Теория и реализация языков программирования.
Задание 1: регулярные языки и автоматы

Сергей Володин, 272 гр.

задано 2013.09.04

Задача 1

1. $\{a, aa\} \cdot \{b, bb\} \overset{\text{def}}{\equiv} \{x \cdot y | x \in \{a, aa\}, y \in \{b, bb\}\} = \{ab, abb, aab, aabb\}$.
2. $\{a, aa\} + \{b, bb\} \overset{\text{def}}{\equiv} \{x | x \in \{a, aa\} \vee x \in \{b, bb\}\} = \{a, aa, b, bb\}$.
3. $\{a, aa\} \times \{b, bb\} \overset{\text{def}}{\equiv} \{(x; y) | x \in \{a, aa\}, y \in \{b, bb\}\} = \{(a; b), (a; bb), (aa; b), (aa; bb)\}$.
4. Так как $(A|B) \supseteq A$, $X \overset{\text{def}}{=} \{((aa|b)^*(a|bb)^*)^*\} \supseteq \{(b^*a^*)^*\}$. Также $b^*a^* \supseteq (a|b)$, поэтому $X \supseteq \{(a|b)^*\}$. Но $\{(a|b)^*\} = \Sigma^*$, откуда $X = \Sigma^*$.
5. $Z \overset{\text{def}}{=} \underbrace{\{a^{3n} | \, n > 0\}}_{\text{X}} \cap \underbrace{\{a^{5n+1} | n \geq 0\}}_{\text{Y}}^*$. $Y \underset{n=0}{\supseteq} \{a\}$, $Y^* \supseteq \{a\}^* \supseteq \{a^{3n} | n > 0\} = X$, поэтому $Z = X = \{a^{3n} | \, n > 0\}$.
6. $\varnothing \cap \{\varepsilon\} \equiv \{\} \cap \{\varepsilon\} = \varnothing$.

Задача 2

$L = \Sigma^* \setminus \underbrace{\{(0^*110^*)^*\}}_{\text{L}_-}$. Для слова w из L_- есть два варианта, в соответствии с количеством повторений N в последней звездочке:

a. ($N = 0$ раз) $w = \varepsilon$

b. ($N > 0$) Докажем по индукции, что w — *строки из четного количества «1», отделенные друг от друга нулями, либо концом/началом слова, причем в слове хотя бы одна единица есть.*
Для $N = 1$ это верно: $w_1 \in \{0^*110^*\} \Rightarrow w_1 = \underbrace{0\ldots0}_{n_1}11\underbrace{0\ldots0}_{n_2}$, n_1 и $n_2 \geqslant 0$. Строка из двух единиц отделена нулями при $n_1, n_2 > 0$, либо концом/началом слова при $n_1 = 0$, либо $n_2 = 0$.
Пусть верно для $N \leqslant n$. Докажем для $n + 1$: $w_{n+1} = w_n \cdot w, w \in \{0^*110^*\} = w_n\underbrace{0\ldots0}_{n_1}11\underbrace{0\ldots0}_{n_2}$. Рассмотрим различные случаи: w_n может заканчиваться на 0, либо на 1; $n_1 = 0$, либо $n_1 > 0$:

1. $(0, n_1 = 0)$ Добавленная строка из единиц отделена слева нулями из w_n.
2. $(0, n_1 > 0)$ Добавленная строка из единиц отделена слева нулями из w.
3. $(1, n_1 = 0)$ Получена строка более, чем из двух единиц, но она четной длины (т.к. строка единиц из w_n имеет четную длину по предположению индукции, и 2 — четно).
4. $(1, n_1 > 0)$ Добавленная строка из единиц отделена слева нулями из w. Строка единиц из w_n отделена теми же нулями.

Очевидно, что под это определение не попадают слова не из L_- (можно построением найти вхождения РВ: найдем все строки из единиц в слове. Рассмотрим их по-очереди, с первой. Если строка длины 2, то единицы и все нули справа и слева — вхождение выражения. Если длина больше 2, то нули слева от первой пары вместе с ней — вхождение, нули справа от последней пары вместе с ней — вхождение. Четное количество единиц между этими парами (если есть) — несколько вхождений. Из этого следует, что слово из L_-).

Таким образом, L — непустые слова, состоящие либо только из нулей, либо из строк единиц, отделенных друг от друга нулями или началом/концом слова, но длина хотя бы одной строки нечетна. Иными словами, непустое слово w:

1. либо состоит из нулей,
2. либо в нем присутствует строка из единиц нечетной длины, отделенная
   a. нулями
   b. началом слова слева и нулями справа
   c. началом слова слева и концом слова справа
   d. нулями слева и концом слова справа.

Тогда $L = \{(\underbrace{00^*}_{1} | \underbrace{(0|1)^*01(11)^*0(0|1)^*}_{2a} | \underbrace{1(11)^*0(0|1)^*)}_{2b} | \underbrace{1(11)^*}_{2c} | \underbrace{(0|1)^*01(11)^*}_{2d}\}$

Задача 3

1. Конкатенация

$$\begin{array}{ccc} & f \bullet^{N} l & \\ / & & \backslash \\ f_\alpha\,\alpha^{N_\alpha}\,l_\alpha & & f_\beta\,\beta^{N_\beta}\,l_\beta \end{array}$$

   В результате будет порождено слово $c = ab$.
   Если $N_\alpha = F$, то a — его префикс, так как слово a всегда непустое. Тогда $f = f_\alpha$. Иначе, если $N_\alpha = T$, либо a, либо b (в случае $a = \varepsilon$) — префикс c, и $f = f_\alpha \cup f_\beta$. Аналогично, если $N_\beta = F$, то b — суффикс c, откуда $l = l_\beta$. Иначе $l = l_\alpha \cup l_\beta$. Всё выражение может быть пустым тогда и только тогда, когда α и β могут быть пустыми. Результат в таблице ниже:

N_α	N_β	f	l	N
F	F	f_α	l_β	F
F	T	f_α	$l_\alpha \cup l_\beta$	F
T	F	$f_\alpha \cup f_\beta$	l_β	F
T	T	$f_\alpha \cup f_\beta$	$l_\alpha \cup l_\beta$	T

2. Объединение

$$\begin{array}{ccc} & f \,|^{N}\, l & \\ / & & \backslash \\ f_\alpha\,\alpha^{N_\alpha}\,l_\alpha & & f_\beta\,\beta^{N_\beta}\,l_\beta \end{array}$$

   В результате будет порождено слово c.
   Во всех случаях c может начинаться как с символов, порожденных первым выражением, так и с символов, порожденных вторым, и ни с каких других. Тогда $f = f_\alpha \cup f_\beta$, $l = l_\alpha \cup l_\beta$, Всё выражение не может быть пустым тогда и только тогда, когда α и β не могут быть пустыми. Результат в таблице ниже:

N_α	N_β	f	l	N
F	F	$f_\alpha \cup f_\beta$	$l_\alpha \cup l_\beta$	F
F	T	$f_\alpha \cup f_\beta$	$l_\alpha \cup l_\beta$	T
T	F	$f_\alpha \cup f_\beta$	$l_\alpha \cup l_\beta$	T
T	T	$f_\alpha \cup f_\beta$	$l_\alpha \cup l_\beta$	T

Задача 4

1.

	$\mathcal{A}$	$\mathcal{B}$
Q	$\{q_0, q_1, q_2\}$	$\{q_0, q_1, q_2\}$
Σ	$\{0, 1\}$	$\{0, 1\}$
q_0	q_0	q_0
δ	$\{((q_0,0),q_0),((q_0,1),q_1),((q_1,1),q_0),$ $((q_1,0),q_2),((q_2,0),q_1),((q_2,1),q_2)\}$	$\{((q_0,0),q_0),((q_0,1),q_1),$ $((q_1,0),\{q_0,q_2\}),((q_2,0),q_1),((q_2,1),q_2)\}$
F	$\{q_1\}$	$\{q_1\}$

2. $\mathcal{A}$ — детерминированный, так как из каждого состояния есть только один переход с определенным символом.
   $\mathcal{B}$ — недетерминированный, так как из состояния q_1 есть два перехода по символу 0: в q_0 и q_2.

3. $(q_0, 101011) \vdash (q_1, 01011) \vdash (q_2, 1011) \vdash (q_2, 011) \vdash (q_1, 11) \vdash (q_0, 1) \vdash (q_1, \varepsilon)$. Принимает, так как $q_1 \in F$.

4. Да: $(q_0, 01001) \vdash (q_0, 1001) \vdash (q_1, 001) \vdash (q_0, 01) \vdash (q_0, 1) \vdash (q_1, \varepsilon)$ и $q_1 \in F$.

5. $\mathcal{A}$ не примет слово 0: $(q_0, 0) \vdash (q_0, \varepsilon)$ и $q_0 \notin F$, но примет 10011: $(q_0, 10011) \vdash (q_1, 0011) \vdash (q_2, 011) \vdash (q_1, 11) \vdash (q_0, 1) \vdash (q_1, \varepsilon)$ и $q_1 \in F$.
   $\mathcal{B}$ не примет пустое слово, так как $q_0 \notin F$, но примет слово 100: $(q_0, 100) \vdash (q_1, 00) \vdash (q_2, 0) \vdash (q_1, \varepsilon)$.

Задача 5

1. Докажем, что $L = T$:

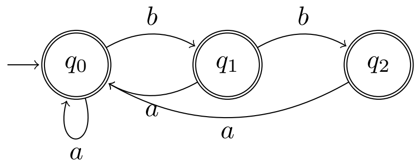
   1. ($L \subseteq T$). Если $w \in L$, то w получено из одного из слов ε, b, bb применением правила (2) $N(w) \geqslant 0$ раз. Действительно, $w \in L \Rightarrow \left[\begin{array}{lll} (1) & w \in \{\varepsilon, b, bb\} & N(w) = 0 \\ (2) & \substack{w \in \{ax, bax, bbax\}, \\ \text{где } x \in L} & N(w) = 1 + N(x) \end{array}\right.$. $N(w) < \infty$, так как в случае (2) $|x| < |w|$.
      Таким образом, определена функция $N(w) : L \longrightarrow \mathbb{N} \cup \{0\}$ — количество применений правила (2) для слова w. Заметим, что значение этой функции также равно количеству букв a в слове w, так как правило (2) добавляет одну букву a. Индукцией по N докажем, что $L \subseteq T$:

      a. ($N = 0$) $w \in \{\varepsilon, b, bb\}$. В w нет трех букв b подряд, поэтому $w \in T$.
      b. (доказано для $N = n - 1$, докажем для $N = n$) $w \in \{ax, bax, bbax\}$, причем $N(x) = n - 1$, так как в w на одну букву a больше, чем в x. Поэтому по предположению индукции в x нет трех букв b подряд. Заметим, что x отделено буквой a от ε, b или bb, поэтому в w нет трех букв b подряд, отсюда $w \in T$.

   2. ($T \subseteq L$). В слове $w \in T$ $M(w)$ букв a. Индукцией по m докажем, что $w \in L$:

      a. ($M = 0$) Букв a нет $\Rightarrow$ w состоит из букв b, причем не более, чем из двух. Тогда $w \in \{\varepsilon, b, bb\} \subset L$.
      b. (доказано для $M = m - 1$, докажем для $M = m$). Разобьем $w = x_1 a x_2$, где в x_1 нет букв a (можно сделать, так как случай, где в w нет букв a разобран выше). В слове x_2 будет $m - 1$ букв a, и, по предположению индукции, $x_2 \in L$. В x_1 только буквы b, поэтому $x_1 \in \{\varepsilon, b, bb\}$. Таким образом, $w \in \{ax_2, bax_2, bbax_2\}$, и $x_2 \in L$. По правилу (2) получаем $w \in L$ $\blacksquare$

2. Докажем, что следующий автомат $\mathcal{A}$ распознает T:

   Определим $N(w) : T \longrightarrow \mathbb{N} \cup \{0\}$ — количество букв a в слове $w \in T$. Индукцией по $N(w)$ докажем, что автомат *принимает* w:

   a. ($N = 0$) слово состоит из букв b. Тогда $w \in T \Rightarrow$ в w не больше 2 букв b подряд $\Rightarrow$ $w \in \{\varepsilon, b, bb\}$. Запишем цепочки конфигураций для этих слов:

      1. ($w = \varepsilon$) (q_0, ε). q_0 — принимающее $\Rightarrow$ автомат принимает w.
      2. ($w = b$) $(q_0, b) \vdash (q_1, \varepsilon)$. q_1 — принимающее $\Rightarrow$ автомат принимает w.
      3. ($w = bb$) $(q_0, bb) \vdash (q_1, b) \vdash (q_2, \varepsilon)$. q_2 — принимающее $\Rightarrow$ автомат принимает w.

   b. (доказано для $N = m - 1$, докажем для $N = m$) Разделим w по последнему символу a (это можно сделать, так как случай, где в w нет символов a разобран выше): $w = x_1 a x_2$, в x_2 нет символов a. Тогда $N(x_1) = m - 1$, откуда следует, что автомат принимает x_1. Поэтому после обработки x_1 он оказывается в одном из состояний. Тогда после обработки следующего символа, a, он окажется в состоянии q_0, так как $\forall q \in Q \hookrightarrow ((q, a), q_0) \in \delta$, где обозначения Q, δ стандартные. x_2 состоит из букв b, поэтому $x_2 \in \{\varepsilon, b, bb\}$. Поскольку автомат находится в состоянии q_0, цепочка конфигураций после конфигурации (q_0, x_2) будет такой же, как в базе индукции.

   Этим доказано, что автомат принимает T, то есть, $L(\mathcal{A}) \supseteq T$.
   Теперь докажем, что $L(\mathcal{A}) \subseteq T \Leftrightarrow w \in L(\mathcal{A}) \Rightarrow w \in T \Leftrightarrow w \notin T \Rightarrow w \notin L(\mathcal{A})$.
   $w \notin T \Rightarrow$ в w есть больше двух букв b подряд. Найдем первый символ b в w, после которого идет больше двух: $w = x_1 \underbrace{b \ldots b}_{n} x_2$, x_1 не заканчивается на b, $x_1 \in T$ (так как там не больше двух b подряд), $n \geqslant 3$.
   $x_1 \in T \Rightarrow$ после обработки x_1 автомат будет в одном из состояний. Также x_1 не заканчивается на $b \Rightarrow$ либо заканчивается на a, либо $x_1 = \varepsilon$. В любом случае, этим состоянием будет q_0. Тогда дальнейшая цепочка конфигураций такая: $(q_0, b^n x_2) \vdash (q_1, b^{n-1} x_2) \vdash (q_2, b^{n-2} x_2)$. Но перехода из q_2 по b нет, поэтому автомат останавливается (и не принимает слово), т.е. $w \notin L(\mathcal{A})$ $\blacksquare$