_i(n) = Y_i(C_i, n)$.

Definicja 4.6 Weźmy derywację $w_0 \to w_1 \to \ldots$ i słowo w_i oraz pozycję k z w_i. Zdefiniujmy funkcję $P_i : \mathbf{N} \to \mathbf{N}$ tak, że $P_i(k)$ jest to numer redukcji w której litera stojąca na tej pozycji (a dokładniej — pozycje $|k|$ i $|k+1|$) została utworzona, bądź $-\infty$ jeśli istnieje ona od początku derywacji. Ściślej:

$$P_i(k) \equiv \min\Big\{\ j \mid \exists l\ \Big(\{k, k+1\} \subset Z_j(l) \land j + l = i\Big)\ \Big\}$$

Liczbę $P_i(k)$ nazywamy datą urodzenia litery na pozycji k. Zwróćmy uwagę, że zachodzi:

$$\forall i\ (P_i(|w_i|) = -\infty)$$

5. I, WB, EI...

Definicja 5.1 Niech $w_0 \to w_1 \to \ldots$ będzie derywacją. Mówimy, że zachodzi $EI(i)$, gdy wszystkie pozycje utworzone w redukcji $w_{i-1} \to w_i$ są zniszczone w dalszym ciągu derywacji. Ściślej:

$$EI(i) \equiv \exists n\ (Z_i(n) = \emptyset)$$

Definicja 5.2 Jeśli zachodzi $EI(i)$, to mówimy, że pozycje utworzone w i-tej redukcji (czyli pochodzące ze zbioru C_i) są niszczone z lewej strony, gdy zachodzi warunek:

$$\exists n\ \Big((\min Z_i(n) \in [(R(i+n), R(i+n) + |s|]) \land (Z_i(n+1) \neq \emptyset)^3\Big),$$

natomiast mówimy, że są niszczone z prawej strony gdy zachodzi warunek:

$$\exists n\ \Big((\max Z_i(n) \in [(R(i+n), R(i+n) + |s|]) \land (Z_i(n+1) \neq \emptyset)\Big)$$

Zauważmy, że ponieważ $|s| < |t|$, to o ile zachodzi $EI(i)$, to zawsze pozycje z C_i będą niszczone z prawej bądź lewej strony.

Definicja 5.3 Jeśli dla żadnej derywacji $w_0 \to w_1 \to \ldots$ nie istnieje i takie, że $EI(i)$, to mówimy, że system jest WB (well-behaved).

Definicja 5.4 Relacja $I(i, j)$ zachodzi jeśli pozycja j w słowie w_i w derywacji $w_0 \to w_1 \to \ldots$ nie jest niszczona w żadnej redukcji po $w_{i-1} \to w_i$. O takiej pozycji mówimy, że jest *niezniszczalna*. Ściślej:

$$I(i, j) \equiv \forall n\ |Y_i(\{j\}, n)| > 0$$

Takie pozycje będziemy czasem otaczać na rysunkach dwoma prostokątami. Pozycje zaś, które ulegają zniszczeniu w trakcie derywacji będziemy oznaczać falkami.

Stwierdzenie 2 $\neg EI(i) \Rightarrow \exists j\ I(i, j)$

Definicja 5.5 Niech $w_0 \to w_1 \to \ldots$ będzie derywacją. Jeśli zachodzi $EI(i)$, to mamy 3 możliwości

- 2A Pozycje z C_i są niszczone i z lewej i z prawej strony
- 2B Pozycje z C_i są niszczone tylko z lewej strony
- 2C Pozycje z C_i są niszczone tylko z prawej strony

[3] Chodzi tu o to, aby nie wszystkie pozycje były niszczone w tej redukcji.

Definicja 5.6 System $s \to t$ spełnia warunek 2A (2B, 2C) jeśli istnieje derywacja $w_0 \to w_1 \to \ldots$ i liczba i taka, że zachodzi $EI(i)$ oraz odpowiednia klauzula z definicji 5.5.

Pisząc: zachodzi 2A, mamy na myśli, że nasz ustalony system spełnia warunek 2A.

Twierdzenie 2 *Problem: czy system jest WB jest rozstrzygalny. Jeśli system nie jest WB, to rozstrzygalne jest stwierdzenie, które z warunków 2A, 2B i 2C zachodzą. Jeśli system jest WB, to problem silnej normalizacji jest rozstrzygalny.*

Dowód [MacNau95]

6. Graf redukcji

Ważnym obiektem, związanym z każdą derywacją $w_0 \to w_1 \to \ldots$ jest jej graf.

Definicja 6.1 Graf derywacji $W = w_0 \to w_1 \to \ldots$ jest skierowany z krawędziami oznaczonymi 1 lub 0. Zbiór wierzchołków V jest to suma rozłączna zbiorów V_i, $i \geq 0$. Elementami zbioru V_i są pozycje w w_i. Krawędzie w grafie są wyłącznie pomiędzy zbiorami V_i i V_{i+1}. Weźmy wierzchołek $v_2 \in V_{i+1}$, $v_2 \geq 0$ odpowiadający pewnemu słowu $\overline{s} \sqsubseteq w_{i+1}$. Są 2 możliwości:

- Słowo $\overline{s}$ pozostaje niezmienione w trakcie redukcji $w_i \to w_{i+1}$. Niech j będzie jego pozycją w w_i. Odpowiada jej pewien wierzchołek $v_1 \in V_i$. Od v_1 do v_2 wiedzie krawędź oznaczona 0.

- Słowo $\overline{s}$ powstaje w trakcie redukcji $w_i \to w_{i+1}$. Niech $v_1 \in V_i$ będzie wierzchołkiem odpowiadającym $R(i)$. Od v_1 do v_2 wiedzie krawędź oznaczona 1. Krawędź tę nazywamy matką $\overline{s}$.

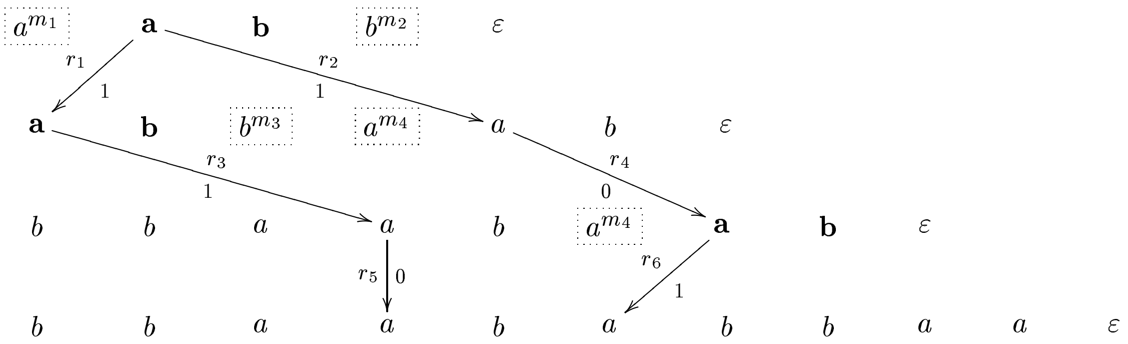
Jeśli k jest krawędzią w grafie redukcji $w_0 \to w_1 \to \ldots$ wiodącą od $v_1 \in V_i$ do $v_2 \in V_{i+1}$ oznaczoną liczbą n, to piszemy:

$$k = (v_1, v_2, i+1, n)$$

Krawędzie zatem utożsamiamy z elementami $\mathbf{N}^4$. Ta definicja daje nam trochę zbyt wiele ścieżek jak na nasze potrzeby. Niech M będzie zbiorem wszystkich nieskończonych ścieżek w grafie takich, że ścieżka L jest elementem zbioru M, jeśli wszystkie krawędzie na niej są oznaczone zerami. Niech N będzie pewną ścieżką ze zbioru M. Jeśli nie istnieje ścieżka N' z M różna od N zawierająca N, to usuwamy z grafu ścieżkę N. Będziemy czasem stosować notację z języków funkcyjnych — na przykład jeśli interesuje nas tylko liczba, która oznacza krawędź k, to będziemy pisać $k = (\_, \_, \_, n)$, aby nie wprowadzać dużej ilości zbędnych oznaczeń. Ścieżki w grafie będziemy oznaczać dużymi literami.

Przykład 2 Weźmy ponownie system $s = ab \to bbaa = t$ i trochę dłuższą derywację:

a**ab**b
a*b*baa*b
*bbaa*ba**ab**
bbaaba*bbaa*
$\vdots$



Rysunek 1.1: Graf derywacji

Graf tej derywacji przedstawiony jest na rysunku 1.1. Zakładamy, że redeksy w słowie *bbaababbaa* zostaną kiedyś zredukowane. Zamiast pozycji p piszemy gwoli większej czytelności literę znajdującą się na pozycji p. Symbole r z indeksami to nazwy odpowiednich krawędzi. Symbole m z indeksami to nazwy podsłów (jednoliterowych) przy których się znajdują. Będziemy się do nich potem odwoływali przy rozmaitych przykładach.

Stwierdzenie 3 *Graf redukcji jest skończonym lasem.*

Stwierdzenie 4 *W grafie redukcji jest tylko skończona liczba rozłącznych skończonych ścieżek.*

Stwierdzenie 5 *Krawędzie w grafie redukcji nie przecinają się.*[4]

Definicja 6.2 Krawędź $k = (\_, \_, i, \_)$ jest *na tym samym poziomie* co $k' = (\_, \_, j, \_)$, jeśli $i = j$.

Definicja 6.3 Krawędź $k = (\_, \_, i, \_)$ jest *ponad* krawędzią $k' = (\_, \_, j, \_)$, jeśli $i < j$.

Definicja 6.4 Krawędź $k = (\_, \_, \_, i)$ jest *aktywna* jeśli $i = 1$.

Definicja 6.5 Krawędź aktywna $k = (v_1, v_2, i, 1)$ jest krawędzią *w lewo* jeśli $v_2 < v_1$. Jeśli $v_2 > v_1$, to k jest krawędzią *w prawo*.

Przykład 3 Krawędzie r_1, r_6 z przykładu 2 są krawędziami w lewo, zaś r_2 oraz r_3 w prawo.

Definicja 6.6 Niech K, M będą różnymi nieskończonymi ścieżkami w grafie zaczynającymi się w tym samym zbiorze V_j. Definiujemy relację $L'(K, M)$ "bycia na lewo od":

$$L'(K, M) \equiv \forall i\ (k_i \in K \wedge m_i \in M \wedge k_i = (c_1, \_, j, \_) \wedge m_i = (d_1, \_, j, \_) \Rightarrow c_1 \leq d_1)$$

Ponieważ jednak na ogół interesować nas będzie ostre "bycie na lewo od", to definiujemy relację $L = L' \setminus \{(x, x) \mid x \text{ jest ścieżką }\}$.

Definicja 6.7 Ścieżka K jest *na lewo* od ścieżki M, jeśli zachodzi $L(K, M)$. W przeciwnym wypadku K jest *na prawo* od M.

Lemat 6 *Relacja L' jest częściowym porządkiem. Ponadto zbiór ścieżek grafu redukcji zaczynających się w V_0 z porządkiem L' jest liniowym porządkiem oraz kratą zupełną.*

[4] Dowód natychmiastowy z własności sof.

Dowód Nietrywialna jest tylko kratowość L'. Weźmy dowolny zbiór ścieżek A zaczynających się w V_0. Niech A_n oznacza wszystkie krawędzie ze ścieżek z A na poziomie n, tzn.

$$A_n = \{k = (\_, \_, n, \_) \mid \exists L \ (k \in L \in A)\}$$

Kresem górnym tego zbioru jest ścieżka powstała w następujący sposób: pierwszą krawędzią na niej jest krawędź najbardziej na lewo ze zbioru A_1. Następnie idziemy w dół zaczynając od tej krawędzi, wybierając na rozstajach zawsze drogę w lewo. Łatwo zobaczyć, że powstała ścieżka będzie kresem górnym A. Istnienia kresu dolnego dowodzimy analogicznie.

Definicja 6.8 Dla krawędzi $k = (\_, \_, i, \_)$ w derywacji $w_0 \to w_1 \to \ldots$ zdefiniujemy jej moment śmierci. Jeśli k jest aktywna, to umiera w i-tej redukcji ($w_{i-1} \to w_i$), jeśli nie to istnieje krawędź $l = (\_, \_, i', \_)$ taka, że l jest najwyższą krawędzią aktywną taką, że k jest przodkiem l. Wtedy k umiera w i'-tej redukcji ($w_{i'-1} \to w_{i'}$).

Stwierdzenie 7 *Jeśli z V_i wychodzi ponad jedna krawędź aktywna, to wychodzą dokładnie dwie — jedna w prawo, jedna w lewo.*[5]

7. Rodziny

Definicja 7.1 Niech $k = (\_, \_, i, 1)$ będzie krawędzią aktywną w lewo (w prawo). Wtedy mamy $w_{i-1} = xsy \to xty = x_1st_1y$ (odpowiednio xt_1sy_1) oraz $x = x_1x_2 \wedge t = t_2t_1$ (odpowiednio $y = y_2y_1 \wedge t = t_1t_2$). Mówimy, że $\overline{x_2}$ (odpowiednio $\overline{y_2}$) jest *mężem* lewym (odpowiednio prawym) k (krawędź potrzebuje tego słowa do stworzenia nowego redeksu). *Małżeństwem* k nazywamy $\overline{x_2t_2}$ (odpowiednio $\overline{t_2y_2}$). *Konsumpcją* małżeństwa nazywamy redukcję w czasie której wszystkie jego pozycje są niszczone.[6]

Rysunek 1.2: Konsumpcja małżeństwa

Przykład 4 W grafie z przykładu 2 zaznaczeni zostali mężowie m_1, m_2, m_3, m_4 krawędzi odpowiednio r_1, r_2, r_3, r_6 (zaznaczanie mężów w trakcie ślubu mija się z celem, więc zaznaczyliśmy ich jeszcze w stanie kawalerskim).

Definicja 7.2 Popatrzmy na słowo $\overline{x} \sqsubseteq w_i$ dla pewnego i. Oznaczmy zbiór jego pozycji bez ostatniej przez P. Jeśli zachodzi:

$$\forall p \in P \ (P_i(p) = -\infty)$$

[5]Dowód natychmiastowy z własności sof systemu.

[6]Poprawność definicji konsumpcji wynika z własności sof systemu.

to $\overline{x}$ nazywamy sierotą[7]. W przeciwnym przypadku definiujemy zbiór dat urodzenia $\overline{x}$:

$$D(\overline{x}) \equiv \{\ P_i(p) \mid P_i(p) \neq -\infty\ \}$$

Zauważmy, że dat urodzenia słowa może być mniej niż ilość liter w tym słowie — niektóre litery mogą być sierotami.

Przykład 5 Słowa m_1 i m_2 z przykładu 2 to sieroty. Natomiast m_3 i m_4 są urodzone w pierwszej redukcji.

Definicja 7.3 Jeśli mąż $\overline{m}$ pewnej krawędzi k nie jest sierotą, to każdej dacie urodzin $\overline{m}$ odpowiada wierzchołek w grafie. Każdy taki wierzchołek nazywamy *świekrą* k. Ściślej, zbiór świekr $\acute{\mathrm{S}}(k)$ definiujemy tak:

$$\acute{\mathrm{S}}(k) \equiv \bigcup_{i \in D(\overline{m})} \{R(i)\}$$

Liczbę $R(i)$ z wzoru traktujemy jako wierzchołek ze zbioru V_i. *Córkami* świekry nazywamy wychodzące z niej krawędzie.

Przykład 6 Świekrą r_6 z przykładu 2 jest pozycja 1 ze zbioru V_0 odpowiadająca redeksowi w pierwszym słowie derywacji.

Definicja 7.4 Matką krawędzi $k = (v_1, \_, i, \_)$ nazywamy krawędź $k_1 = (\_, v_1, i-1, \_)$ (matka jest tylko jedna). Krawędź k nazywamy córką krawędzi k_1. Niech M będzie przechodnim domknięciem relacji "bycia matką". Krawędź k jest nazywamy przodkiem l, gdy zachodzi $M(k,l)$. Krawędź $k_2 = (\_, \_, 1, \_)$ będącą przodkiem k nazywamy praprzodkiem k.

Definicja 7.5 Wierzchołek v w grafie derywacji nazywamy przodkiem krawędzi k jeśli istnieje krawędź wychodząca z v będąca przodkiem k.

Stwierdzenie 8 *Dla każdego $i > 1$ krawędź $k = (\_, \_, i, \_)$ ma matkę oraz praprzodka.*

Definicja 7.6 *Rodem* krawędzi k nazywamy jedyną ścieżkę od praprzodka do k.

Definicja 7.7 Krawędź l jest *uogólnioną córką* krawędzi k jeśli k jest w rodzie l.

Lemat 9 *Niech m będzie pewnym mężem, dla ustalenia uwagi prawym. Niech $\dot{z}$ będzie jego żoną. Niech $k = (p, \_, i, \_)$ będzie córką pewnej świekry m. Niech $l = (n, \_, i, \_)$ będzie przodkiem $\dot{z}$ położonym na poziomie i. Wtedy $n \leq p$.*

Dowód W przeciwnym przypadku ścieżki zawierające l i k musiałyby się przeciąć, co jest niemożliwe ze względu na stwierdzenie 5.

Lemat 10 *Niech m będzie pewnym mężem, dla ustalenia uwagi prawym. Niech $\dot{z} = (\_, \_, j, \_)$ będzie jego żoną. Niech $\acute{s}$ będzie najniższą świekrą $\dot{z}$ należącą do zbioru V_{i-1} dla pewnego i i niech p będzie pozycją męża najbardziej na lewo utworzoną w trakcie i-tej redukcji ($w_{i-1} \to w_i$). Wtedy nie istnieje krawędź $k = (\_, v, i, \_)$ położona na lewo od p ($v \leq p$) posiadająca potomka na poziomie j.*

Dowód Łatwo zobaczyć, że gdyby istniała taka krawędź, to ścieżki w grafie musiałyby się przeciąć, bo uogólnione córki k oddzielałyby męża od rodu jego żony.

[7]Chodzi tu oczywiście o to, że $\overline{x}$ nie ma matki — istnieje bowiem od początku derywacji.

8. Jednowątkowość i niezależność

Definicja 8.1 Ścieżka w grafie jest *jednowątkowa*, gdy wszystkie krawędzie na niej są oznaczone jedynkami. Derywacja jest jednowątkowa, gdy jej graf jest ścieżką (jednowątkową). System jest *jednowątkowy*, gdy ma nieskończoną jednowątkową derywację.

Twierdzenie 3 *Jeśli system jest WB i ma nieskończoną derywację, to jest jednowątkowy.*

Dowód Wynika bezpośrednio z twierdzenia 7.4 w [MacNau95].

Od tej pory zajmujemy się wyłącznie systemami, które nie są WB.

Definicja 8.2 Niech k będzie krawędzią aktywną w grafie derywacji $w_0 \to w_1 \to \dots$ położoną na pewnej ścieżce S. Zdefiniujemy pojęcie niezależności k (od krawędzi aktywnych powyżej niej położonych poza S). Krawędź k jest niezależna jeśli wszystkie świekry k leżą na S, w szczególności gdy jej mąż jest sierotą.

Definicja 8.3 Ścieżka w grafie jest niezależna, jeśli każda krawędź aktywna na niej jest niezależna.

Oczywiście każda podścieżka niezależnej ścieżki jest też niezależna.

Stwierdzenie 11 *Każda ścieżka jednowątkowa jest niezależna.*

Definicja 8.4 Niech $k = (\_, \_, i, 1)$ będzie krawędzią aktywną na pewnej ścieżce S. Niech $k_1 = (\_, \_, \_, 1)$ będzie następną krawędzią aktywną na S. Krawędź k_1 *można od razu wykonać* po k, gdy wszystkie świekry k_1 znajdują się powyżej i. Ściślej, gdy zachodzi warunek:

$$(v \in \acute{S}(k_1) \land v \in V_l) \Rightarrow l < i$$

Stwierdzenie 12 *Jeśli ścieżka jest niezależna, to dla każdej krawędzi aktywnej k na niej, można od razu wykonać następną krawędź aktywną po k.*

Możemy zatem na niezależnej ścieżce "wypchnąć do góry" każdą krawędź aktywną. Ściślej: Niech $W = w_0 \to w_1 \to \dots$ będzie derywacją. Niech S będzie ścieżką w jej grafie. Niech $k = (\_, \_, i, 1)$ będzie krawędzią aktywną na S i niech $k_1 = (\_, \_, j, 1)$ będzie następną krawędzią aktywną na S. Dla ustalenia uwagi, niech $w_i = x_1 x_2 s y$ (oznaczenia jak przy definicji męża). Z własności sof systemu i ze stwierdzenia 12 redukcje $w_i \to w_{i+1} \to \dots \to w_{j-1}$ mają miejsce albo w x_1 albo w y. Zatem możemy zdefiniowac zmodyfikowany ciąg redukcji $w_i \to w'_{i+1} \to w'_{i+2} \to \dots \to w'_j$, w której wykonujemy najpierw redukcję, której odpowiada k_1, a potem dopiero wszystkie redukcje pomiędzy Co więcej, $w'_j = w_j$. Niech

$$W' = w_0 \to w_1 \to \dots \to w_i \to w'_{i+1} \to \dots \to (w'_j = w_j) \to w_{j+1} \to \dots$$

Łatwo zobaczyć iż ścieżka otrzymana z S jest wciąż niezależna.

Lemat 13 *Jeśli w grafie derywacji $w_0 \to w_1 \to \dots$ istnieje nieskończona niezależna ścieżka S, to istnieje też derywacja, w której grafie jest ścieżka jednowątkowa dowolnej długości.*

Dowód Bierzemy ścieżkę otrzymaną z S po odpowiedniej ilości "wypchnięć w górę" krawędzi aktywnych.

Lemat 14 *Niech w_0 będzie takie, że inicjuje nieskończoną derywację $w_0 \to w_1 \to \ldots$. Jeśli dla dowolnego $i > 0$ istnieje słowo v oraz derywacja $W_1 = w_0 \to^i v$ i w grafie W_1 jest jednowątkowa ścieżka długości i (czyli on sam jest tą ścieżką), to istnieje derywacja $W_2 = w_0 \to w'_1 \to w'_2 \to \ldots$, w grafie której jest nieskończona jednowątkowa ścieżka.*

Dowód Zbudujmy drzewo, w którym wierzchołkami będą skończone derywacje. Korzeniem jest zerowa derywacja zawierająca w_0. Kolejne poziomy w drzewie definiujemy indukcyjnie. Derywacja W zawierająca n redukcji zaczynająca się od słowa w_0: $W = w_0 \to w_1 \to \ldots \to w_n$ jest na poziomie n, gdy na poziomie $n-1$ istnieje derywacja $V = w_0 \to w_1 \to \ldots \to w_{n-1}$ oraz gdy graf derywacji W jest jednowątkową ścieżką. Wtedy derywacja V jest ojcem derywacji W. Z założenia drzewo jest nieskończone i ma ścieżki dowolnej długości. Ponieważ zbiór wszystkich skończonych derywacji długości n zaczynających się w w_0 jest skończony dla każdego n, to każdy wierzchołek ma skończenie wiele synów. Z lematu Königa dostajemy nieskończoną ścieżkę, która odpowiada jednowątkowej nieskończonej derywacji.

Lemat 15 *Jeśli w grafie derywacji $w_0 \to w_1 \to \ldots$ istnieje nieskończona niezależna ścieżka S, to istnieje derywacja, w grafie której jest nieskończona jednowątkowa ścieżka.*

Dowód Natychmiastowy z lematów 13 i 14.
Zatem pojęcia niezależności i jednowątkowości są bardzo blisko związane. Pojęcie jednowątkowości, jak się potem okaże, jest wygodne przy badaniu własności systemów, natomiast niezależność pozwoli nam scharakteryzować jednowątkowość.

9. Charakteryzacja jednowątkowości

Definicja 9.1 Niech $w_0 \to w_1 \to \ldots$ będzie derywacją. Redukcja *i rozgałęzia się* jeśli z V_{i-1} wychodzą dokładnie dwie krawędzie aktywne. Mówimy wtedy, że zachodzi warunek $B(i)$.

Przykład 7 Redukcja 1 z przykładu 2 rozgałęzia się.

Lemat 16 *Jeśli w grafie derywacji $W = w_0 \to w_1 \to \ldots$ jest tylko jedna nieskończona ścieżka S, to prawie każda aktywna krawędź albo leży na S, albo ma przodka leżącego na S.*

Dowód Przez sprzeczność. Załóżmy, że istnieje nieskończenie wiele krawędzi aktywnych leżących poza S, których żaden przodek nie leży na S. Ponieważ tych krawędzi jest nieskończenie wiele, to możemy znaleźć je dowolnie nisko w grafie redukcji. Każda taka krawędź ma praprzodka w w_0. Mamy więc ścieżki dowolnej długości w grafie redukcji rozłączne z S, więc mamy też nieskończoną ścieżkę rozłączną z S, sprzeczność z założeniem kończy dowód.

Lemat 17 $EI(i) \land B(i) \Rightarrow 2A$

Dowód Ponieważ zachodzi $B(i)$ oraz s nie jest podsłowem t, to:

$$w_i = xty = x_1st_1sy_1, s = x_2t_2 = t_3y_2, x = x_1x_2, y = y_2y_1, t = t_2t_1t_3.$$

Z definicji grafu oba $\overline{s}$-y (na pozycjach $|x_1|$ i $|x_1st_1|$) będą kiedyś zredukowane. Dla ustalenia uwagi, niech lewy zredukuje się pierwszy. W trakcie jego redukcji C_i jest niszczone z lewej strony. Niech p będzie pozycją sy_1 ($|x_1st_1|$). Z własności sof systemu, pozycja $p+1$ będzie zniszczona dopiero w trakcie redukcji $\overline{s}$ z prawej strony — w czasie której C_i będzie niszczone z prawej strony. Zatem zachodzi 2A, bo pozycje z C_i są niszczone raz z lewej raz z prawej strony.

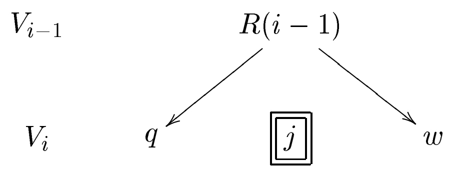
<table>
<tr><td colspan="2">x</td><td colspan="3">t</td><td colspan="2">y</td></tr>
<tr><td>x_1</td><td colspan="2">s</td><td>t_1</td><td colspan="2">s</td><td>y_1</td></tr>
</table>

Rysunek 1.3: Ilustracja do lematu 17

Lemat 18 *Jeśli zachodzi $B(i)$, to z $R(i-1)$ (traktowanego jako wierzchołek w V_{i-1}) wychodzą 2 krawędzie: $k_1 = (\_, q, \_, \_)$, $k_2 = (\_, w, \_, \_)$. Dla ustalenia uwagi, niech $q < w$. Wtedy:*

$$B(i) \wedge \neg 2A \Rightarrow \exists j \ (q + |s| \leq j \leq w \wedge I(j,i))$$

Dowód Łatwo wynika z lematu 17.



Rysunek 1.4: Ilustracja do lematu 18

Definicja 9.2 Niech L, M będą ścieżkami w grafie redukcji. *M wpływa* na L jeśli pewien wierzchołek z M jest świekrą pewnej krawędzi z S. Analogicznie definiujemy wpływanie drzew na drzewa.

Stwierdzenie 19 *Jeśli na ścieżkę L nie wpływa żadna inna ścieżka, to jest ona niezależna.*

Stwierdzenie 20 *Niech i, j będą takie, że zachodzi $I(i,j)$ w pewnej derywacji $w_0 \to w_1 \to \ldots$ Niech $k = (v, \_, i, \_)$ będzie pewną krawędzią. Jeśli k jest na lewo od j (czyli $v \leq j$), to wszystkie uogólnione córki k są również na lewo od j (a ściślej — od jedynego elementu zbioru $Y_i(\{j\}, n)$ dla odpowiedniego n). Analogiczny fakt zachodzi, gdy k jest na prawo od j.*

Lemat 21 *Popatrzmy na dowolną żonę $\dot{z} = (v'_1, v'_2, j, \_)$ w pewnej derywacji $w_0 \to w_1 \to \ldots$. Niech $k = (v_1, v_2, i, \_)$ należy do rodu $\dot{z}$ dla pewnego i i niech v'_1 będzie pozycją $\dot{z}$. Dla ustalenia uwagi, niech $v_1 \leq v'_1$. Wtedy zachodzi:*

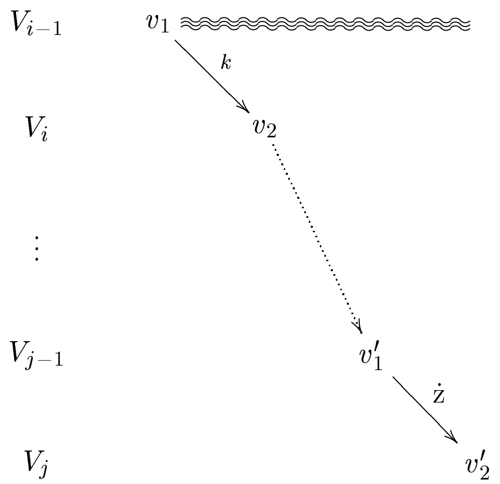
$$\forall j \ (v_1 < j < v'_1 \Rightarrow \neg I(i,j))$$

Lemat 22 *Popatrzmy na dowolną żonę $\dot{z}$ i pewną jej świekrę $\acute{s}$. w pewnej derywacji $w_0 \to w_1 \to \ldots$. Niech $k = (v_1, v_2, i, \_)$ należy do rodu $\dot{z}$ i niech $l = (v'_1, v'_2, i, \_)$ należy do rodu $\acute{s}$ dla pewnego i. Dla ustalenia uwagi, niech $v_1 \leq v'_1$. Wtedy zachodzi:*

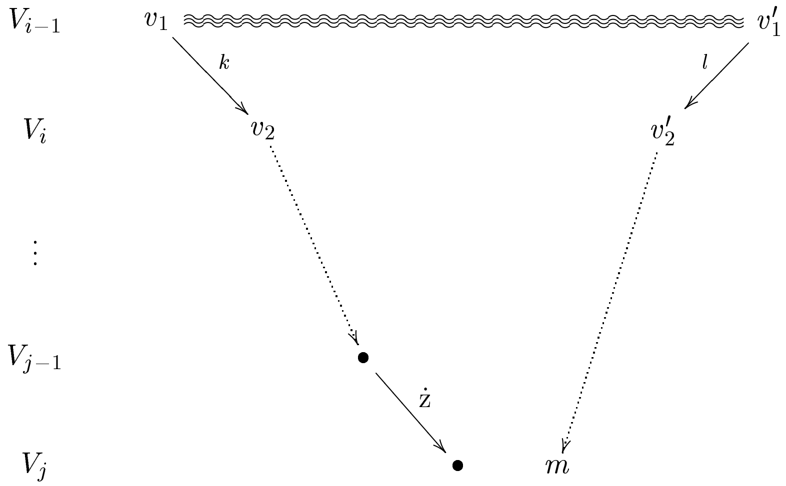
$$\forall j \ (v_1 < j < v'_1 \Rightarrow \neg I(i,j))$$

Dowód lematów łatwy ze stwierdzenia 20.

Lemat 23 *Niech S będzie ścieżką i niech v spoza S będzie świekrą pewnej krawędzi l z S. Jeśli v i l mają wspólnego przodka, to zachodzi warunek 2A.*



Rysunek 1.5: Ilustracja do lematu 21



Rysunek 1.6: Ilustracja do lematu 22

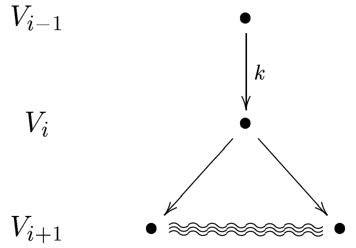
Dowód Popatrzmy na najniższego wspólnego przodka $k = (\_,\_,i,\_)$. Musi zachodzić $B(i+1)$ a ponadto wszystkie pozycje utworzone w redukcji $w_i \to w_{i+1}$ muszą zostać potem zniszczone z lematu 22. Z lematu 17 dostajemy tezę.

Definicja 9.3 Jeśli L jest ścieżką, to przez $PL(L)$ oznaczamy zbiór ścieżek na lewo od L, a przez $PR(L)$ oznaczamy zbiór wszystkich ścieżek na prawo od L.

Stwierdzenie 24 *Jeśli K wpływa na L, to zachodzi jeden z warunków:*

- $K = L$
- $K = \sup PL(L)$
- $K = \inf PR(L)$

Dowód Wynika ze stwierdzenia 5 oraz z lematu 10.



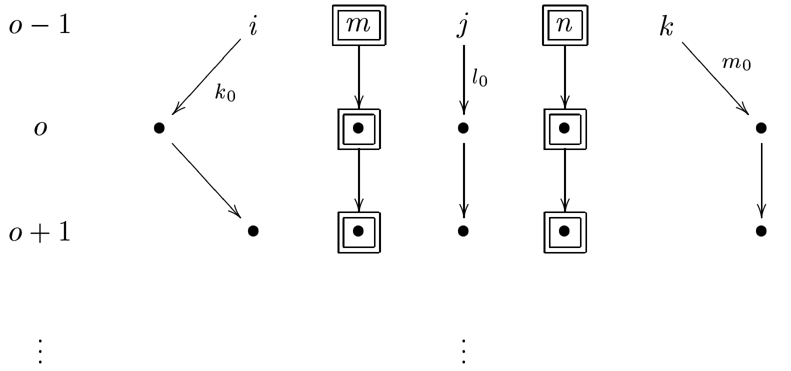
Rysunek 1.7: Ilustracja do lematu 23

Lemat 25 *Niech L będzie ścieżką w grafie. Niech $K = \sup PL(L)$ i $M = \inf PR(L)$. Niech $k_0 = (\_, i, o, \_)$, $l_0 = (\_, j, o, \_)$, $m_0 = (\_, k, o, \_)$ będą najwyższymi krawędziami na ścieżkach odpowiednio K, L i M. Jeśli:*

$$\exists m, n \ (i < m < j < n < k \wedge I(o, m) \wedge I(o, n))$$

to nie istnieje ścieżka wpływająca na L.

Dowód Pokażemy, że M i K nie wpływają na L co z stwierdzenia 24 da tezę. Załóżmy przeciwnie, że K (dla ustalenia uwagi) wpływa na L. Czyli istnieje krawędź k z L taka, że pewien przodek jej świekry to k_0. Z lematu 22 dostajemy sprzeczność.



Rysunek 1.8: Ilustracja do lematu 25

Lemat 26 *Jeśli graf derywacji (nieskończonej) jest drzewem i nie zachodzi warunek 2A, to system jest jednowątkowy.*

Dowód Weźmy dowolną nieskończoną ścieżkę S. Załóżmy, że mamy k spoza S, która jest świekrą pewnej krawędzi l z S. Oczywiście k i l mają wspólnego przodka, co daje sprzeczność z lematu 23. Zatem każda świekra każdej krawędzi aktywnej z S też leży na S, co ze stwierdzenia 8.2 i lematu 15 daje tezę.

Pozostaje rozpatrzeć przypadek, kiedy w grafie jest więcej niż jedno drzewo. Pokażemy, że dla dowolnego nieskończonego drzewa, inne drzewa wpływają na nie co najwyżej skończoną ilość razy.

Wybierzmy więc drzewo nieskończone F. Mogą na nie wpływać jedynie drzewa położone najbliżej F z lewej (nazwijmy je E) i z prawej strony (nazwijmy je G). Dla ustalenia uwagi zajmiemy się drzewem G. Bez straty ogólności możemy przyjąć, że G jest nieskończone.

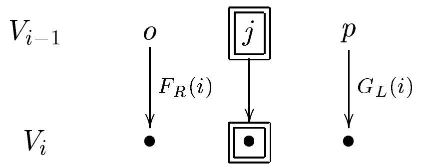
Lemat 27 *Jeśli G i E wpływają na F tylko skończoną ilość razy, to istnieje derywacja, której graf jest drzewem.*

Dowód Zaczynamy redukcję od poziomu, poniżej którego G i E nie wpływają na F i wykonujemy tylko redukcje z drzewa F.

Lemat 28 *Niech T będzie drzewem. Oznaczamy przez $T_R(i)(T_L(i))$ krawędź z T na poziomie i, położoną najbardziej na prawo (lewo). Załóżmy, że zachodzi:*

$$\exists i, j\ (F_R(i) = (o, \_, i, \_) \wedge G_L(i) = (p, \_, i, \_) \wedge o < j < p \wedge I(i, j))$$

Wtedy G poniżej i nie wpływa na F.



Rysunek 1.9: Ilustracja do lematu 28

Dowód Łatwy z lematu 22.

Lemat 29 *Jeśli istnieje niezniszczalna pozycja p taka, że prawie wszystkie krawędzie G znajdują się na prawo od niej, a drzewo F znajduje się na lewo od niej, to G wpływa na F tylko skończenie wiele razy.*

Dowód Poniżej najniższej krawędzi z G na prawo od p widzimy założenie lematu 28.

Lemat 30 *Jeśli pewien wierzchołek z G jest świekrą krawędzi k z F, to k jest krawędzią w prawo.*

Dowód Gdyby k była w lewo, to mąż byłby na lewo od rodu k, więc ród hipotetycznej świekry zawierałby co najmniej jedną krawędź na lewo od rodu k, co jest sprzeczne z wyborem G.

Przyszedł czas na główny lemat.

Lemat 31 *Jeśli nie zachodzi 2A, to drzewo G wpływa na F tylko skończenie wiele razy.*

Dowód Niech l będzie krawędzią z F taką, że pewna jej świekra jest w drzewie G. Niech $v \in V_i$ będzie najniżej położoną świekrą krawędzi l, która znajduje się w drzewie G. Niech $\overline{m}$ będzie mężem l i niech j będzie jego pozycją w słowie w_i. Rozpatrzymy oddzielnie 2 przypadki:

- Żaden z przodków v nie rozgałęzia się. Wtedy v musi mieć nieskończenie wiele uogólnionych córek, bo G jest nieskończone. Niech $k = (v, v_1, i + 1, 1)$ będzie krawędzią wychodzącą z v. Słowo $\overline{m}$ jest sufiksem s, bo l jest krawędzią w prawo z lematu 31. Niech p będzie pozycją jednej z liter męża utworzoną w czasie redukcji $w_i \to w_{i+1}$ (definicja jest dobra, bo v jest świekrą). Rozpatrzmy kilka podprzypadków:
  - $p \geq v_1$ i k jest krawędzią w prawo (mąż nakłada się na redeks na pozycji v_1 w w_{i+1}). Ten redeks nie może zostać zredukowany przed konsumpcją małżeństwa l, bo wtedy v_1 byłaby świekrą niżej położoną od v. Łatwo zauważyć, że w momencie konsumpcji małżeństwa zostałaby naruszona jakaś sofa[8], więc ten przypadek jest niemożliwy.

[8]dokładniej własność sof.

- $p < v_1$ i k jest krawędzią w prawo. Wtedy część pozycji z C_{i+1} będzie zniszczona z lewej strony (w czasie konsumpcji małżeństwa), a część z prawej (w czasie śmierci k). Zachodziłby więc 2A, więc i ten przypadek jest niemożliwy.
- $p < v_1$ i k jest krawędzią w lewo. Ten przypadek jest po prostu niemożliwy z definicji świekry.
- $p \geq v_1$ i k jest krawędzią w lewo. Ponieważ k ma nieskończenie wiele potomków to dostajemy sprzeczność z lematu 10.

Zauważmy, że w istocie tutaj wykorzystywaliśmy jedynie własność, iż k inicjuje nieskończone drzewo. Możemy zatem w dowodzie następnego podpunktu przyjąć iż k ma tylko skończoną liczbę potomków (bo jeśli ma ich nieskończenie wiele, to przejdzie właśnie wykonany dowód).

- Pozycja v ma rozgałęziającego się przodka. Niech i będzie maksymalne takie, że zachodzi $B(i)$ i niech $v' = R(i-1) \in V_{i-1}$ będzie taki, że jedna z córek v' jest przodkiem v. Oczywiście musi być ona krawędzią w lewo. Popatrzmy na słowo t w w_i na pozycji $R(i-1)$. Z lematu 18 dostajemy istnienie pozycji w t, która nie jest niszczona w dalszym ciągu derywacji. Z lematu 29 dostajemy tezę.

Twierdzenie 4 *Jeśli nie zachodzi warunek 2A, to system ma nieskończoną derywację wtedy i tylko wtedy, gdy jest jednowątkowy.*

Dowód Natychmiastowy z lematów 31, 27 i 26.

10. Wnioski z jednowątkowości

Na czas tego rozdziału wszystkie derywacje są jednowątkowe i nieskończone, chyba że explicite stwierdzimy inaczej.

Definicja 10.1 Dla derywacji $w_0 \to w_1 \to \ldots$ definiujemy w naturalny sposób ciąg jej krawędzi $\{k_n\}_{n\geq 1}$ tak, że k_n jest krawędzią odpowiadającą redukcji $w_{n-1} \to w_n$.

Lemat 32 *W ciągu krawędzi jest nieskończenie wiele krawędzi w lewo i nieskończenie wiele krawędzi w prawo.*

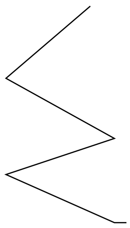
Dowód Załóżmy, dla ustalenia uwagi, że jest tylko skończenie wiele krawędzi w lewo. Wtedy istnieje N takie, że k_n jest krawędzią w prawo dla każdego $n > N$. Niech $a_n = |w_n| - R(n)$. Oczywiście $a_n \geq 0$ dla każdego n,. Ponadto łatwo zobaczyć, że a_n jest ciągiem ostro malejącym dla $n > N$, co daje sprzeczność.

Zatem graf derywacji jednowątkowej wygląda mniej więcej tak jak na rysunku 1.10.

Stwierdzenie 33 *System ma nieskończoną derywację wtedy i tylko wtedy, gdy ma nieskończoną derywację, w której pierwsza krawędź jest w prawo.*

Od tej chwili wszystkie rozważane derywacje zaczynają się od krawędzi w prawo.

Stwierdzenie 34 *Załóżmy, że dla derywacji jednowątkowej $w_0 \to w_1 \to \ldots$ istnieje taka liczba i, że zachodzi $EI(i)$. Wtedy jeśli k_i jest krawędzią w lewo, to pozycje z C_i są niszczone tylko z lewej strony, a jeśli jest krawędzią w prawo, to pozycje z C_i są niszczone tylko z prawej strony.*



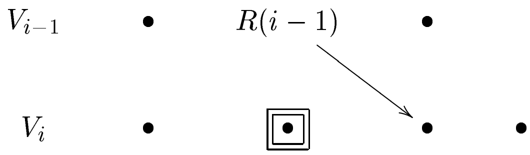
Rysunek 1.10: Graf jednowątkowej derywacji

Warunek 2A zatem nie ma szansy objawić się w trakcie jednowątkowej derywacji. Nie znam odpowiedzi na pytanie, czy istnieje jednowątkowy system spełniający warunek 2A.

Zajmijmy się bliżej warunkami 2B i 2C z definicji 5.5. Są one symetryczne, skupimy więc naszą uwagę na warunku 2B. Od tej pory rozważany system spełnia warunek 2B i nie spełnia warunku 2C.

Lemat 35 *Niech $k = (v, \_, i, 1)$ będzie krawędzią w prawo w grafie W. Wtedy pozycja v w zbiorze V_{i-1} jest niezniszczalna.*

Dowód Przez sprzeczność. Załóżmy, że istnieje krawędź w prawo $k = (v, \_, i, 1)$ taka, że zachodzi $EI(i)$ i pozycja v nie jest niezniszczalna. Wtedy ze stwierdzenia 34, pozycje z C_i będą zniszczone od prawej strony. Czyli, z definicji 5.5, zachodzi warunek 2C, co daje szukaną sprzeczność.
Intuicyjnie, po ruchu w prawo blokujemy pozycję, na której był redeks.



Rysunek 1.11: Ilustracja lematu 35

Definicja 10.2 Podciąg ciągu krawędzi jest w lewo (odpowiednio w prawo), jeśli wszystkie jego elementy są krawędziami w lewo (odpowiednio w prawo).

Definicja 10.3 Dowolny maksymalny podciąg krawędzi w lewo (odpowiednio w prawo) nazywamy partią lewicową (odpowiednio prawicową).

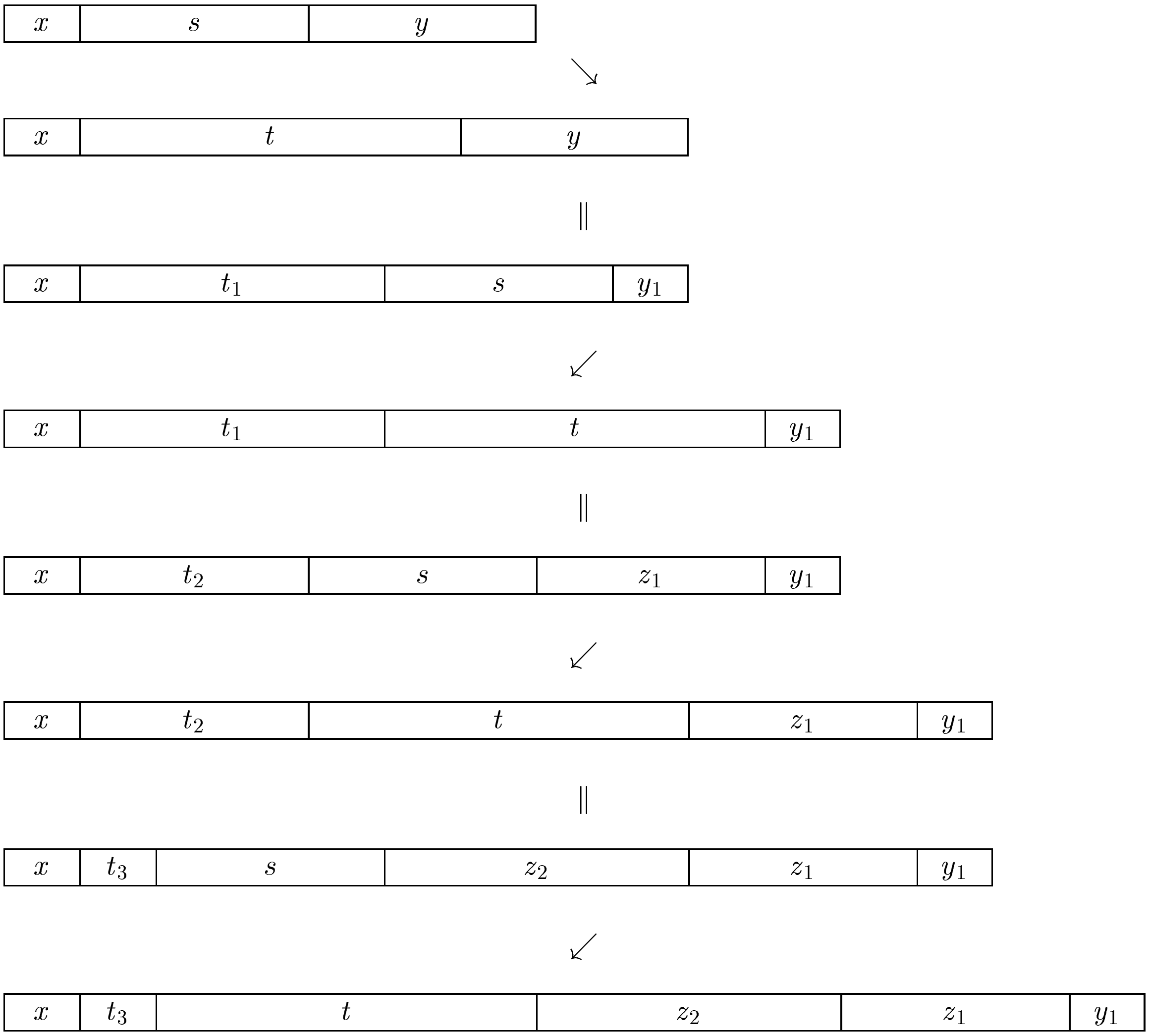
Definicja 10.4 Poprzedniczką partii lewicowej nazywamy najniższą krawędź w prawo $k = (v, \_, i, 1)$, która znajduje się ponad nią. Pozycję v w zbiorze V_i nazywamy założycielką partii.

Lemat 36 *Wszyscy mężowie członkiń dowolnej partii lewicowej mają tylko jedną matkę*[9] *i jest nią założycielka tej partii.*

Dowód Łatwy przez indukcję korzystając z lematu 35.

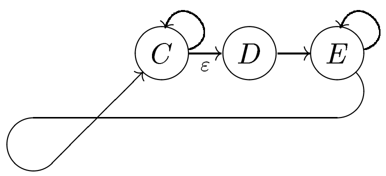
Ilustracją do tego lematu jest rysunek 1.12 (strzałki oznaczają kierunek odpowiednich krawędzi) — mężowie krawędzi w lewo są brani ze słowa $\overline{t}$ powstałego w pierwszym kroku redukcyjnym.

[9] Tzn. zbiór świekr członkiń ma tylko jeden element.



Rysunek 1.12: Pewien ciąg redukcji

Kluczową obserwacją jest to, że możemy traktować słowo na prawo od redeksu jako stos. Kroki w prawo zdejmują kawałek stosu (np. pierwszy krok zdejmuje y_2, gdzie $y = y_2y_1$), zaś każdy członek partii lewicowej wkłada coś na stos (na rysunku słowa z_1 i z_2). Co więcej, istnieje skończona ilość możliwych partii lewicowych z powodu lematu 36 (albo z powodu lematu 35). Zdefiniujemy zatem automat ze stosem, który będzie odpowiadał danemu systemowi, a właściwie to automat z dwoma stosami, A i B, gdzie długość stosu A jest ograniczona z góry przez $|t|$. Dla uproszczenia konstrukcji, przyjmujemy, że automat może wykonywać kroki typu: "zdejmij ze stosu słowo x" i "włóż na stos słowo y". Mało ściśle, automat będzie działał tak: krawędzi w prawo, która nie jest poprzedniczką partii lewicowej będzie odpowiadał krok zdejmujący ze stosu B dowolną prawą s-frazę systemu i ustawiający zawartość stosu A na słowo t. Poprzedniczce partii lewicowej odpowiada krok ustawiający stos A na słowo t, po czym zdejmujący naraz słowa ze stosów A i B, przy czym ze stosu A zdejmuje on dowolną prawą nakładkę systemu, a ze stosu B odpowiadającą jej prawą s-frazę, po czym wkłada na stos B słowo t. Wreszcie dowolnej krawędzi z partii lewicowej odpowiada ta sama sekwencja, co poprzedniczce partii, pomijając pierwszy krok, czyli zdejmuje on naraz słowa ze stosów A i B, przy czym ze stosu A zdejmuje dowolną prawą nakładkę systemu, a ze stosu B



Rysunek 1.13: Automat

odpowiadającą jej prawą s-frazę, po czym wkłada na stos B słowo t. A oto bardziej ścisła definicja.

Definicja 10.5 Automat z dwoma stosami A i B, gdzie długość A jest ograniczona z góry przez $|t|$, odpowiadający naszemu systemowi ma stany C, D i E. Stanem początkowym jest stan C, odpowiadający krawędzi w prawo nie będącej poprzedniczką partii lewicowej, stan D odpowiada krawędzi w prawo będącej poprzedniczką partii lewicowej, zaś stan E dowolnej krawędzi w lewo. Automat porusza się w następujący sposób:

- Od stanu C może przejść nie robiąc nic do stanu D. Decyduje on w ten sposób, że następna krawędź będzie poprzedniczką partii lewicowej.
- Od stanu C może przejść do stanu C, wykonując przy tym po kolei następujące czynności (przejście to odpowiada krawędzi w prawo):
  - Ustawić zawartość stosu A na słowo t. Odpowiada to samej redukcji. Dzięki stwierdzeniu 36 wiemy, że wszystko na lewo od redeksu już nie może brać udziału w dalszym ciągu derywacji.
  - Zdjąć ze stosu B dowolną s-frazę (czyli męża odpowiadającej temu przejściu krawędzi).
- Od stanu D może przejść do stanu E wykonując po kolei następujące czynności (przejście to odpowiada poprzedniczce partii lewicowej):
  - Ustawić zawartość stosu A na słowo t.
  - Zdjąć ze stosu A dowolną lewą nakładkę systemu.
  - Zdjąć ze stosu B s-frazę tej nakładki (czyli męża odpowiadającej temu przejściu krawędzi).
  - Wstawić na stos B słowo t.
- Od stanu E może przejść do stanu C lub stanu E wykonując po kolei następujące czynności (przejście to odpowiada krawędzi w lewo):
  - Zdjąć ze stosu A dowolną lewą nakładkę systemu.
  - Zdjąć ze stosu B s-frazę tej nakładki (czyli męża odpowiadającej temu przejściu krawędzi).
  - Wstawić na stos B słowo t.

Lemat 37 *Niech $\{a_n\}$ będzie ciągiem kroków automatu (np. $a_0 = C$, $a_1 = C$, $a_2 = D$, ...) przy pewnym stanie początkowym stosu B (stan początkowy stosu A, jak łatwo zauważyć, nie ma znaczenia). Dla każdego takiego ciągu i stanu początkowego B istnieje derywacja $\{w_n\}$ o takiej samej długości jak ciąg kroków (w szczególności nieskończona, gdy ten ciąg jest nieskończony).*

Dowód Wprost z definicji automatu. Derywacja zaczynać się będzie od słowa sx, gdzie x jest początkowym stanem stosu B.

Lemat 38 *Dla każdej derywacji istnieje stan początkowy stosu i ciąg kroków automatu o tej samej długości co derywacja.*

Dowód Niech derywacja zaczyna się od słowa $w_0 = xsy \to xty = w_1$. Automat zaczynał będzie ze stosem B zawierającym słowo y. Reszta dowodu jest łatwa.

11. Główne twierdzenie

Twierdzenie 5 *Jeśli system $s \to t$ ma własność sof i spełnia dokładnie jeden z warunków 2B i 2C i nie spełnia warunku 2A, to problem silnej normalizacji tego systemu jest rozstrzygalny.*

Dowód Z lematów 37 i 38 wiemy, że odpowiadający systemowi automat ze stosami terminuje wtedy i tylko wtedy gdy terminuje ten system. Ponieważ problem istnienia początkowego stanu stosu takiego, że automat ma nieskończony ciąg kroków rozpoczynający się od tego stanu jest rozstrzygalny, to rozstrzygalny jest też problem silnej normalizacji systemu $s \to t$.

12. Uwagi końcowe

Trudno ocenić użyteczność dowiedzionego twierdzenia, jako iż nie jest mi znany żaden nieterminujący nietrywialny system spełniający warunek 2B bądź 2C i nie spełniający warunku 2A. Podstawową trudnością w konstrukcji tego przykładu, jak i w całej gamie innych nasuwających się pytań (czy koniunkcja warunków 2B i 2C implikuje warunek 2A, czy istnieje nietrywialny nieterminujący system, w którym z pewnego wierzchołka grafu wychodzą ponad 2 krawędzie, czy redeks w jednowątkowej derywacji może "wędrować" od jednej strony słowa do drugiej — czyli czy istnieje nieskończona derywacja jednowątkowa zawierająca nieskończenie wiele krawędzi $k = (v, \_, \_, \_)$ takich, że $v = 0$) jest to, że każde takie pytanie wiąże się z dużym układem równań na słowach — które zawsze mają trywialne rozwiązania składające się np. z samych liter a, a znalezienie innych rozwiązań wydaje się być bardzo trudnym zadaniem. Na przykład, układ odpowiadający samemu warunkowi $2B$ wygląda tak (nazwy niewiadomych pochodzą od angielskich słów beginning, middle i end):

$$t = e_1 e_2 \ldots e_n m_1 = x_1 B_1 = x_2 b_2 = x_3 e_3 = \ldots = x_n e_n$$
$$s = B_1 m_1 E_1 = b_1 e_1 = b_2 e_2 = \ldots = b_n e_n$$

Łatwo wyobrazić sobie horror powstający z dodania warunku 2C i próbie dowodu czegokolwiek, w szczególności warunku 2A.

Wydaje mi się jednak, iż przedstawiona konstrukcja automatu ze stosem może mieć pewne zastosowania. Np. trywialnie wynika z niej, że nieskończona jednowątkowa derywacja istnieje wtedy i tylko wtedy, gdy istnieje nieskończona jednowątkowa derywacja, w której $w_0 = sx$ lub $w_0 = ys$ dla pewnych x, y. Dalsze badania mogłyby skoncentrować się na dokładniejszej charakteryzacji działań automatu dla systemów spełniających naraz warunki 2B i 2C — jako iż wtedy długości stosu A nie można ograniczyć z góry, to mogłoby się wydawać, iż jest to pełnoprawny automat z dwoma stosami. Ograniczenia nałożone na definicję automatu są jednak niemałe i być może dałoby się wyciągnąć z nich pewne wnioski, które potem przeniosłyby się na najtrudniejszy przypadek 2A. Prawdopodobnie jest to jednak narzędzie za słabe do dostarczenia odpowiedzi na wciąż otwarte pytanie o rozstrzygalność problemu silnej normalizacji.

Literatura

[BoOt93] R. V. Book, F. Otto, *String rewriting systems. Texts and monographs in Computer Science*, Springer-Verlag, 1993.

[Davis58] M. Davis, *Computability and unsolvability*, Mc Graw-Hill, 1958.

[Ges97] Alfons Geser, *Decidability of termination in a large class of one-rule string rewriting systems*, Technical report, Wilhelm-Schickard-Institut, Universität Tübingen, 1997.

[Ges98] Alfons Geser, *Termination of one-rule string rewriting systems $\ell \to r$ where $|r| \leq 9$*, Technical report, Wilhelm-Schickard-Institut, Universität Tübingen, 1998.

[Ges01] Alfons Geser, *Is termination decidable for string rewriting with only one rule?*, praca habilitacyjna, 2001.

[Her71] Gabor T. Herman, *Strong computability and variants of the uniform halting problem*, Z. Math. Logik Grundl. Math., 17:115-131, 1971.

[Kur90] Winfried Kurth, *Termination und Konfluenz von Semi-Thue-Systemem mit nur einer Regel*, praca doktorska, Technische Universität Clausthal, 1990.

[MacNau95] Robert MacNaughton, *Well behaved derivations in one-rule semi-Thue systems*, Report 95-15, Department of Computer Science, Rensselaer Polytechnic Institute, 1995.

[MatSen96] Jurij Matiasiewicz, Géraud Sénizergues , *Decision problems for Semi-Thue Systems with a Few Rules*, Logic in Computer Science, strony 532-531, 1996.

[STKK97] Kayoko Shikishima-Tsuji, Masashi Katsura, Yuji Kobayashi *On termination of confluent one-rule string rewriting systems*, Inform. Process. Lett., 61(2):91-96, 1997.

[Thue10] Axel Thue, *Die Lösung eines Spezialfalles eines generellen logischen Problems*, Kra. Vidensk. Selsk. Skrifter. I. Mat. - Nat. Kl., Christiana 1910, Nr. 8.

[Wra92] Celia Wrathall, *Confluence of one-rule Thue systems*, K. U. Schulz, redaktor, *Word equations and related topics*, LNCS 572. Springer, 1992.

[ZanGes95] Hans Zantema, Alfons Geser, *A complete characterization of termination of* $0^p1^q \to 1^r0^s$, Jieh Hsiang, redaktor, *Proc. 6th Int. Conf. Rewriting Techniques and Applications*, LNCS 914. Springer, 1995.