Języki regularne i automaty

Łukasz Kamiński
l.kaminski5@student.uw.edu.pl

9 lipca 2020r

Podstawowe definicje

Alfabet - skończony zbiór 'symboli'
Słowo (nad danym alfabetem) - skończony ciąg symboli danego alfabetu
Język - zbiór słów nad danym alfabetem (może być nieskończony)

Podstawowe definicje

Alfabet - skończony zbiór 'symboli'
Słowo (nad danym alfabetem) - skończony ciąg symboli danego alfabetu
Język - zbiór słów nad danym alfabetem (może być nieskończony)

Przykład: alfabetem może być $A = \{1, q, \clubsuit\}$ zaś językiem nad A $L = \{11, q, 1q1, \clubsuit\clubsuit 1qq\}$

Podstawowe definicje

Alfabet - skończony zbiór 'symboli'
Słowo (nad danym alfabetem) - skończony ciąg symboli danego alfabetu
Język - zbiór słów nad danym alfabetem (może być nieskończony)

Konwencja: jeżli w pewnym słowie pewien symbol a występuje n razy pod rząd to zamiast pisać $aa\dots a$ można napisać a^n.
(Przykład $aaaba = a^3ba$)

Wyrażenia regularne

Niech L, K będą językami. Rozważmy następujące operacje

- $L + K$ to **suma** języków L, K (tak będziemy tutaj oznaczać $L \cup K$)
- $LK = \{vw | v \in L, w \in K\}$ to **konkatenacja** języków L, K. vw to konkatenacja (czyli sklejenie) słów v, w. Np. dla $v = aab$ oraz $w = ab$ mamy $vw = aabab$.
- $L^* = \{l_1 l_2 \ldots l_n | n \in \mathbb{N}, l_1, l_2, \ldots, l_n \in L\}$ to **gwiazdka Kleene'go**. (czyli każde ze słów w L^* to kilka sklejonych ze sobą słów z L)

Konwencja: jeśli języki są jednoelementowe to można pominąć nawiasy w zapisie. Np. $\{a\} + \{ab\} = a + ab$ oraz $\{a\}^* = a^*$.

Wyrażenia regularne cd.

Wyrażenie reularne definiuje się następująco

- słowo puste ε jest wyrażeniem regularnym
- symbole danego alfabetu są wyrażeniami regularnymi
- jeśli A, B są wyrażeniami regularnymi to $A + B$, AB, A^* też są wyrażeniami regularnymi. (Oczywiście dla jednoznaczności napisu możemy używać też nawiasów)

Wszystkie wyrażenia regularne spełniają pewien z powyższych warunków.

Wyrażenie reularne definiuje się następująco

- słowo puste ε jest wyrażeniem regularnym
- symbole danego alfabetu są wyrażeniami regularnymi
- jeśli A, B są wyrażeniami regularnymi to $A + B$, AB, A^* też są wyrażeniami regularnymi. (Oczywiście dla jednoznaczności napisu możemy używać też nawiasów)

Wszystkie wyrażenia regularne spełniają pewien z powyższych warunków.

Każde wyrażenie regularne definiuje pewien język przy interpretacji takiej jak na poprzednim slajdzie.

Przykłady wyrażeń regularnych

- a^* definiuje język $\{\varepsilon, a, aa, aaa, \dots\}$
- $(a + b + \varepsilon)a$ definiuje język $\{aa, ba, a\}$
- $(0 + 1)^*0$ definiuje język liczb zapisanych w systemie binarnym i podzielnych przez 2.

Ćwiczenie 1.

Ćwiczenie

Znaleźć wyrażenie regularne opisujące liczby zapisane w systemie binarnym i podzielne przez 3.

Ćwiczenie

Znaleźć wyrażenie regularne opisujące liczby zapisane w systemie binarnym i podzielne przez 3.

$$(1(00)^*1 + 10(1 + 00)^*01 + 0)^*$$

Język regularny to język, który może być zdefiniowany przez pewne wyrażenie regularne.

Ćwiczenie 2.

Ćwiczenie

Udowodnij, że każdy skończony język jest regularny.

Ćwiczenie 3.

Treść

Niech L będzie językiem regularnym. Udowodnij, że regularny jest też język słów z L zapisanych od tyłu.

Własnośći

Czy umiemy odpowiedzieć na następujące pytania:

- Niech L, K będą językami regularnymi (nad alfabetem A). Czy $L \cap K$ jest regularny?
- Czy dopełnienie języka regularnego jest językiem regularnym?
- Czy różnica języków regularnych jest językiem regularnym?
- Czy język $\{a^n b^n | n \in \mathbb{N}\}$ jest regularny?
- Czy język liczb podzielnych przez k zapisanych w systemie o podstawie n jest regularny?
- Jak stworzyć algorytm, który stwierdza czy dane słowo należy do danego języka regularnego?

Automaty

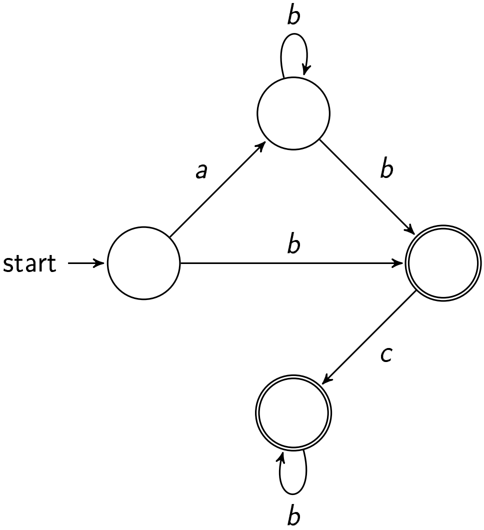
Automat to abstrakcyjna maszyna która składa się z:

- skończonego zbioru **stanów** spośród których niektóre są **początkowe** a niektóre są **akceptujące**.
- przejść pomiędzy stanami

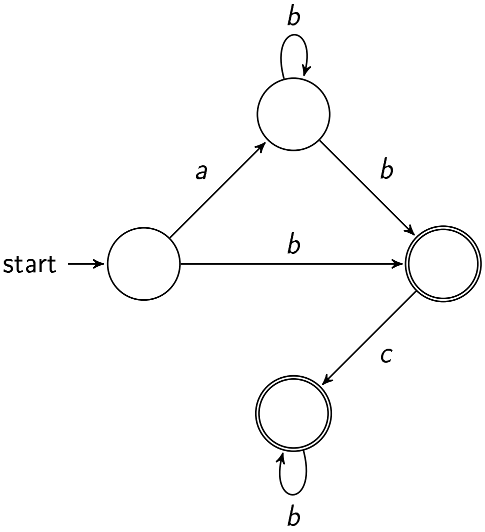
Na początku automat jest w jednym ze stanów początkowych. Automat wczytuje słowo symbol po symbolu i za każdym razem przechodzi do któregoś ze stanów. Jeśli nie istnieje odpowiednie przejście to automat się zawiesza. Jeśli po wczytaniu całego słowa automat znajdzie się w stanie akceptującym, to mówimy, że **akceptuje** to słowo.

Język rozpoznawany przez dany automat to zbiór słów, które dany automat akceptuje.

Przykład automatu



Przykład automatu



$$(b + ab^* b)(\varepsilon + cb^*)$$

Automat deterministyczny - automat który ma najwyżej jeden bieg po każdym słowie. Dokładniej, automat taki ma dokładnie jeden stan początkowy oraz dla z każdego stanu dla danego wczytanego symbolu jest najwyżej jedno przejście.

Automat niedeterministyczny - automat w którym dozwolone są rzeczy wymienione powyżej.

Ćwiczenia

Znajdź automat rozpoznający język

- a^*
- $(ab^* + c)^*$
- słów nad alfabetem $\{a, b, c\}$ które mają dokładnie 3 litery a
- liczb zapisanych w systemie dziesiętnym podzielnych przez 3

Twierdzenie

Dany jest język L. Następujące warunki są równoważne:

- L jest regularny
- L jest rozpoznawany przez automat niedeterministyczny
- L jest rozpoznawany przez automat deterministyczny

Niech L, K będą językami regularnymi (nad alfabetem A).
Udowodnij, że regularne są

- $L \cap K$
- dopełnienie języka L (czyli $A^* \setminus L$)
- język słów z L zapisanych od tyłu

Twierdzenie

Niech L będzie nieskończonym językiem regularnym. Wówczas istnieje taka stała n, że dla dowolnego słowa w o długości większej niż n istnieją x,y,z, że $w = xyz$ oraz

- $y \neq \varepsilon$
- $|xz| \leq n$ ($|xy|$ to długość xy)
- dla każdego $k \geq 0$ mamy $xy^k z \in L$

Lemat o pompowaniu

Twierdzenie

Niech L będzie nieskończonym językiem regularnym. Wówczas istnieje taka stała n, że dla dowolnego słowa w o długości większej niż n istnieją x, y, z, że $w = xyz$ oraz

- $y \neq \varepsilon$
- $|xz| \leq n$ ($|xy|$ to długość xy)
- dla każdego $k \geq 0$ mamy $xy^k z \in L$

Rozważmy deterministyczny automat rozpoznający L. Niech n będzie liczbą jego stanów. Niech w będzie słowem długości większej niż n. Bieg automatu po słowie w zawiera pewien stan S dwukrotnie. Przyjmujemy za x fragment w wczytany aż pierwszy raz doszliśmy do S, za z przyjmiemy fragment w wczytany po ostatnim dojściu do stanu S zaś y to reszta słowa. Aby zagwarantować również drugi punkt twierdzenia przyjmujemy za S pierwszy z powtarzających się stanów w biegu po w.

Zastosowanie lematu o pompowaniu

Ćwiczenie

Udowodnij, że język $\{a^n b^n | n \in \mathbb{N}\}$ nie jest regularny.

Ćwiczenie

Udowodnij, że język nad alfabetem $\{a, b\}$ słów w których jest tyle samo liter a oraz b nie jest regularny.

Literatura

1. `https://www.mimuw.edu.pl/~szymtor/jao/skrypt.pdf`
2. Damian Niwiński, Wojciech Rytter: *200 Problems in Formal Languages and Automata Theory*