УДК 519.111

И. О. Коряков

СИЛЬНО ПРЕФИКСНЫЕ КОДЫ И ОБОБЩЕННАЯ ФОРМУЛА РЭЙНИ*

Джордж Рэйни в своей работе [1], исследуя свойства языка Лукасевича, выводит перечислительную формулу, на которой строит элегантное комбинаторное доказательство известной формулы Лагранжа–Бюрмана. Сама перечислительная формула (именно ее я называю «формулой Рэйни») не привлекла внимания математиков, хотя, в силу транзитивности импликации, результаты, получаемые с помощью формулы Лагранжа–Бюрмана (а это наиболее мощный и популярный инструмент в комбинаторных приложениях производящих функций), формально вытекают уже из формулы Рэйни.

В настоящей заметке вводится класс сильно префиксных кодов (включающий язык Лукасевича) и доказывается некоторое обобщение формулы Рэйни. В качестве иллюстрации возможностей новой формулы приводятся краткие доказательства нескольких известных результатов.

Зафиксируем используемые обозначения: $\mathbb{Z}$ – аддитивная группа целых чисел; $\mathbb{N} = \{n \in \mathbb{Z} \mid n \geqslant 0\}$ – множество натуральных чисел; $\langle n \rangle = \{1, 2, \ldots, n\}$; X – конечный или счетный алфавит; X^* – множество всех слов над X. Если $w \in X^*$, $a \in X$, то $d_a(w)$ обозначает число вхождений буквы a в слово w; если буквы пронумерованы, то вместо $d_{a_i}(w)$ пишем $d_i(w)$. Если u – префикс слова w (т. е. $w \in uX^*$), то пишем $u \leqslant w$. Пустое слово обозначается через e; $X^+ = X^* \setminus \{e\}$.

Слово $p \in X^*$ называется *примитивным*, если равенство $p = w^k$, где $k \in \mathbb{N}$, влечет $k = 1$. Слова $u, v \in X^*$ (циклически) *сопряжены* (мы пишем $u \sim v$), если одно получается из другого циклической перестановкой букв. Иначе говоря, $u \sim v$, если $u = xy$, $v = yx$ для некоторых $x, y \in X^*$. Класс сопряженности слова w обозначим через $C(w)$.

1. Сильно префиксные коды

Лемма 1. *Пусть $u, v \in X^*$. Тогда*
(1) *если $uv = vu$, то $u = p^m, v = p^n$ для некоторых $p \in X^*$, $m, n \in \mathbb{N}$;*
(2) *любое непустое слово является степенью примитивного слова;*
(3) *если u – степень примитивного слова p, то $|C(u)| = |C(p)| = |p|$.*

---

*Работа выполнена при поддержке программы «Развитие научного потенциала высшей школы», проект № 2.1.1/3537.

Доказательство. (1) и (2) см. в [2, гл. 11]. Докажем (3). Пусть $u = p^k$, где p примитивно, и $v \sim u$. Тогда v получается из u перестановкой из начала в конец слова u нескольких экземпляров слова p (что не меняет u) и некоторого префикса p_1 слова $p = p_1p_2$. Положив $q = p_2p_1$, получим $v = q^k$; поэтому отображение $q \mapsto q^k$ – биекция $C(p)$ на $C(u)$.

Пусть $p = a_1a_2 \dots a_m$, $a_i \in X$. Положим $p_i = a_i \dots a_m a_1 \dots a_{i-1}$, $i \in \langle m \rangle$. Все слова из $C(p)$ находятся в списке $(p_1, \dots, p_m)$. Допустим, среди них есть равные: $p_i = p_j$, $i < j$. Положив $v = a_i \dots a_{j-1}$, получим $p_i = vw = wv = p_j$. В силу (1) слова v, w и, значит, p_i будут степенями некоторого слова r. Если $p_i = r^k$, то $p = q^k$, где $q \sim r$, причем $|q| = |r| \leqslant |v| = j - i < m = |p|$, т. е. $k > 1$, что противоречит примитивности p. Следовательно, $|C(p)| = m = |p|$.

Пусть $L \subseteq X^+$ – код, т. е. базис свободного подмоноида L^*. Тогда любое слово $w \in L^*$ можно рассматривать как слово в алфавите L. Поэтому можно ввести его длину $|w|_L$ над L ($|w|_L = k$, если $w \in L^k$), понятия примитивности и сопряженности над L. Через $C_L(w)$ обозначим класс сопряженности слова w над L: если $w = w_1 \dots w_n$, где все $w_i \in L$, то

$$C_L(w) = \{w_i \dots w_n w_1 \dots w_{i-1} \mid i \in \langle n \rangle\}.$$

Естественно, лемма 1 верна и при замене X алфавитом L.

Рассмотрим морфизм $\delta : X^* \to \mathbb{Z}$ такой, что $\delta \geqslant -1$ на X, причем $\delta^{-1}(-1) \cap X \neq \varnothing$ (впредь буква δ всегда будет обозначать именно такой морфизм). Обозначим через L язык, состоящий из всех слов $w \in X^*$, удовлетворяющих условиям

(δ1) $\delta(w) = -1$;

(δ2) $u < w \Rightarrow \delta(u) \geqslant 0$.

Очевидно, L – префиксный код (т. е. $L \cap LX^+ = \varnothing$). Назовем его *сильно префиксным кодом относительно* δ или просто δ*-кодом.*

Примеры. (1) Пусть $X = \{x, \overline{x}, a_1, \dots, a_r\}$, $r \in \mathbb{N}$, а $\delta : X^* \to \mathbb{Z}$ задается равенствами $\delta(x) = 1$, $\delta(\overline{x}) = -1$, $\delta(a_1) = \ldots = \delta(a_r) = 0$; таким образом, $\delta(w) = d_x(w) - d_{\overline{x}}(w)$ для всех $w \in X^*$. Назовем r*-цве́тным языком Моцкина* язык $\mathcal{M}(r)$, определяемый следующими двумя условиями на слова $w \in X^*$: $\delta(w) = 0$; $u \leqslant w \Rightarrow \delta(u) \geqslant 0$. Язык $\mathcal{D} = \mathcal{M}(0)$ называется (ограниченным) *языком Дика.* Очевидно, δ-кодом является язык $\mathcal{M}(r)\overline{x}$.

(2) Пусть $X = \{x_n \mid n \in \mathbb{N}\}$ и δ определен равенствами $\delta(x_n) = n - 1$, $n \in \mathbb{N}$. Соответствующий сильно префиксный код $\mathcal{L}$ называется *языком Лукасевича.* Этот язык полностью характеризуется равенством $\mathcal{L} = \bigcup_{n \geqslant 0} x_n \mathcal{L}^n$, где $\mathcal{L}^0 = \{e\}$, означающим, что если интерпретировать x_n как символ n-арной

операции, то $\mathcal{L}$ описывает все типы правильных алгебраических выражений в бесскобочной префиксной форме Лукасевича.

Лемма 2. *Пусть $L \subseteq X^+$ есть δ-код, $w \in X^+$, $k > 0$. Тогда $w \in L^k$, если и только если $(\delta(w) = -k)$ & $(u < w \;\Rightarrow\; \delta(u) > -k)$.*

Доказательство. Пусть $w = w_1 \dots w_k$, где все $w_i \in L$. Очевидно, $\delta(w) = -k$. Если $u < w$, то $u = w_1 \dots w_{i-1}v$, где $i \leqslant k$ и $v \leqslant w_i$, причем $v < w_i$ при $i = k$. Тогда $\delta(u) = -i + 1 + \delta(v)$. Если $i < k$, то $\delta(v) \geqslant -1$ и $\delta(u) \geqslant -i > -k$; если $i = k$, то $\delta(v) \geqslant 0$ и снова $\delta(u) \geqslant -k + 1 > -k$.

Обратное утверждение докажем индукцией по k. При $k = 1$ – это определение δ-кода. Пусть $k > 1$ и слово $w = ua$, где a – буква, удовлетворяет условиям леммы. Тогда $-k = \delta(w) = \delta(u) + \delta(a)$, т. е. $\delta(a) < 0$, откуда $\delta(a) = -1$, $\delta(u) = -k+1$. Пусть $w = w_{k-1}v_k$, где w_{k-1} – кратчайший префикс со значением $\delta = -k + 1$. Тогда $\delta(v_k) = -1$, и если $v < v_k$, то

$$\delta(w_{k-1}v) = -k + 1 + \delta(v) \geqslant -k + 1,$$

т. е. $\delta(v) \geqslant 0$. Таким образом, $v_k \in L$. По индуктивной гипотезе имеем $w_{k-1} = v_1 \dots v_{k-1}$, где все $v_i \in L$; поэтому $w = v_1 \dots v_k \in L^k$.

Доказательства следующих двух лемм в основном следуют общей методологии Рэйни.

Лемма 3. *Пусть $L \subseteq X^+$ – сильно префиксный код, $k > 0$ и $w \in L^k$. Тогда*
(1) *$C_L(w) = C(w) \cap L^k$;*
(2) *если w – степень слова $p \in X^+$, то $p \in L^+$;*
(3) $\dfrac{|C_L(w)|}{|C(w)|} = \dfrac{|w|_L}{|w|}$.

Доказательство. (1) Включение слева направо очевидно. Докажем обратное включение. Пусть $w = w_1 \dots w_k$, где все $w_i \in L$, и $w' \in C(w) \cap L^k$. Тогда

$$w' = uw_{i+1} \dots w_k w_1 \dots w_{i-1}v,$$

где $i \in \langle k \rangle$, $w_i = vu$, $v \neq e$. Если допустить $u \neq e$, то получим $\delta(v) \geqslant 0$, $\delta(u) \leqslant -1$, откуда $\delta(uw_{i+1} \dots w_{i-1}) = \delta(u) - (k-1) \leqslant -k$, что противоречит лемме 2. Поэтому $u = e$, $v = w_i$ и $w' \in C_L(w)$.

(2) Пусть снова $w = w_1 \dots w_k, w_i \in L$, и $w = p^m$, где $p \notin L^+$. Тогда $p < w$ и $p = w_1 \dots w_{i-1}u$, где $i \in \langle k \rangle$, $w_i = uv$, причем $u, v \neq e$; в частности, $\delta(u) \geqslant 0$. Имеем

$$\delta(w) = m\delta(p) = -k, \quad \delta(p) = -(i-1) + \delta(u).$$

Рассмотрим слово $w' = p^{m-1}w_1 \dots w_{i-1} < p^m = w$. Тогда

$$\delta(w') = (m-1)\delta(p) - (i-1) = -k - \delta(p) - (i-1) = -k - \delta(u) \leqslant -k,$$

что противоречит лемме 2. Поэтому $u = e$ или $v = e$, т. е. $p \in L^{i-1}$ или $p \in L^i$.

(3) Пусть $|w| = n$ и $w = p^m$, где p примитивно (лемма 1(2)). Тогда, по (2), $p \in L^+$ и p примитивно над L. Имеем

$$|w| = n = m|p|, \quad |w|_L = k = m|p|_L.$$

В силу леммы 1(3) $|C(w)| = |p|$ и $|C_L(w)| = |p|_L$. Поэтому

$$\frac{|C_L(w)|}{|C(w)|} = \frac{|p|_L}{|p|} = \frac{k}{n} = \frac{|w|_L}{|w|}.$$

Замечание 1. Лемма 3 неверна в случае произвольного префиксного кода. Например, для бипрефиксного кода $L = \{a^2, ab^2, aba, b^2ab\}$ нарушается п. (1): слова $w = (a^2)(b^2ab)$ и $w' = (ab^2)(aba)$ лежат в $C(w) \cap L^2$, но $w' \notin C_L(w)$. А для бипрефиксного кода $L = \{b, aba\}$ нарушаются пп. (2) и (3): $w = baba \in L^2$, но $ba \notin L^+$; $C(w) = C_L(w)$, но $|w| = 4, |w|_L = 2$.

Замечание 2. Лемму 3(3) можно слегка обобщить. Пусть R – кольцо, расширяющее поле рациональных чисел, и $h : X^* \to R$ – отображение, постоянное на каждом классе сопряженности. Тогда $|w| \cdot h(C_L(w)) = |w|_L \cdot h(C(w))$, где $h(K) = \sum_{w \in K} h(w)$ для конечного $K \subset X^*$.

Лемма 4. *Пусть $L \subseteq X^+$ – δ-код, $w \in X^+$ и $\delta(w) = -k$, где $k > 0$. Тогда $C(w) \cap L^k \neq \varnothing$.*

Доказательство. Пусть наименьшее значение δ на префиксах слова w будет $-m$, а w_m – кратчайший префикс с этим значением δ. Рассуждая по индукции, предположим, что для каждого i, $\ell < i \leqslant m$, найден кратчайший префикс w_i с $\delta(w_i) = -\ell$, причем $w_{\ell+1} < \ldots < w_m \leqslant w$. Если $w_{\ell+1} = w'a$, где $a \in X$, то $\delta(w') = -\ell$. В самом деле, $\delta(w') \geqslant -\ell$; будь неравенство строгим, получили бы $\delta(a) = \delta(w_{\ell+1}) - \delta(w') < -(\ell+1) + \ell = -1$, что невозможно по определению δ. Значит, существует и кратчайший префикс $w_\ell < w_{\ell+1}$ с $\delta(w_\ell) = -\ell$. Таким образом, $w = u_1 \dots u_m v$, где $u_1 \dots u_i = w_i$, $i \in \langle m \rangle$, и все $u_i \in L$. Осталось показать, что $u = vu_1 \dots u_{m-k+1} \in L$; тогда лемма будет доказана:

$$w \sim w' = u_{m-k+2} \dots u_m u \in L^k.$$

Заметим сначала, что $\delta(v) = \delta(w) - \delta(u_1 \dots u_m) = -k + m$, откуда $\delta(u) = \delta(v) - (m-k+1) = -1$. Если $u' < u$, то либо $u' \leqslant v$ (а тогда $\delta(u') \geqslant 0$, что завершает доказательство), либо $u' = vu_1 \ldots u_{i-1}u_i'$, где $i \leqslant m - k + 1$, $u_i' \leqslant u_i$, причем $u_i' < u_i$, если $i = m-k+1$. При $i = m-k+1$ будет $\delta(u_i') \geqslant 0$, а при меньших i будет $\delta(u_i') \geqslant -1$. В любом случае

$$\delta(u') = \delta(v) + \delta(u_1 \ldots u_{i-1}) + \delta(u_i') = m - k - i + 1 + \delta(u_i') \geqslant 0.$$

Назовем язык $Q \subseteq X^*$ *устойчивым относительно сопряжения,* если для любых $u, v \in X^*$

$$(u \in Q) \& (v \sim u) \Rightarrow v \in Q.$$

Положим $Q_n = Q \cap X^n$.

Теорема (обобщенная формула Рэйни, ОФР). *Пусть $n, k \geqslant 1$, $L \subseteq X^+$ есть δ-код, а язык $Q \subseteq X^*$ устойчив относительно сопряжения. Тогда*

$$|L^k \cap Q_n| = \frac{k}{n} |\delta^{-1}(-k) \cap Q_n|.$$

Доказательство. Язык $\delta^{-1}(-k) \cap Q_n$ устойчив относительно сопряжения, ибо таковыми являются $\delta^{-1}(-k)$, Q и X^n. По лемме 4 каждый класс сопряженности $C \subseteq \delta^{-1}(-k) \cap Q_n$ содержит класс сопряженности $C_L \subseteq L^k \cap Q_n$, единственный в силу леммы 3(1). Поэтому представители w_i, $i \in I$, классов $C_L \subseteq L^k \cap Q_n$ образуют множество представителей и классов $C \subseteq \delta^{-1}(-k) \cap \cap Q_n$, т. е.

$$L^k \cap Q_n = \bigcup_{i \in I} C_L(w_i), \quad \delta^{-1}(-k) \cap Q_n = \bigcup_{i \in I} C(w_i).$$

Учитывая, что $|w_i| = n$, $|w_i|_L = k$, из леммы 3(3) получаем

$$|L^k \cap Q_n| = \sum_{i \in I} |C_L(w_i)| = \frac{k}{n} \sum_{i \in I} |C(w_i)| = \frac{k}{n} |\delta^{-1}(-k) \cap Q_n|.$$

Замечание 3. Используя замечание 2, для указанного там отображения h получаем формулу

$$h(L^k \cap Q_n) = \frac{k}{n} h(\delta^{-1}(-k) \cap Q_n).$$

Это и есть исходная формула Рэйни, если положить $L = \mathcal{L}$ (язык Лукасевича) и $Q = X^*$.

2. Иллюстрации

Приведем несколько примеров, иллюстрирующих перечислительные возможности ОФР, сопровождая их полными, но (в отличие от оригиналов!) короткими доказательствами.

Следующее утверждение доказали Кори и Вокелэн с помощью методов (но не формулы) Рэйни.

Предложение 1 [3]. *Пусть $\mathcal{D}$ – язык Дика. Число k-списков $(w_1, \ldots, w_k)$ слов из $\mathcal{D}$ суммарной длины $2m$ равно* $\dfrac{k}{2m+k}\dbinom{2m+k}{m}$.

Доказательство. Имеется биекция $(w_1, w_2, \ldots, w_k) \mapsto w_1\overline{x}w_2\overline{x}\ldots w_k\overline{x}$ перечисляемого множества на $(\mathcal{D}\overline{x})^k \cap X^{2m+k}$. Согласно ОФР

$$|(\mathcal{D}\overline{x})^k \cap X^{2m+k}| = \frac{k}{2m+k}|\delta^{-1}(-k) \cap X^{2m+k}|.$$

Чтобы получить слово из $\delta^{-1}(-k) \cap X^{2m+k}$, нужно из $2m+k$ мест выбрать m мест для буквы x, т. е.

$$|\delta^{-1}(-k) \cap X^{2m+k}| = \binom{2m+k}{m}.$$

Применение биекций в комбинаторных моделях расширяет возможности ОФР. Приведем один пример.

Напомним, что корневое дерево называется *упорядоченным*, если сыновья каждого узла линейно упорядочены; оно называется *полным n-арным*, если каждый отец имеет n сыновей ($n > 0$). Обозначим через $\mathcal{T}$ множество всех конечных упорядоченных деревьев, а через $\mathcal{CT}(n,k)$ – множество всех полных n-арных деревьев с k отцами.

Путь длины m в целочисленной плоской решетке $\mathbb{Z} \times \mathbb{Z}$ – это список $\pi = (P_0, P_1, \ldots, P_m)$ точек решетки; векторы $\pi_i = P_i - P_{i-1}$, $i \in \langle m \rangle$, – *шаги*. Если точка P_0 фиксирована (обычно это $(0,0)$), то путь π однозначно определен списком шагов, и мы пишем $\pi = (\pi_1, \ldots, \pi_m)$ в виде слова над алфавитом допустимых шагов. Обозначим через $\mathcal{P}(n)$ множество путей с началом $(0,0)$, шагами $(1,0)$ и $(0,1)$, лежащих не выше прямой $y = (n-1)x$ и заканчивающихся на ней. Через $\mathcal{P}(n,k)$ обозначим множество путей из $\mathcal{P}(n)$, заканчивающихся в точке $(k, (n-1)k)$.

Пусть $\mathcal{L}$ – язык Лукасевича, $Y_n = \{x_0, x_n\}$, $\mathcal{L}(n,k)$ – множество всех слов из $\mathcal{L}(n) = \mathcal{L} \cap Y_n^+$, содержащих k вхождений буквы x_n.

Наконец, положим $F(n,k) = \frac{1}{nk+1}\binom{nk+1}{k}$; это *число Фусса*, введенное Н. И. Фуссом в 1791 г. Заметим, что $F(2,k) = C_k$ – k-е число Каталана.

Предложение 2. *Справедливы равенства*

$$|\mathcal{CT}(n,k)| = |\mathcal{P}(n,k)| = |\mathcal{L}(n,k)| = F(n,k).$$

Доказательство. 1) Известна биекция $\mathcal{T} \to \mathcal{L}$: каждый узел дерева $T \in \mathcal{T}$, имеющий p сыновей, помечается буквой x_p; результат сканирования меток сверху вниз (при обходе слева направо) – слово из $\mathcal{L}$. Сужение этой биекции на $\mathcal{CT}(n,k)$ является биекцией на $\mathcal{L}(n,k)$.

2) Пусть $r = (1,0)$, $u = (0,1)$. Определим морфизм δ из $\{r,u\}^*$ в $\mathbb{Z}$: $\delta(r) = n-1$, $\delta(u) = -1$. Легко видеть, что $\mathcal{P}(n)$ есть множество всех слов w из $\{r,u\}^*$, удовлетворяющих условию $(\delta(w) = 0)$ & $(v \leqslant w \Rightarrow \delta(v) \geqslant 0)$. Язык $\mathcal{P}(n)u$ является δ-кодом, а соответствие $r \mapsto x_n, u \mapsto x_0$ определяет биекцию из $\mathcal{P}(n,k)u$ на $\mathcal{L}(n,k)$.

3) Пусть Q – множество всех слов $w \in Y_n^*$ таких, что $d_n(w) = k$. Слова $w \in \delta^{-1}(-1) \cap Q$ определяются условиями

$$\delta(w) = -1 = d_0(w) + (n-1)k, \quad |w| = d_0(w) + k,$$

т. е. $d_0(w) = (n-1)k+1$, $|w| = nk+1$. Чтобы сформировать такое слово, нужно из $nk+1$ мест выбрать k мест для x_n. Отсюда, пользуясь ОФР, получаем

$$|\mathcal{L}(n,k)| = |\mathcal{L}(n) \cap Q_{nk+1}| = \frac{1}{nk+1}|\delta^{-1}(-1) \cap Q| = \frac{1}{nk+1}\binom{nk+1}{k}.$$

Утверждение о $\mathcal{CT}(n,k)$ хорошо известно; обычно его доказывают с помощью формулы Лагранжа–Бюрмана. Утверждение о $\mathcal{P}(n,k)$ доказано в [4] (двойным) применением той же формулы. Утверждение об $\mathcal{L}(n,k)$ можно вывести из следующей теоремы, которую Рэйни объявляет основным результатом своей работы [1] и доказывает независимо от формулы Рэйни (она неприменима).

Пусть $\mathcal{L}$ – язык Лукасевича, $n_1, \ldots, n_m$ – натуральные числа, $\mathcal{L}^k(n_1, \ldots, n_m)$, где $k > 0$, – множество всех слов $w \in \mathcal{L}^k$ таких, что $d_i(w) = n_i$ для $i \in \langle m \rangle$, и $d_i(w) = 0$ для $i > m$.

Предложение 3 [1]. *Справедливо равенство*

$$|\mathcal{L}^k(n_1, \ldots, n_m)| = \frac{k}{n}\binom{n}{n_0, n_1, \ldots, n_m},$$

где $n_0 = k + \sum_{i=1}^m (i-1)n_i$, $n = k + \sum_{i=1}^m i n_i$.

Доказательство. Обозначим через Q множество всех слов из X^*, удовлетворяющих условиям, сформулированным перед предложением. Таким образом, $\mathcal{L}^k(n_1,\ldots,n_m)=\mathcal{L}^k\cap Q$. Для слов $w\in\delta^{-1}(-k)\cap Q$ легко вычисляются $n_0=d_0(w)$ и $n=|w|$:

$$\delta(w)=-k=-n_0+\sum_{i=1}^{m}(i-1)n_i,\quad n=\sum_{i=0}^{m}n_i=k+\sum_{i=1}^{m}in_i.$$

Чтобы сформировать слово из $\delta^{-1}(-k)\cap Q=\delta^{-1}(-k)\cap Q_n$, нужно из n мест выбрать n_i мест для x_i, $0\leqslant i\leqslant m$, откуда

$$|\delta^{-1}(-k)\cap Q_n|=\binom{n}{n_0,n_1,\ldots,n_m}.$$

ОФР завершает доказательство.

Замечание 4. Язык Лукасевича $\mathcal{L}$ является в некотором смысле универсальным для широкого класса δ-кодов. Если M – δ-код над алфавитом Y и сужение $\delta|Y$ инъективно, то существует единственный морфизм $\varphi:Y^*\to X^*$, сохраняющий длины слов и значения δ. При этом φ биективно отображает M на δ-код $\mathcal{L}\cap\varphi(Y)^*$; скажем, что δ-код M *вложим* в $\mathcal{L}$. Таковым, например, является δ-код $\mathcal{D}\overline{x}$; поэтому предложение 1 вытекает уже из исходной формулы Рэйни. Однако язык Лукасевича редко используется для кодирования комбинаторных объектов. Он уступает языку Дика и родственным языкам (таким, как r-цветные языки Моцкина, смеси языков Дика и т. п.). Заметим, что кодирующие возможности языка Дика не отличаются от таковых возможностей языка Лукасевича — существует очень простая биекция $\mathcal{D}$ на $\mathcal{L}$:

$$x^{k_1}\overline{x}x^{k_2}\overline{x}\ldots x^{k_n}\overline{x}\ \mapsto\ x_{k_1}x_{k_2}\ldots x_{k_n}x_0.$$

Приведем два примера, где работают δ-коды, невложимые в $\mathcal{L}$.

Рассмотрим множество P путей в плоской целочисленной решетке с шагами $a=(0,1)$, $b=(0,-1)$, $c=(1,0)$, начинающихся в точке $(0,0)$, заканчивающихся на диагонали $y=x$ и не опускающихся ниже этой диагонали. Очевидно, длины таких путей четны. В доказательстве следующего результата Гессель (см. также [4]) использовались формальные ряды Лорана и формула Лагранжа–Бюрмана.

Предложение 4 [5]. *Число путей из P длины $2k$, заканчивающихся в точке (m,m), равно*

$$\frac{1}{2k+1}\binom{2k+1}{k}\binom{k}{m}.$$

Доказательство. Легко видеть, что P – множество всех слов w в алфавите $X = \{a, b, c\}$, удовлетворяющих условиям

(1) $d_a(w) = d_b(w) + d_c(w)$;

(2) $u \leqslant w \Rightarrow d_a(u) \geqslant d_b(u) + d_c(u)$.

Определим морфизм $\delta\colon X^* \to \mathbb{Z}$, положив $\delta(a) = 1$, $\delta(b) = \delta(c) = -1$. Соответствующим δ-кодом служит язык $L = Pb \cup Pc$. Пусть $P_k(m)$ – перечисляемое множество, $L_k(m) = P_k(m)b \cup P_k(m)c$; тогда $|L_k(m)| = 2|P_k(m)|$. Определим язык $Q \subseteq X^*$ условием

$$w \in Q \Leftrightarrow (d_c(w) = m \vee d_c(w) = m + 1).$$

Таким образом, $L_k(m) = L \cap Q_{2k+1}$ и, в силу ОФР,

$$|P_k(m)| = \frac{1}{2}|L \cap Q_{2k+1}| = \frac{1}{2(2k+1)}|\delta^{-1} \cap Q_{2k+1}|.$$

Если $w \in \delta^{-1}(-1) \cap Q_{2k+1}$, то

$$d_a(w) - d_b(w) - d_c(w) = -1, \quad d_a(w) + d_b(w) + d_c(w) = 2k + 1,$$

откуда $d_a(w) = k$. Чтобы сформировать такое w, нужно из $2k+1$ мест выбрать k мест для a, из первых k среди оставшихся мест выбрать m мест для c и на последнее свободное место поставить b или c. Поэтому

$$|\delta^{-1}(-1) \cap Q_{2k+1}| = 2\binom{2k+1}{k}\binom{k}{m}.$$

Пусть $\mathcal{M}$ – двухцветный язык Моцкина над $X = \{x, \overline{x}, a, b\}$.

Предложение 5. *Число слов $w \in \mathcal{M}$ длины n, в которых $d_x(w) + d_a(w) = k$, равно числу Рюньона*

$$R(n+1, k+1) = \frac{1}{n+1}\binom{n+1}{k+1}\binom{n+1}{k}.$$

Доказательство. Определим язык $Q \subseteq X^*$ условием

$$w \in Q \Leftrightarrow d_x(w) + +d_a(w) = k.$$

Тогда перечисляемое множество равномощно $\mathcal{M}\overline{x} \cap Q_{n+1}$. Чтобы сформировать слово из $\delta^{-1}(-1) \cap Q_{n+1}$, нужно из $n + 1$ мест выбрать k мест для x и a, из этих k мест выбрать $k - i$ мест для a (т. е. i мест для x), из оставшихся

$n+1-k$ мест выбрать $i+1$ мест для $\overline{x}$ и просуммировать по i от 0 до $n-k$. Тогда с помощью тождества Вандермонда получим

$$|\delta^{-1}(-1) \cap Q_{n+1}| = \binom{n+1}{k} \sum_{i=0}^{n-k} \binom{k}{k-i} \binom{n+1-k}{i+1} = \binom{n+1}{k} \binom{n+1}{k+1}.$$

ОФР завершает доказательство.

Разумеется, в этом предложении буквы x и $\overline{x}$, а также a и b взаимозаменяемы.

1. Raney G. N. Functional composition patterns and power series reversion // Trans. Amer. Math. Soc. 1960. Vol. 94, № 5. P. 441–451.
2. Лаллеман Ж. Полугруппы и комбинаторные приложения. М.: Мир, 1985.
3. Cori R., Vauquelin B. Planar maps are well labeled trees // Canad. J. Math. 1981. Vol. 33, № 5. P. 1023–1042.
4. Гульден Я., Джексон Д. Перечислительная комбинаторика. М.: Наука, 1990.
5. Gessel I.M. A factorization for formal Laurent series and lattice path enumeration // J. Combin. Theory Ser. A. 1980. Vol. 28, № 3. P. 321–337.

Статья поступила 06.12.2007