Określenie regularności języka $L = \{ww^R : w \in \{0,1\}^+, w \neq \varepsilon\}$

Zofia Tarchalska, indeks: 279699

1 Treść zadania

Rozważmy język

$$L = \{\, ww^R x \mid w, x \in \{0,1\}^* \,\},$$

gdzie w^R oznacza słowo w odwrócone. Należy ustalić, czy L jest językiem regularnym.

2 Idea dowodu

Twierdzenie Myhilla-Nerode'a orzeka, że język L jest regularny wtedy i tylko wtedy, gdy relacja sufiksowej nieodróżnialności generuje nam skończenie wiele klas abstrakcji. Pokażemy, że L nie jest regularny poprzez znalezienie sufiksu, który rozróżnia dwa słowa dla każdej pary indeksów co daje nam nieskończenie wiele klas.

3 Dowód

Dla $n \geq 0$ zdefiniujmy

$$s_n = (01)^n, \qquad t_n = (10)^n 1.$$

Zauważmy najpierw, że słowa $s_n t_n$ należą do L, bo

$$s_n t_n = (01)^n (10)^n 1 = w\, w^R x$$

dla $w = (01)^n$ i $x = 1$. Weźmy dwie różne liczby naturalne n i m i bez utraty ogólności załóżmy $n < m$. Pokażemy, że sufiks t_n rozróżnia s_n i s_m, tzn.

$$s_n t_n \in L \quad \text{oraz} \quad s_m t_n \notin L.$$

Załóżmy przeciwnie, że $s_m t_n \in L$. Wtedy istnieją słowa $u, v \in \{0,1\}^*$ takie, że

$$s_m t_n = u\, u^R v.$$

Niech $k = |u|$. Rozważymy pozycję u względem prefiksu s_m.

3.1 Przypadek 1: $k < |s_m|$

W tym przypadku część u kończy się wewnątrz naprzemiennego wzoru $s_m = (01)^m$. Ostatni znak u musi być zgodny z naprzemiennym wzorcem s_m, a pierwszy znak u^R jest równy temu ostatniemu znakowi. To prowadzi do wymogu, by w odpowiednich pozycjach, gdzie łączy się u i u^R znajdowały się dwa takie same bity pod rząd. Jednak po każdym 0 w s_m występuje 1, a po każdym 1 występuje 0. To równocześnie nie może zajść. Zatem nie może zajść $k < |s_m|$.

3.2 Przypadek 2: $k \geq |s_m|$

Jeśli $k \geq |s_m|$, to u zawiera cały prefiks s_m. W takim wypadku u^R musi znaleźć się w pozostałej części słowa $s_m t_n$. Jednak ponieważ $n < m$, pozostała część t_n ma długość mniejszą niż k, więc nie wystarczy miejsca na pełne u^R. To również prowadzi do sprzeczności.

Z powyższych przypadków wynika, że założenie $s_m t_n \in L$ jest fałszywe, stąd $s_m t_n \notin L$.

4 Wniosek

Dla dowolnych dwóch różnych indeksów $n \neq m$ istnieje sufiks (np. t_n), który rozróżnia s_n i s_m. Oznacza to, że słowa s_n należą do nieskończenie wielu różnych klas równoważności Myhilla–Nerode. W konsekwencji L nie ma skończonej liczby klas równoważności i nie jest akceptowany przez żaden deterministyczny automat skończony. Zatem L nie jest językiem regularnym.