Języki automaty i obliczenia

Definicje

1. Przez **alfabet** rozumiemy dowolny zbiór, którego elementy nazywane są **literami** bądź **symbolami**, podczas tego wykładu dodatkowo zakładamy, że zbiór ten musi być *skończony.*

2. **Słowo** nad alfabetem Σ jest to skończony ciąg liter z alfabetu Σ.
   Przykładowo: 0101011 jest słowem nad alfabetem $\{0, 1\}$, $ababa$ jest słowem nad alfabetem $\{a, b, c\}$.

3. Jeśli słowo w ma długość n, to piszemy $|w| = n$, mówimy też, że słowo w ma n pozycji, liczonych od 1.
   Jeśli $1 \leqslant i \leqslant |w|$, to przez $w[i]$ oznaczamy i-tą literę słowa w.
   Jest tylko jedno słowo długości 0, nazywane słowem pustym i oznaczone symbolem ε.

4. Konkatenacja słów $u = a_1 \ldots a_k$ oraz $v = b_1 \ldots b_l$ to słowo $a_1 \ldots a_k b_1 \ldots b_l$, oznaczane $u \cdot v$ lub uv.

5. Przez Σ^* oznaczamy zbiór wszystkich słów (skończonych) nad alfabetem Σ.

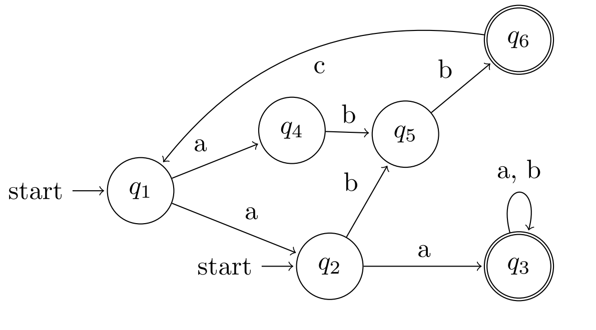
6. **Językiem** nad alfabetem Σ nazywamy dowolny podzbiór $L \subseteq \Sigma^*$.

7. Dla języków L i K definiujemy następujące operacje:
   (a) $L + K \coloneqq L \cup K$
   (b) $LK \coloneqq \{v \cdot w \mid v \in L, w \in K\}$
   (c) $L^* \coloneqq \bigcup_{n \geqslant 0} L^n$, gdzie $L^n = L \ldots L$ oraz $L^0 = \varepsilon$

8. **Wyrażeniami regularnymi** są: $\varnothing$, ε, $a \in \Sigma$.
   Jeśli e, f są wyrażeniami regularnymi to również $e^*, ef, e + f, (e)$ są wyrażeniami regularnymi.
   Wszystkie wyrażenia regularne są tej postaci.
   W naturalny sposób wyrażeniu $\mathcal{E}$ przypisujemy język oznaczany $L(\mathcal{E})$.

9. Język L nazywamy **językiem regularnym** jeśli jest generowany przez pewne wyrażenie regularne, to znaczy, że istnieje takie wyrażenie regularne $\mathcal{E}$, że zachodzi $L = L(\mathcal{E})$.

10. **Automat niedeterministyczny** (*NFA* - nondeterministic finite automaton) jest to model prostego urządzenia mającego skończoną liczbę stanów, reagującego na bodźce w postaci liter alfabetu Σ. Będąc w danym stanie, otrzymując daną literę automat wykonuje tranzycję do jakiegoś stanu (być może tego samego). Wyróżniamy ponadto dwa zbiory stanów (niekoniecznie rozłączne): stany początkowe oraz stany akceptujące. Automaty możemy przedstawiać za pomocą diagramów:



Formalnie jest to piątka $\langle \Sigma, Q, I, F, \delta \rangle$, gdzie Σ to alfabet liter na tranzycjach, Q to zbiór stanów, $I \subseteq Q$ to zbiór stanów początkowych, $F \subseteq Q$ to zbiór stanów końcowych, a $\delta \subseteq Q \times \Sigma \times Q$ to relacja przejścia.

11. **Automat deterministyczny** (*DFA* - deterministic finite automaton) to szczególny przypadek *NFA*, o dodatkowych warunkach:
    (1) relacja możliwych przejść δ jest funkcją $\delta : (Q \times \Sigma) \to Q$
    (2) $|I| = 1$

    Warto odnotować, że *DFA* może mieć liczbę stanów akceptujących różną od 1, a z każdego stanu wychodzi dokładnie $|\Sigma|$ tranzycji, po jednej dla każdej litery alfabetu Σ.
    Dla *DFA* definiujemy $\delta(q, w)$, gdzie $w \in \Sigma^*$ jako stan który otrzymamy idąc z q po słowie w.
12. **Biegiem** po słowie w w automacie $\mathcal{A}$ nazywamy ścieżkę w diagramie automatu $\mathcal{A}$ zaczynającą się w jednym ze stanów początkowych która składa się z tranzycji po kolejnych literach słowa w.
13. **Biegiem akceptującym** nazywamy bieg który kończy się w jednym ze stanów akceptujących automatu.
14. **Językiem automatu** $\mathcal{A}$ oznaczanym przez $L(\mathcal{A})$ nazywamy zbiór wszystkich słów $\omega \in \Sigma^*$ po których $\mathcal{A}$ ma co najmniej jeden bieg akceptujący.

Twierdzenia

1. Dla każdego automatu niedeterministycznego istnieje równoważny mu automat deterministyczny.

   Szkic dowodu:
   Automat potęgowy automatu $\mathcal{A}$ jest automatem deterministycznym równoważnym $\mathcal{A}$.
   Automat potęgowy konstruujemy biorąc jako stany wszystkie podzbiory zbioru stanów, a jako tranzycje przejścia pomiędzy tymi zbiorami po wszystkich literach alfabetu Σ
   (stan q jest w zbiorze docelowym jeśli istnieje tranzycja z jakiegoś stanu zbioru z którego idziemy do q).
2. Wyrażenia regularne, NFA i DFA rozpoznają (opisują) tę samą klasę języków.

   Szkic dowodu:
   Z automatu niedeterministycznego korzystając z pierwszego twierdzenia można skonstruować automat deterministyczny, który jest szczególnym przypadkiem automatu niedeterministycznego, więc te modele są sobie równoważne.

   Z wyrażenia regularnego budujemy automat postępując zgodnie z drzewem parsowania, dla pojedynczej operacji konstrukcja jest prosta.
   Z automatu deterministycznego aby uzyskać wyrażenie regularne wprowadzamy automaty z wyrażeniami regularnymi na krawędziach. Najpierw automat sprowadzamy do takiego o jednym stanie końcowym, a następnie usuwamy kolejne stany (modyfikując krawędzi) dochodząc do automatu z jedną krawędzią i dwoma stanami.
3. Lemat o pompowaniu: jeśli L jest regularny, to istnieje stała N taka, że dla każdego słowa $w \in L$ dłuższego niż N, istnieje podział $w = w_1 w_2 w_3$ taki, że $w_2 \neq \varepsilon$, $|w_1 w_2| \leqslant N$ oraz $w_1 w_2^k w_3 \in L$ dla każdego $k \geqslant 0$.

   Szkic dowodu:
   Bierzemy automat danego języka, jeśli ma on N stanów, to dla słów dłuższych niż N bieg ma cykl, z tego cyklu wynika postulowana własność.
4. Klasa języków regularnych jest zamknięta na następujące operacje:
   - sumę (suma rozłączna automatów)
   - przecięcie (automat iloczynowy)
   - dopełnienie $(DFA + F \leftrightarrow Q \setminus F)$
   - $L^r = \{a_n a_{n-1} \dots a_1 \mid a_1 a_2 \dots a_n \in L\}$
5. Relacja Myhill-Nerode'go: możemy połączyć stany z których dojść możemy tylko po tych samych słowach do stanów akceptujących - stworzymy tak automat minimalny - deterministyczny danego języka, jeśli klas abstrakcji jest nieskończenie wiele to język nie jest regularny.

Automaty a półgrupy

1. **Półgrupa** jest to zbiór S z łącznym działaniem $\cdot$ nazywanym mnożeniem, co rozumiemy przez:

$$\forall_{a,b,c\in S} \quad (a \cdot b) \cdot c = a \cdot (b \cdot c)$$

2. **Monoid** jest to półgrupa z elementem neutralnym (oznaczanym przez 1 lub e), czyli takim, który spełnia:

$$\forall_{s\in S} \quad 1 \cdot s = s \cdot 1 = s$$

3. Przykładami półgrup są:

   - $\mathbb{N}$ z dodawaniem
   - $\mathbb{N}$ z działaniem $a \cdot b = \max(a, b)$
   - Dla dowolnego zbioru X zbiór X^X wszystkich funkcji z $X \to X$ z operacją składania
   - Σ^* z operacją konkatenacji słów
   - Dowolna grupa $(G, \cdot)$

   Wszystkie powyższe przykłady półgrup są monoidami.
   Ważnym przykładem półgrupy która nie jest monoidem jest $\Sigma^* \setminus \{\varepsilon\}$ z operacją konkatenacji.

4. Dla danych monoidów $(S, \cdot)$ i $(M, \bullet)$, **homomorfizm** z S do M to funkcja $h : S \to M$ spełniająca:

$$\forall_{s_1,s_2\in S} \quad h(s_1) \bullet h(s_2) = h(s_1 \cdot s_2), \quad h(e_S) = e_M$$

   Odnotujmy, że w ogólności homomorfizm nie musi być ani iniekcją, ani suriekcją.

Twierdzenia

1. Niech $L \subseteq \Sigma^*$, następujące warunki są równoważne:

   (a) L jest językiem regularnym.

   (b) Istnieje monoid **skończony** S oraz homomorfizm $h : \Sigma^* \to S$ i podzbiór $K \subseteq S$ taki, że $L = h^{-1}(K)$.

   O monoidzie S mówimy, że jest skończony, gdy moc zbioru jego elementów jest skończona.

   Dowód:

$(a) \to (b)$ L jest regularny, więc istnieje DFA rozpoznający L, nazwijmy go $\mathcal{A}$. Nazwijmy jego zbiór stanów Q.
Rozważmy monoid S funkcji Q^Q, z działaniem lewostronnego składania funkcji $(f \cdot g)(a) = g(f(a))$.
Niech $h(w) = f$, dla takiej funkcji f, że $\forall_{q\in Q}\ f(q) = \delta(q, w)$.
Oczywiście h jest homomorfizmem, ponieważ wymagane warunki zachodzą.
Istotnie $h(\varepsilon) = \lambda x.x$ oraz $h(w_1)h(w_2) = h(w_1w_2)$ jest równoważne $\forall_{q\in Q}\ \delta(q, w_1w_2) = \delta(\delta(q, w_1), w_2)$.
Niech K będzie zbiorem funkcji $f : Q \to Q$ o własności $f(q_0) \in F$, gdzie q_0, F to stan początkowy i zbiór stanów końcowych automatu $\mathcal{A}$. $h(w) \in K \iff \delta(q_0, w) \in F \iff w \in L$, więc $h^{-1}(K) = L$.

$(b) \to (a)$ Niech S będzie monoidem skończonym. Ustalmy $K \subseteq S$ oraz $h : \Sigma^* \to S$ będące homomorfizmem.
Niech $L = h^{-1}(K)$, pokażemy, ze L jest regularny poprzez konstrukcję DFA rozpoznającego L.
Zdefiniujmy automat $\mathcal{B}$ w następujący sposób: zbiór stanów to zbiór elementów S, funkcja tranzycji zadana jest poprzez $\delta(s, a) = s \cdot h(a)$, stan początkowy $q_0 = h(\varepsilon)$, zbiór stanów akceptujących $F = K$.
Na mocy indukcji otrzymujemy, że $h(w) = \delta(q_0, w)$.

$$L(\mathcal{B}) = \{w \in \Sigma^* \mid \delta(q_0, w) \in K\} = \{w \in \Sigma^* \mid h(w) \in K\} = h^{-1}(K) = L$$

Więc L jest regularny.

$\blacksquare$

2. Niech S będzie monoidem skończonym. Dla każdego $s \in S$ istnieje $m \in \mathbb{N}$, takie, że s^m jest idempotentem, czyli $s^m s^m = s^m$. Co więcej, jeśli $k \in \mathbb{N}$ i $s^k s^k = s^k$, to $s^k = s^m$ (element jest unikalny).