S©MR ISSN 1813-3304

СИБИРСКИЕ ЭЛЕКТРОННЫЕ
МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ

Siberian Electronic Mathematical Reports
http://semr.math.nsc.ru

Том 3, стр. 304–311 (2006) УДК 519.1
MSC 68R05

О ПЕРЕСТАНОВКАХ, ПОРОЖДЕННЫХ БЕСКОНЕЧНЫМИ БИНАРНЫМИ СЛОВАМИ

М.А. МАКАРОВ

Abstract. Let $w = w(1)w(2)\dots w(n)\dots$ be an arbitrary non-periodic infinite word on $\{0,1\}$. For every $i \in \mathbb{N}$ we may consider the binary real number $R_w(i) = 0, w(i)w(i+1)\dots$. For all $n \in \mathbb{N}$ the numbers $R_w(1), \dots, R_w(n)$ generate some permutation π_w^n of length n such that for all $i, j \in \{1, \dots, n\}$ the inequalities $\pi_w^n(i) < \pi_w^n(j)$ and $R_w(i) < R_w(j)$ are equivalent. A permutation is said to be *valid* if it is generated by some word. In this paper we investigate some properties of valid permutations. In particular, we prove a precise formula for the number of valid permutations of a given length. Also we consider a problem of continuability of valid permutations to the left.

1. Введение

В последнее время стало появляться много работ, посвящённых бесконечным перестановкам. Исследуются вопросы, связанные с избеганием паттернов, по аналогии с тем, как для символьных последовательностей рассматривается избегание некоторого набора подслов. Так например, в работе [1] рассмотрены перестановки, арифметически избегающие длинных монотонных паттернов.

Изучаются и другие свойства перестановок, так или иначе аналогичные соответствующим свойствам символьных последовательностей. Например, в статье [3] исследуется периодичность и низкая комбинаторная сложность перестановок.

Следует заметить, что символьные последовательности и бесконечные перестановки имеют много общего. В некоторых случаях удаётся каждой символьной последовательности сопоставить по некоторой схеме бесконечную перестановку, которая бы имела свойства, похожие на свойства самой символьной

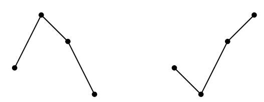
Makarov, M.A., On permutations generated by infinite binary words.

Поступила 23 ноября 2005 г., опубликована 25 июля 2006 г.

последовательности. В настоящей статье рассматривается схема, возникшая в связи с практическими задачами поиска подстроки в строке и, в частности, с так называемыми суффиксными массивами. Так, в статье [4] в качестве открытого вопроса сформулирована проблема комбинаторной характеризации перестановок, соответствующих суффиксным массивам. Эта проблема была решена в работе [6]. Однако в обоих указанных статьях, во-первых, рассматривались лишь конечные слова и перестановки, что существенно упрощает задачу, а, во-вторых, в них акцент сделан прежде всего на поиск оптимальных алгоритмов определения принадлежности перестановки рассматриваему классу и другие практические задачи. А такие характеристики, как количество перестановок заданной длины в указанном классе, продолжаемость перестановок и ряд других чисто комбинаторных свойств остались неисследованными.

Итак, пусть $w = w(1)w(2)\ldots w(n)\ldots$ — произвольное непериодическое бесконечное вправо слово над алфавитом $\{0, 1\}$. Каждому $i \in \mathbb{N}$ может быть сопоставлено действительное число $R_w(i) = 0{,}w(i)w(i+1)\ldots$ (в двоичной системе счисления). В силу непериодичности слова w все эти действительные числа будут различны. Поэтому для любого $n \in \mathbb{N}$ первые n этих чисел порождают некоторую перестановку $\pi_w^n\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$ такую, что для любых $i, j \in \{1, \ldots, n\}$ неравенство $\pi_w^n(i) < \pi_w^n(j)$ выполнено тогда и только тогда, когда выполнено неравенство $R_w(i) < R_w(j)$.

Например, для слова $w = 01101001\ldots$ имеем $R_w(1) = 0{,}01101001\ldots$, $R_w(2) = 0{,}1101001\ldots$, $R_w(3) = 0{,}101001\ldots$, $R_w(4) = 0{,}01001\ldots$. Ясно, что $R_w(4) < R_w(1) < R_w(3) < R_w(2)$. Поэтому слово w порождает перестановку $\pi_w^4 = 2431$, которая схематично изображена на рис. 1 слева.



2431 2134

Рис. 1. Допустимая и недопустимая перестановки

Оказывается, что не все перестановки могут быть порождены некоторым бесконечным словом над алфавитом $\{0, 1\}$. Например, для перестановки $\tau = 2134$ не существует слова w такого, что $\pi_w^4 = \tau$. В самом деле, если бы такое слово w существовало, то первый его символ был бы 1, ибо иначе было бы выполнено неравенство $R_w(1) < R_w(2)$, противоречащее неравенству $\tau(1) > \tau(2)$ (подробнее это рассуждение будет проведено в доказательстве леммы 1, где будет показано, что если $w(i) = 0$, то $R_w(i) < R_w(i+1)$). Аналогично получаем, что $w(3) = 0$. Но тогда $R_w(1) > R_w(3)$, однако $\tau(1) < \tau(3)$.

Итак, перестановку $\pi^n\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$ длины n назовём *допустимой*, если $\pi^n = \pi_w^n$ для некоторого слова w. Из предыдущих примеров следует, что перестановка 2431 допустима, а перестановка 2134 — нет. В статье исследуются свойства допустимых перестановок, в частности, для любого n найдена точная формула для $P(n)$ — количества всех допустимых перестановок длины n. Мы приведём два доказательства этой формулы, причём в одном из них будет параллельно рассмотрен вопрос о продолжаемости допустимых перестановок влево.

Оказалось, что данная задача тесно связана с так называемыми *примитивными словами*, т.е. конечными словами, не представляющимися в виде β^k ни для какого слова β и натурального числа $k > 1$. Как известно (см. например [5, стр. 9]), примитивных слов длины l над двухсимвольным алфавитом существует в точности $\psi(l) = \sum_{d|l} \mu(l/d)2^d$, где μ — функция Мёбиуса, которая определяется следующим образом:

$$\mu(n) = \begin{cases} 0, & \text{если } n \text{ делится на квадрат целого числа, большего } 1; \\ (-1)^k, & \text{если } n = p_1 \cdot \ldots \cdot p_k, \text{ где } p_i \text{ — различные простые числа.} \end{cases}$$

2. Допустимые перестановки

Лемма 1. *Пусть слова u и v порождают одну и ту же перестановку длины l, т.е. $\pi_u^l = \pi_v^l$. Тогда эти слова совпадают в первых $l-1$ символах, т.е. $u(i) = v(i)$ для $i \in \{1, \ldots, l-1\}$.*

Доказательство. Легко видеть, что если $w(i) = 0$, то $R_w(i) < R_w(i+1)$. В самом деле, пусть $m = \min\{j | j \geq i, w(j) = 1\}$. Тогда $2^{m-i-1} \cdot R_w(i) = 0,1w(m+1)w(m+2)\ldots < 1$, а $2^{m-i-1} \cdot R_w(i+1) = 1, w(m+1)w(m+2)\ldots > 1$, и, тем самым, $R_w(i) < R_w(i+1)$. Аналогично, если $w(i) = 1$, то $R_w(i) > R_w(i+1)$.

Предположим теперь, что слова u и v отличаются в p-ом символе, $p \leq l-1$. Тогда, не ограничивая общности, можно считать, что $u(p) = 0$ и $v(p) = 1$. Но тогда $R_u(p) < R_u(p+1)$ и $R_v(p) > R_v(p+1)$. Поэтому $\pi_u^l \neq \pi_v^l$, что противоречит условию. □

Следствие 1. *Для любого $n \in \mathbb{N}$ выполнено неравенство $P(n) \geq 2^{n-1}$.*

Доказательство. Рассмотрим все слова длины $n-1$ над алфавитом $\{0,1\}$. Продолжим каждое из них до бесконечного вправо слова произвольным образом. Все 2^{n-1} получившихся слов порождают некоторые перестановки длины n, причём согласно лемме 1 все эти перестановки различны. Тем самым, будет порождено 2^{n-1} допустимых перестановок. □

Для перестановки π и $i \neq j$ через γ_{ij}^π (или просто через γ_{ij}, если понятно, о какой перестановке идёт речь) будем обозначать один из символов $<$ или $>$, соответствующий неравенствам $\pi(i) < \pi(j)$ и $\pi(i) > \pi(j)$. Иными словами,

$$\gamma_{ij} = \begin{cases} <, & \text{если } \pi(i) < \pi(j), \\ >, & \text{если } \pi(i) > \pi(j). \end{cases}$$

Например, для перестановки $\pi = 132$ имеем $\gamma_{12}^\pi =<$, $\gamma_{23}^\pi =>$, $\gamma_{31}^\pi =>$, $\gamma_{13}^\pi =<$, $\gamma_{32}^\pi =<$.

Лемма 2. *Допустимая перестановка длины l полностью определяется символами $\gamma_{1l}, \gamma_{2l}, \ldots, \gamma_{(l-1)l}, \gamma_{12}, \gamma_{23}, \ldots, \gamma_{(l-2)(l-1)}$.*

Доказательство. Понятно, что перестановка полностью определяется всеми $l(l-1)/2$ символами γ_{ij} для $i, j \in \{1, \ldots, l\}$. Покажем, как в случае допустимой перестановки определить их через указанные в условии символы. Пусть π_w^l — допустимая перестановка, порождённая словом w. Пусть $1 \leq i < j \leq l$ и мы хотим определить символ γ_{ij}^π. Если $j = l$, то этот символ уже определён. Поэтому далее считаем, что $1 \leq i < j < l$. Если $\gamma_{m(m+1)} =<$, то, повторив рассуждения из доказательства леммы 1, легко получить, что $w(m) = 0$. Аналогично, если $\gamma_{m(m+1)} =>$, то $w(m) = 1$. Применяя этот результат к $m = i$ и $m = j$, получим, что если $\gamma_{i(i+1)} \neq \gamma_{j(j+1)}$, то и $w(i) \neq w(j)$ и, тем самым, символ γ_{ij} определён. Если же $\gamma_{i(i+1)} = \gamma_{j(j+1)}$, то и $w(i) = w(j)$, а значит, $\gamma_{ij} = \gamma_{(i+1)(j+1)}$. Продолжая рассуждать таким образом, на t-ом шаге мы либо определим $\gamma_{(i+t)(j+t)}$, либо получим, что $\gamma_{(i+t)(j+t)} = \gamma_{(i+t+1)(j+t+1)}$. Но при $t = l - j$ символ $\gamma_{(i+t)(j+t)} = \gamma_{(i+l-j)l}$ определён по условию. Поэтому процесс завершится и символ γ_{ij} будет определён. □

Следствие 2. *Для любого $n \geq 2$ выполнено неравенство $P(n) \leq 2^{2n-3}$.*

Доказательство. Непосредственно следует из леммы 2. □

Теорема 1. *Каждую допустимую перестановку длины $l \geq 2$ можно продолжить влево до допустимой перестановки длины $l + 1$ либо двумя, либо тремя способами, причём перестановок второго типа существует в точности $\psi(l) = \sum_{d|l} \mu(l/d)2^d$ штук, где μ — функция Мёбиуса.*

Доказательство. Пусть π_w^l — допустимая перестановка длины l, порождённая словом w. Тогда слова $0w$ и $1w$ порождают перестановки длины $l + 1$, продолжающие π_w^l слева, причём согласно лемме 1 эти перестановки различны. Тем самым, любая допустимая перестановка продолжается влево не менее чем двумя способами.

Покажем, что больше чем тремя способами ни одну перестановку продолжить влево не удастся. Для перестановки $\tau = \pi_{\alpha w}^{l+1}$, порождённой словом αw, где $\alpha \in \{0, 1\}$, символы γ_{ij}^τ для $i, j \in \{2, \ldots, l+1\}$ определены перестановкой π_w^l. Поэтому, согласно лемме 2, для определения всей перестановки τ достаточно определить символы γ_{12}^τ и $\gamma_{1(l+1)}^\tau$. Однако символ γ_{12}^τ, очевидно, определяется символом α. Покажем теперь, что хотя бы в одном из случаев $\alpha = 0$ или $\alpha = 1$ оставшийся символ $\gamma_{1(l+1)}^\tau$ определяется однозначно. В самом деле, если $\pi_w^l(1) < \pi_w^l(l)$, то при $\alpha = 0$ имеем $\gamma_{1(l+1)}^\tau =<$, поскольку $R_{0w}(1) < R_{0w}(2) = R_w(1) < R_w(l) = R_{0w}(l + 1)$; если же $\pi_w^l(1) > \pi_w^l(l)$, то при $\alpha = 1$ имеем $\gamma_{1(1+l)}^\tau =>$, поскольку $R_{1w}(1) > R_{1w}(2) = R_w(1) > R_w(l) = R_{1w}(l + 1)$. Тем самым, никакую допустимую перестановку нельзя продолжить влево более чем тремя способами.

Подсчитаем теперь более точно количество допустимых перестановок длины l, продолжаемых влево тремя способами. Для этого рассмотрим всевозможные слова α длины l над алфавитом $\{0, 1\}$. Каждое из них при продолжении вправо до бесконечного слова порождает какие-то перестановки длины $l + 1$. Среди них могут встречаться пары перестановок, имеющие одинаковый конец (при удалении первого члена). Например, слово 101 длины 3 при двух продолжениях вправо $1011010\ldots$ и $1011011\ldots$ порождает перестановки 3142 и 2143, имеющие одинаковый конец 132.

Количество таких пар перестановок, имеющих одинаковый конец, очевидно и равно числу допустимых перестановок длины l, продолжаемых влево тремя способами. В самом деле, каждой такой паре соответствует случай продолжения влево перестановки длины l до перестановки длины $l + 1$, в котором символ $\gamma_{1(1+l)}$ не определяется однозначно по остальным символам; обратно если перестановка продолжаема влево тремя способами, то из приведённых выше рассуждений следует, что это возможно только в случае, если два из этих трёх продолжений отличаются лишь символами $\gamma_{1(1+l)}$.

Мы собираемся доказать, что каждому примитивному слову α соответствует в точности одна пара перестановок, имеющих одинаковый конец, а каждому непримитивному слову α — ни одной. Если мы это докажем, то результат леммы будет очевиден, поскольку, как известно, примитиных слов длины l существует ровно $\psi(l) = \sum_{d|l} \mu(l/d)2^d$.

Сначала рассмотрим непримитивные слова. Пусть α — произвольное непримитивное слово длины l, $\alpha = \beta^m$ $(m > 1)$, где β — слово длины t. Пусть слово α при некотором продолжении до бесконечного вправо слова порождает перестановку τ длины $l + 1$. Поскольку $\alpha(i) = \alpha(i + t)$ для $i \in \{1, \dots, l - t\}$, то $\gamma^\tau_{1(1+t)} = \gamma^\tau_{2(2+t)} = \dots = \gamma^\tau_{(1+t)(1+2t)} = \dots = \gamma^\tau_{(l+1-t)(l+1)}$. Тем самым, последовательность $\tau(1)$, $\tau(1 + t)$, $\dots$, $\tau(l + 1)$ монотонная. Поэтому символ $\gamma^\tau_{1(l+1)}$ однозначно определяется по символу $\gamma^\tau_{(1+t)(l+1)}$ (и равен ему). Стало быть, слову α не может соответствовать пара перестановок длины $l + 1$ с одинаковым концом.

Пусть теперь α — произвольное примитивное слово длины l. Поскольку для $l \geq 2$, то в слове α встречаются и единицы, и нули (если бы это было не так и α состояло бы только из нулей или только из единиц, то оно было бы непримитивным). Пусть α_0 — слово, полученное из α заменой одного из нулей на единицу, а α_1 — слово, полученное из α заменой одной из единиц на ноль. Продолжим α вправо до бесконечных слов w_0 и w_1 следующим образом: $w_i = \alpha^2\alpha_i\beta$, $i = 0, 1$, где β — произвольное непериодическое бесконечное вправо слово над алфавитом $\{0, 1\}$. Положим $\tau_i = \pi^{l+1}_{w_i}$.

Рассмотрим два слова $\overline{\alpha_i} = \alpha(i)\alpha(i+1)\dots\alpha(l)\alpha(1)\dots\alpha(i-1)$ и $\overline{\alpha_j} = \alpha(j)\alpha(j+1)\dots\alpha(l)\alpha(1)\dots\alpha(j-1)$, являющихся циклическими перестановками слова α. Согласно предложению 1.3.2 из книги [5, стр. 8] следует, что все циклические перестановки любого примитивного слова различны. Поэтому слова $\overline{\alpha_i}$ и $\overline{\alpha_j}$ не могут совпадать. Стало быть, в словах w_0 и w_1 все символы $\gamma^{\tau_0}_{ij}$ и $\gamma^{\tau_1}_{ij}$, кроме $\gamma^{\tau_0}_{1(l+1)}$ и $\gamma^{\tau_1}_{1(l+1)}$, полностью определяются префиксом α^2. Поэтому перестановки τ_0 и τ_1 имеют одинаковый конец. Осталось лишь показать, что они различны. Но это очевидно, поскольку слова w_0 и w_1 конструировались так, чтобы символы $\gamma^{\tau_0}_{1(l+1)}$ и $\gamma^{\tau_1}_{1(l+1)}$ были различны.

Итак, для каждого примитивного слова α мы указали пару соответствующих перестановок, имеющих одинаковый конец. Покажем, что такая пара только одна. Предположим противное. Пусть (π_0, π_1) и (τ_0, τ_1) — две такие пары. Перестановки π_0 и π_1, а также τ_0 и τ_1 отличаются друг от друга только символом $\gamma_{1(l+1)}$. Поэтому далее под π будем понимать одну из перестановок π_0 и π_1 (или и ту и другую сразу), а под τ — одну из перестановок τ_0 и τ_1; при этом символы $\gamma^\pi_{1(l+1)}$ и $\gamma^\tau_{1(l+1)}$ будем считать неопределёнными.

Поскольку перестановки π и τ различны, то существуют такие i, j, $i < j$, что $\gamma^\pi_{ij} \neq \gamma^\tau_{ij}$. Поскольку слова, порождающие перестановки π и τ, имеют общий префикс α, то это может означать только то, что $\alpha(i) = \alpha(j)$. В самом деле, если $\alpha(i) = 0$ и $\alpha(j) = 1$, то $\gamma^\pi_{ij} = \gamma^\tau_{ij} =<$; если же $\alpha(i) = 1$ и $\alpha(j) = 0$, то $\gamma^\pi_{ij} = \gamma^\tau_{ij} =>$.

Итак, $\alpha(i) = \alpha(j)$. Поэтому $\gamma^{\pi}_{ij} = \gamma^{\pi}_{(i+1)(j+1)}$ и $\gamma^{\tau}_{ij} = \gamma^{\tau}_{(i+1)(j+1)}$. Проводя такие рассуждения и далее, придём к тому, что $\gamma^{\pi}_{(l+1-d)(l+1)} \neq \gamma^{\tau}_{(l+1-d)(l+1)}$, где $d = j - i$.

Для удобства дальнейшего изложения будем называть число d *хорошим*, если существуют такие i, j, что $j - i = d$ и $\gamma^{\pi}_{ij} \neq \gamma^{\tau}_{ij}$. Мы только что показали, что если d — хорошее число, то $\gamma^{\pi}_{(l+1-d)(l+1)} \neq \gamma^{\tau}_{(l+1-d)(l+1)}$.

Легко видеть, что $\alpha^{\pi}_{i1} = \alpha^{\pi}_{i(l+1)}$ и $\alpha^{\tau}_{i1} = \alpha^{\tau}_{i(l+1)}$ для $i \in \{2, \ldots, l\}$. Поэтому если d — хорошее число, то $l - d$ — тоже хорошее число, причём $\alpha^{\pi}_{(l+1-(l-d))(l+1)} \neq \alpha^{\pi}_{(l+1-d)(l+1)}$ и $\alpha^{\tau}_{(l+1-(l-d))(l+1)} \neq \alpha^{\tau}_{(l+1-d)(l+1)}$. Поэтому хорошим является также число $l - 2d$. Продолжая рассуждать так и далее, придём либо к тому, что l делится на d, либо к тому, что остаток от деления l на d также является хорошим числом. В последнем случае мы можем снова повторять все эти рассуждения, и, в конце концов, получим, что некоторый делитель числа l (возможно единица) является хорошим числом.

Итак, пусть $l = md$, $m > 1$ и d — хорошее число. Тогда из предыдущих рассуждений следует, что хорошими будут также числа $2d$, $3d$, $\ldots$, $(m-1)d$. Но тогда $\gamma^{\pi}_{1(1+d)} = \gamma^{\pi}_{(1+d)(1+2d)} = \cdots = \gamma^{\pi}_{(1+(m-1)d)(1+l)}$ и $\gamma^{\tau}_{1(1+d)} = \gamma^{\tau}_{(1+d)(1+2d)} = \ldots = \gamma^{\tau}_{(1+(m-1)d)(1+l)}$, а значит последовательности $\pi(1), \pi(1+d), \pi(1+2d), \ldots, \pi(1+l)$ и $\tau(1), \tau(1+d), \tau(1+2d), \ldots, \tau(1+l)$ монотонные. Стало быть, $\gamma^{\pi_0}_{1(1+l)} = \gamma^{\pi}_{1(1+d)} = \gamma^{\pi_1}_{1(1+l)}$ и $\gamma^{\tau_0}_{1(1+l)} = \gamma^{\tau}_{1(1+d)} = \gamma^{\tau_1}_{1(1+l)}$, однако пары перестановок (π_0, π_1) и (τ_0, τ_1) выбирались таким образом, что $\gamma^{\pi_0}_{1(1+l)} \neq \gamma^{\pi_1}_{1(1+l)}$ и $\gamma^{\tau_0}_{1(1+l)} \neq \gamma^{\tau_1}_{1(1+l)}$. Противоречие. □

Теорема 2. *Допустимых перестановок длины $l + 1 \geq 2$ существует в точности $P(l+1) = \sum\limits_{t=1}^{l} \psi(t) \cdot 2^{l-t}$ штук.*

Первое доказательство теоремы 2. Из теоремы 1 очевидным образом следует, что $P(n+1) = 2P(n) + \psi(n)$. В самом деле, все $P(n)$ допустимых перестановок длины n можно продолжить двумя способами влево, а для $\psi(n)$ из них существует дополнительное третье продолжение. Ясно также, что $P(2) = 2$. Избавляясь от рекуррентности, легко находим, что $P(n+1) = \sum_{t=1}^{n} \psi(t) \cdot 2^{n-t}$. □

Для второго доказательства теоремы 2, не использующего теорему 1, нам понадобится следующая лемма.

Лемма 3. *Количество перестановок длины $l + 1$, порождаемых различными продолжениями слова α длины l, равно количеству примитивных суффиксов слова α.*

Доказательство. Пусть бесконечное вправо слово β продолжает слово α до слова $\alpha\beta$. Будем непрерывно менять β от последовательности нулей до последовательности единиц в лексикографическом порядке. Во время этого процесса при некоторых β перестановка $\pi^{l+1}_{\alpha\beta}$ будет не определена. Ясно, что это возможно, только если β есть периодическое слово, периодом которого является некоторый примитивный суффикс слова α. Пусть α имеет k примитивных суффиксов. Точки неопределённости порождаемой перестановки $\pi^{l+1}_{\alpha\beta}$ делят весь интервал изменения β на k промежутков, на каждом из которых порождается одна и та же перестановка, а на разных промежутках — разные. В самом деле, при лексикографическом увеличении β символ $\gamma_{i(l+1)}$ может измениться только в точке неопределённости, соответствующей суффиксу $\alpha(i+1)\dots\alpha(l)$ в случае примитивности этого суффикса, либо наименьшему периоду этого суффикса в случае его непримитивности. Доказывать больше нечего. □

Второе доказательство теоремы 2. Учитывая лемму 3, для нахождения $P(l+1)$ нам осталось лишь просуммировать количество примитивных суффиксов по всем словам длины l, поскольку согласно лемме 1 разным словам будут соответствовать разные допустимые перестановки. Но это просто: ко всем $\psi(t)$ примитивным словам длины t приписываем слева $n-t$ произвольных символов алфавита $\{0,1\}$, а затем суммируем всё по t от 1 до l. В итоге получаем требуемую формулу $P(l+1) = \sum_{t=1}^{l} \psi(t) \cdot 2^{l-t}$. □

Интересно отметить, что та же самая последовательность $P(n)$ возникла совершенно в другой задаче. В работе [2] доказано, что количество регулярных языков, порождённых конечными детерминированными автоматами над односимвольным алфавитом с количеством состояний не больше n, равно $\sum\limits_{t=1}^{n} \psi(t) \cdot 2^{n-t}$, т.е. в точности $P(n+1)$. Можно построить биекцию между допустимыми перестановками и регулярными языками. Действительно, как было показано во втором доказательстве теоремы 2, каждой допустимой перестановке π длины $n+1$ соответствует слово α длины n с выделенным примитивным суффиксом. Дальнейшая биекция легко восстанавливается по лемме 2 и теореме 1 из статьи [7], а также по теоремам 2, 3 и 4 из [2]. Каждому слову α длины n с выделенным примитивным суффиксом сопоставим детерминированный автомат A с n состояниями по следующему правилу. Состояниями автомата $s_1, \dots, s_n$ объявим символы $\alpha(1), \dots, \alpha(n)$ слова α, причём если символ равен нулю, то состояние объявим неконечным, а если единице, то — конечным. Функцию перехода определим таким образом, чтобы из состояния s_i можно было перейти в состояние s_{i+1}; из состояния s_n переход в состояние, соответствующее первому символу выделенного примитивного суффикса слова α. Нетрудно понять, что любой минимальный детерминированный автомат над односимвольным алфавитом не больше чем с n состояниями имеет ровно один цикл, поскольку из каждого состояния есть ровно один переход. Кроме того, возможен ещё какой-то предцикл. Поэтому по каждому такому автомату можно построить слово длины n с выделенным примитивным суффиксом. Таким образом, построеное соответствие действительно является биекцией.

Кроме того, в теореме 5 из [2] найдена асимптотическая оценка для последовательности $P(n)$. А именно,

$$P(n+1) = 2^n(n - \alpha + O(n2^{-n/2})),$$

где

$$\alpha = \sum_{d=2}^{\infty} \frac{\mu(d)}{1-2^{d-1}} \approx 1.38271445540239628547.$$

Автор благодарен С.В. Августиновичу за внимание к работе и содержательные замечания, позволившие дать второе доказательство теоремы 2, а также А.Э. Фрид за внимание к работе и помощь при оформлении статьи.

Список литературы

[1] J.A. Davis, R.C. Entringer, R.L. Graham, and G.J. Simmons. *On permutations containing no long arithmetic progressions.* Acta Arithmetica 34 (1977), 81–90.

[2] M. Domaratzki, D. Kisman, J. Shallit. *On the number of distinct languages accepted by finite automata with n states.* J. Autom. Lang. Comb. 7 (2002), 469–486.

[3] D.G. Fon-Der-Flaass and A.E. Frid. *On periodicity and low complexity of infinite permutations.* Discrete Mathematics and Theoretical Computer Science proc. AE, 2005, 267–272.

[4] R. Grossi and J.S. Vitter. *Compressed suffix arrays and suffix trees with applications to text indexing and string matching.* In Proceedings of the 32nd Annual ACM Symposium on Theory of Computing, pages 397–406, 2000. http://www.cs.duke.edu/~jsv/Papers/GrV00.text_indexing.ps.gz.

[5] M. Lothaire. *Combinatorics on Words*, Vol. 17 of Encyclopedia of Mathematics and Its Applications. Addison-Wesley, 1983.

[6] M. He, J.I. Munro, S.S. Rao. *A Categorization Theorem on Suffix Arrays with Applications to Space Efficient Text Indexes.* To appear in Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms, January 2005.

[7] C. Nicaud. *Average state complexity of operations on unary automata.* In M. Kutylowski, L. Pacholski, and T. Wierzbicki, editors, Proc. 24th Symposium, Mathematical Foundations of Computer Science 1999, Vol. 1672 of Lecture Notes in Computer Science, pp. 231–240. Springer-Verlag, 1999.

Михаил Александрович Макаров
Новосибирский государственный университет,
ул. Пирогова, 2,
630090, Новосибирск, Россия
E-mail address: mike_mak@ngs.ru