Упражнение 6

1 Минимален автомат

Дефиниция 1. Нека $L \subseteq \Sigma^*$ е език, и нека $x, y \in \Sigma^*$. Казваме, че x и y са **еквивалентни по отношение на L**, означено $x \approx_L y$, ако за всяка дума $z \in \Sigma^*$, следното е вярно: $xz \in L \iff yz \in L$. Тази релация наричаме **релация на Нероуд за езика L**.

Забележете, че $\approx_L$ е релация на еквивалентност над Σ^* за всеки език L над Σ. Класовете на еквивалентност на тази релация, неформално казано, са множествата от думи, такива че слепването на коя да е фиксирана дума към края на кои да е две от тях води до получаването на две думи, за които или и двете са в L, или и двете са извън L. С $[x]_L$ означаваме класът на думата x по отношение на релацията на Нероуд за L.

Пример 1. Да намерим класовете на еквивалентност за езика $L = (ab + ba)^*$. Не е трудно да се съобрази, че този език има точно *четири* класа на еквивалентност по отношение на релацията на Нероуд, а именно:
(1) $[\epsilon]_L = L$,
(2) $[a]_L = La$,
(3) $[b]_L = Lb$,
(4) $[aa]_L = L(aa + bb)\Sigma^*$.

За (1) можем да отбележим, че не всеки език има свойството, че $[\epsilon]_L = L$. Разгледайте например езика $(ab)^*(a + b)$. (2) и (3) са класовете на думите, които представляват конкатенация на дума от L с a или b съответно. Тези думи могат да се продължат до дума в L *единствено* с представител на bL^* и aL^* съответно. (4) е класът на думите с дължина ≥ 2, които не са в L. Каквото и да конкатенираме към произволни два представителя на този клас, резултатната дума ще бъде извън езика.

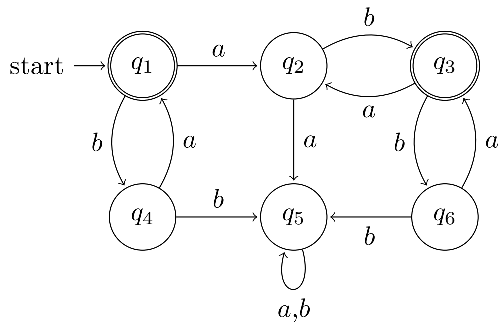
Дефиниция 2. Нека $A = (Q, \Sigma, \delta, s, F)$ е ДКА. Казваме, че две думи $x, y \in \Sigma^*$ са **еквивалентни по отношение на A**, означено с $x \sim_A y$, ако

$$\hat{\delta}(s, x) = \hat{\delta}(s, y).$$

Тази релация наричаме **релация на Нероуд по автомата A**.

Отново, $\sim_L$ е релация на еквивалентност над Σ^* за всеки език L над Σ. Класовете ѝ на еквивалентност се идентифицират с достижимите от s състояния в A. Означаваме съответстващия на състоянието $q \in Q$ клас с E_q.

Пример 2. Да разгледаме отново езика $L = (ab + ba)^*$. Следният автомат A разпознава точно L.



Класовете на еквивалентност на $\sim_A$ са
(1) $E_{q_1} = (ba)^*$,
(2) $E_{q_2} = La$,
(3) $E_{q_3} = (ba)^* abL$,
(4) $E_{q_4} = b(ab)^*$,
(5) $E_{q_5} = L(bb + aa)\Sigma^*$,
(6) $E_{q_6} = (ba)^* abLb$.

Връзката между двете релации, които дефинирахме е следната.
Теорема 1. За всеки ДКА $A = (Q, \Sigma, \delta, s, F)$ и всеки две думи $x, y \in \Sigma^*$, ако $x \sim_A y$, то $x \approx_{L(A)} y$.

Друг начин да изкажем тази връзка е да кажем, че релацията $\sim_L$ *прецизира* релацията $\approx_{L(A)}$. В общия случай, за две релации на еквивалентност R и R' казваме, че R **прецизира** R', ако за всеки x, y xRy влече $xR'y$. Обратно, ще казваме, че R' **апроксимира** R. Ако релацията на еквивалентност $\sim$ прецизира релацията на еквивалентност $\approx$, то всеки от класовете на еквивалентност на $\sim$ се съдържа в някой от класовете на $\approx$. С други думи, всеки клас на еквивалентност на $\approx$ е обединение на един или повече класове на еквивалентност на $\sim$.

Теорема 1 ни дава, че всеки автомат разпознаващ език L има поне толкова състояния, колкото са класовете на еквивалентност на $\approx_L$. Тоест, броят на класовете на еквивалентност на тази релация е *долна граница* за броя на състоянията на кой да е автомат, разпознаващ езика L. Но дали тази долна граница е достижима? Следната теорема ни дава отговор на този въпрос.

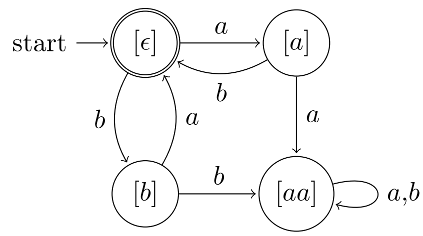
Теорема 2(на Майхил-Нероуд). Нека $L \subseteq \Sigma^*$ е регулярен език. Тогава съществува ДКА с точно толкова на брой състояния, колкото е броят на класовете на еквивалентност на $\approx_L$, който разпознава L.

Идеята е най-естествена. Използвайки само релацията $\approx_L$ ще конструираме ДКА $A = (Q, \Sigma, \delta, s, F)$, такъв че $L(A) = L$. A е дефиниран както следва:

$Q = \{[w] \mid w \in \Sigma^*\}$, множеството от класовете на еквивалентност на $\approx_L$.
$s = [\epsilon]$, класът на еквивалентност на ϵ спрямо $\approx_L$.
$F = \{[w] \mid w \in L\}$, множествтото от класовете на еквивалентност на думите в L спрямо $\approx_L$.
Накрая, за всеки клас $[w] \in Q$ и всяка буква $a \in \Sigma$, дефинираме $\delta([w], a) = [wa]$.

Разбира се, такъв автомат може да се конструира само ако $\approx_L$ има краен брой класове на еквивалентност. Как сме сигурни, че това е така? Езикът L е регулярен. Значи съществува ДКА A', такъв че $L(A') = L$. Според **Теорема 2** конструираният от нас автомат A също разпознава L. Но броя на състоянията на A е точно броят на класовете на еквивалентност на $\approx_L$. Този брой на свой ред е по-малък от броя на класовете на еквивалентност на $\sim_{A'}$, съгласно **Теорема 1**. Сега е достатъчно да забележим, че броят на класовете на еквивалентност на $\sim_{A'}$ е точно броя на състоянията на A', който е краен. Значи броят на състоянията на A е ограничен отгоре от крайно число.

Пример 3. Минималния автомат за езика $L = (ab + ba)^*$ можем да получим от **Пример 1** и **Пример 2**.



Изведеното до тук не предоставя алгоритъм с който по даден ДКА A да конструираме минималния автомат за $L(A)$. Следва да опишем един такъв алгоритъм. Първо да дефинираме релация $\equiv_A \subseteq Q \times Q$ по следния начин. За всеки две състояния $q, p \in Q$

$$q \equiv_A p \iff (\forall w \in \Sigma^*)[\hat{\delta}(q, w) \in F \iff \hat{\delta}(p, w) \in F].$$

Лесно се забелязва, че тази релация е релация на еквивалентност. Нейните класове на еквивалентност са точно тези множества от състояния, които трябва да се "слеят" в A, за да получим минималния автомат за $L(A)$.

Алгоритъмът ни ще трябва да изчислява класовете на еквивалентност на $\equiv_A$. За целта дефинираме следната редица от апроксимации на релацията $\equiv_A$. За всеки две състояния $q, p \in Q$

$$q \equiv_A^n p \iff (\forall w \in \Sigma^*)[|w| \leq n \implies (\hat{\delta}(q, w) \in F \iff \hat{\delta}(p, w) \in F)].$$

Очевидно, всяка от релациите $\equiv_A^0, \equiv_A^1, \equiv_A^2, ...$ е апроксимация на предшественика си в редицата. Освен това, $q \equiv_A^0 p$ е вярно тстк q и p са едновременно крайни или едновременно некрайни състояния. Тоест $\equiv_A^0$ има точно два класа на еквивалентност: F и $Q \setminus F$. Сега остана само да покажем как $\equiv_A^n$ зависи от $\equiv_A^{n-1}$ за всяко $n \geq 1$. Връзката е следната

$$(\forall q \in Q)(\forall p \in Q)[q \equiv_A^n p \iff q \equiv_A^{n-1} p \;\&\; (\forall a \in \Sigma)[\delta(q, a) \equiv_A^{n-1} \delta(p, a)]].$$

Алгоритъмът за изчисление на $\equiv_A$ има следния вид.

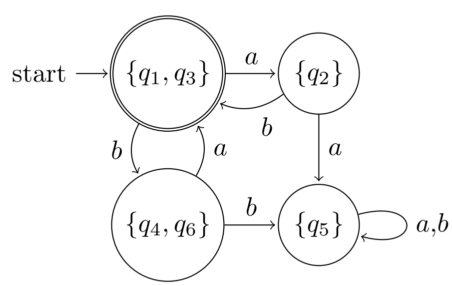
(1) По начало класовете на $\equiv_A^0$ са F и $Q \setminus F$;
(2) За всяко $n = 1, 2, ...$ изчисли класовете на $\equiv_A^n$ чрез класовете на $\equiv_A^{n-1}$ докато тези на $\equiv_A^n$ не станат същите като тези на $\equiv_A^{n-1}$.
(3) Върни $\equiv_A^n$ спрямо текущото достигнатото n.

Пример 4. Да приложим алгоритъма върху автомата от **Пример 2**. Очакваме естествено да получим като резултат автомата от **Пример 3**.

По начало, класовете на $\equiv_A^0$ са $\{q_1, q_3\}$ и $\{q_2, q_4, q_5, q_6\}$.

След първата итерация на (2), класовете на $\equiv_A^1$ са $\{q_1, q_3\}$, $\{q_2\}$, $\{q_4, q_6\}$ и $\{q_5\}$.

След втората итерация на (2), класовете не се разбиват допълнително. Алгоритъмът терминира и връща горните четири множества като класове на еквивалентност на $\equiv_A$. Това са именно състоянията на минималния автомат за $L = (ab + ba)^*$. Крайни са тези от тях, които са получени от разбиването на класа на крайните състояния (тоест тези, които съдържат само крайни състояния). Преход от състояние Q с буквата σ има към състоянието $\{p \in Q \mid \delta(q, \sigma) = p$, за някое $q \in Q\}$. Резултатният автомат естествено е *изоморфен* на този от **Пример 3**.



2 Задачи

Задача 1. (a) Намерете класовете на еквивалентност спрямо $\approx_L$ за всеки от следните езици:

(i) $L = (aab + ab)^*$.
(ii) $L = \{w \mid w$ има поддумата $aababa\}$.
(iii) $L = \{ww^{rev} \mid w \in \{a, b\}^*\}$.
(iv) $L = \{ww \mid w \in \{a, b\}^*\}$
(v) $L_n = (a + b) * a(a + b)^n$, където $n \in N^+$.
(б) За тези езици от (а), за които отговорът е краен, дайте минимален ДКА, разпознаващ съответния език.

3 Решения