ТРЯП. Домашнее задание № 9

Шарапов Денис, Б05-005

Задача 1

Доказать, что класс КС-языков замкнут относительно операции пересечиния с регулярным языком.

Решение.

Пусть A — КС-язык, B — регулярный язык. Пусть тогда $\mathcal{A}$ — МП-автомат, допускающий язык A по принимающему состоянию, $\mathcal{B}$ — ДКА, построенный для языка B. Запишем описания автоматов.

$\mathcal{A} = \langle Q_{\mathcal{A}},\ \Sigma,\ \Gamma,\ q^0_{\mathcal{A}},\ \delta_{\mathcal{A}},\ Z^0,\ F_{\mathcal{A}}\rangle;$

$\mathcal{B} = \langle Q_{\mathcal{B}},\ \Sigma,\ q^0_{\mathcal{B}},\ \delta_{\mathcal{B}},\ F_{\mathcal{B}}\rangle.$

Воспользуемся конструкцией пересечения (автомат $\mathcal{C}$):

$$\mathcal{C} = \langle (Q_{\mathcal{A}} \times Q_{\mathcal{B}}),\ \Sigma,\ \Gamma,\ (q^0_{\mathcal{A}}, q^0_{\mathcal{B}}),\ \delta,\ Z^0,\ (F_{\mathcal{A}} \times F_{\mathcal{B}})\rangle,$$

$$\delta((p, q), a, S) = (\delta_{\mathcal{A}}(p, a, S),\ \delta_{\mathcal{B}}(q, a)),$$

$$(p, q) \in F_{\mathcal{C}} \Longleftrightarrow (p \in F_{\mathcal{A}}) \wedge (q \in F_{\mathcal{B}}).$$

Построенный автомат принимает пересечение языков $A \cap B$: для того, чтобы он попал в принимающее состояние, необходимо, чтобы оба автомата попали в принимающее состояние (индукция по числу переходов). $\square$

Задача 2

Являются ли следующие языки КС-языками?

а) $SQ = \{ww \mid w \in \{a, b\}^*\}$;

б) $\Sigma^* \setminus SQ$;

в) $\{a^{3^n} \mid n > 0\}$.

Решение.

а) Нет. Докажем от противного. Воспользуемся леммой о накачке:

$$L \in \text{CFL} \Rightarrow \exists p \geq 1 : \forall w \in L\ \exists x, y, z, u, v : w = xuyvz :$$

1) $|uyv| \leq p$,
2) $|uv| \geq 1$,
3) $\forall i \geq 0 \hookrightarrow xu^iyv^iz \in L$.

Рассмотрим слово $w = a^p b^p a^p b^p$. Пусть выполняется лемма о начке. Тогда в силу первого утверждения слово uyv либо состоит только из одного символа (a или b), либо из последовательностей каждого из символов ($a^s b^t$ или $b^t a^s$). В первом случае выберем $i = 2$ и получим противоречие с третим утверждением (получится слово, не принадлежащее языку). Во втором случае выберем произвольное i из утверждения 3 и получим слово длины $4p-(k+l)+ki+li$, где $|u| = k$, $|v| = l$. В силу определения языка L, в слове $w \in L$ должны совпадать символы на позициях s_1 и s_2:

$$s_2 = k + \frac{4p + (k+l)(i-1)}{2},$$

где $s_1 = k$ — позиция, с которой начинается слово u. В силу произвольности i получаем противоречие, т. к. на рассматриваемой позиции может стоять произвольный символ (зависит от выбора i, потому что изменяется длина слова: например, достаточно взять i так, чтобы $s_1 = k$ было меньше половины длины слова). $\square$

б) Да. Заметим, что все слова нечетной длины принадлежат заданному языку. Построим грамматику $G = \langle N, T, P, S\rangle$ с правилами:

$$S \to AB \mid BA \mid A \mid B \mid \varepsilon,$$

$$A \to aAa \mid aAb \mid bAa \mid bAb \mid a,$$

$$B \to aBa \mid aBb \mid bBa \mid bBb \mid b.$$

Правила $S \to A$ и $S \to B$ задают все слова нечетной длины (по индукции длины вывода или длины слова).

Утверждение. $L(G) \subseteq \Sigma^* \setminus SQ$:

Слова нечетной длины принадлежат языку $\Sigma^* \setminus SQ$.

По построению грамматики из неё выводятся слова вида $\mathcal{T}_1$ или $\mathcal{T}_2$, где

$$\mathcal{T}_1 = T^n a T^n T^m b T^m,$$

$$\mathcal{T}_2 = T^n b T^n T^m a T^m$$

для некоторых натуральных m и n.

Видно, что слова $u \in \mathcal{T}_1$ и $v \in \mathcal{T}_2$ лежат в языке $L(G)$, т. к. они имеют разные символы на «симметричных» позициях.

Утверждение. $\Sigma^* \setminus SQ \subseteq L(G)$:

По индукции длины вывода (или слова) получаем, что любое слово нечетной длины из языка $\Sigma^* \setminus SQ$ выводится грамматикой G.

Для вывода слов четной длины докажем от противного. Допустим слово $w \in \Sigma^* \setminus SQ$ не имеет вид $\mathcal{T}_1$ или $\mathcal{T}_2$. Тогда

$$\forall i:\ 1 \le i \le n+m+1 \hookrightarrow w[i] = w[n+m+i+1].$$

Но тогда $w \in SQ$ — противоречие. Откуда заключаем, что $w \in L(G)$. $\square$

в) Нет. От противного. Пусть выполняется лемма о накачке. Тогда найдётся такое $p \ge 1$ что, для произвольного слова из языка найдётся разбиение, удовлетворяющее трём условиям. Выберем слово $w = a^{3^p}$. Для него по предположению выполняются все три условия. В силу условия (1) верно неравенство $|uyv| \le p$: Тогда выберем $i = 2$:

$$|xuyvz| = 3^p < |xu^2yv^2z| \le 3^p + p < 3^{p+1},$$

$$3^p < |xu^2yv^2z| < 3^{p+1},$$

откуда и получим противоречие. $\square$

Задача 3

Для языка

$$L = \{w : w = xcy,\ x, y \in \{a, b\}^*,\ c \in T,\ |x| = |y|\}$$

а) построить КС-грамматику G, порождающую язык L;

б) построить недерминированный МА, эквивалентный этой грамматике;

в) продемонстрировать работу построенного МА на слове $acab$ (все варианты поведения).

Решение.

а) $G = \langle N, T, P, S\rangle = \langle \{S\},\ \{a, b, c\},\ P,\ S\rangle$, где множество P задается как

$$S \to aSa \mid aSb \mid bSa \mid bSb \mid c.$$

Язык, порожденный грамматикой G, обозначим за $R(G)$.

Утверждение. $R(G) \subseteq L$:

Рассмотрим произвольное слово $w \in R(G)$. По построению грамматики оно имеет структуру

$$w \in \mathcal{T} = \{a, b\}^n \cdot c \cdot \{a, b\}^n,$$

что в точности является словом из L. Т. е. верно утверждение

$$\forall w \in R(G) \hookrightarrow w \in L.$$

$\square$

Утверждение. $L \subseteq R(G)$:

Докажем по индукции n длины слова $x \sqsubseteq w \in L$.

База индукции $n = 0$:

Используем правило $S \to c$.

Предположение индукции $n = k$:

Пусть из грамматики G выводимы все слова вида $u = \xi \cdot c \cdot \zeta$, где $\xi, \zeta \in \{a, b\}^*,\ |\xi| = |\zeta| = k$.

Переход индукции $n = k + 1$:

Рассмотрим вывод произвольного слова $u \in L$ из предположения индукции:

$$S \Rightarrow^* \xi \cdot S \cdot \zeta.$$

Если применить правило $S \to c$, то получится в точности слово u. Заметим, что на этом шаге верны включения $\xi, \zeta \in \{a, b\}^k$. Теперь воспользуемся одним из правил из множества P (отличного от $S \to c$) и получим следующий шаг вывода

$$S \Rightarrow^* \xi \cdot S \cdot \zeta \Rightarrow \alpha \cdot S \cdot \beta,$$

где $\alpha, \beta \in \{a, b\}^{k+1}$. Осталось применить последнее правило $S \to c$ и получить вывод произвольного слова $w \in L$:

$$S \Rightarrow^* \xi \cdot S \cdot \zeta \Rightarrow \alpha \cdot S \cdot \beta \Rightarrow \alpha \cdot c \cdot \beta.$$

$\square$

б) Воспользуемся алгоритмом КС-грамматика G — МП-автомат M.

Описание МП-автомата $M = \langle Q, T, \Gamma, \delta, q_0, Z_0, F \rangle$:

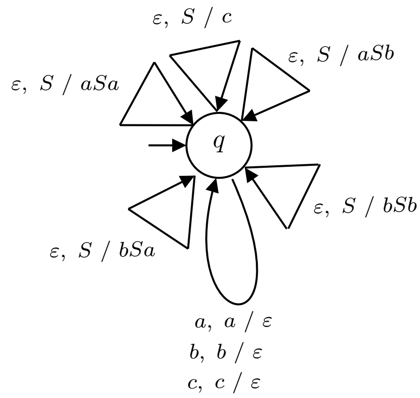
1) $Q = \{q\}$;
2) $T = \{a, b, c\}$;
3) $\Gamma = N \cup T$;
4) $q_0 = q$;
5) $Z_0 = S$;
6) $F = \varnothing$;

Описание функции $\delta : Q \times (T \cup \{\varepsilon\}) \times \Gamma \to 2^{Q \times \Gamma^*}$:

$$(S \to aSa) \in P \Rightarrow (q, aSa) \in \delta(q, \varepsilon, S),$$

$$(S \to aSb) \in P \Rightarrow (q, aSb) \in \delta(q, \varepsilon, S),$$
$$(S \to bSa) \in P \Rightarrow (q, bSa) \in \delta(q, \varepsilon, S),$$
$$(S \to bSb) \in P \Rightarrow (q, bSb) \in \delta(q, \varepsilon, S),$$
$$(S \to c) \in P \Rightarrow (q, c) \in \delta(q, \varepsilon, S),$$
$$\delta(q, a, a) = \{(q, \varepsilon)\},\ \delta(q, b, b) = \{(q, \varepsilon)\},\ \delta(q, c, c) = \{(q, \varepsilon)\}.$$

Диаграмма:



в) Демонстрация работы на слове $w = acab$. Демонстрация приведена на рисунке 1.

Задача 5

Задана грамматика G с правилами:

$S \to AB \mid C,$

$A \to aE \mid a,$

$E \to aE \mid \varepsilon,$

$B \to bB \mid Bb \mid b,$

$C \to CD,$

$F \to ab,$

$D \to aba.$

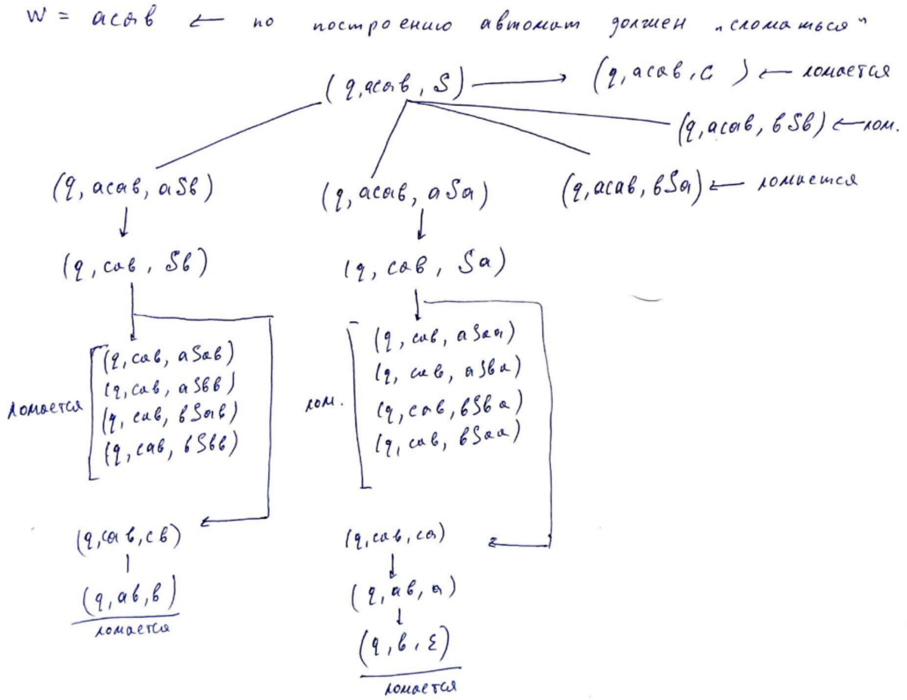
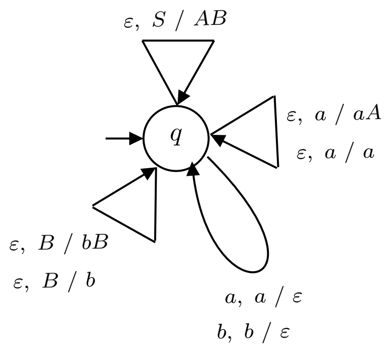


Рис. 1: Демонстрация к задаче 3(в)

Задан автомат M:



1. Эквивалентны ли грамматика G и автомат M?

2. Однозначна ли грамматика G? Если нет, постройте эквивалентную ей однозначную грамматику.

Решение.

1. Для начала удалим бесполезные символы. Воспользуемся алгоритмами удаления бесплодных и недостижимых символов.

   После удаления бесплодных символов множество правил имеет вид:

$$S \to AB,$$
$$A \to aE \mid a,$$
$$E \to aE \mid \varepsilon,$$
$$B \to bB \mid Bb \mid b,$$

$F \to ab,$

$D \to aba.$

После удаления недостижимых символов множество правил имеет вид:

$S \to AB,$

$A \to aE \mid a,$

$E \to aE \mid \varepsilon,$

$B \to bB \mid Bb \mid b,$

Удалим ε-продукции:

$S \to AB,$

$A \to aE \mid a,$

$E \to aE \mid a,$

$B \to bB \mid Bb \mid b.$

Удалим цепную продукцию, содержащую E:

$S \to AB,$

$A \to aA \mid a,$

$B \to bB \mid Bb \mid b.$

Теперь видно, что язык, порождаемый заданной грамматикой G эквивадентен языку, порождаемому автоматом M (в последнем правиле порядок подстановки bB или Bb неважен). □

2. Грамматика не является однозначной:

$$S \Rightarrow abB \Rightarrow abb,$$

$$S \Rightarrow aBb \Rightarrow abb.$$

Заметим, что язык, задаваемый грамматикой G, при выводе которого раскрывается только правый нетерминал B эквивалентен языку, при выводе которого раскрывается только левый нетерминал B. В свою очередь, они оба эквивалентны языку, порождаемому грамматикой G (этим утверждением и пользовались в пункте 1). Поэтому получаем окончательный вид грамматики:

$S \to AB,$

$A \to aA \mid a,$

$B \to bB \mid b.$

□

Замечание.

В решении задачи 5 нарушен порядок безопасного преобразования грамматик. В учебнике Д. Хопкрофта на странице 280 приведён порядок применения преобразований: 1) удалить ε-продукции; 2) удалить цепные продукции; 3) удалить бесполезные символы. В общем случае, если нарушить порядок, в грамматике могут остаться удаляемые элементы. Но в данном случае грамматика осталась без удаляемых элементов (за исключением правила $S \to AB$).

Задача 6

Язык L задан КС-грамматикой с правилами:

$$S \to aSa \mid aSb \mid bSa \mid bSb \mid a.$$

1. Является L ли регулярным языком?
2. Является ли дополнение L регулярным языком?
3. Является ли L КС-языком?
4. Является ли дополнение L КС-языком?

Решение.

Язык L аналогичен языку из задачи 3

$$L = \{w : w = xay,\ x, y \in \{a, b\}^*,\ a \in T,\ |x| = |y|\}.$$

Чтобы получить его грамматику, заменим правило $S \to c$ на правило $S \to a$ в ответе на задачу 3:

$$S \to aSa \mid aSb \mid bSa \mid bSb \mid a.$$

1. $L \notin \mathrm{REG}$, т. к. он имеет бесконечно много классов L-эквивалентности (к примеру, слова вида b^+a).
2. Из первого пунтка следует, что $\Sigma^* \setminus L \notin \mathrm{REG}$.
3. Да, является, т. к. он задан КС-грамматикой.
4. $T^* \setminus L = L_1 \cup L_2$, где
$$L_1 = \{w : |w| = 2n, n \geq 0\},$$
$$L_2 = \{w : w = xby,\ x, y \in \{a, b\}^*,\ b \in T,\ |x| = |y|\}.$$
   Язык L_2 задаётся той же грамматикой, что и язык L, если заменить правило $S \to a$ на правило $S \to b$. Поэтому язык $L_2 \in \mathrm{CFL}$.

   Язык L_1 задаётся той же грамматикой, что и язык L, если заменить правило $S \to a$ на правило $S \to \varepsilon$. Поэтому язык $L_1 \in \mathrm{CFL}$.

   КС-языки замкнуты относительно операции объединения, поэтому $T^* \setminus L \in \mathrm{CFL}$.

Ответ: 1) $L \notin \mathrm{REG}$; 2) $\Sigma^* \setminus L \notin \mathrm{REG}$; 3) $L \in \mathrm{CFL}$; 4) $T^* \setminus L \in \mathrm{CFL}$.

Задача 7

Язык L задан КС-грамматикой с правилами:

$$S \to aSb \mid A \mid B \mid \varepsilon, \quad A \to aAa \mid \varepsilon, \quad B \to bBb \mid \varepsilon.$$

1. Является L ли регулярным языком?
2. Является ли дополнение L регулярным языком?
3. Является ли L КС-языком?
4. Является ли дополнение L КС-языком?

Решение.

Рассмотрим каждое правило грамматики в отдельности. После k шагов (для некоторого n) применения привила

1. $S \to A$ получается слово $w = a^{n+2k}b^n$;

2. $S \to B$ получается слово $w = a^n b^{n+2k}$;

3. $S \to \varepsilon$ получается слово $a^n b^n$.

В зависимости от чётности n получим 4 случая:

а) $n = 2q,\ q \geq 0$

   1) $w = a^{2(q+k)}b^{2q}$;
   2) $w = a^{2q}b^{2(q+k)}$;

б) $n = 2q + 1,\ q \geq 0$

   3) $w = a^{2(q+k)} \cdot ab \cdot b^{2q}$;
   4) $w = a^{2k} \cdot ab \cdot b^{2(q+k)}$;

Тогда можно записать РВ, задающее язык $L(G)$:

$$R = (aa)^*(bb)^* \mid (aa)^* \cdot ab \cdot (bb)^*.$$

1. Язык L является регулярным, т. к. его задаёт регулярное выражение R.

2. Дополнение языка L является регулярным, т. к. регулярные языки замкнуты относительно операции дополнения.

3. Язык L является КС-языком, т. к. он задаётся грамматикой.

4. Да, является, поскольку дополнение языка L является регулярным языком, значит является и КС-языком.

Ответ: 1) $L \in \mathrm{REG}$; 2) $\Sigma^* \setminus L \in \mathrm{REG}$; 3) $L \in \mathrm{CFL}$; 4) $\Sigma^* \setminus L \in \mathrm{CFL}$.