Egzamin z JAiO, 23 czerwca 2009

Komplet punktów za przedmiot to $17\frac{1}{2}$ punktu, z czego 40% to kolokwium (7 punktów), a 60% to egzamin ($10\frac{1}{2}$ punktu). Zadanie 2(*) jest nieobowiązkowe, pozwala zdobyć więcej niż komplet punktów.

1. **[3 punkty]** Dla słów $w, v \in A^*$ napiszemy $w \sim v$ jeśli jeśli słowo w można otrzymać z v po skończonej ilości zamian, gdzie dowolną literę $x \in A$ zamieniamy na xx, czy też xx zamieniamy na x. Na przykład

$$aabbbaaabba \sim ababa \qquad abbaa \sim aaba \ .$$

   Dla $L \subseteq A^*$, niech $L_\sim = \{w : w \sim v \text{ dla pewnego } v \in L\}$.

   (a) Pokazać, że jeśli L jest regularny, to $L_\sim$ też;

   (b) Czy jeśli $L_\sim$ jest regularny, to L też?

2. Dla języka $L \subseteq A^*$ oraz liczby $k \in \mathbb{N}$, niech $sito_k(L)$ oznacza język

$$\{a_1 \cdots a_n : a_1, \ldots, a_n \in A, a_1 w_1 \cdots a_n w_n \in L \text{ dla pewnych } w_1, \ldots, w_n \in A^k\}.$$

   (a) **[$2\frac{1}{2}$ punktu]** Pokazać, że jeśli L jest bezkontekstowy, to dla dowolnego $k \in \mathbb{N}$, język $sito_k(L)$ też jest bezkontekstowy.

   (*) **[4 punkty]** Pokazać, że jeśli L jest regularny, to $\bigcup_k sito_k(L)$ też.

3. **[$2\frac{1}{2}$ punktu]** Czy bezkontekstowy jest zbiór skończonych prefiksów słowa

$$a^1 b^2 a^3 b^4 \cdots \quad ?$$

4. **[$2\frac{1}{2}$ punktu]** Pokazać, że następujący problem jest nierozstrzygalny:

   Dana: gramatyka bezktontekstowa $\mathcal{G}$ nad alfabetem A.
   Pytanie: czy $\mathcal{G}$ generuje wszystkie, oprócz najwyżej skończenie wielu, słowa z A^*?

   Można skorzystać z nierozstrzygalności poniższego problemu:

   Dana: gramatyka bezktontekstowa $\mathcal{G}$ nad alfabetem A.
   Pytanie: czy $\mathcal{G}$ generuje wszystkie słowa z A^*?

Zadanie 1

[**3 punkty**] Dla słów $w, v \in A^*$ napiszemy $w \sim v$ jeśli jeśli słowo w można otrzymać z v po skończonej ilości zamian, gdzie dowolną literę $x \in A$ zamieniamy na xx, czy też xx zamieniamy na x. Na przykład

$$aabbbaaabba \sim ababa \qquad abbaa \sim aaba \ .$$

Dla $L \subseteq A^*$, niech $L_\sim = \{w : w \sim v \text{ dla pewnego } v \in L\}$.

1. Pokazać, że jeśli L jest regularny, to $L_\sim$ też;
2. Czy jeśli $L_\sim$ jest regularny, to L też?

Rozwiązanie

Zacznijmy od punktu (b). Odpowiedź brzmi nie. Na przykład dla nieregularnego języka $L = \{a^n b^n : n \in \mathbb{N}\}$, język $L_\sim$, czyli a^*b^*, jest językiem regularnym.

Zajmijmy się teraz punktem (a). Pokażemy, jak z automatu $\mathcal{A}$ rozpoznającego język L można skonstruować automat $\mathcal{A}_\sim$ rozpoznający język $L_\sim$. Niech Q to zbiór stanów automatu $\mathcal{A}$. Poniższa konstrukcja automatu $\mathcal{A}_\sim$ działa nawet wtedy, gdy automat $\mathcal{A}$ jest niedeterministyczny.

Stany. Stany automatu $\mathcal{A}_\sim$ to $Q \times (A \cup \{\epsilon\})$.

Przejścia. Jeśli automat $\mathcal{A}_\sim$ jest w stanie (q, ϵ) i czyta literę $a \in A$, to może przejść do dowolnego stanu (p, a), gdzie w automacie $\mathcal{A}$ istniało przejście z q do p po literze a. Jeśli jest w stanie (p, a), to może przejśc po literze a z powrotem do stanu (p, a), albo zrobić ϵ-przejście do dowolnego stanu (r, a) lub (r, ϵ), o ile $p = r$ lub w automacie $\mathcal{A}$ istniało przejście z p do r po literze a.

Stany początkowe i akceptujące. Załóżmy, że w automacie $\mathcal{A}$ stanami początkowymi były stany $I \subseteq Q$, a stanami akceptującymi były stany $F \subseteq Q$. Stanami początkowymi w $\mathcal{A}_\sim$ są stany $I \times \{\epsilon\}$, a stanami akceptującymi są stany $Q \times \{\epsilon\}$.

Poprawność konstrukcji. To, że automat $\mathcal{A}_\sim$ rozpoznaje język $L_\sim$, wynika z następującego niezmiennika, którego można dowieść przez indukcję ze względu na długość słowa $w \in A^*$.

> Niech $w \in A^*$, $p, q \in Q$ oraz $x, y \in A \cup \{\epsilon\}$. W automacie $\mathcal{A}_\sim$ można przejśc ze stanu (q, x) do stanu (p, y) po słowie w wtedy i tylko wtedy, gdy w automacie $\mathcal{A}$ można przejśc ze stanu q do stanu p po pewnym słowie $x^i v y^j$, dla pewnych $i, j \in \mathbb{N}, v \in A^*, v \sim w$.

Zadanie 2

Dla języka $L \subseteq A^*$ oraz liczby $k \in \mathbb{N}$, niech $sito_k(L)$ oznacza język

$\{a_1 \cdots a_n : a_1, \ldots, a_n \in A, a_1 w_1 \cdots a_n w_n \in L$ dla pewnych $w_1, \ldots, w_n \in A^k\}$.

1. **[2$\frac{1}{2}$ punktu]** Pokazać, że jeśli L jest bezkontekstowy, to dla dowolnego $k \in \mathbb{N}$, język $sito_k(L)$ też jest bezkontekstowy.

(*) **[4 punkty]** Pokazać, że jeśli L jest regularny, to $\bigcup_k sito_k(L)$ też.

Rozwiązanie 2a

Pokażemy jak z automatu ze stosem $\mathcal{A}$ rozpoznającego język L stworzyć automat ze stosem $\mathcal{A}_k$ rozpoznający język $sito_k(L)$. Konstrukcja działa nawet wtedy, gdy automat $\mathcal{A}$ jest niedeterministyczny. Niech Q to stany automatu $\mathcal{A}$, a jego zbiór przejść to

$$\delta \subseteq Q \times (A \cup \{\epsilon\}) \times I \times Q$$

gdzie I to zbiór instrukcji operujących na stosie. Poniżej opiszemy automat $\mathcal{A}_k$.

Stany. W automacie $\mathcal{A}_k$ stany to $Q \times \{0, \ldots, k\}$. Idea jest taka, że stany $Q \times \{0\}$ służą do czytania wejścia, a stany $Q \times \{1, \ldots, k\}$ służą do zgadywania, za pomocą ϵ-przejść, tych liter, które zostały skasowane przez operację $sito_k$.

Przejścia. Jeśli w automacie $\mathcal{A}$ jest przejście (q, a, i, p), dla $a \in A \cup \{\epsilon\}$, to w automacie $\mathcal{A}_k$ jest przejście (które jest ϵ-przejściem dla $a = \epsilon$)

$$((q, 0), a, i, (p, 1))$$

oraz następujące przejścia (które są ϵ-przejściami niezależnie od a)

$$((q, 1), \epsilon, i, (p, 2)), \cdots, ((q, k-1), \epsilon, i, (p, k)), ((q, k), \epsilon, i, (p, 0)) \ .$$

Stany początkowe i akceptujące. Jeśli stanem początkowym automatu $\mathcal{A}$ jest q, to stanem początkowym automatu $\mathcal{A}_k$ jest $(q, 0)$. Załóżmy, że automat $\mathcal{A}$ akceptuje przez stan akceptujący, którym jest p. Wówczas automat $\mathcal{A}_k$ też akceptuje przez stan akceptujący, którym jest $(p, 0)$.

Poprawność konstrukcji. To, że automat $\mathcal{A}_k$ rozpoznaje język $sito_k(L)$, wynika z następującego niezmiennika, którego można dowieść przez indukcję ze względu na n.

> Niech $a_1 \cdots a_n \in A^*$ oraz $q \in Q$. Automat $\mathcal{A}_k$ po wczytaniu słowa $a_1 \cdots a_n$ może być w stanie $(q, 0)$ ze stosem s wtedy i tylko wtedy gdy istnieją słowa $w_1, \ldots, w_n \in A^k$, takie że po wczytaniu słowa $a_1 w_1 \cdots a_n w_n$ automat $\mathcal{A}$ może być w stanie q ze stosem s.

Rozwiązanie 2*

Załóżmy, że język L jest rozpoznawany przez automat deterministyczny $\mathcal{A}$, o stanach Q i o przejściach

$$\{\delta_a \subseteq Q \times Q\}_{a \in A}$$

Skonstruujmy najpierw automat $\mathcal{A}_k$ dla języka $sito_k(L)$. Jego stany są takie same jak w $\mathcal{A}$, takie same ma też stany początkowe i akceptujące. Jedyna zmiana to relacja przejść, oznaczmy ją γ_a, która wygląda tak:

$$\gamma_a = \bigcup_{a_1, \ldots, a_k \in A} \delta_a \circ \delta_{a_1} \circ \cdots \circ \delta_{a_k}$$

Kluczowa obserwacja jest taka: skoro każdy automat $\mathcal{A}_k$ ma stany Q i alfabet A, to różnych automatów $\mathcal{A}_k$ jest skończenie wiele. W związku z tym, różnych języków $sito_k(L)$ jest skończenie wiele, a więc

$$\bigcup_k sito_k(L)$$

jest regularny, jako skończona suma języków regularnych.

Zadanie 3

[2$\frac{1}{2}$ punktu] Czy bezkontekstowy jest zbiór skończonych prefiksów słowa

$$a^1 b^2 a^3 b^4 \cdots \quad ?$$

Rozwiązanie

Język ten, nazwijmy go L, nie jest bezkontekstowy. Udowodnimy to korzystając z lematu o pompowaniu. Niech M to stała z lematu. Załóżmy, że M jest parzysta, dla nieparzystej dowód jest taki sam. Rozważmy słowo

$$a^1 b^2 a^3 \cdots a^{M-1} b^M \in L \ .$$

Poniżej będziemy mówić o blokach, są to maksymalne podsłowa składające się z tych samych liter. W powyższym słowie mamy bloki $a^1, b^2, \ldots, a^{M-1}, b^M$. Zgodnie z lematem o pompowaniu, istnieje podział

$$u_1 v_1 u v_2 u_2 = a^1 b^2 a^3 \cdots a^{M-1} b^M$$

taki, że po i-krotnym pompowaniu słowo pozostaje w L, czyli

$$u_1 (v_1)^i u (v_2)^i u_2 \in L \qquad \text{dla każdego } i \in \{0, 1, \ldots\}.$$

Pokażemy, że powyższa własność nie zachodzi dla $i = 3$, która to sprzeczność dowodzi, że L nie mógł być bezkontekstowy. Jeśli v_1 albo $v_1 v_2$ zawiera same litery a (czy też same litery b), to 3-krotne (2-krotne też) pompowanie wydłuża jeden blok, powodując wypadnięcie z L. W przeciwnym przypadku, jedno ze słów v_1 czy v_2 zawiera zarówno a jak i b. Wówczas 3-krotne pompowanie tworzy dwa bloki o tej samej długości, powodując wypadnięcie z L.

Zadanie 4

[2$\frac{1}{2}$ punktu] Pokazać, że następujący problem jest nierozstrzygalny:

> **Dana**: gramatyka bezktontekstowa $\mathcal{G}$ nad alfabetem A.
> **Pytanie**: czy $\mathcal{G}$ generuje wszystkie, oprócz najwyżej skończenie wielu, słowa z A^*?

Można skorzystać z nierozstrzygalności poniższego problemu:

> **Dana**: gramatyka bezktontekstowa $\mathcal{G}$ nad alfabetem A.
> **Pytanie**: czy $\mathcal{G}$ generuje wszystkie słowa z A^*?

Rozwiązanie

Przedstawimy redukcję z drugiego problemu do pierwszego problemu. Inaczej mówiąc, dla każdej gramatyki $\mathcal{G}$ obliczmy gramatykę $\mathcal{G}'$ taką, że gramatyka $\mathcal{G}$ generuje wszystkie słowa wtedy i tylko wtedy gdy gramatyka $\mathcal{G}'$ generuje prawie wszystkie słowa. Taka redukcja pokazuje, że gdyby pierwszy problem był rozstrzygalny, to drugi też, a więc pierwszy jest nierozstrzygalny.

Niech A będzie alfabetem gramatyki $\mathcal{G}$, i niech x nowa litera $x \notin A$. Gramatyka $\mathcal{G}'$ będzie tak zaprojektowana, by generowała słowa postaci

$$x^{n_0}a_1x^{n_1}a_2x^{n_2}\cdots x^{n_{k-1}}a_kx^{n_k} \qquad \text{dla } a_1\cdots a_k \in A^* \text{ generowanego przez } \mathcal{G}.$$

Inaczej mówiąc, do $\mathcal{G}'$ trafiają słowa z $\mathcal{G}$, z powstawianymi w dowolnych miejscach literami x. Łatwo z gramatyki $\mathcal{G}$ obliczyć gramatykę $\mathcal{G}'$, na przykład zamieniając w $\mathcal{G}$ każde wystąpienie terminala $a \in A$ na XaX, gdzie X to nowy nieterminal o regułach $X \to xX|\epsilon$.

By dowieść że przekształcenie, które gramatyce $\mathcal{G}$ przyporządkowuje gramtykę $\mathcal{G}'$, jest redukcją drugiego problemu do pierwszego, należy pokazać, że $\mathcal{G}$ generuje wszystkie słowa w A^* wtedy i tylko wtedy, gdy $\mathcal{G}'$ generuje prawie wszystkie słowa w $(A \cup \{x\})^*$. Jeśli $\mathcal{G}$ generuje wszystkie słowa w A^*, to $\mathcal{G}'$ generuje wszystkie słowa w $(A \cup \{x\})^*$. Jeśli $\mathcal{G}$ nie generuje pewnego słowa $w \in A^*$, to $\mathcal{G}'$ nie generuje żadnego z nieskończenie wielu słów postaci x^*w.