О сложности представления коллекции языков в конечных автоматах

М. А. Кибкало

Определение сложности представления языков — одна из традиционных задач теории автоматов. В статье рассматривается случай совместной представимости семейства языков в одном автомате. Существуют различные определения сложности регулярного языка основанные на характеристиках самого языка или представляющего его автомата. В данной работе под сложностью языка (семейства непересекающихся конечных языков) понимается число состояний в представляющем его (их) приведенном автомате. В работе решена задача о нахождении точного значения максимальной сложности семейства конечных языков в зависимости от максимальной длины слова в нем.

1. Начальные сведения

Пусть A, B, Q — конечные алфавиты, $|A| = N, |B| = M, |Q| = T$. Без ограничения общности можем считать, что $A = 0, 1, \ldots, N-1$, $B = 0, 1, \ldots, M-1, Q = 0, 1, \ldots, T-1$. Определим согласно [1] понятия конечного автомата (КА), инициального конечного автомата (ИКА), приведенного автомата, приводимого автомата, преставимости конечного языка в ИКА, отличимости, неотличимости и достижимости состояний в КА и ИКА, регулярного языка.

Определение 1.1. Конечный автомат $V = (A, Q, B, \varphi, \psi)$ назовем монотонным, если $\forall a \in A, \forall q \in Q, \varphi(q, a) \geqslant q$. Монотонный автомат назовем строго монотонным, если $\forall a \in A, \forall q \in Q \setminus \{T-1\}, \varphi(q, a) > q$ и $\forall a \in A$ $\varphi(T-1, a) = T-1$.

Определение 1.2. Пусть C — произвольный конечный алфавит. Обозначим через C^* множество всех слов конечной длины над алфавитом C. Определим:

1) $C^k = \{\alpha \in C^* || \alpha| = k\}$;
2) $C^{\leqslant k} = \{\alpha \in C^* || \alpha| \leqslant k\}$.

Определение 1.3. Конечным языком над алфавитом C назовем любое непустое конечное подмножество $L \subseteq C^*$. Языком длины k назовем произвольное $L \subseteq C^k$, языком длины не больше k — произвольное $L \subseteq C^{\leqslant k}$.

Определение 1.4. Класс всех языков длины k в алфавите A обозначим $\mathcal{L}_k(A)$, класс всех языков длины не больше k — $\mathcal{L}_{\leqslant k}(A)$. Если понятно, о каком алфавите идет речь, будет использовать обозначения $\mathcal{L}_k$ и $\mathcal{L}_{\leqslant k}$, соответственно.

Определение 1.5. Пусть $B' \subseteq B$, $V_{q_0} = (A, Q, B, \varphi, \psi, q_0)$ — инициальный конечный автомат. Язык $L = \{\alpha \in A^* | \ \psi(q_0, \alpha) \in B'\}$ назовем языком, представимым в конечном автомате V_{q_0} с помощью подмножества выходных символов $B' \subseteq B$. Обозначим представимость языка L в автомате V_{q_0} через $V_{q_0} \sim L$.

Определение 1.6. Пусть $L \subseteq A^*$ — регулярный язык, $s \geqslant 2$, $s \in \mathbb{N}$. **s-коллекцией языка** L назовем семейство языков $\mathcal{T}(L, s) = \{L_0, \ldots, L_{s-1}\}$ таких, что:

1) $L_i \cap L_j = \varnothing$, $i \neq j$, $i, j = 0, \ldots, s-1$;
2) $\bigcup_{i=1}^{s-1} L_i = L$;
3) $L_0 \stackrel{def}{=} A^* \backslash L$.

Если ясно, о каких L, s идет речь, будем использовать обозначение $\mathcal{T}$ вместо $\mathcal{T}(L, s)$.

Определение 1.7. Конечный инициальный автомат $V_{q_0} = (A, B, Q, \varphi, \psi, q_0)$ представляет s-коллекцию языка L $\mathcal{T}(L, s)$ ($V_q \sim \mathcal{T}(L, s)$) с помощью системы подмножеств выходного алфавита $\{B_0, \ldots, B_{s-1}\}, B_i \subset B$, $B_i \cap B_j = \varnothing$, $i \neq j$, $i, j = 0, \ldots, s-1$, если

$$\forall \alpha \in L_i \quad \psi(q_0, \alpha) \in B_i, \quad i = 0, \ldots, s-1.$$

В этом случае говорим также, что s-коллекция $\mathcal{T}(L,s)$ представима в автомате V_{q_0} с помощью системы подмножеств выходного алфавита $\{B_0,\ldots,B_{s-1}\}$. Представимость коллекции $\mathcal{T}$ в автомате V_{q_0} обозначим через $V_{q_0} \sim \mathcal{T}$.

Определение 1.8. s-коллекцию языка L, $\mathcal{T}(L,s)$, назовем представимой, если найдется система подмножеств выходного алфавита $\{B_0,\ldots,B_{s-1}\}, B_i \subset B$, $B_i \cap B_j = \varnothing$, $i \neq j$, $i,j = 0,\ldots,s-1$ и ИКА $V_{q_0} = (A,B,Q,\varphi,\psi,q_0)$ такие, что $V_{q_0} \sim \mathcal{T}(L,s)$ с помощью системы подмножеств выходного алфавита $\{B_0,\ldots,B_{s-1}\}$.

Очевидно, что автомат, представляющий язык $L \subseteq A^*$ с помощью множества $B' \subseteq B$, представляет 2-коллекцию $\{A^* \setminus L, L\}$ с помощью системы множеств $\{B \setminus B', B'\}$.

Теорема 1.1. *1) Если $M \geqslant s \geqslant 2$, $L \subseteq A^*$, $|L| = \infty, V_{q_0} \sim L$, то существует s-коллекция $\mathcal{T}(L,s)$, не представимая ни в каком ИКА $\widetilde{V_q} = (A,\widetilde{Q},B,\widetilde{\varphi},\widetilde{\psi},q)$;*

2) Если $M \geqslant s \geqslant 2$, $L \subseteq A^*$, $|L| < \infty, V_{q_0} \sim L$, то любая s-коллекция $\mathcal{T}(L,s)$, представима в ИКА с выходным алфавитом B.

Теорема 1.2. *1) Если ИКА $V_{q_0} = (A,Q,B,\varphi,\psi,q_0)$ представляет s-коллекцию $\mathcal{T}(L,s)$, $L \subseteq A^*$ с помощью системы подмножеств выходного алфавита $\{B_0,\ldots,B_{s-1}\}$, то существует автомат ${V'}_{q'} = (A,Q',B',\varphi',\psi',q'), B' = \{0,\ldots,s-1\}, |Q| = |Q'|$, представляющий $\mathcal{T}(L,s)$ с помощью системы подмножеств $\{\{0\},\ldots,\{s-1\}\}$;*

2) Если ИКА ${V'}_{q'} = (A,Q',B',\varphi',\psi',q'), B' = \{0,\ldots,s-1\}$ представляет s-коллекцию $\mathcal{T}(L,s), L \subseteq A^*$ с помощью системы подмножеств $\{\{0\},\ldots,\{s-1\}\}$, то для любого конечного алфавита B и системы его подмножеств $\{B_0,\ldots,B_{s-1}\}, B_i \subset B$, $B_i \cap B_j = \varnothing$, $i \neq j$, $i,j = 0,\ldots,s-1$ существует ИКА $V_{q_0} = (A,Q,B,\varphi,\psi,q_0), |Q| = |Q'|$, такой что $V_{q_0} \sim \mathcal{T}(L,s)$ с помощью системы подмножеств $\{B_0,\ldots,B_{s-1}\}$.

Определение 1.9. Сложностью $S_a(V_q)$ ИКА V_q назовем $|Q|$ — число состояний в автомате V_q.

Понятно, что в задаче о представлении коллекции $\{L_0, \dots, L_{s-1}|\}$ автоматами минимум сложности достигается при $B = \{0, \dots, s-1\}$ и $B_i = \{i\}, i = 0, \dots, s-1$.

Определение 1.10. Пусть $N \geqslant 2$, $M \geqslant 2$. Сложностью M-коллекции $\mathcal{T}(L, M)$ языка $L \subseteq A^*$ назовем величину

$$S_{\mathcal{T}}(\mathcal{T}(L, M), N) = \min_{V_q \sim \mathcal{T}(L,M)} S_a(V_q).$$

Определение 1.11. Пусть $N \geqslant 2$, $M = 2$. Сложностью языка $L \subseteq A^*$ назовем величину

$$S_l(L, N) = \min_{V_q \sim L} S_a(V_q).$$

Определение 1.12. Пусть $N \geqslant 2$, $M \geqslant 2$. $c-$сложностью языка $L \subseteq A^*$ назовем величину

$$S_l^c(L, N, M) = \max_{\mathcal{T}(L,M)} S_{\mathcal{T}}(\mathcal{T}(L, M), N).$$

Определение 1.13. Пусть $N \geqslant 2$, K — класс языков над алфавитом A. Сложностью класса K назовем величину

$$S(K, N) = \max_{L \in K} S_l(L, N).$$

Определение 1.14. Пусть $N \geqslant 2$, $M \geqslant 2$, K — класс языков над алфавитом A. c-сложностью класса K назовем величину

$$S^c(K, N, M) = \max_{L \in K} S_l^c(L, N, M).$$

В статье рассматривается асимптотическое поведение функций $S(K, N)$, $S^c(K, N, M)$ для классов $\mathcal{L}_k(A)$ и $\mathcal{L}_{\leqslant k}(A)$, строятся ИКА, имеющие максимальную сложность среди автоматов, представляющих коллекции языков из данных классов, а также сравниваются значения функций $S(K, N), S^c(K, N, M)$ для $\mathcal{L}_k(A)$ и $\mathcal{L}_{\leqslant k}(A)$.

2. Основные результаты

Теорема 2.1. *$\forall N \geqslant 2$, $M \geqslant 2$, $\forall k \geqslant 1$ существует $p \geqslant 0$, конечный язык $L \in \mathcal{L}_k(A)$, коллекция $\mathcal{T}(L, M)$ и строго монотонный ИКА $V_q(k, N, M) \sim \mathcal{T}(L, M)$, такие что*

$$S_a(V_q(k, N, M)) = S_l^c(L, N, M) = S^c(\mathcal{L}_k(A), N, M) = \\ = \frac{N^{k-p} - 1}{N - 1} + \sum_{i=1}^{p} (M^{N^i} - p + 1).$$

Следствие 2.1. *$\forall N \geqslant 2$, $M \geqslant 2$ для $S^c(\mathcal{L}_k(A), N, M)$ выполнено*

$$\frac{1}{N-1} \cdot \frac{N^k \cdot \log_N M}{k} \lesssim S^c(\mathcal{L}_k(A), N, M) \lesssim \frac{N}{N-1} \cdot \frac{N^k \cdot \log_N M}{k}.$$

Замечание 2.1. Ясно, что в задаче распознавания представимости языка L (как единого множества) в ИКА достаточно рассматривать $B = \{0, 1\}$. Случай $M = 2$ позволяет непосредственно оценить сложность ИКА, необходимого для представления L. При этом $S_l(L, N) = S_l^c(L, N, 2)$ и $S(K, N) = S^c(K, N, 2)$, и

$$\frac{1}{N-1} \cdot \frac{N^k \cdot \log_N 2}{k} \lesssim S^c(\mathcal{L}_k(A), N, 2) = \\ = S(\mathcal{L}_k(A), N) \lesssim \frac{N}{N-1} \cdot \frac{N^k \cdot \log_N 2}{k}.$$

Из оценок для $S^c(\mathcal{L}_k(A), N, M)$ и $S^c(\mathcal{L}_k(A), N, 2)$ видно, что можно представить язык вместе с его произвольным разбиением на непересекающиеся подмножества (случай $M > 2$) без существенного усложнения автомата по сравнению со случаем $M = 2$. Сложность будет отличаться лишь умножением на величину, растущую логарифмически.

Следствие 2.2. *Для p выполнено:*

$$p = \max\left(0, \log_N k + \log_N \log_M N - \frac{\log_M (k \cdot \log_M N)}{k \cdot \ln N \cdot \log_M N} + O\left(\frac{\log^2 k}{k^2}\right)\right).$$

Замечание 2.2. Значение p однозначно вычисляется с помощью алгоритма:
$P_{\min} = 1;\ P_{\max} = M^N;\ Q = 1;$
$for\ (s = 1, p = 0;\ s \leqslant k;\ s++)\ \{$
 $if\ (Q \geqslant P_{\max}/N)\ \{$
  $P_{\min} = P_{\max};\ P_{\max} = P_{\max}^N;\ p++$
 $\}\ else\ \{$
  $Q = Q \cdot N;$
 $\}$
$\}$

Теорема 2.2. *$\forall N \geqslant 2$, $M \geqslant 2$, $\forall k \geqslant 1$ существует $p \geqslant 0$, конечный язык $L \in \mathcal{L}_{\leqslant k}(A)$, коллекция $\mathcal{T}(L, M)$ и строго монотонный ИКА $V_q(k, N, M) \sim \mathcal{T}(L, M)$, такие что*

$$S_a(V_q(k, N, M)) = S_l^c(L, N, M) = S^c(\mathcal{L}_{\leqslant k}(A), N, M) = \frac{N^{k-p} - 1}{N - 1} + M^{\frac{N^{p+1} - N}{N-1}}.$$

Следствие 2.3. *$\forall N \geqslant 2$, $M \geqslant 2$ для $S^c(\mathcal{L}_{\leqslant k}(A), N, M)$ выполнено*

$$\frac{N}{(N-1)^2} \cdot \frac{N^k \cdot \log_N M}{k} \lesssim S^c(\mathcal{L}_{\leqslant k}(A), N, M) \lesssim \frac{N^2}{(N-1)^2} \cdot \frac{N^k \cdot \log_N M}{k}.$$

Следствие 2.4. *Для p выполнено:*

$$p = \max\Big(0, \log_N k + \log_N \log_M N + \log_N \tfrac{N-1}{N} + \tfrac{N}{N-1} \cdot \tfrac{1}{k \cdot \ln N \cdot \log_M N} - \tfrac{\log_M \frac{N-1}{N}}{k \cdot \ln N \cdot \log_M N} - \tfrac{\log_M (k \cdot \log_M N)}{k \cdot \ln N \cdot \log_M N} + O(\tfrac{\log^2 k}{k^2})\Big).$$

Замечание 2.3. Значение p однозначно вычисляется с помощью алгоритма:
$P_{\min} = 1;\ P_{\max} = M^N;\ Q = 1;$
$for\ (s = 1, p = 0;\ s \leqslant k;\ s++)\ \{$
 $if\ (Q \geqslant P_{\max}/N)\ \{$
  $P_{\min} = P_{\max};\ P_{\max} = (M \cdot P_{\max})^N;\ p++$

```
    } else {
    Q = Q · N;
    }
}
```

3. Основные утверждения и доказательства

3.1. Основные определения

В соответствии с [1] определим понятия неотличимости (эквивалентности) и достижимости состояний в ИКА V_{q_0}. Эквивалентность $q, \tilde{q} \in Q$ обозначим $q \sim q$, отличимость - через $q \nsim q$. Согласно [1] распространим функции перехода ($\varphi : Q \times A \to Q$) и выхода ($\psi : Q \times A \to B$) ИКА V_{q_0} на множество $Q \times A^*$, сохранив за ними те же обозначения.

Определение 3.1.1. Состояние $q \in Q$ автомата V назовем тупиковым, если $\forall a \in A\ \varphi(q, a) = q$.

Определение 3.1.2. Состояние $q \in Q$ автомата V назовем бесполезным, если $\forall \alpha \in A^*\ \psi(q, \alpha) = q$.

Определение 3.1.3. Состояние $q \in Q$ автомата V_{q_0} назовем финальным, если $\forall \alpha \in A^* | \varphi(q_0, \alpha) = q\ \psi(q, \alpha) \in B'$. Финальное состояние $q \in Q$ автомата V_{q_0} назовем i-финальным, если $\forall \alpha \in A^*\ \varphi(q_0, \alpha) = q\ \psi(q, \alpha) = i$.

Начальное состояние монотонного автомата имеет номер 0. Строго монотонный автомат имеет одно тупиковое состояние $T - 1$. Все тупиковые состояния автомата, представляющего коллекцию конечных языков, являются бесполезными.

Определение 3.1.4. Уровнем достижимого состояния q автомата V_{q_0} назовем минимальную длину слова $\alpha \in A^*$ такого, что $\varphi(q_0, \alpha) = q$:

$$level(q) = \min\{|\alpha| \mid \alpha \in A^*, \varphi(q_0, \alpha) = q\}.$$

Уровень начального состояния равен 0 по определению.

Определение 3.1.5. Высотой состояния $q \in Q$ автомата V_{q_0} назовем максимальную длину слова $\alpha \in A^*$ такого, что состояние $\varphi(q, \alpha)$ является финальным.

Определение 3.1.6. Подавтоматом КА $V = (A, Q, B, \varphi, \psi)$ назовем автомат $V' = (A, P, B, \varphi, \psi)$, где $P \subseteq Q$.

Несколько подавтоматов (одного автомата) $V_i = (A, B, Q_i, \varphi_i, \psi_i)$, $i = 1, \ldots, n$ с непересекающимися множествами состояний Q_i можно доопределить до конечного автомата $V = (A, B, Q, \varphi, \psi)$, где $Q = \bigcup_{i=1}^{n} Q_i$, $\psi : A \times Q \to B$, $\psi(q, a) = \psi_i(q, a)$, $q \in Q_i$, $a \in A$ и $\varphi : Q \times A \to Q$, $\begin{cases} \varphi(q, a) = \varphi_i(q, a), & q \in P_i, a \in A \\ \varphi(q, a) \in Q \backslash Q_i, & q \in Q_i \backslash P_i, a \in A \end{cases}$.

Любому подмножеству Q' множества состояний конечного автомата $V = (A, B, Q, \varphi, \psi)$ соответствует подавтомат $V' = (A, B, Q', \varphi', \psi')$, $\varphi' : (\varphi^{-1}(A \times Q') \cap Q') \times A \to Q'$, $\varphi'(q, a) = \varphi(q, a)$ и $\psi' : A \times Q' \to B$, $\psi'(q, a) = \psi(q, a)$. В этом случае автомат V будем называть объемлющим по отношению к подавтомату V'.

Определение 3.1.7. Назовем подавтомат (полным) (N, l)-деревом, если его диаграмма есть (полное) N-арное дерево высоты l, а корень дерева соответствует начальному состоянию объемлющего автомата.

Определение 3.1.8. Назовем подавтомат обратным (N, p)-деревом, если для любого состояния ν-го яруса дерева $R_i \varphi(\nu, j) \in R_{\min(i-1, 0)}$, $i = 0, \ldots, p$, $j = 0, \ldots, N-1$ и $\forall \nu \in R_0 \psi(\nu, j) = 0$, $j = 0, \ldots, N-1$.

Определение 3.1.9. Назовем обратное (N, p)-дерево, полным (N, M, p)-деревом первого рода, если для $\forall$ M-разбиения множества A^i ($T_r \subset A^i$, $T_r \cap T_j = \varnothing$, $r \neq j$, $r, j = 0, \ldots, M-1$) $\exists \nu \in R_i | \psi(\nu, \alpha) = r$ $\forall \alpha \in T_r$ и $\psi(\nu, \alpha) = 0$ при $|\alpha| < i$, $i = 1, \ldots, p$, $r = 0, \ldots, M-1$.

Легко строится полное обратное (N, M, p)-дерево такое, что $|R_i| = M^{N^i}$, $i = 0, \ldots, p$. Очевидно, что после минимизации объемлющего автомата оно будет содержать не более $\sum_{i=1}^{p} (|R_i| - 1) + 1 = \sum_{i=1}^{p} M^{N^i} - p + 1$ попарно отличимых состояний.

Определение 3.1.10. Назовем обратное (N, p)-дерево, полным (N, M, p)-деревом второго рода, если для $\forall$ M-разбиения T множества $A^{\leqslant i} \exists \nu \in R_i | \psi(\nu, \alpha) = r$ $\forall \alpha \in T_r$, $i = 1, \ldots, p$, $r = 0, \ldots, M-1$.

Минимальное полное обратное (N, M, p)-дерево второго рода содержит $M^{\frac{N^{p+1)}-N}{N-1}}$ состояний. Будем дальше считать, что в автомат, содержащий обратное полное дерево (первого или второго рода) включается уже минимизированное обратное дерево.

3.2. Сложность коллекций языков длины k

Утверждение 3.2.1. *$\forall L \subseteq \mathcal{L}_k$ и $M \geqslant 2$ любая M-коллекция $\mathcal{T}(L, M)$ представима монотонным автоматом в виде полного (N, k)-дерева, соединенного с тупиковым состоянием q_t, так что $\forall q \in R_k$, $a \in A$ $\psi(q, a) = q_t$ и $\psi(q, a) = 0$.*

Один из вариантов минимизации этого автомата может идти следующим путем: объединяются i-финальные листовые состояния для слов, входящих в языки L_i, $i = 0, \ldots, M-1$. Все состояния, соответствующие префиксам длины от 0 до k, отсутствующим в языках L_i, $i = 1, \ldots, M-1$ эквивалентны тупиковому состоянию. Из оставшихся состояний неотличимыми могут быть только пары состояний из одного яруса дерева.

Утверждение 3.2.2. *В любом конечном автомате, содержащем обратное полное (N, M, p)-дерево первого рода, любая пара состояний этого дерева (не эквивалентных тупиковому) будет отличима.*

Утверждение 3.2.3. *Состояния обратного полного (N, M, p)-дерева первого рода можно пронумеровать так, что для любого состояния q кроме одного (тупикового) и $\forall a \in A$ выполнено $\varphi(q, a) > q$.*

Теорема 3.2.1. *Любая M-коллекция $\mathcal{T}(L, M)$, $L \subseteq \mathcal{L}_k$ представима строго монотонным автоматом в виде полного (N, l)-дерева, соединенного с обратным полным (N, M, p)-деревом первого рода для $p = k - l - 1$, $l = 0, \ldots, k-1$.*

Лемма 3.2.1. *Длину префикса, для которой построено (N, l)-дерево, можно выбрать таким образом, что*

$$P_p'/N \leqslant Q_l < P_p^N, \quad k = l + p + 1 \tag{1}$$

где $Q_l = N^l$ *— число состояний в последнем уровне* (N, l)*-дерева префиксов,* $P_p = |R_p| = M^{N^p}$ *— число состояний в последнем ярусе обратного* (N, M, p)*-дерева первого рода,* $P'_p = \begin{cases} 1, & p = 0 \\ P_p, & \textit{иначе} \end{cases}$*, (*$P'_p$ *— число состояний в p-м ярусе обратного дерева после его минимизации).*

Формальной записью доказательства леммы 3.2.1. является алгоритм, приведенный в теореме 2.1.

Следствие 3.2.1. *Для любой* M*-коллекции* $\mathcal{T}(L, M)$*,* $L \subseteq \mathcal{L}_k$*,* $N \geqslant 2$*,* $M \geqslant 2$ *существует представляющий ее приведенный строго монотонный автомат в виде (возможно, неполного) префиксного* (N, l)*-дерева, соединенного с обратным полным* (N, M, p)*-деревом первого рода, из которого удалены недостижимые состояния.*

Утверждение 3.2.4. *Любая* M*-коллекция* $\mathcal{T}(L, M)$ *языков из слов длины* k *представима строго монотонным автоматом сложности* $\frac{N^{k-p}-1}{N-1} + \sum_{i=1}^{p} M^{N^i} - p + 1$*, где* k, p *удовлетворяют условию* (1)*.*

Следовательно, для любой M-коллекции языков из $\mathcal{L}_k$ будет существовать приведенный автомат, сложности не более, чем $\frac{N^{k-p}-1}{N-1} + \sum_{i=1}^{p} M^{N^i} - p + 1$.

Все инициальные автоматы, представляющие произвольную M-коллекцию конечных языков над алфавитом A, неотличимы друг от друга. Поэтому, по утверждению о единственности с точностью до изоморфизма автомата приведенного вида в классе эквивалентных инициальных автоматов, для каждой M-коллекции $\mathcal{T}(L, M)$, $L \subseteq \mathcal{L}_k$, $M \geqslant 2$ среди всех представляющих его инициальных автоматов существует единственный с точностью до изоморфизма приведенный автомат. Следовательно, приведенные автоматы, представляющие данную коллекцию языков, построенные различными способами, будут иметь одинаковое число состояний. Следовательно, приведенный автомат для коллекции T, построенный каким-либо другим способом, будет содержать столько же состояний, как и автомат в виде (N, l)-дерева, соединенного с обратным полным (N, M, p)-деревом первого рода, минимизированный согласно следствию 3.2.1. Чтобы построить коллекцию, представляемую построенным автоматом, сопоставим всем $\alpha_r \in A^{k-p}$, $r = 0, \ldots, N^{k-p}-1$ различные M-разбиения $\{A^p_{r,j},\ j = 0, \ldots, M-1\}$ множества A^p. Языки $L_j = \cup_{k=0}^{N^{k-p}-1}\{\alpha_r b \mid b \in A^p_{r,j}\}$ составляют искомую коллекцию.

Теорема 3.2.2. *Существует язык $L' \in \mathcal{L}_k$ и коллекция $\mathcal{T}(L', M)$, $\cup_{i=1}^{M-1} L'_i = L'$, представимая в сторого монотонным приведенном автомате сложности $\frac{N^{k-p}-1}{N-1} + \sum_{i=1}^{p} M^{N^i} - p + 1$, где p удовлетворяет условию* (1)*.*

Следствие 3.2.2. *Справедливы оценки $\frac{1}{N-1} \cdot \frac{N^k \cdot \log_N M}{k} \lesssim S(k) \lesssim \frac{N}{N-1} \cdot \frac{N^k \cdot \log_N M}{k}$, причем, верхняя оценка достигается при $l = N^p \cdot \log_N M - 1$ или $l = N^{p+1} \cdot \log_N M - 1$, а нижняя — при $l \sim N^p \cdot \log_N M + p$, $k = l + p + 1$.*

Следствие 3.2.3. *Для таких p и k, что $P_p \sim N \cdot Q_{k-p-1}$, выполнено*

$$p = \log_N k + \log_N \log_M N - \frac{\log_N e \cdot \log_M (k \cdot \log_M N)}{k \cdot \log_M N} + O\left(\frac{\log^2 k}{k^2}\right)$$

3.3. Сложность языков длины не больше k

Утверждение 3.3.1. *$\forall L \subseteq \mathcal{L}_{\leqslant k}$ и $M \geqslant 2$ любая M-коллекция $\mathcal{T}(L, M)$ представима строго монотонным автоматом в виде полного (N, k)-дерева, соединенного с тупиковым состоянием q_t, так что $\forall q \in R_k$, $a \in A$ $\varphi(q, a) = q_t$. Функция выхода $\psi(q, a) = i$ при переходах в состояния, соответствующие словам из L_i, $i = 0, \ldots, M-1$. В таком автомате некоторые состояния будут финальными и все состояния k-го уровня будут эквивалентны тупиковому.*

Утверждение 3.3.2. *В любом автомате, содержащем обратное полное (N, M, p)-дерево второго рода, любая пара состояний этого дерева (не эквивалентных тупиковому) будет отличима.*

Утверждение 3.3.3. *Состояния обратного полного (N, M, p)-дерева второго рода можно пронумеровать так, что для любого состояния q кроме одного (тупикового) выполнено $\varphi(q, a) > q \forall a \in A$.*

Теорема 3.3.1. *Любая M-коллекция $\mathcal{T}(L, M)$, $L \in \mathcal{L}_{\leqslant k}$ представима строго монотонным автоматом в виде полного (N, l)-дерева, соединенного с обратным полным (N, M, p)-деревом второго рода для $p = k - l - 1$, $l = 0, \dots, k-1$. При этом для каждого слова $\widetilde{a} \in L_i$ длины не больше l состояние дерева, соотвутствующее префиксу $\widetilde{a}$ является i-финальным, $i = 1, \dots, M-1$.*

Лемма 3.3.1. *Длину префикса, для которой построено (N, l)-дерево, можно выбрать таким образом, что*

$$P_p/N \leqslant Q_l \leqslant (M \cdot P_p)^N, \; k - l + p + 1, \tag{2}$$

где $Q_l = N^l$ — число состояний в последнем уровне (N, l)-дерева префиксов, $P_p = |R_p| = M^{\frac{N^{p+1}-N}{N-1}}$ — число состояний в последнем ярусе обратного (N, M, p)-дерева второго рода.

Формальной записью доказательства леммы 3.3.1. является алгоритм, приведенный в теореме 2.2.

Следствие 3.3.1. *При условии (2) можно соединить полное префиксное дерево и обратное полное (N, M, p)-дерево второго рода так, что все состояния (N, l)-дерева будут отличимы и все состояния последнего яруса обратного дерева высоты меньше p будут достижимы.*

Следствие 3.3.2. *Для любой M-коллекции $\mathcal{T}(L, M)$, $L \in \mathcal{L}_{\leqslant k}$, $N \geqslant 2$, $M \geqslant 2$ существует представляющий ее приведенный, строго монотонный автомат в виде (возможно, неполного) префиксного (N, l)-дерева, соединенного с обратным полным (N, M, p)-деревом второго рода, из которого удалены недостижимые состояния.*

Утверждение 3.3.4. *Любая коллекция $\mathcal{T}(L, M)$ языков длины не больше k представима строго монотонным автоматом сложности $\frac{N^{k-p}-1}{N-1} + M^{\frac{N^{p+1}-N}{N-1}}$, где k, p удовлетворяют условию (1).*

Следовательно, для любой коллекции языков из $\mathcal{L}_{\leqslant k}$ будет существует приведенный автомат сложности не более, чем $\frac{N^{k-p}-1}{N-1} + M^{\frac{N^{p+1}-N}{N-1}}$, представляющий данную коллекцию.

Все инициальные автоматы, представляющие произвольную коллекцию конечных языков над алфавитом A с помощью системы одноэлементных подмножеств алфавита B, неотличимы друг от друга. Поэтому, в силу единственности (с точностью до изоморфизма) автомата приведенного вида в классе эквивалентных ИКА, для каждой коллекции $\mathcal{T}(L, M)$, $L \in \mathcal{L}_{\leqslant k}$, $M \geqslant 2$ среди всех представляющих ее ИКА существует единственный (с точностью до изоморфизма) автомат приведенного вида.

Следовательно, приведенные автоматы, представляющие данную коллекцию языков, построенные различными способами, будут иметь одинаковое число состояний. Следовательно, приведенный автомат для коллекции $\mathcal{T}$, построенный каким-либо другим способом, будет содержать столько же состояний, сколько и автомат в виде (N, l)-дерева, соединеного с обратным полным (N, M, p)-деревом второго рода, минимизированный согласно следствию 3.3.1.

Для того, чтобы построить коллекцию языков, представимую в построенным автомате, соспоставим всем $\alpha_r \in A^{k-p}$, $r = 0, \ldots, N^{k-p} - 1$ различные M-разбиения $\{A^p_{r,j},\ j = 0, \ldots M-1\}$ множества A^{k-p-1}. Языки

$$L_j = A_j^{k-p-1} \cup \bigcup_{r=0}^{N^{k-p}-1} \{\alpha_r b \mid b \in A^p_{r,j}\}$$

составляют искомую коллекцию.

Теорема 3.3.2. *Существует язык* $L' \in \mathcal{L}_{\leqslant k}$ *и коллекция* $\mathcal{T}(L', M)$, $\bigcup_{i=1}^{M-1} L'_i = L'$, *представимая в строго монотонном приведенном автомате сложности* $\frac{N^{k-p}-1}{N-1} + M^{\frac{N^{p+1}-N}{N-1}}$, *где* p *удовлетворяет условию* (1).

Следствие 3.3.3. *Справедливы оценки*

$$\frac{N}{(N-1)^2} \cdot \frac{N^k \cdot \log_N M}{k} \lesssim S^c(\mathcal{L}_{\leqslant k}(A), N, M) \lesssim \frac{N^2}{(N-1)^2} \cdot \frac{N^k \cdot \log_N M}{k},$$

причем верхняя оценка достигается при $l = \frac{N^{p+1}-N}{N-1} \cdot \log_N M - 1$ *или* $l = \frac{N^{p+2}-N}{N-1} \cdot \log_N M - 1$, *а нижняя — при* $l \sim \frac{N^{p+1}-N}{N-1} \cdot \log_N M + p$, $k = l + p + 1$.

Следствие 3.3.4. *Для таких* p *и* k, *что* $P_p \sim N \cdot Q_{k-p-1}$, *выполнено*

$$p = \max\Big(0, \log_N k + \log_N \log_M N + \log_N \tfrac{N-1}{N} + \tfrac{N}{N-1} \cdot \tfrac{1}{k \cdot \ln N \cdot \log_M N} - \\ - \tfrac{\log_M \frac{N-1}{N}}{k \cdot \ln N \cdot \log_M N} - \tfrac{\log_M (k \cdot \log_M N)}{k \cdot \ln N \cdot \log_M N} + O(\tfrac{\log^2 k}{k^2})\Big).$$

Список литературы

[1] Кудрявцев В. Б., Алёшин С. В., Подколзин А. С. Введение в теорию автоматов. М.: Наука, 1985.

[2] Кудрявцев В. Б., Подколзин А. С., Ушчумлич Ш. Введение в теорию абстрактных автоматов. М.: Изд-во Моск. ун-та, 1985.

[3] Campeanu C., Santean N., Yu S. Minimal Cover-Automata for Finite Languages / Proceedings of the Third International Workshop on Implementing Automata (WIA'98). 1998. P. 32–42.