Języki, Automaty i Obliczenia

Zadania

Część druga

Damian Niwiński i Wojciech Rytter

Styczeń 2006

1 Języki bezkontekstowe

1.1 Gramatyki bezkontekstowe

1. Podać gramatyki bezkontekstowe generujące następujące języki:

   (a) zbiór słów nad alfabetem $\{a, b\}$, które zawierają tyle samo a co b;

   (b) zbiór słów nad alfabetem $\{a, b\}$, które zawierają dwa razy więcej a niż b;

   (c) zbiór słów nad alfabetem $\{a, b\}$ o długości parzystej, w których liczba wystąpień litery b na pozycjach parzystych jest równa liczbie wystąpień tej litery na pozycjach nieparzystych;

   (d) zbiór wyrażeń arytmetycznych nad alfabetem $\{0, 1, (,), +, \cdot\}$, które, przy zwykłej interpretacji działań dla liczb naturalnych, mają wartość 3;

   (e) zbiór wyrażeń arytmetycznych w notacji polskiej (nad alfabetem $\{0, 1, +, \cdot\}$) o wartości 4;

   (f) zbiór poprawnie zbudowanych formuł rachunku zdań ze zmienną zdaniową p i stałymi logicznymi **true**, **false** (alfabet: $\{p, \mathbf{true}, \mathbf{false}, \wedge, \vee, \neg, (,)\}$);

   (g) zbiór tych formuł z poprzedniego punktu, które przy każdym wartościowaniu zmiennej p mają wartość logiczną *prawda* (tzn. tautologii);

   (h) $\{a^i b^j c^k \ : \ i \neq j \ \vee \ j \neq k\}$;

   (i) $\{a^i b^j a^k \ : \ i + k = j\}$.

2. Dla danych gramatyk bezkontekstowych G, H, skonstruować gramatyki generujące języki $L(G) \cup L(H)$, $L(G)L(H)$, $(L(G))^*$, $(L(G))^R$ (=lustrzane odbicie).

3. Wykazać, że zbiór palindromów nad ustalonym alfabetem jak również jego dopełnienie są językami bezkontekstowymi.

4. Skonstruować gramatykę bezkontekstową z jednym symbolem nieterminalnym generującą zbiór $\{x \in (a+b)^* \ : \ \#_a(x) = \#_b(x) \ \}$, gdzie $\#_s(w)$ oznacza liczbę wystąpień symbolu s w słowie w.

5. Dowieść, że następujące warunki są równoważne dla języka $L \subseteq \Sigma^*$:

   (a) L jest regularny,

   (b) L jest generowany przez gramatykę bezkontekstową, w której każda reguła jest postaci $X \to \epsilon$, $X \to Y$, lub $X \to \sigma Y$, $\sigma \in \Sigma$,

   (c) L jest generowany przez gramatykę bezkontekstową, w której każda reguła jest postaci $X \to \epsilon$, $X \to Y$, lub $X \to Y\sigma$, $\sigma \in \Sigma$,

   (d) L jest generowany przez gramatykę bezkontekstową, w której każda reguła jest postaci $X \to \alpha$ lub $X \to \beta Y$, $\alpha, \beta \in \Sigma^*$.

6. Podać przykład gramatyki bezkontekstowej w której każda reguła jest postaci $X \to \epsilon$, $X \to Y$, $X \to \sigma Y$ lub $X \to Y\sigma$, $\sigma \in \Sigma$, ale język generowany przez gramatykę nie jest regularny.

7. Czy każdy język bezkontekstowy jest generowany przez gramatykę w postaci z poprzedniego zadania?

8. Powiemy, że gramatyka G ma własność *właściwego samozapętlenia*, jeśli dla pewnej zmiennej X zachodzi $X \stackrel{G}{\to}^* \alpha X \beta$, gdzie $\alpha, \beta \neq \epsilon$. Udowodnij, że gramatyka bezkontekstowa nie mająca własności właściwego samozapętlenia generuje język regularny.

9. Udowodnij, że każdy język bezkontekstowy nad jednoliterowym alfabetem jest regularny.

10. Udowodnij, że jeśli L jest bezkontekstowy to język $\{a^{|w|} \ : \ w \in L\}$ jest regularny.

11. Niech G będzie gramatyką bezkontekstową, z m zmiennymi i niech, dla każdej reguły $Y \stackrel{G}{\to} w$, $|w| \leq \ell$. Dowieść, że jeśli $X_I \stackrel{G*}{\to} \epsilon$, to istnieje wyprowadzenie o długości $1 + \ell + \ell^2 + \ldots + \ell^{m-1}$. Czy to oszacowanie jest optymalne?

12. Dowieść, że dla każdej gramatyki G istnieje stała C, taka, że dla dowolnego $w \neq \epsilon$, jeśli $X_I \stackrel{G*}{\to} w$, to istnieje wyprowadzenie o długości $\leq C \cdot |w|$.

13. Załóżmy, że mamy pewną skończoną liczbę reguł wymazujących postaci $\alpha \to \epsilon$. Stosując takie reguły możemy w danym słowie zastępować słowo α przez słowo puste. Niech L będzie zbiorem słów, które możemy przekształcić na słowo puste stosując reguły wymazywania.

    Czy zawsze istnieje gramatyka bezkontekstowa generująca L i mająca tylko jeden symbol nieterminalny ?

14. Zaprojektować algorytm, który, dla danej gramatyki G, odpowiada na pytanie, czy język $L(G)$ jest nieskończony.

15. Udowodnij, że każdy język bezkontekstowy może być generowany przez gramatykę w której każdy symbol nieterminalny (poza być może symbolem początkowym) generuje nieskończenie wiele słów terminalnych.

1.2 Bezkontekstowy czy nie ? — lematy o pompowaniu

1. Udowodnić, że żaden nieskończony podzbiór języka $L = \{a^n b^n c^n \; : \; n \geq 1\}$ nie jest językiem bezkontekstowym.

2. Udowodnić, że dopełnienie języka z poprzedniego zadania jest językiem bezkontekstowym.

3. Udowodnić, że jeśli alfabet Σ ma co najmniej dwie litery, to język $L = \{ww \; : \; w \in \Sigma^*\}$ nie jest bezkontekstowy, natomiast jego dopełnienie $\Sigma^* - L$ jest językiem bezkontekstowym.

4. Dowieść, że dla każdego ustalonego k dopełnienie języka $\{w^k : w \in \Sigma^*\}$ jest językiem bezkontekstowym.

5. Udowodnić, że język $L = \{xcx \; : \; x \in (a{+}b)^* \}$ nie jest językiem bezkontekstowym.

6. Udowodnić, że dopełnienie języka z poprzedniego zadania jest językiem bezkontekstowym.

7. Pokazać, że język dopasowywania wzorca $L = \{xcy \; : \; x, y \in (a + b)^*, \; y \in Subwords(x)\}$ nie jest bezkontekstowy. Czy dopelnienie tego języka jest bezkontekstowe ?

8. Pokazać, że język $L = \{xcy^R \; : \; x, y \in (a + b)^*, \; y \in Subwords(x)\}$ jest bezkontekstowy.

9. (*) Pokazać, że dopełnienie języka z poprzedniego zadania nie jest językiem bezkontekstowym.

10. Udowodnić, że język $L = \{a^i b^j c^k \; : i \neq j, \; i \neq k, \; j \neq k \}$ nie jest bezkontekstowy.

    Czy jego dopełnienie jest bezkontekstowe ?

11. Czy język $L = \{a^i b^j a^i b^j \; : i, j \geq 1 \}$ jest bezkontekstowy ?

    Czy jego dopełnienie jest językiem bezkontekstowym ?

12. Udowodnić, że język $L = \{ww^R w : w \in (a + b)^*\}$ nie jest bezkontekstowy.

    Czy jego dopełnienie jest bezkontekstowe ?

13. Pokaż, że język $\{x\#y^R \; : \; x, y \in \{0,1\}^+, \; [x]_2 + 1 = [y]_2\}$ jest bezkontekstowy.

14. Pokaż, że język $\{x\#y \; : \; x, y \in \{0,1\}^+, \; [x]_2 + 1 = [y]_2\}$ nie jest bezkontekstowy.

15. Które z następujących języków są bezkontekstowe ?

(a) $\{a^m b^n : m < n < 2m\}$

(b) $(a+b)^* - \{(a^n b^n)^n : n \geq 1\}$

(c) $\{ww^R w : w \in (a+b)^*\}$

(d) $\{a^x b a^y b a^z : x + y = z\}$

(e) $\{a^x b a^y b a^z : x \cdot y = z\}$

16. Dowiedź, że następujące języki nie są bezkontekstowe:

    (a) $\{a^i b^j a^k : j = \max\{i, k\}$

    (b) $\{a^i b^i c^k : k \neq i\}$.

17. Czy język
$$\{\mathrm{bin}(n)\,\$\,\mathrm{bin}(n^2)^R \;:\; n \in N\},$$
gdzie $\mathrm{bin}(n) \in \{0,1\}^*$ jest binarnym przedstawieniem liczby n, jest bezkontekstowy?

18. (a) Udowodnij, że zbiór tautologii nad ustalonym skończonym zbiorem zmiennych jest bezkontekstowy (stanowi to uogólnienie zadania (1g) z sekcji 1.1).

    (b) Formuły nad przeliczalnym zbiorem zmiennych można przedstawić jako język nad skończonym alfabetem, przyjmując indeksowanie zmiennych. Dokładniej, przyjmijmy, że zbiór wszystkich formuł jest generowany przez gramatykę
$$\begin{array}{rcl} F & \to & \mathbf{true} \mid \mathbf{false} \mid V \mid (F \vee F) \mid (F \wedge F) \mid (\neg F) \\ V & \to & xI \\ I & \to & 0 \mid 1J \\ J & \to & J0 \mid J1 \mid \epsilon \end{array}$$
Np. $((x101 \vee (\neg x0)) \wedge (\neg(\mathbf{false} \vee x101)))$ jest formułą.
Udowodnij, że zbiór wszystkich tautologii nie jest bezkontekstowy[1].

19. Sformuluj $uvwxy$-lemat o pompowaniu dla języków liniowych w którym $|uvxy| = O(1)$.
Dowiedź, że język $\{a^i b^i c^j d^j : i, j \in \mathbf{N}\}$ nie jest liniowy.

20. Dowiedź, że $L = \{x \in (a+b)^*, \#_a(x) = \#_b(x)\}$ nie jest liniowym językiem bezkontekstowym.

21. Udowodnij, że zbiór słów w nad alfabetem $\{a, b\}$ mających tę samą liczbę liter a co b nie jest liniowym językiem bezkontekstowym.

[1] Stwierdzenie to wynika łatwo z hipotezy $\mathrm{P} \neq \mathrm{NP}$, ale należy je dowieść bez tej hipotezy.

1.3 Automaty ze stosem

1. Skonstruować automaty ze stosem rozpoznające poznane wcześniej języki bezkontekstowe: zbiór palindromów, zbiór poprawnie uformowanych ciągów nawiasów, zbiór słów, które mają dwa razy więcej b niż a, zbiór ciągów, które nie są postaci ww.

2. Dla liczby naturalnej n, niech $\mathrm{bin}(n) \in \{0,1\}^*$ będzie binarnym przedstawieniem liczby n. Skonstruować automat ze stosem rozpoznający język
$$\{\mathrm{bin}(n)\,\$\,\mathrm{bin}(n+1)^R \;:\; n \in N\}$$

3. Skonstruować automat ze stosem rozpoznający język
$$\{\mathrm{bin}(n)\,\$\,\mathrm{bin}(3*n)^R \;:\; n \in N\}$$
Uogólnić tezę zadania[2].

4. Dowieść, że dla każdego automatu ze stosem A, można skonstruować automat ze stosem *o dwóch stanach* A', taki że $L(A) = L(A')$.

5. Dowieść, że automatowi A' z poprzedniego zadania można postawić dalszy wymóg, że każde przejście jest postaci
$$q, a, Z \;\rightarrow_{A'}\; q', \alpha$$
gdzie $|\alpha| \leq 2$ (q, q' dowolne).

6. Dowieść, że dla każdego automatu ze stosem A, można skonstruować równoważny mu automat ze stosem A'', w którym każde przejście jest postaci *push* lub *pop*, tzn.
$$q, a, Z \;\rightarrow_{A''}\; q', YZ$$
lub
$$q, a, Z \;\rightarrow_{A''}\; q', \epsilon$$
Czy dla takich automatów można nadal ograniczyć liczbę stanów?

7. Mając dany automat ze stosem akceptujący język L, skonstruować automaty ze stosem akceptujące następujące języki:
   - $\mathrm{Prefix}(L) = \{w \;:\; (\exists v) wv \in L\}$
   - $\mathrm{Suffix}(L) = \{w \;:\; (\exists u) uw \in L\}$
   - $\mathrm{Subword}(L) = \{w \;:\; (\exists u, v) uwv \in L\}$
   - $L^R = \{w^R \;:\; w \in L\}$
   - (*) $\mathrm{Cycle}(L) = \{vw \;:\; wv \in L\}$

[2] Można zacząć od przykładu $\{\mathrm{dec}(n)\,\$\,\mathrm{dec}(2006*n)^R \;:\; n \in N\}$, gdzie $\mathrm{dec}(n) \in \{0,1,\ldots,9\}^*$ jest dziesiętnym przedstawieniem liczby n.

8. Mając dany automat ze stosem akceptujący język L i automat skończony akceptujący język R, skonstruować automaty ze stosem akceptujące następujące języki:
   - LR^{-1}
   - $R^{-1}L$
   - $L \cap R$

9. Dla danych językow regularnych L i M, skonstruować automat ze stosem rozpoznający język $\bigcup_{i \in N}(L^i \cap M^i)$. Uwaga: ten zbiór nie musi być regularny.

10. Dowieść, że dla każdego automatu ze stosem A istnieje stała C (zależna od automatu), taka, że dla każdego słowa $w \in Z(A)$, istnieje obliczenie akceptujące (przez pusty stos) o długości $\leq C|w|$. Wskazówka: oszacować wysokość stosu w obliczeniu akceptującym.

11. Niech A będzie automatem ze stosem. Dowieść, że zbiór słów, które są możliwymi zawartościami stosu automatu A, jest językiem regularnym. Formalnie, mamy na myśli zbiór

$$\{\alpha \in \Gamma^* \,:\, (\exists w, v \in \Sigma^*)(\exists q \in Q)\, q_I, w, Z_I \vdash_A \, q, v, \alpha\}$$

    Wywnioskować stąd, że zbiór słów, które są możliwymi zawartościami stosu automatu A w jakimś obliczeniu *akceptującym*, jest językiem bezkontekstowym.

12. Automat ze stosem $A = (\Sigma, \Gamma, Q, q_I, Z_I, \delta, F)$ nazywamy *deterministycznym*, jeśli w każdej sytuacji możliwy jest co najwyżej jeden ruch. Dokładniej:
    - jeśli, dla pewnej pary q, Z, zachodzi $q, \epsilon, Z \rightarrow_A p, \alpha$ przy pewnych p, α, to dla żadnego $\sigma \in \Sigma$, nie zachodzi $q, \sigma, Z \rightarrow_A p', \alpha'$, dla żadnych p', α';
    - dla każdych q, σ, Z, istnieje co najwyżej jedna para p, α, taka że $q, \sigma, Z \rightarrow_A p, \alpha$.

    Język bezkontekstowy jest *deterministyczny* jeśli jest rozpoznawany przez pewien deterministyczny automat ze stosem (w sensie stanów akceptujących, tzn. $L = L(A)$). Dowieść, że język $\{a^n b^n \,:\, n \in \mathbf{N}\} \cup \{a^n b^{2n} \,:\, n \in \mathbf{N}\}$ nie jest deterministycznym językiem bezkontekstowym.

13. Wykazać, że zbiór palindromów nad alfabetem dwuelementowym nie jest deterministycznym językiem bezkontekstowym.

1.4 Własności języków bezkontekstowych

1. Podaj przykład języka bezkontekstowego L t.że język $L = \{x \,:\, (\exists y)\, |x| = |y| \wedge xy \in L\}$ nie jest bezkontekstowy.

2. Udowodnić, że przecięcie języka (deterministycznego) bezkontekstowego z regularnym jest też językiem (deterministycznym) bezkontekstowym.

3. Przeplotem słów w i v nazwiemy dowolne słowo długości $|w|+|v|$, które można rozbić na rozłączne podciągi w i v. L i M jest zbiór wszystkich możliwych przeplotów słów $w \in L$, $v \in M$. Język ten oznaczamy $L \parallel M$. Udowodnij, ze przeplot języka bezkontekstowego i regularnego jest językiem bezkontekstowym. Podaj przykład języków bezkontekstowych L i M, dla których $L \parallel M$ nie jest językiem bezkontekstowym.

4. Domknięcie przeplotne języka L określamy przez $L^\sharp = L \cup (L \parallel L) \cup (L \parallel L \parallel L) \cup \ldots$. Wykazać, że operacja domknięcia przeplotnego języka skończonego może dać w wyniku język który nie jest bezkontekstowy.

5. Udowodnić, że jeśli X, Y są językami regularnymi to język
$$L \;=\; \textstyle\sum_{n\geq 1} X^n \cap Y^n$$
jest bezkontekstowy.

6. Podać przykład języków regularnych X, Y t.że
$$\textstyle\sum_{n\geq 1}(X^n \cap Y^n) \;=\; \{a^n b^n \;:\; n \geq 1\}$$

7. Podaj języki regularne X, Y, Z t.że język
$$\textstyle\sum_{n\geq 1}(X^n \cap Y^n \cap Z^n)$$
nie jest bezkontekstowy.

8. Wskaż język bezkontekstowy L, dla którego $\sqrt{L} = \{w : ww \in L\}$ nie jest językiem bezkontekstowym.

9. Podaj przykład języka bezkontekstowego takiego. że $\{x \;:\; x^k \in L$ dla pewnego $k\}$ nie jest bezkontekstowy.

10. Rozważmy następujący morfizm:
$$h(a) = a,\; h(b) = b,\; h(a') = a,\; h(b') = b$$
oraz $h(x) = \epsilon$ dla symboli $x \in \{a, b, a', b'\}$ dla których morfizm nie został zdefiniowany powyżej. Wykazać, że język $EQ(h, g) = \{w : h(w) = g(w)\}$ nie jest bezkontekstowy.

11. Udowodnij, że jeśli U jest regularny to następujący język jest bezkontekstowy
$$\{xy^R \;:\; x \neq y,\; xy \in U\}$$

12. Język ma własność prefiksową, gdy dla każdych dwóch słow z tego języka jedno z nich jest prefiksem drugiego. Wykaż, że jeśli język bezkontekstowy ma własność prefiksową to jest on regularny.

13. Niech $L \subseteq \{a, b\}^*$ będzie językiem regularnym oraz h, g będą morfizmami. Udowodnić, że następujący język jest liniowym językiem bezkontekstowym
$$\{h(u)c(g(u))^R \;:\; u \in L\}$$

14. Udowodnij, że przecięcie liniowego języka bezkontekstowego z regularnym jest językiem liniowym.

15. Dla języka L określamy $\mathrm{Min}(L)$ jako zbiór słów w L minimalnych ze względu na porządek bycia prefiksem. (Zatem $u \in \mathrm{Min}(L) \Leftrightarrow$ nie istnieją $v \in L$ oraz $w \neq \epsilon$ takie, że $vw \in L$.) Udowodnij, że dla deterministycznego języka bezkontekstowego L język $\mathrm{Min}(L)$ jest również deterministycznym językiem bezkontekstowym.

16. Niech $L = \{a^i b^j c^k : k \geq i \textit{ lub } k \geq j\}$. Pokaż, że $Min(L)$ nie jest językiem bezkontekstowym.

17. Zbiór $\mathrm{Max}(L)$ określamy analogicznie jak zbiór $\mathrm{Min}(L)$ w zadaniu 15. Podaj przykład języka bezkontekstowego L, dla którego $\mathrm{Max}(L)$ nie jest językiem bezkontekstowym.

18. Niech h_1, h_2 będą morfizmami takimi, że alfabet wyjściowy nie zawiera \$. Wykaż, że języki $\{x\$y^R : h_1(x) = h_2(x)\}$ i $\{x\$y^R : h_1(x) \neq h_2(x)\}$ są liniowymi językami bezkontekstowymi.

19. Podzbiór $M \subseteq \mathbf{N}^k$ nazywamy *liniowym* jeśli można go przedstawić $M = \{\vec{a} + n \cdot \vec{b} : n \in \mathbf{N}\}$, dla pewnych wektorów $\vec{a}, \vec{b} \in \mathbf{N}^k$, a *semi-liniowym*, jeśli jest sumą skończenie wielu zbiorów semi-lniowych.

    (a) Udowodnij, że zbiór długości słów języka bezkontekstowego jest semi-liniowy.

    (b) (*) Udowodnij *twierdzenie Parikha* głoszące, że dla języka bezkontekstowego $L \subseteq \Sigma^*$ zbiór ($\subseteq \mathbf{N}^{|\Sigma|}$) wektorów liczby wystąpień liter z Σ w słowach z L jest semiliniowy.