Конечные преобразователи. Часть 2.
Алексей Сорокин

1 Свойства замкнутости.

Теорема 1.

1. *Автоматные языки замкнуты относительно конечных преобразований.*
2. *Контекстно-свободные языки замкнуты относительно конечных преобразований.*

Доказательство.

1. Пусть L — автоматный язык, задаваемый конечным автоматом $A = \langle Q, \Sigma, \Delta, q_0, F\rangle$ с однобуквенными переходами. Пусть $T = \langle Q', \Sigma, \Sigma', \Delta', q'_0, F'\rangle$ — конечный преобразователь, ψ — задаваемое им преобразование, причём $L' = \psi(L)$. Докажем автоматность языка L'. Можно считать, что для всякого перехода $(\langle q_1, x\rangle \to \langle q_2, x'\rangle) \in \Delta'$ выполняется соотношение $|x| + |x'| = 1$.

   Рассмотрим автомат $M = \langle Q \times Q', \Sigma', \Delta'', (q_0, q'_0), F \times F'\rangle$, где множество Δ'' задаётся следующим образом:

   - Если $(\langle q_1, x\rangle \to q_2) \in \Delta$ и $(\langle q'_1, x\rangle \to \langle q'_2, \varepsilon\rangle) \in \Delta',\ x \in \Sigma$, то $(\langle (q_1, q'_1), \varepsilon\rangle \to (q_2, q'_2)) \in \Delta''$.
   - Если $(\langle q'_1, \varepsilon\rangle \to \langle q'_2, x'\rangle \in \Delta',\ x' \in \Sigma'$, то для всех $q \in Q$ выполняется $(\langle (q, q'_1), x'\rangle \to (q, q'_2)) \in \Delta''$.

   Утверждение 1. $\langle (q_1, q'_1), u'\rangle \vdash_M \langle (q_2, q'_2), \varepsilon\rangle \leftrightarrow \exists u \in \Sigma^*(\langle q_1, u\rangle \vdash_A \langle q_2, \varepsilon\rangle \wedge \langle q'_1, u, \varepsilon\rangle \vdash_T \langle q'_2, \varepsilon, u'\rangle)$

   Доказательство. $\Rightarrow$ Индукция по числу переходов. База очевидна. Пусть последний переход имел вид $\langle (q_3, q'_3), \varepsilon\rangle \to (q_2, q'_2)$, тогда имеем $\langle (q_1, q'_1), u'\rangle \vdash_M \langle (q_3, q'_3), \varepsilon\rangle$. По предположению индукции получаем, что найдётся слово v, такое что $\langle q_1, v\rangle \vdash_A \langle q_3, \varepsilon\rangle$ и $\langle q'_1, v, \varepsilon\rangle \vdash_T \langle q'_3, \varepsilon, u'\rangle$. Кроме того, по построению имеем, что найдётся $a \in \Sigma$, такое что $(\langle q_3, a\rangle \to q_2) \in \Delta$ и $(\langle q'_3, a\rangle \to \langle q'_2, \varepsilon\rangle) \in \Delta'$. Отсюда вытекает, что $\langle q_1, va\rangle \vdash_A \langle q_2, \varepsilon\rangle$ и $\langle q'_1, va, \varepsilon\rangle \vdash_T \langle q'_2, \varepsilon, u'\rangle$, то есть в качестве v можно взять ua.

   Пусть теперь последний переход имел вид $\langle (q_3, q'_3), x'\rangle \to (q_2, q'_2)$, тогда по построению имеем $q_3 = q_2$ и $\langle (q_1, q'_1), v'\rangle \vdash_M \langle (q_2, q'_3), \varepsilon\rangle$, где $v'x' = u'$, а также $(\langle q'_3, \varepsilon\rangle \to \langle q'_2, x'\rangle) \in \Delta'$. По предположению индукции получаем, что найдётся слово v, такое что $\langle q_1, v\rangle \vdash_A \langle q_2, \varepsilon\rangle$ и $\langle q'_3, v, \varepsilon\rangle \vdash_T \langle q'_2, \varepsilon, v'\rangle$. Отсюда следует, что $\langle q_1, v\rangle \vdash_A \langle q_2, \varepsilon\rangle$ и $\langle q'_1, v, \varepsilon\rangle \vdash_T \langle q'_2, \varepsilon, v'x'\rangle$, то есть можно взять $u = v$.

   $\Leftarrow$ Индукция по $|u| + |v|$ и $|u|$. База индукции очевидна, на шаге индукции нужно рассмотреть последний переход в преобразователе T, после чего воспользоваться предположением индукции и определением автомата M. Детали предоставляются читателю. $\square$

Таким образом, условие $\exists(q,q')\in F\times F'(\langle(q_0,q_0'),u'\rangle \vdash_M \langle(q,q'),\varepsilon\rangle)$ равносильно конъюнкции условий $\exists q\in F((\langle q_0,u\rangle \vdash_A \langle q,\varepsilon\rangle)$ и $\exists q'\in F'((\langle q_0',u,\varepsilon\rangle \vdash_A \langle q',\varepsilon,u'\rangle)$ для некоторого слова u. Отсюда очевидным образом следует доказываемое утверждение.

2. Пусть L —контекстно-свободный язык, задаваемый контекстно-свободной грамматикой $G=\langle N,\Sigma,P,S\rangle$ в нормальной форме Хомского. Пусть $T=\langle Q,\Sigma,\Sigma',\Delta, q_0,F\rangle$ — конечный преобразователь, ψ — задаваемое им преобразование, причём $L'=\psi(L)$. Докажем, что L' является контекстно-свободным языком. Можно считать, что для всякого перехода $(\langle q_1,x\rangle\to\langle q_2,x'\rangle)\in\Delta'$ выполняется соотношение $|x|+|x'|=1$.

   Положим $G'=\langle Q\times N\times Q\cup S',\Sigma',P',S'\rangle$, где P' задаётся следующим определением:

   - Для каждого правила $(A\to BC)\in P$ и произвольных состояний q_1,q_2,q_3 множество P' содержит правило $\langle q_1,A,q_2\rangle\to\langle q_1,B,q_3\rangle\langle q_3,C,q_2\rangle$.
   - Для каждого правила $(A\to a), a\in\Sigma$ и каждого перехода $\langle q_1,a\rangle\to\langle q_2,\varepsilon\rangle$ множество P' содержит правило $\langle q_1,A,q_2\rangle\to\varepsilon$.
   - Для каждого перехода $\langle q_1,\varepsilon\rangle\to\langle q_2,a'\rangle$ и произвольных нетерминала A и состояния q множество P' содержит правило $\langle q,A,q_2\rangle\to\langle q,A,q_1\rangle a'$
   - Для всех завершающих состояний $q\in F$ множество P' содержит правило $S'\to\langle q_0,S,q\rangle$.

   Утверждение 2. $\langle q_1,A,q_2\rangle\vdash_{G'}u'\Leftrightarrow\exists u((A\vdash_G u)\wedge\langle q_1,u,\varepsilon\rangle\vdash_T\langle q_2,\varepsilon,u'\rangle$.

   Доказательство. $\Rightarrow$ Индукция по длине вывода. База индукции следует из определения грамматики G'. На шаге индукции нужно разобрать самое верхнее правило вывода в дереве вывода u' из $\langle q_1,A,q_2\rangle$.

   $\Leftarrow$ Индукция по длине слов u и u'. База индукции для случая $|u|=|u'|=0$ легко проверяется. Теперь докажем шаг индукции, пусть вначале $|u|=1$, тогда при $|u'|=0$ утверждение следует из определения грамматики G'. Пусть $|u|=1, |u'|>0$, тогда без ограничения общности можно считать, что $u'=v'a'$ и $\langle q_1,u,\varepsilon\rangle\vdash_T\langle q_3,\varepsilon,v'\rangle$, а также $(\langle q_3,\varepsilon\rangle\to\langle q_2,a'\rangle)\in\Delta$. По предположению индукции и построению грамматики G' получаем, что $\langle q_1,A,q_3\rangle\vdash_{G'}v'$ и $\langle q_1,A,q_2\rangle\to\langle q_1,A,q_3\rangle a'$, что приводит к требуемому утверждению. В случае $|u|\geqslant 2$ имеем, что найдётся такое представление $u=u_1u_2, u'=u_1'u_2'$, состояние q_3 и правило $(A\to BC)\in P$, что $B\vdash_G u_1, C\vdash_G u_2, \langle q_1,u_1,\varepsilon\rangle\vdash_T\langle q_3,\varepsilon,u_1'\rangle$ и $\langle q_3,u_2,\varepsilon\rangle\vdash_T\langle q_2,\varepsilon,u_2'\rangle$. Отсюда следует, что $(\langle q_1,A,q_2\rangle\to\langle q_1,B,q_3\rangle\langle q_3,C,q_2\rangle)\in P'$, а также $\langle q_1,B,q_3\rangle\vdash_{G'}u_1'$ и $\langle q_3,C,q_2\rangle\vdash_{G'}u_2'$. Вывод $\langle q_1,A,q_2\rangle\vdash\langle q_1,B,q_3\rangle\langle q_3,C,q_2\rangle\to v_1v_2$ даёт требуемое утверждение. □

   Теперь утверждение теоремы следует из четвёртого пункта определения грамматики G' и определения контекстно-свободной грамматики и рационального преобразования. Данная теорема может быть доказана и проще с использованием теоремы Нива.

$\square$

Обозначим через $\mathcal{B}_k$ язык правильных скобочных последовательностей с k типами скобок. Формально, язык $\mathcal{B}_k$ задаётся грамматикой $S \to a_1 S \overline{a}_1 S \mid \ldots \mid a_k S \overline{a}_k S \mid \varepsilon$.

Теорема 2 (Хомский-Щютценберже, слабая форма)**.** *Всякий контекстно-свободный язык можно получить рациональным преобразованием из языка $\mathcal{B}_k$ для некоторого натурального k.*

Доказательство. Пусть L — контекстно-свободный язык, распознаваемый МП-автоматом $M = \langle Q, \Sigma, \Gamma, \Delta, q_0, F\rangle$. Можно считать, что для всякого перехода $(\langle q_1, x, \alpha\rangle \to \langle q_2, \beta\rangle) \in \Delta$ выполняется условие $|\alpha| + |\beta| = 1$. Положим $k = |\Gamma|$ и будем считать, что $\Gamma = \{A_1, \ldots, A_k\}$. Обозначим $\overline{\Gamma} = \{\overline{A}_i \mid A_i \in \Gamma\}$.

Построим конечный преобразователь $T = \langle Q, \Gamma \cup \overline{\Gamma}, \Sigma, \Delta', q_0, F\rangle$, где Δ' строится по следующему алгоритму:

- Если $(\langle q_1, x, A_i\rangle \to \langle q_2, \varepsilon\rangle) \in \Delta$, то $(\langle q_1, \overline{A}_i\rangle \to \langle q_2, x\rangle) \in \Delta'$.
- Если $(\langle q_1, x, \varepsilon\rangle \to \langle q_2, A_i\rangle) \in \Delta$, то $(\langle q_1, A_i\rangle \to \langle q_2, x\rangle) \in \Delta'$.

Нетрудно доказать, что условия $\langle q_1, u, \varepsilon\rangle \vdash_M \langle q_2, \varepsilon, \varepsilon\rangle$ и $\exists w! \in \mathcal{B}_k$ $(\langle q_1, w, \varepsilon\rangle \vdash_T \langle q_2, \varepsilon, u\rangle)$ являются равносильными. Отсюда получаем, что $S_T(\mathcal{B}_k) = L(M)$, что и требовалось. $\square$

Теорема 3 (Хомский-Щютценберже, сильная форма)**.** *Всякий контекстно-свободный язык $L \in \Sigma^*$ представим в виде $L = \mathcal{B}_k \cap R$, где R — автоматный язык, для некоторого натурального k.*

Доказательство.

Лемма 1. *Пусть $\phi_1 \colon \Sigma^* \to \Sigma'^*, \phi_2 \colon \Sigma'^* \to \Sigma''^*$ — гомоморфизмы, $R_1 \subseteq \Sigma^*, R_2 \subseteq \Sigma'^*$ — автоматные языки. Тогда найдутся автоматный язык $R \subseteq \Sigma^*$ и гомоморфизм $\phi \colon \Sigma^* \to \Sigma''^*$, такие что для всякого языка $L \subseteq \Sigma^*$ выполняется соотношение $\phi(L \cap R) = \phi_2(\phi_1(L \cap R_1) \cap R_2)$.*

Доказательство. Достаточно положить $\phi = \phi_2 \circ \phi_1, R = R_1 \cap \phi_1^{-1}(R_2)$. $\square$

Введём скобочный алфавит $\mathbb{A} = \{a_1, \overline{a}_1, \ldots, a_k, \overline{a}_k\}$. В силу теоремы Нива и слабой формы теоремы Хомского-Шютценберже найдутся алфавит Γ, автоматный язык $R' \subseteq \Gamma^*$ и неудлиняющие гомоморфизмы $\psi_1 \colon \Gamma^* \to \mathbb{A}^*, \psi_2 \colon \Gamma^* \to \Sigma^*$, такие что $L = \psi_2(\psi_1^{-1}(\mathcal{B}_k) \cap R')$. По доказанной лемме достаточно подобрать такое натуральное число l, автоматный язык R'' и гомоморфизм ψ, что $\psi_1^{-1}(\mathcal{B}_k) = \psi(\mathcal{B}_l) \cap R''$.

Пусть $\Gamma = \{a_{11}, \ldots, a_{1m_1}, b_{11}, \ldots, b_{1n_1}, \ldots, a_{k1}, \ldots, a_{km_k}, b_{k1}, \ldots, b_{kn_k}, c_1, \ldots, c_r\}$ и $\psi_1(a_{ij}) = a_i, \psi_1(b_{ij}) = \overline{a}_i, \psi_1(c_j) = \varepsilon$. Положим $\mathbb{A}' = \{a_{ijk}, \overline{a}_{ijk} \mid i \leqslant k, j \leqslant m_i, k \leqslant n_i\} \cup \{d_i, \overline{d}_i \mid i \leqslant r\}$, $\psi(a_{ijk} = a_{ij}, \psi(\overline{a_i jk}) = b_i k, \psi(d_i) = c_1, \psi(\overline{d_i}) = \varepsilon$. Пусть язык R состоит из всех слов в алфавите $\mathbb{A}'$, содержащих буквы d_i и $\overline{d}_i$ только в подсловах вида $d_i \overline{d}_i$, очевидно, его можно задать конечным автоматом.

Обозначим через $\mathcal{B}_l$ язык скобочных последовательностей над алфавитом $\mathbb{A}'$ и докажем, что $\psi(\mathcal{B}_l \cup R) = \psi_1^{-1}(\mathcal{B}_k)$. Включение слева направо очевидно следует из того, что при композиции $\psi_1 \circ \phi$ скобки одного и того же типа переходят либо в скобки одного и того же типа, либо одновременно переходят в пустое слово. Обратное включение проверяется непосредственно.

$\square$