Теория и реализация языков программирования. Задание 2: НКА и алгоритмы поиска подстрок

Сергей Володин, 272 гр.

задано 2013.09.18

Упражнение 1

Пусть $\sim \subset X \times X$. $C(x) = \{z \in X | x \sim z\}$, $C(y) = \{w \in X | y \sim z\}$.
Пусть $\exists z \in C(x) \cap C(y)$. Тогда $x \sim z, y \sim z$, и $w \in C(x) \overset{\text{def}}{\Leftrightarrow} x \sim w \underset{\substack{\text{транз.} \\ \text{симм.}}}{\overset{z \sim x}{\Leftrightarrow}} z \sim w \overset{y \sim z}{\Leftrightarrow} y \sim w \overset{\text{def}}{\Leftrightarrow} w \in C(y)$, то есть, $C(x) = C(y)$.
В противном случае $\rceil(\exists z \in C(x) \cap C(y)) \Leftrightarrow C(x) \cap C(y) = \varnothing$. Получаем, что возможны два случая:

1. $C(x) \cap C(y) = \varnothing$ *(не пересекаются)*
2. $C(x) = C(y)$ *(совпадают)*

Упражнение 2

Пусть $\varphi \colon \Sigma^* \supseteq X \longrightarrow \Delta^*$. $\varphi(\sigma_i) \overset{\text{def}}{=} \delta_i \in \Delta^*$, $|\sigma_i| = 1$.

1. *(единственность)* Предположим, что существует такое φ — морфизм. Тогда $\forall w = w_1 ... w_n \in X, |w_i| = 1 \hookrightarrow \varphi(w) \equiv \varphi(w_1 ... w_n) = \varphi(w_1) \cdot \varphi(w_2 ... w_n) = ... = \varphi(w_1) \cdot ... \cdot \varphi(w_n) \in \Delta^*$. Для $w = \varepsilon$ получаем $\varphi(\varepsilon) = \varepsilon$, так как φ — морфизм: $w_0 \overset{\text{def}}{=} \varphi(\varepsilon) = \varepsilon$. $\varphi(\varepsilon) \equiv \varphi(\varepsilon\varepsilon) = \varphi(\varepsilon)\varphi(\varepsilon) = w_0 w_0 \Rightarrow w_0 = w_0 w_0 \Rightarrow |w_0| = |w_0||w_0| \Rightarrow w_0 = \varphi(\varepsilon) = \varepsilon$.

   Таким образом, получаем, что такой морфизм единственный (если существует).

2. *(существование)* Докажем, что определенное выше отображение φ — морфизм: пусть $x, y \in X$. Рассмотрим случаи:

   a. $|x| = 0, |y| = 0 \Rightarrow \varphi(xy) = \varphi(\varepsilon\varepsilon) = \varphi(\varepsilon) = \varepsilon = \varepsilon\varepsilon = \varphi(\varepsilon)\varphi(\varepsilon)$
   b. $|x| = 0, |y| > 0 \Rightarrow \varphi(xy) = \varphi(y) = \varepsilon\varphi(y) = \varphi(x)\varphi(y)$
   c. $|x| > 0, |y| = 0 \Rightarrow \varphi(xy) = \varphi(x) = \varphi(x)\varepsilon = \varphi(x)\varphi(y)$
   d. $|x| > 0, |y| > 0 \Rightarrow \varphi(xy) = \varphi(x_1 ... x_m y_1 ... y_n) = \underbrace{\varphi(x_1) ... \varphi(x_m)}\underbrace{\varphi(y_1) ... \varphi(y_n)} = \varphi(x)\varphi(y)$.

Таким образом, если заданы значения $\varphi(\sigma_i), \sigma_i \in X \subset \Sigma$, то морфизм $\varphi \colon \Sigma^* \supseteq X \longrightarrow \Delta^*$ с этими значениями существует и единственнен.

Задача 1

Определим $R_3 : \mathsf{REG} \ni X \longrightarrow \mathbb{N} \cup \{0\}$ — количество применений правила 3 из определения регулярности X. В случае $X = AB$ или $X = A|B, A, B \in \mathsf{REG}$ $R_3(X) \overset{\text{def}}{=} 1 + R_3(A) + R_3(B)$. В случае $X = A^*, A \in \mathsf{REG}$, определим $R_3(X) \overset{\text{def}}{=} 1 + R_3(A)$. В случае $X = \varnothing$ или $X = \{\sigma\}$ определим $R_3(X) \overset{\text{def}}{=} 0$. Функция $R_3(X)$ определена корректно, так как определение регулярного языка коррентное.
Пусть $\varphi \colon \Sigma^* \supset X \longrightarrow Y \subset \Delta^*$ — морфизм, $X \in \mathsf{REG}$. Докажем, что $Y \equiv \varphi(X) \in \mathsf{REG}$ индукцией по $R_3(X)$:
$P(i) = (\forall X \in \mathsf{REG} \colon R_3(X) \leqslant i \; \forall \varphi$ — морфизм $\hookrightarrow \varphi(X) \in \mathsf{REG})$.

1. Докажем $P(0)$: пусть $X \in \mathsf{REG} \colon R_3(X) = 0$. Тогда X получен без применения третьего правила. Значит, $\forall \varphi$ — морфизм либо $X = \varnothing \Rightarrow \varphi(X) = \varnothing$, либо $X = \{\sigma\} \Rightarrow \varphi(X) = \{\varphi(\sigma)\} = \{w\}, w \in \Delta^*$.

   Докажем, что $\Delta^* \supset \{w\} \in \mathsf{REG}$. $\{w\} \equiv \{\sigma_1 ... \sigma_n\} \equiv \{\sigma_1\} \cdot ... \cdot \{\sigma_n\}$. Поскольку $\{\sigma_i\} \in \mathsf{REG}$, и регулярные языки замкнуты относительно конкатенации (по определению), получаем требуемое.

   Итак, $\varphi(X) \in \mathsf{REG}$ ■

2. Пусть $P(n)$. Докажем $P(n+1)$. Пусть $\mathsf{REG} \ni X \colon R_3(X) \leqslant n + 1$. Если $R_3(X) < n + 1$, $P(n) \Rightarrow X \in \mathsf{REG}$.
   $\measuredangle X \colon R_3(X) = n + 1$. Возможны случаи:

   a. $X = WZ$, $W, Z \in \mathsf{REG}$. Тогда $\varphi(X) \equiv \varphi(WZ) = \{\varphi(wz) | w \in W, z \in Z\} = \{\varphi(w)\varphi(z) | w \in W, z \in Z\} = \{\varphi(w) | w \in W\} \cdot \{\varphi(z) | z \in Z\} = \varphi(W)\varphi(Z)$. $R_3(X) = 1 + R_3(W) + R_3(Z) = n + 1 \Rightarrow R_3(W), R_3(Z) \leqslant n \overset{P(n)}{\Rightarrow} \varphi(W), \varphi(Z) \in \mathsf{REG} \Rightarrow \varphi(X) = \varphi(W)\varphi(Z) \in \mathsf{REG}$.

b. $X = W|Z,\ W, Z \in \mathsf{REG}$. Тогда $\varphi(X) \equiv \varphi(W|Z) \equiv \varphi(W)|\varphi(Z)$. Аналогично $R_3(W), R_3(Z) \leqslant n \overset{P(n)}{\Rightarrow} \varphi(W), \varphi(Z) \in \mathsf{REG} \Rightarrow \varphi(X) = \varphi(W)|\varphi(Z) \in \mathsf{REG}$.

c. $X = W^*, W \in \mathsf{REG}$. Тогда $R_3(X) = 1 + R_3(W) = n + 1 \Rightarrow R_3(W) = n \overset{P(n)}{\Rightarrow} \varphi(W) \in \mathsf{REG} \Rightarrow \varphi(W^*) = \varphi(\varepsilon|W|WW|...) = \varphi(\varepsilon)|\varphi(W)|\varphi(WW)... \overset{\varphi(\varepsilon)=\varepsilon}{=} \varepsilon|\varphi(W)|\varphi(WW)... = \varphi(W)^* \in \mathsf{REG}$.

Получаем $\forall i \geqslant 0 \hookrightarrow P(i) \Rightarrow \forall X \in \mathsf{REG}\, \forall \varphi$ — морфизм $\hookrightarrow \varphi(X) \in \mathsf{REG}\, \blacksquare$

Задача 2

1. Нет. Пусть $\Sigma = \{0, 1\}$, $L = \Sigma^*$. Определим $\varphi\colon L \longrightarrow L\colon \forall w \in L \hookrightarrow \varphi(w) = \varepsilon$. В этом случае φ — морфизм, так как $\forall x \in L\, \forall y \in L \hookrightarrow \varphi(xy) = \varepsilon = \varepsilon\varepsilon = \varphi(x)\varphi(y)$. Тогда $\forall \varnothing \neq X \subset L \hookrightarrow \varphi(X) = \{\varepsilon\}$, так как $\forall w \in L \hookrightarrow \varphi(w) = \varepsilon$. Поскольку $\varphi(\varepsilon) = \varepsilon \in L$, $\varphi^{-1}(\varepsilon) \ni \varepsilon \Rightarrow \varphi^{-1}(L) \supset \{\varepsilon\} \neq \varnothing \Rightarrow \varphi^{-1}(L) \neq \varnothing \Rightarrow \varphi(\varphi^{-1}(L)) = \{\varepsilon\} \neq L$.
Таким образом, $\exists L \subseteq \Sigma^*\, \exists \varphi$ — морфизм: $\varphi(\varphi^{-1}(L)) \neq L$.

2. Нет. Пусть $\Sigma = \{a, b\}$, $L = \{b\}^*$, $\varphi(a) \overset{\text{def}}{=} \varphi(b) \overset{\text{def}}{=} a$. Доопределим φ так, чтобы оно было морфизмом (это возможно, см. упражнение 2). Тогда $\varphi(L) \equiv \varphi(\{b^*\}) \ni \varphi(b) = a \Rightarrow \varphi^{-1}(\varphi(L)) \supset \varphi^{-1}(a) \ni a \notin L \Rightarrow \varphi^{-1}(\varphi(L)) \nsubseteq L \Rightarrow \varphi^{-1}(\varphi(L)) \neq L$.
Таким образом, $\exists L \subseteq \Sigma^*\, \exists \varphi$ — морфизм: $\varphi^{-1}(\varphi(L)) \neq L$.

3. Нет. Пусть $\Sigma = \{a, b\}$, $L = \{ab\}$, морфизм $\varphi\colon \Sigma^* \longrightarrow \Sigma^*$ — из предыдущего пункта. Тогда $\varphi(L) = \{\varphi(ab)\} = \{\varphi(a)\varphi(b)\} = \{aa\}$, $\varphi^{-1}(L) = \{x \in \Sigma^* | \varphi(x) \in \{ab\}\} = \{x \in \Sigma^* | \varphi(x) = ab\} = \varnothing$, так как $\varphi(\Sigma^*) = \varphi((a|b)^*) \overset{1.2.c}{=} (\varphi(a|b))^* = \{\varphi(a), \varphi(b)\}^* = \{a\}^* = a^* \not\ni ab$. Тогда $\varphi(\varphi^{-1}(L)) = \varphi(\varnothing) = \varnothing \not\ni aa \in \varphi^{-1}(aa) = \varphi^{-1}(\varphi(L))$.
Таким образом, $\exists L \subseteq \Sigma^*\, \exists \varphi$ — морфизм: $\varphi(\varphi^{-1}(L)) \neq \varphi^{-1}(\varphi(L))$.

Задача 3

$L \in \mathsf{REG} \Rightarrow \exists \mathcal{A}$ — ДКА: $L(\mathcal{A}) = L$. Построим ДКА $\mathcal{A}'$ для $L^{-1} \overset{\text{def}}{=} \varphi^{-1}(L)$. Для этого каждый переход по σ в $\mathcal{A}$ заменим на переход по $\varphi^{-1}(\sigma)$, а именно, переход по множеству слов понимается как множество переходов по словам, переход по слову — с введением дополнительных состояний.

Задача 4

Пусть языки $\Sigma^* \supset X, Y \in \mathsf{REG}$. Докажем, что

1. $X \cup Y \in \mathsf{REG}$: из определения регулярности $\forall X, Y \in \mathsf{REG} \hookrightarrow X \cup Y \in \mathsf{REG}\, \blacksquare$
2. $\overline{X} \overset{\text{def}}{=} \Sigma^* \backslash X \in \mathsf{REG}$: $X \in \mathsf{REG} \Rightarrow \exists$ полный ДКА $\mathcal{A}$: $L(\mathcal{A}) = X$. $F' \overset{\text{def}}{=} Q \backslash F$, $\mathcal{A}'$ — автомат $\mathcal{A}$ с множеством принимающих состояний F'. Докажем, что $L(\mathcal{A}') = \Sigma^* \backslash X$: $w \in \Sigma^*$, $(q_0, w) \vdash^* (q_w, \varepsilon)$ *(здесь используется полнота)*. $w \in X \Leftrightarrow w \in L(\mathcal{A}) \Leftrightarrow q_w \in F \Leftrightarrow \neg(q_w \in Q \backslash F) \Leftrightarrow \neg(q_w \in F') \Leftrightarrow \neg(w \in L(\mathcal{A}'))$. Но $w \in X \Leftrightarrow \neg(w \in \Sigma^* \backslash X)$, откуда $\neg(w \in \Sigma^* \backslash X) \Leftrightarrow \neg(w \in L(\mathcal{A}'))$ и Получаем ДКА $\mathcal{A}'$: $L(\mathcal{A}') = \Sigma^* \backslash X \underset{\text{семинаре}}{\overset{\text{на}}{\Rightarrow}} \Sigma^* \backslash X \in \mathsf{REG}\, \blacksquare$
3. $X \cap Y \in \mathsf{REG}$: $X \cap Y = \overline{\overline{X} \cup \overline{Y}}$. $X, Y \in \mathsf{REG} \overset{(2)}{\Rightarrow} \overline{X}, \overline{Y} \in \mathsf{REG} \overset{(1)}{\Rightarrow} \overline{X} \cup \overline{Y} \in \mathsf{REG} \overset{(2)}{\Rightarrow} \overline{\overline{X} \cup \overline{Y}} \in \mathsf{REG}\, \blacksquare$
$$w \in X \cap Y \Leftrightarrow \begin{cases} w \in X \\ w \in Y \end{cases} \Leftrightarrow \begin{cases} \neg(w \in \overline{X}) \\ \neg(w \in \overline{Y}) \end{cases} \Leftrightarrow \neg \left[\begin{array}{l} w \in \overline{X} \\ w \in \overline{Y} \end{array} \right. \Leftrightarrow \neg(w \in \overline{X} \cup \overline{Y}) \Leftrightarrow w \in \overline{\overline{X} \cup \overline{Y}} \text{ (подразумевается } w \in \Sigma^*) \blacksquare$$
4. $X \backslash Y \in \mathsf{REG}$: $X \backslash Y = X \cap \overline{Y}$. $Y \in \mathsf{REG} \overset{(2)}{\Rightarrow} \overline{Y} \in \mathsf{REG} \overset{(3)}{\Rightarrow} X \cap \overline{Y} \in \mathsf{REG}\, \blacksquare$
$$w \in X \cap \overline{Y} \Leftrightarrow \begin{cases} w \in X \\ w \in \overline{Y} \end{cases} \Leftrightarrow \begin{cases} w \in X \\ \neg(w \in Y) \end{cases} \Leftrightarrow w \in X \backslash Y \text{ (подразумевается } w \in \Sigma^*) \blacksquare$$

Задача 5

$\Sigma \overset{\text{def}}{=} \{a\}$. Предположим, что $\Sigma^* \subset L = \{a^{2^n} | n \geqslant 0\} \in \mathsf{REG} \underset{\text{о накачке}}{\overset{\text{по лемме}}{\Rightarrow}} \exists p = p_0 \geqslant 1\colon \forall w \in L \hookrightarrow (w = xyz, |y| \geqslant 1, |xy| \leqslant p, (\forall i \geqslant 0 \hookrightarrow xy^iz \in L))$. Фиксируем $n = n_0 = p, w_0 = a^{2^p} \in L$. Получаем $w_0 = x_0y_0z_0, |y0| \geqslant 1, |x_0y_0| \leqslant p$. Поскольку $L \subset a^*, y \in a^*$, откуда $y = a^j$, $j \geqslant 1$. Аналогично $x = a^i, z = a^k \Rightarrow w_0 = a^{2^p} = xyz = a^{i+j+k} \Rightarrow i + j + k = 2^p$. По лемме должно выполняться $xy^2z = a^{i+2j+k} \in L \Rightarrow a^{i+2j+k} = a^{2^q}$, откуда $i + 2j + k = 2^q \Rightarrow j = 2^q - 2^p \geqslant 2^{p+1} - 2^p = 2^p(2-1) = 2^p$. Но $|x_0y_0| \leqslant p \Rightarrow |y_0| \leqslant p$. Получаем $p \geqslant |y_0| = j \geqslant 2^p$ — противоречие, т.к. $\forall p \geqslant 1 \hookrightarrow p < 2^p$.

Значит, предположение неверно, и $L \notin \mathsf{REG}\, \blacksquare$

Задача 6

1. Да. $L_1 = \{a^{2013n+5} | n \in \mathbb{N} \cup \{0\}\} \cap \{a^{509k+29} | k \in \mathbb{N}, k \geqslant 401\}$. $w \in L_1 \Leftrightarrow \exists n_w \in \mathbb{N} \cup \{0\}, 401 \leqslant k_w \in \mathbb{N} \colon w = a^{2013n_w+5} = a^{509k_w+29}$.

   Решим в целых числах $2013n + 5 = 509k + 24 \Leftrightarrow 2013n - 509k = 24 \Leftrightarrow (*)$ — линейное диофантово уравнение, $24 \vdots 1 = gcd(2013, 509) \Rightarrow$ решение существует, и $(*) \Leftrightarrow \begin{Vmatrix} n \\ k \end{Vmatrix} = \begin{Vmatrix} x_0 \\ y_0 \end{Vmatrix} + t \begin{Vmatrix} x \\ y \end{Vmatrix}$; $x_0, y_0,\ x, y,\ t \in \mathbb{Z}$, $x_0, y_0,\ x, y$ — фиксированные, $t \geqslant t_0$ — параметр. Тогда $2013n + 5 = 509k + 29 = 2013(x_0 + xt) + 29 = pt + q$, $p, q \in \mathbb{Z}$ — фиксированные, $\mathbb{Z} \ni t \geqslant 0$ — параметр.

   Получаем $L_1 = \{a^{pt+q} | \mathbb{Z} \ni t \geqslant 0\} = \{a^{pt} | \mathbb{Z} \ni t \geqslant 0\} \cdot \{a^q\} = \{(a^p)^t | \mathbb{Z} \ni t \geqslant 0\} \cdot \{a^q\} = (a^p)^* a^q \equiv (\underbrace{a...a}_{p})^* \underbrace{a...a}_{q}$ — задается регулярным выражением ■

2. Нет. Предположим, что $L_2 = \{a^{200n^2+1} | \mathbb{Z} \ni n \geqslant 1000\} \in \mathsf{REG} \overset{\text{по лемме}}{\underset{\text{о накачке}}{\Rightarrow}} \exists p \geqslant 1 \colon \forall w \in L_2 \hookrightarrow (w = xyz, |y| \geqslant 1, |xy| \leqslant p, (\forall i \geqslant 0 \hookrightarrow xy^i z \in L_2))$. Выберем $\mathbb{Z} \ni n = \max\{p, 1000\} \geqslant 1000 \Rightarrow w \overset{\text{def}}{=} a^{200n^2+1} \in L_2$. Получаем $\exists x, y, z \colon |y| \geqslant 1, |xy| \leqslant p \colon w = xyz$, откуда $x = a^i, y = a^j, z = a^k, i + j + k = 200n^2 + 1$. Также получаем $xy^2z \in L_2$. Но $xy^2z = a^{i+2j+k} = a^{200m^2+1} \Rightarrow i + 2j + k = 200m^2 + 1 \geqslant 200(n+1)^2 + 1 \Rightarrow j \geqslant [200(n+1)^2 + 1] - [200n^2 + 1] = 200 + 400n \geqslant 200 + 400p$. С другой стороны, $|xy| \leqslant p \Rightarrow j = |y| \leqslant p \Rightarrow p \geqslant j \geqslant 200 + 400p \Rightarrow 399p + 200 \leqslant 0$ при $p \geqslant 1000$ — противоречие.

   Значит, предположение неверно, и $L_2 \notin \mathsf{REG}$ ■