Об однозначности алфавитного декодирования

П. С. Дергач

Целью этой работы является установление алгоритмической разрешимости проблемы однозначности алфавитного декодирования регулярных текстов.

Ключевые слова: однозначное декодирование, регулярный язык, алфавитное кодирование.

1. Введение

А. А. Марковым было показано, что проблема однозначности алфавитного декодирования всех текстов над заданным конечным алфавитом A сводится к декодированию конечного числа слов над алфавитом A длины не большей некоторой вычислимой величины, зависящей от длины схемы кодирования, мощности алфавита A и др. параметров [1]. В предлагаемой работе исследован случай, когда кодируемое множество слов является любым регулярным множеством. Показывается, что результат Маркова А. А. может быть обобщен на случай алфавитного декодирования любого регулярного множества слов над алфавитом A.

2. Основные понятия и результаты

Абстрактным конечным автоматом называется набор $V = (A, Q, B, \varphi, \psi)$, где A, Q, B — конечные множества, φ — функция, определенная на множестве $Q \times A$ и принимающая значения из Q, ψ — функция, определенная на множестве $Q \times A$ и принимающая

значения из B. Множества A, Q, B называются соответственно *входным алфавитом*, *алфавитом состояний* и *выходным алфавитом* автомата V. Функция φ называется *функцией переходов*, а функция ψ — *функцией выходов* автомата V. *Входными словами* автомата V, $V = (A, Q, B, \varphi, \psi)$ называем произвольные конечные последовательности символов алфавита A. Для удобства рассматриваем при этом также «пустое» слово, не имеющее ни одного символа и обозначаемое Λ. *Выходными словами* алфавита V называем конечные последовательности символов алфавита B, *словами состояний* — конечные последовательности символов алфавита Q (в обоих случаях допускается и пустое слово Λ). Для каждого состояния автомата V можно рассмотреть набор $(A, Q, B, \varphi, \psi, q)$, определяющий автомат V с выделенным начальным состоянием q. Такие наборы $(A, Q, B, \varphi, \psi, q)$ называются *инициальными абстрактными конечными автоматами*; для них используется обозначение V_q.

Введем ряд понятий, связанных со словами. Пусть C — некоторое конечное множество. Если $\gamma = c(1) \dots c(n)$ — конечная последовательность символов $c(1), \dots, c(n)$ алфавита C, то говорим, что γ есть *слово в алфавите* C. Число n называем *длиной* слова γ и обозначаем через $|\gamma|$. Длина пустого слова равна 0. Если γ и δ — слова, причем $\gamma = \delta\delta'$ для некоторого слова δ', то говорим, что δ — *начало* слова γ, δ' — *конец* слова γ. Множество всех слов в алфавите C обозначаем C^*. Начало слова γ, имеющее длину l, обозначаем $[_l(\gamma)$; окончание слова γ, имеющее длину l, обозначаем через $]_l(\gamma)$. Обозначим через $\gamma_{l,m} :=]_{m-l}([_m(\gamma))$, где $|\gamma| \geqslant m > l \geqslant 1$.

Функции переходов и выходов алфавита $V = (A, Q, B, \varphi, \psi)$ доопределим на множестве $Q \times A^*$ (сохраним за ними те же обозначения). Именно, полагаем по определению

$$\varphi(q, \Lambda) = q, \ \varphi(q, \alpha a) = \varphi(\varphi(q, \alpha), a),$$

где $q \in Q, \ \alpha \in A^*, \ a \in A$. Аналогично,

$$\psi(q, \Lambda) = \Lambda, \ \psi(q, \alpha a) = \psi(\varphi(q, \alpha), a).$$

Пусть $V_q = (A, Q, B, \varphi, \psi, q)$ — инициальный абстрактный конечный автомат, $B' \subseteq B$. Множество $M = \{\alpha \mid \alpha \in A^*, \ \psi(q, \alpha) \in B'\}$ называем *представимым в конечном автомате* V_q с помощью подмножества

B' выходных символов. Говорим также, что автомат V_q *представляет* M *посредством* B'. Пусть $M \subseteq A^* \setminus \{\Lambda\}$. Если существует конечный автомат V_q, представляющий событие M посредством некоторого подмножества $B' \subseteq B$, то событие M называем *представимым*.

Введем понятие *обобщенного источника*. *Обобщенным источником* в алфавите A назовем конечный ориентированный граф G, у которого выделены начальная и финальная вершины $v, w, v \neq w$, причем каждому ребру приписано пустое слово Λ либо символ алфавита A. Допускается наличие в графе петель и кратных ребер. *Путем* в обобщенном источнике G будем называть последовательность $\pi = (v_1, \rho_1, v_2,\ \rho_2, \ldots, \rho_n, v_{n+1})$, где $v_1, v_2, \ldots, v_{n+1}$ — вершины графа G, ρ_i — ребро графа G, ведущее от вершины v_i к вершине v_{i+1}, $i = 1, \ldots, n,\ n \geqslant 1$. Пути π сопоставляем слово $[\pi] = a_1 \ldots a_n$, где a_i — символ алфавита A либо пустое слово Λ, приписанное ребру ρ_i, $i = 1, \ldots, n$. Говорим, что путь π *ведет от вершины* v_1 *к вершине* v_{n+1}. Пусть $\alpha \in A^* \setminus \{\Lambda\}$, u — вершина обобщенного источника G, множество всех вершин u' обобщенного источника G, для которых существует путь π, ведущий от u к u' и такой, что $[\pi] = \alpha$, обозначим $\theta(u, \alpha)$. Каждый обобщенный источник G с начальной вершиной v и финальной вершиной w определяет событие $|G| = \{\alpha \mid \alpha \in A^* \setminus \{\Lambda\}, w \in \theta(v, \alpha)\}$.

Пусть $A = \{a_1, \ldots, a_r\}$ — произвольный конечный непустой алфавит. Пусть P_1, P_2 — непустые множества слов в алфавите A. Здесь и далее для удобства пустое слово за элемент множества A^* не считается. Определим следующие операции над P_1 и P_2:

1. *Произведение* множеств P_1 и P_2 (обозначаем $P_1 \cdot P_2$) есть множество всех слов вида $\alpha_1\alpha_2$, где $\alpha_1 \in P_1,\ \alpha_2 \in P_2$.

2. *Итерация* множества P_1 (обозначаем $(P_1)^*$) есть множество всех слов вида $\alpha_1 \ldots \alpha_k$, где $\alpha_1 \in P_1,\ \ldots, \alpha_k \in P_1,\ k \geqslant 1$.

Введем понятие регулярного множества в алфавите A. Множество P, $P \subseteq A^*$, называем *регулярным в алфавите* A, если его можно получить из множеств вида $\{a\}$, $a \in A$, применением конечного числа операций $\cup, \cdot, ()^*$. Более подробно, определение регулярных множеств таково:

1. $\{a\}$, где a — произвольная буква алфавита A, — регулярное множество в алфавите A;

2. Если P_1, P_2 — регулярные множества в алфавите A, то $P_1 \cup P_2,\ P_1 \cdot P_2,\ (P_1)^*$ — регулярные множества в алфавите A;

3. Регулярность произвольного множества в алфавите A устанавливается в соответствиями с пп. 1, 2 за конечное число шагов. Введем понятие регулярного выражения в алфавите A. Регулярное выражение в алфавите A представляет собой слово в алфавите $A \cup \{\vee, \cdot, (,), *\}$, определяемое следующим образом:

1. Буквы алфавита A — регулярные выражения в алфавите A;

2. Если $\alpha,\ \beta$ — регулярные выражения в алфавите A, то $(\alpha \vee \beta),\ (\alpha \cdot \beta),\ (\alpha)^*$ — регулярные выражения в алфавите A;

3. Регулярность произвольного выражения в алфавите A устанавливается в соответствиями с пп. 1, 2 за конечное число шагов.

Сопоставим индуктивно каждому регулярному выражению $\mathfrak{P}$ в алфавите A регулярное множество $|\mathfrak{P}|$ в алфавите A:

1. Множество $\{a\}$ — в случае $\mathfrak{P} = a,\ a \in A$;
2. Множество $|\mathfrak{P}_1| \cup |\mathfrak{P}_2|$ — в случае $\mathfrak{P} = (\mathfrak{P}_1 \vee \mathfrak{P}_2)$;
3. Множество $|\mathfrak{P}_1| \cdot |\mathfrak{P}_2|$ — в случае $\mathfrak{P} = (\mathfrak{P}_1 \cdot \mathfrak{P}_2)$;
4. Множество $(|\mathfrak{P}_1|)^*$ — в случае $\mathfrak{P} = (\mathfrak{P}_1)^*$.

Зафиксируем два конечных непустых алфавита A и B.

Пусть есть какое-то отображение $f : A \to B^*$:

$$\begin{aligned} f(a_1) &= \beta_1, \\ f(a_2) &= \beta_2, \\ &\dots \\ f(a_r) &= \beta_r. \end{aligned}$$

Это соотношение называется *схемой кодирования*. Доопределим отображение f до отображения $\tilde{f} : A^* \to B^*$ следующим образом:

$$\tilde{f}(a_{i_1} a_{i_2} \dots a_{i_n}) = \beta_{i_1} \beta_{i_2} \dots \beta_{i_n}.$$

Это отображение $\tilde{f}$ будем называть *алфавитным кодированием*.

Пусть есть некоторое регулярное множество P в алфавите A и некоторое алфавитное кодирование f. Пусть $\beta \in \tilde{f}(P)$. Тогда $\alpha \in P$ называется *расшифровкой β при алфавитном кодировании $\tilde{f}$ на регулярном множестве P* или *расшифровкой β*, если $f(\alpha) = \beta$. Таких

расшифровок может быть несколько. Если для любых различных $\alpha_1, \alpha_2 \in P$ выполняется $\tilde{f}(\alpha_1) \neq \tilde{f}(\alpha_2)$, то говорим, что *декодирование однозначно на P по $\tilde{f}$.*

Теорема 1. *Существует алгоритм проверки однозначности алфавитного декодирования для любого регулярного текста.*

3. Доказательство вспомогательных утверждений

Лемма 1. *Если R — регулярное множество в алфавите A и представимо регулярным выражением $\mathfrak{P}$ в алфавите A с k операциями $\cdot, \vee$, то существует такой обобщенный источник G в алфавите A, что $R = |G|$, причём в G не более, чем $4k + 2$ вершин.*

Доказательство. Будем доказывать утверждение индукцией по количеству операций в $\mathfrak{P}$.

Если $R = \{a\}$, где $a \in A$, то искомый обобщенный источник G указан на рис. 1.

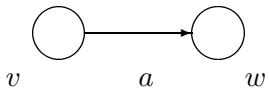

Рис. 1.

Здесь через v обозначена начальная вершина источника G, w — финальная вершина источника G. Здесь $k = 0$, и в G две вершины. Для этого случая утверждение доказано.

Пусть $\mathfrak{P} = \mathfrak{P}_1 \vee \mathfrak{P}_2$, в $\mathfrak{P}_1$ — k_1 операций $\cdot, \vee$, в $\mathfrak{P}_2$ — k_2 операций $\cdot, \vee$, и G_1, G_2 — обобщенные источники в алфавите A такие, что $|\mathfrak{P}_1| = |G_1|, |\mathfrak{P}_2| = |G_2|$. По предположению индукции можно считать, что в G_1 не более, чем $4k_1 + 2$ вершин, а в G_2 не более $4k_2 + 2$ вершин. Можно считать, что множества вершин графов G_1, G_2 не пересекаются. Объединим графы G_1, G_2 и введем две новые вершины v, w. От v проведем ребра с отметкой Λ к начальным вершинам обобщенных источников G_1, G_2; к w проведем ребра с отметкой Λ

от финальных вершин этих обобщенных источников. Получим обобщенный источник G с начальной вершиной v и финальной вершиной w, для которого $|G| = |\mathfrak{P}_1| \vee |\mathfrak{P}_2|$ (см. рис. 2).

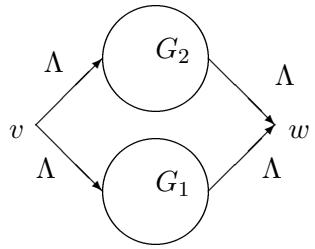

Рис. 2.

В обобщенном источнике G не более $(4k_1 + 2) + (4k_2 + 2) + 2 = 4(k_1 + k_2 + 1) + 2$ вершин, а в $\mathfrak{P}$ $k_1 + k_2 + 1$ операций $\cdot, \vee$.

Пусть $\mathfrak{P} = \mathfrak{P}_1 \cdot \mathfrak{P}_2$, в $\mathfrak{P}_1$ — k_1 операций $\cdot, \vee$, в $\mathfrak{P}_2$ — k_2 операций $\cdot, \vee$, и G_1, G_2 — обобщенные источники в алфавите A такие, что $|\mathfrak{P}_1| = |G_1|, |\mathfrak{P}_2| = |G_2|$. По предположению индукции можно считать, что в G_1 не более, чем $4k_1 + 2$ вершин, а в G_2 не более $4k_2 + 2$ вершин. Можно считать, что множества вершин графов G_1, G_2 не пересекаются. Проведем от финальной вершины обобщенного источника G_1 ребро с отметкой Λ к начальной вершине обобщенного источника G_2. Получим обобщенный источник G, начальной вершиной которого служит начальная вершина v графа G_1, а финальной — финальная вершина w графа G_2 (см. рис. 3).

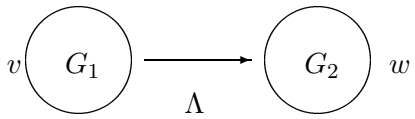

Рис. 3.

При этом $|G| = |\mathfrak{P}_1| \cdot |\mathfrak{P}_2|$. В обобщенном источнике G не более $(4k_1 + 2) + (4k_2 + 2) + 2 = 4(k_1 + k_2 + 1) + 2$ вершин, а в $\mathfrak{P}$ $k_1 + k_2 + 1$ операций $\cdot, \vee$.

Пусть $\mathfrak{P} = (\mathfrak{P}_1)^*$, в $\mathfrak{P}_1$ k_1 операций и G_1 — обобщенный источник такой, что $|\mathfrak{P}_1| = |G_1|$. По предположению индукции можно считать,

что в G_1 не более, чем $4k_1 + 2$ вершин. Проведем от финальной вершины w обобщенного источника G_1 ребро с отметкой Λ к начальной вершине v этого обобщенного источника. Получили обобщенный источник G с теми же начальной и финальной вершинами, что и G_1, причем $|G| = |(\mathfrak{P}_1)^*|$ (см. рис. 4).

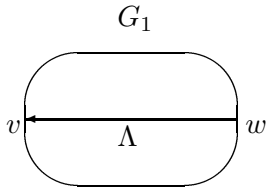

Рис. 4.

В обобщенном источнике G не более $(4k_1 + 2)$ вершин, а в $\mathfrak{P}$ k_1 операций $\cdot, \vee$.

Таким образом, утверждение индукции, а с ним и лемма 1 доказана.

Лемма 2. *Пусть G — обобщенный источник с k вершинами. Тогда событие $|G|$ представимо автоматом свходным алфавитом A, выходным алфавитом 0,1 и алфавитом состояний мощности 2^k.*

Доказательство леммы изложено в [2].

Лемма 3. *Пусть A — конечный алфавит, $\delta_1, \delta_2, \xi_1, \xi_2, \beta$ — слова (возможно, пустые) в алфавите A, B_1, B_2 — регулярные множества в алфавите A, и $\delta_1\beta\delta_2 \in B_1$, $\xi_1\beta\xi_2 \in B_2$, $V_1 = (A, Q_1, \{0,1\}, \varphi_1, \psi_1, q_1)$, $V_2 = (A, Q_2, \{0,1\}, \varphi_2, \psi_2, q_2)$ — инициальные абстрактные конечные автоматы, представляющие по множеству $\{1\}$ множества B_1, B_2 соответственно, тогда существует слово β' в алфавите A, $|\beta'| \leqslant |Q_1||Q_2|$, такое, что $\delta_1\beta'\delta_2 \in B_1$, $\xi_1\beta'\xi_2 \in B_2$.*

Доказательство. Допустим, что $|\beta| > |Q_1||Q_2|$. Рассмотрим множество T пар состояний автоматов V_1, V_2 такое, что $T = \{(\varphi_1(q_1, \delta_1[_l(\beta)), \varphi_2(q_2, \xi_1[_l(\beta))),\ 1 \leqslant l \leqslant |\beta|\}$. Так как $|T| < |Q_1||Q_2| + 1$, то существует $1 \leqslant l_1 < l_2 \leqslant |Q_1||Q_2| + 1$ такие, что $\varphi_1(q_1, \delta_1[_{l_1}(\beta)) = \varphi_1(q_1, \delta_1[_{l_2}(\beta))$, $\varphi_2(q_2, \xi_1[_{l_1}(\beta)) = \varphi_2(q_2, \xi_1[_{l_2}(\beta))$. Пусть β_1 и β_2 — такие, что $\beta_1 = [_{l_1}(\beta),\ \beta = [_{l_2}(\beta)\beta_2$. Тогда

$$\psi_1(q_1, \alpha_1\beta_1\beta_2\alpha_2) = \psi_1(\varphi_1(q_1, \alpha_1\beta_1), \beta_2\alpha_2) =$$
$$= \psi_1(\varphi_1(q_1, \alpha_1[_{l_2}(\beta)), \beta_2\alpha_2) = \psi_1(\varphi_1(q_1, \delta_1[_{l_2}(\beta)\beta_2), \delta_2) =$$
$$= \psi_1(\varphi_1(q_1, \delta_1\beta), \delta_2) = \psi_1(q_1, \delta_1\beta\delta_2) = 1.$$

Аналогично $\psi_2(q_2, \xi_1\beta_1\beta_2\xi_2) = 1$. Таким образом, существует $\beta'' = \beta_1\beta_2$ такое, что $\delta_1\beta''\delta_2 \in B_1, \xi_1\beta''\xi_2 \in B_2$ и $|\beta''| < |\beta|$. Проведя конечное число раз процедуру сокращения β, получим β' длины не больше $|Q_1||Q_2|$ такое, что $\delta_1\beta'\delta_2 \in B_1, \xi_1\beta'\xi_2 \in B_2$. Лемма 3 доказана.

Лемма 4. *Пусть α — слово в алфавите A, B — регулярное множество в алфавите A, представимое с помощью инициального абстрактного конечного автомата $V_{q_0} = (A, Q, \{0,1\}, \varphi, \psi, q_0)$ с помощью множества $\{1\}$, $\alpha \in B$, $1 \leqslant s < |\alpha|$, тогда существуют регулярные множества B_1 и B_2 в алфавите A, представимые с помощью инициальных абстрактных конечных автоматов с входным алфавитом A, выходным алфавитом $\{0,1\}$ и алфавитом состояний $|Q|$ с помощью множества $\{1\}$ такие, что $B_1 \cdot B_2 \subseteq B$, $\alpha \in B_1 \cdot B_2, [_s(\alpha) \in B_1,]_{|\alpha|-s}(\alpha) \in B_2$.*

Доказательство. По леммам 1 и 2, существует инициальный абстрактный конечный автомат $V = (A, Q, \{0,1\}, \varphi, \psi, q_0)$, представляющий регулярное множество B с помощью множества $\{1\}$. Пусть $\varphi(q_0, [_s(\alpha)) = q_1$. Рассмотрим два инициальных абстрактных конечных автомата $V_1 = (A, Q, \{0,1\}, \varphi, \psi_1, q_0)$, $V_2 = (A, Q, \{0,1\}, \varphi, \psi, q_1)$, где $\psi_1(q, a) = 1$ при $\varphi(q, a) = q_1$ и $\psi_1(q, a) = 0$ в противном случае. По теореме Клини (доказательство теоремы Клини изложено в [2]) эти два автомата по множеству $\{1\}$ представляют регулярные множества B_1 и B_2. Так как $\varphi(q_0, [_s(\alpha)) = q_1$, то $\psi_1(q_0, [_s(\alpha)) = 1$, то есть $[_s(\alpha) \in B_1$. Далее, так как $\alpha \in B$, то

$$1 = \psi(q_0, \alpha) = \psi(\varphi(q_0, [_s(\alpha)),]_{|\alpha|-s}(\alpha)) = \psi(q_1,]_{|\alpha|-s}(\alpha)),]_{|\alpha|-s}(\alpha) \in B_2.$$

Значит, $\alpha \in B_1 \cdot B_2$.

Пусть $\alpha_1 \in B_1, \alpha_2 \in B_2$. Тогда $\varphi(q_0, \alpha_1) = q_1$, $\psi(q_1, \alpha_2) = 1$. Значит, $\psi(q_0, \alpha_1\alpha_2) = \psi(\varphi(q_0, \alpha_1), \alpha_2) = \psi(q_1, \alpha_2) = 1$, то есть $\alpha_1\alpha_2 \in B$. Поэтому $B_1 \cdot B_2 \subseteq B$. Лемма 4 доказана.

Лемма 5. *Пусть f — алфавитное кодирование из алфавита A в алфавит B, $R_f = \max\limits_{1\leqslant i\leqslant r} l(\beta_i)$, где $\beta_i = f(a_i)$, $A = \{a_1, a_2, \ldots, a_r\}$, $\mathfrak{P}$ — регулярное выражение в алфавите A и в $\mathfrak{P}$ k операций $\cdot$, $\vee$, тогда в $\mathfrak{P}$ не более $k+1$ вхождений (с повторениями) букв алфавита A и в регулярном выражении $\tilde{f}(\mathfrak{P})$ в алфавите B, получаемом из $\mathfrak{P}$ заменой каждой буквы a_i на слово β_i, не более $(k+1)R_f - 1$ операций $\cdot, \vee$.*

Доказательство. Докажем первую часть утверждения леммы индукцией по количеству операций s в $\mathfrak{P}$. Если $s = 0$, то $\mathfrak{P} = a$ и в $\mathfrak{P}$ $k = 0$ операций $\cdot, \vee$. В $\mathfrak{P}$ входит ровно одна буква алфавита A и $1 \leqslant k + 1 = 1$.

Пусть утверждение доказано для $m \leqslant s$ и в $\mathfrak{P}$ $s+1$ операция. Рассмотрим три случая.

1. Последняя операция в $\mathfrak{P}$ — $*$, то есть $\mathfrak{P} = (\mathfrak{P}_0)^*$ и пусть в $\mathfrak{P}$ k вхождений букв алфавита A. Тогда и в $\mathfrak{P}_0$ k вхождений букв алфавита A и по предположению индукции в $\mathfrak{P}_0$, а, значит, и в $\mathfrak{P}$, входит не более $k+1$ операций $\cdot$, $\vee$.

2. Последняя операция в $\mathfrak{P}$ — $\cdot$, то есть $\mathfrak{P} = \mathfrak{P}_1 \cdot \mathfrak{P}_2$ и пусть в $\mathfrak{P}$, $\mathfrak{P}_1, \mathfrak{P}_2$ входит k, k_1, k_2 операций $\cdot$, $\vee$ соответственно, $k = k_1 + k_2 + 1$. По предположению индукции в $\mathfrak{P}_1$ входит не более $k_1 + 1$ букв алфавита A, в $\mathfrak{P}_2$ входит не более $k_2 + 1$ букв алфавита A. Значит, в $\mathfrak{P}$ входит не более $(k_1 + 1) + (k_2 + 1) = (k_1 + k_2 + 1) + 1 = k + 1$ букв алфавита A.

3. Последняя операция в $\mathfrak{P}$ — $\vee$, то есть $\mathfrak{P} = \mathfrak{P}_1 \vee \mathfrak{P}_2$ и пусть в $\mathfrak{P}$, $\mathfrak{P}_1, \mathfrak{P}_2$ входит k, k_1, k_2 операций $\cdot$, $\vee$ соответственно, $k = k_1 + k_2 + 1$. По предположению индукции в $\mathfrak{P}_1$ входит не более $k_1 + 1$ букв алфавита A, в $\mathfrak{P}_2$ входит не более $k_2 + 1$ букв алфавита A. Значит, в $\mathfrak{P}$ входит не более $(k_1 + 1) + (k_2 + 1) = (k_1 + k_2 + 1) + 1 = k + 1$ букв алфавита A. Первая часть утверждения доказана.

Докажем вторую часть утверждения леммы. Так как в $\mathfrak{P}$ не более $k+1$ вхождений (с повторениями) букв алфавита A и в регулярном выражении для β_i не больше $R_f - 1$ операций $\cdot$, $\vee$, то в регулярном выражении $\tilde{f}(\mathfrak{P})$ не более $k + (k+1)(R_f - 1) = (k+1)R_f - 1$ операций $\cdot$, $\vee$. Лемма 5 доказана.

Лемма 6. *Пусть $\tilde{f}$ — алфавитное кодирование, P — регулярное множество в алфавите A, соответствующее регулярному выражению $\mathfrak{P}$ в алфавите A с k операциями $\cdot, \vee, \alpha_1, \alpha_2 \in P$, $\alpha_1 \neq \alpha_2$ и $\tilde{f}(\alpha_1) = \tilde{f}(\alpha_2)$, тогда существуют такие $\alpha_1', \alpha_2' \in P$, $\alpha_1' \neq \alpha_2'$, что $\tilde{f}(\alpha_1') = \tilde{f}(\alpha_2')$ и $|\alpha_1'|, |\alpha_2'| \leqslant R_f^2 2^{8k+5}$.*

Доказательство. По лемме 1 существует обобщенный источник G в алфавите A такой, что $P = |G|$ и в нем не более $4k+2$ вершин. По лемме 2 множество $|G|$ представимо автоматом с входным алфавитом A и алфавитом состояний мощности не более 2^{4k+2}. Введем обозначения: пусть α — некоторое слово в алфавите A и $n \geqslant 2$. Пусть $\Delta := [|\alpha|/n]$. Тогда

$$(\alpha)_{n,1} := [_{\Delta}(\alpha),\ (\alpha)_{n,2} :=]_{|\alpha|-\Delta}(\alpha).$$

Пусть $p := 2R_f$. Возможны два случая: $|\tilde{f}\,((\alpha_1)_{p,1})\,| \leqslant |\tilde{f}\,((\alpha_2)_{p,1})\,|$ и $|\tilde{f}\,((\alpha_1)_{p,1})\,| > |\tilde{f}\,((\alpha_2)_{p,1})\,|$.

Разберем, например, первый случай. Второй случай разбирается аналогично.

Заметим, что

$$|\tilde{f}\,((\alpha_1)_{p,1})\,| \leqslant R_f|(\alpha_1)_{p,1}| \leqslant |\alpha_1|/2 \leqslant |\tilde{f}(\alpha_1)|/2,$$
$$|\tilde{f}\,((\alpha_2)_{p,1})\,| \leqslant R_f|(\alpha_2)_{p,1}| \leqslant |\alpha_2|/2 \leqslant |\tilde{f}(\alpha_2)|/2,$$

при этом $|\tilde{f}(\alpha_1)| = |\tilde{f}(\alpha_2)|$. Значит, части слова $\tilde{f}(\alpha_1) = \tilde{f}(\alpha_2)$, соответствующие словам $\tilde{f}\,((\alpha_1)_{p,1})$ и $\tilde{f}\,((\alpha_2)_{p,1})$ находятся в первой его половине.

По лемме 4 существуют регулярные множества B_1, B_2, C_1, C_2 в алфавите A такие, что $(\alpha_1)_{p,i} \in B_i, (\alpha_2)_{p,i} \in C_i$ для $i = 1, 2$, $B_1 \times B_2 \subseteq P$, $C_1 \times C_2 \subseteq P$, и эти множества представимы инициальными абстрактными конечными автоматами с входным алфавитом A, выходным алфавитом 0,1 и алфавитом состояний мощности не более 2^{4k+2} с помощью множества 1. Заменим в этих автоматах входной алфавит A на $\tilde{f}(A)$ и соответственно поправим функции перехода и выхода, заменив в них a_i на $\tilde{f}(a_i)$. Выходной алфавит, алфавит состояний и начальное состояние оставим теми же. Тогда регулярные множества $\tilde{f}(B_1)$,

$\tilde{f}(B_2)$, $\tilde{f}(C_1)$, $\tilde{f}(C_2)$ представимы этими автоматами. Мощность их алфавитов состояний попрежнему не больше 2^{4k+2}.

Теперь мы можем применить лемму 3 к паре регулярных множеств $\tilde{f}(B_1)$, $\tilde{f}(C_1)$ и к словам

$$\tilde{f}\left((\alpha_1)_{p,1}\right), \tilde{f}\left((\alpha_2)_{p,1}\right),$$

где

$$\beta = \tilde{f}\left((\alpha_1)_{p,1}\right), \delta_1 = \delta_2 = \gamma_1 = \Lambda, \gamma_2 = \tilde{f}(\alpha_2)_{|\tilde{f}((\alpha_1)_{p,1})|,|\tilde{f}((\alpha_2)_{p,1})|}.$$

Из нее следует, что существует слово $\beta' \in B^*$, $|\beta'| \leqslant 2^{8k+4}$ такое, что

$$\beta' \in \tilde{f}(B_1), \beta'\gamma_2 \in \tilde{f}(C_1).$$

Из построения автомата, представляющего $\tilde{f}(B_1)$, делаем вывод, что существует слово $\beta'' \in B_1$ такое, что $\tilde{f}(\beta'') = \beta'$. Обозначим через λ слово $\beta''(\alpha_1)_{p,2} \in B_1 \times B_2 \subseteq P$. Пусть различие в расшифровках α_1, α_2 слова $\tilde{f}(\alpha_1) = \tilde{f}(\alpha_2)$ есть в его второй половине. Тогда и у $\tilde{f}(\lambda)$ есть различные расшифровки, так как во второй половине слова расшифровки не изменились. Если же это не так, то различие в расшифровках α_1, α_2 слова $\tilde{f}(\alpha_1) = \tilde{f}(\alpha_2)$ есть в его первой половине. В этом случае такие же рассуждения применяются к записанному в обратном порядке слову $\tilde{f}(\alpha_1) = \tilde{f}(\alpha_2)$.

Итак, можно считать, что $|(\alpha_1)_{p,1}| \leqslant 2^{8k+4}$. Но $|(\alpha_1)_{p,1}| \geqslant |\alpha_1|/2R_f$. Поэтому $|\alpha_1| \leqslant R_f 2^{8k+5}$ и, значит, $|\alpha_2| \leqslant R_f^2 2^{8k+5}$. Утверждение леммы 6 доказано.

Лемма 7. *Существует алгоритм, определяющий по произвольной паре $(\tilde{f}, \mathfrak{P})$, где $\tilde{f}$ — алфавитное кодирование из алфавита A в алфавит B, а P — регулярное выражение в алфавите A, существуют ли $\alpha_1, \alpha_2 \in |\mathfrak{P}|$, $\alpha_1 \neq \alpha_2$, такие, что выполняется $\tilde{f}(\alpha_1) = \tilde{f}(\alpha_2)$.*

Доказательство. Проверим конечным перебором слов из $|\mathfrak{P}|$ длины не больше $R_f^2 2^{8k+5}$, существуют ли два слова $\alpha_1, \alpha_2 \in |\mathfrak{P}|$, $\alpha_1 \neq \alpha_2$, такие, что $\tilde{f}(\alpha_1) = \tilde{f}(\alpha_2)$ и $|\alpha_1|, |\alpha_2| \leqslant R_f^2 2^{8k+5}$. Если такие слова есть, то $\tilde{f}$ не однозначно декодируемо на $|\mathfrak{P}|$. Если таких слов нет, то по лемме 6 $\tilde{f}$ однозначно декодируемо на $|\mathfrak{P}|$. Алгоритм проверки

однозначности алфавитного декодирования, определяющий по произвольной паре $(\tilde{f}, \mathfrak{P})$, однозначно ли декодируемо $\tilde{f}$ на $|\mathfrak{P}|$, построен. Утверждение леммы 7 доказано. Нетрудно видеть, что из леммы 7 вытекает утверждение нашей теоремы 1.

Список литературы

[1] Яблонский С. В. Введение в дискретную математику. М.: Наука, 1979.

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