УДК 519.713

ИТЕРАЦИИ ЯЗЫКОВ И МАКСИМАЛЬНЫЕ ПРЕФИКСНЫЕ КОДЫ

С. Ю. Корабельщикова[1], Б. Ф. Мельников[2]

[1] *Северный (Арктический) федеральный университет им. М. В. Ломоносова,*
[2] *Самарский государственный университет*

Поступила в редакцию 17.02.2014 г.

Аннотация: в статье рассматривается связь максимальных префиксных кодов с бесконечными итерациями языков. Приведено необходимое условие коммутирования в глобальном надмоноиде свободного моноида и рассмотрены частные случаи такого коммутирования: когда один из языков состоит из одного или двух слов. В терминах максимальных префиксных кодов сформулирован критерий эквивалентности пары конечных языков. Произведён подсчёт числа максимальных префиксных кодов с ограничением длины слов над алфавитом произвольной мощности. Приведён ряд нерешённых задач, сформулированы гипотезы необходимых условий коммутирования, требующие проверки в будущем.

Ключевые слова: итерации конечных языков, максимальные префиксные коды.

ITERATIONS OF LANGUAGES AND MAXIMAL PREFIX CODES

S. Y. Korabelshchikova, B. F. Melnikov

Abstract: in this paper we consider the relationship between maximal prefix codes and iterations of finite languages. There is obtained the necessary conditions of commutation in the global supermonoid of the free monoid and considered some special cases such commutation: i.e., when one of considered languages consists of one or two words. In the terms of maximal prefix codes, we formulated the criterium for equivalence of two finite languages. We counted the number of maximal prefix codes having the limitation for the length of words over the alphabet of arbitrary power. We consider some unsolved problems, and formulated necessary conditions of commutation, which requires verification in the future.

Keywords: iterations of finite languages, maximal prefix codes.

1. ВВЕДЕНИЕ. МАКСИМАЛЬНЫЕ ПРЕФИКСНЫЕ КОДЫ И ПРОБЛЕМА РАВЕНСТВА В РАЗНЫХ КЛАССАХ ЯЗЫКОВ

Максимальные префиксные коды рассматриваются в нескольких разных разделах широко известной монографии [1]. В настоящей статье мы опишем их связь с некоторыми вопросами теории формальных языков, в частности – с бесконечными итерациями языков, рассмотренными одним из авторов настоящей статьи в [2] и др.

В качестве «мотивации» (т.е. *почему* мы рассматриваем эти задачи) отметим следующее. Определяемое нами далее отношение эквивалентности $A \equiv B$, а также приведённые в [2] необходимые и достаточные условия его выполнения (прежде всего – в префиксном случае),

© Корабельщикова С. Ю., Мельников Б. Ф., 2015

могут быть применены при решении разных задач, относящихся ко многим областям теории формальных языков. Опишем очень кратко только небольшую часть этих областей.

- На основе условий выполнения отношения $A \equiv B$ можно сформулировать некоторые необходимые и достаточные условия коммутирования в глобальном надмоноиде (супермоноиде) свободного моноида и некоторых его подмоноидах – [3], [4]. Иными словами – сформулировать некоторые критерии для выполнения равенства $A \cdot B = B \cdot A$, где $A, B \subseteq \Sigma^*$. В настоящей статье этим вопросам посвящены разделы 3. и 4.

- В некоторых подклассах класса контекстно-свободных (КС) языков разрешима проблема эквивалентности — в отличие от всего этого класса; см. [5], [6], [7]. Важно отметить, что мы при этом имеем в виду *не* «пресловутый» класс *детерминированных* контекстно-свободных языков, проблема эквивалентности для которого была сформулирована ещё в конце 1960-х, [1)] и впоследствии была решена. [2)]

  Итак, мы имеем в виду подклассы, не совпадающие с классом детерминированных КС-языков. Однако с помощью этих подклассов могут быть описаны некоторые реальные языки программирования – см. примеры в [7]. Для описания этих подклассов мы часто рассматриваем пары языков, удовлетворяющих определяемому далее отношению эквивалентности $A \equiv B$.

- Задачи, рассматриваемые в настоящей статье, возникают при *графическом* описании класса КС-языков и некоторых его подклассов – см. [9], [13], [14], [23], [24], [25] — в частности, при решении проблем эквивалентности в этих подклассах. [3)]

- Возможным *практическим* приложением этой теории является применение префиксных кодов при обработке файлов для их защиты при частичном повреждении. См. теоретические аспекты данной проблемы в [15], [16], некоторые вопросы, демонстрирующие связь этой проблемы с материалом настоящей статьи, – в [17], [18], а описание практического программного продукта – в [19].

- «Бесконечный случай» – т.е. когда мы в (1) допускаем возможность бесконечных языков A или B – по-видимому, менее интересен, чем «конечный». Однако всё-таки в некоторых задачах необходимость рассмотрения бесконечных языков A и B возникает – например, в случае когда рассматривается задача описания условий эквивалентности морфических образов т.н. скобочных языков. См. [6], [7], [18], [20], [21]; отметим, что в этих работах мы рассматривали и другие варианты применения морфизмов бесконечных языков при описании некоторых подклассов класса КС-языков.

2. ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ

Как уже отмечалось, бесконечные итерации конечных языков были впервые рассмотрены одним из авторов настоящей статьи в [2]. Иными словами, рассматривались ω-языки вида A^ω, где A – конечный язык над заданным алфавитом Σ. (Мы предполагаем, что $|\Sigma| \geqslant 2$, а также – если специально не сказано иного – что $A \not\ni \varepsilon$.) Для двух конечных языков A и B мы рассматривали их равенство $A^\omega = B^\omega$, которое равносильно выполнению специального

[1)] См., например, [8, разд. 4.2].

[2)] Частичные решения см. в [9], [10], [11], а полное – в [12].

[3)] Более того, развитие этого графического подхода даёт возможность описать – также графически – языки, имеющие в иерархии Хомского типы 1 и 0.

отношения эквивалентности, также определённого в [2]: $A^* \equiv_\infty B^*$ тогда и только тогда, когда

$$\begin{cases} (\forall\, u \in A^*)\ (\exists\, v \in B^*)\ (u \in \mathrm{Pref}\,(v)) \\ (\forall\, v \in B^*)\ (\exists\, u \in A^*)\ (v \in \mathrm{Pref}\,(u))\,. \end{cases} \qquad (1)$$

Здесь $\mathrm{Pref}\,(u)$ – множество всех префиксов слова u, включая ε и u. Ниже записью $\mathrm{Opref}\,(u)$ будем обозначать множество всех собственных префиксов слова u (также включая ε). Аналогично для языков:

$$\mathrm{Pref}\,(A) = \{\, u \,|\, u \in \mathrm{Pref}\,(v),\ v \in A \,\}\,,$$
$$\mathrm{Opref}\,(A) = \{\, u \,|\, u \in \mathrm{Opref}\,(v),\ v \in A \,\}\,.$$

Для языка $A \subseteq \Sigma^*$ обозначим

$$\mathrm{pv}\,(A) = \{\, u \in A \,|\, (\forall v \in A)\,(u \notin \mathrm{Opref}\,(v)) \,\}\,.$$

Для некоторых языка $A \subseteq \Sigma^*$ и слова $u \in \Sigma^*$ будем писать

$$A\,|^{\,u} = \{\, v \in \Sigma^* \,|\, uv \in A \,\}\,.$$

В [2] мы рассматривали равенство ω-языков и определённую с помощью (1) эквивалентность – в случае *произвольных* ω-языков и языков соответственно. А в настоящей работе мы будем рассматривать их *только для ω-языков типа A^ω и языков типа A^**. Вследствие этого ниже в данной статье мы будем писать $A \equiv B$ – вместо использовавшегося в [2] обозначения $A^* \equiv_\infty B^*$; итак, условие $A \equiv B$ выполняется тогда и только тогда, когда выполняются оба условия (1).

В [2] были получены необходимые и достаточные условия эквивалентности $A \equiv B$; стоит отметить, что существенно проще эти условия могут быть получены в т.н. префиксном случае, т.е. когда

$$(\forall\, u, v \in A,\ u \neq v)\ (u \notin \mathrm{Pref}\,(v))\,,$$

Отметим также, что в большинстве рассматриваемых нами задач случай *бесконечных* итерируемых языков менее интересен – это отражено в названии статьи [2]. [4)]

Итак, мы повторили основные обозначения, применявшиеся, прежде всего, в [2], [3], [4], [6] и др. Очень важно упомянуть также статью [21], одним из основных результатов которой является следующее утверждение (необходимое условие коммутирования в глобальном надмоноиде свободного моноида): если

$$A \cdot B = B \cdot A, \qquad (2)$$

то $A \equiv B$.

А основной доказанный в [2] факт ([2, Th. 1]) можно эквивалентно сформулировать следующим образом.

Критерий эквивалентности $A \equiv B$.

Для некоторой пары конечных языков A и B, таких что $A \equiv B$, существует конечный язык D (пусть $D = \{\, u_1, \ldots, u_n \,\}$), для которого выполнено следующее условие. Для некоторого нового алфавита $\Delta = \{\, c_1, \ldots, c_n \,\}$ существуют два языка $A', B' \subseteq \Delta^*$, такие что:

[4)] Можно провести аналогию между рассматриваемыми нами проблемами и т.н. примером М.-П. Шютценберже ([1, гл. 5, пр. 3.3.1]): над *конечным* алфавитом подобных примеров не существует.

Отметим ещё, что приведённая в этой же главе ([1, гл. 5]) теория, относящаяся к максимальным префиксным кодам, связана и с проблемами, рассматриваемыми далее в настоящей статье.

- оба языка (A' и B') в качестве подмножеств содержат максимальные префиксные коды над алфавитом Δ;[5)]
- для морфизма

$$h : \Delta^* \to \Sigma^*, \quad \text{где} \quad h(c_1) = u_1\,,\ \ldots,\ h(c_n) = u_n\,,$$

выполнены условия $h(A') = A$ и $h(B') = B$. $\square$

Обозначения, связанные с этим критерием (употреблявшиеся в [2] и др.) таковы. Всё множество максимальных префиксных кодов над алфавитом Δ будет обозначаться $\mathrm{mp}\,(\Delta)$. Для применяемых в формулировке критерия обозначений морфизма h и языков (D, A' и др.) множества, полученные из $\mathrm{mp}\,(\Delta)$ с помощью применения морфизма h, будем обозначать $\mathrm{mp}\,(D)$. Кроме того, будем писать $A', B' \in \mathrm{mp}^+\,(\Delta)$ и $A, B \in \mathrm{mp}^+\,(D)$. Заметим, что для любого $A \in \mathrm{mp}^+\,(\Delta)$ выполнено условие

$$\mathrm{pv}\,(A) \in \mathrm{mp}\,(\Delta)\,.$$

Итак, согласно [2], рассмотренное нами в [21] необходимое условие коммутирования (2) может быть заменено на сформулированный критерий эквивалентности $A \equiv B$ – также являющийся для (2) необходимым условием.

В двух следующих разделах мы рассмотрим простые частные случаи такого коммутирования. Специально отметим, что в некоторых из исследуемых нами далее задач достаточно рассматривать только случай префиксных языков – в котором доказательство необходимых и достаточных условий эквивалентности $A \equiv B$ существенно проще, чем в рассматривавшемся в [2] общем случае.

3. КОММУТИРОВАНИЕ В СЛУЧАЕ, КОГДА ОДИН ИЗ ЯЗЫКОВ СОСТОИТ ИЗ 1 СЛОВА

В этом и следующем разделах на множества A и B не накладывается требование префиксности, однако одно из множеств (всюду пусть A) должно содержать не более 2 элементов.[6)]

Будем считать, что к языкам A и B *уже был применён* специальный *инверсный морфизм*, который может быть построен на основе изложенного в разделе 2. . Таким образом, пусть

$$A, B \in \mathrm{mp}^+\,(\Sigma)\,.$$

При этом:

- употребление одного алфавита – вместо рассматривавшихся ранее двух – недоразумений не вызовет;
- а критерии равенства (2) для произвольных множеств, одно из которых содержит не более 2 элементов, могут быть получены из приведённых ниже путём применения к A и B произвольного морфизма, причём не обязательно инъективного.

Кроме этого, пусть всюду в данном разделе алфавит Σ будет равен либо $\{0\}$, либо $\{0, 1\}$, либо $\{0, 1, 2\}$ (в зависимости от числа $|\Sigma|$, которое не будет превышать 3).

Сначала рассмотрим случай $|A| = 1$. Если $A = \{\varepsilon\}$, то условие (2) выполняется для поизвольного B. Иначе из условий

$$|A| = 1 \quad \text{и} \quad A \in \mathrm{mp}^+\,(\Sigma)$$

[5)] Эти коды, вообще говоря, различны.

[6)] При этом, конечно, в некоторых частных случаях условие префиксности можно получить в качестве тривиального следствия.

получается следующее:

$$|\Sigma| = 1 \quad \text{и} \quad A = \{0^k\}$$

для некоторого положительного k. Кроме того, $B \subseteq \Sigma^*$, а, как несложно убедиться, любое подмножество B множества $\{0\}^*$ при любом k удовлетворяет условию

$$\{0^k\} \cdot B = B \cdot \{0^k\}.$$

Таким образом, приведённое в разделе 2. необходимое условие равенства (2)

$$A, B \in \mathrm{mp}^+(\Sigma)$$

в случае $|A| = 1$ является необходимым и достаточным.

4. КОММУТИРОВАНИЕ В СЛУЧАЕ, КОГДА ОДИН ИЗ ЯЗЫКОВ СОСТОИТ ИЗ 2 СЛОВ

Теперь пусть $|A| = 2$.

Несложно показать, что согласно приведённому выше критерию эквивалентности $A \equiv B$, а также определению mp^+, возможны только следующие два варианта (и только они):

- $|\Sigma| = 1$, $A = \{\, 0^k, 0^l \,\}$ при некоторых различных $k, l \geqslant 0$;
- $|\Sigma| = 2$, $A = \Sigma$.

Очевидно, что в первом случае условие

$$B \in \mathrm{mp}^+(\Sigma)$$

является, как и при $|A| = 1$, не только необходимым, но и достаточным; оно для однобуквенного алфавита Σ уже было рассмотрено. Далее будем исследовать второй вариант.

Обозначим

$$\widetilde{B} = \mathrm{pv}(B).$$

Несложно убедиться, что, поскольку $A = \{0, 1\}$, из (2) следует равенство

$$\{0, 1\} \cdot \widetilde{B} = \widetilde{B} \cdot \{0, 1\} \tag{1}$$

(ср. это равенство с примером, рассмотренным в разделе 6.).

Докажем, что

$$\widetilde{B} = \Sigma^n$$

для некоторого $n \in \mathbb{N}_0$. Это утверждение является следствием такого факта: если

$$k \leqslant \left\| \widetilde{B} \right\|_{\min},$$

то для каждого слова длины k ($u \in \Sigma^k$) числа $\left| \widetilde{B}|^u \right|$ равны между собой [7]; докажем сформулированный факт индукцией по k.

База индукции (случай $k = 0$) очевидна, поскольку при этом есть только одно число

$$\left| \widetilde{B}|^{\varepsilon} \right| = \left| \widetilde{B} \right|;$$

[7] Т.е. число слов языка $\{\, w \mid uw \in \widetilde{B} \,\}$ не зависит от выбранного слова $u \in \Sigma^k$.

докажем шаг индукции. Согласно сделанным обозначениям,

$$\widetilde{B} = \bigcup_{u \in \Sigma^k} u\widetilde{B}|^u. \tag{2}$$

Пусть $k+1 \leqslant \left\|\widetilde{B}\right\|_{\min}$; предположение индукции –

$$\left(\forall u \in \Sigma^k\right) \left(|B|^u| = \mathrm{const}\right).$$

Значит, вследствие (1),

$$\left(\forall u \in \Sigma^k,\ a \in \{0,1\}\right) \left(\left|\left(\{0,1\}\widetilde{B}\right)|^{au}\right| = \mathrm{const}\right),$$

что даёт

$$\left(\forall v \in \Sigma^{k+1}\right) \left(\left|\left(\{0,1\}\widetilde{B}\right)|^{v}\right| = \mathrm{const}\right),$$

а из последнего, согласно (1) и (2), получаем

$$\left(\forall v \in \Sigma^{k+1}\right) \left(\left|\widetilde{B}|^{v}\right| = \mathrm{const}\right);$$

тем самым доказан шаг индукции.

Итак,

$$\mathrm{pv}\,(B) = \Sigma^{\|B\|_{\min}}.$$

Обозначив

$$B_1 = B \setminus \mathrm{pv}\,(B),$$

можно таким же способом доказать, что

$$\mathrm{pv}\,(B_1) = \Sigma^{\|B_1\|_{\min}},$$

и т.д. Поэтому выполняется следующая

Теорема 1. *Если $A = \Sigma = \{0,1\}$ и $A \cdot B = B \cdot A$, то*

$$B = \bigcup_{i \in I} \Sigma^i$$

для некоторого $I \subseteq \mathbb{N}_0$. □

Заметим, что мы не пользовались какими-либо ограничениями на $|B|$, поэтому в формулировке утверждения 1 множество индексов $I \subseteq \mathbb{N}_0$ может быть и бесконечным.

Случаи, когда A содержит больше двух элементов, даже, к примеру, $|A| = 3$, сложны для детального исследования. Действительно, после применения упомянутого инверсного морфизма только для одного множества A имеется бесконечно много вариантов, даже если дополнительно потребовать выполнение условия $A \not\ni \varepsilon$, а именно:

- $|\Sigma| = 1$, множество A произвольное (только этот случай исследуется легко);
- $|\Sigma| = 2$, $A = \{\,0, 10, 11\,\}$;
- $|\Sigma| = 2$, $A = \{\,0, 1, u\,\}$ для произвольного слова u такого, что $|u| \geqslant 2$ (здесь имеется бесконечно много «подслучаев»);
- $|\Sigma| = 3$, $A = \Sigma$.

5. НЕКОТОРЫЕ ВОПРОСЫ ПОДСЧЁТА ЧИСЛА МАКСИМАЛЬНЫХ ПРЕФИКСНЫХ КОДОВ

Для некоторых алфавитов

$$\Delta = \{\, b_1, b_2, \ldots, b_n \,\} \qquad (1)$$

и Σ рассмотрим морфизм

$$h : \Delta^* \to \Sigma^*, \quad \text{где} \quad h(b_1) = u_1\,,\ h(b_2) = u_2\,,\ \ldots,\ h(b_n) = u_n\,,$$

(в терминах теории кодирования – [22] и др. – это отображение является алфавитным кодированием, задаваемым схемой h).

Код называется *однозначно декодируемым*, если любое слово в кодирующем алфавите либо не является кодовым, либо является кодом ровно одного сообщения – см. [1, стр. 134] или [22, стр. 260]. В терминах теории формальных языков то же самое формулируется следующим образом: инверсный морфизм h^{-1} является инъективной функцией. При т.н. помехоустойчивом кодировании алфавит сообщений совпадает с кодирующим, и выбирается равным множеству элементов некоторого конечного поля, рассматриваемого нами, например, в [17]. При этом однозначная декодируемость кода зависит от множества

$$C = \{u_1, u_2, \ldots, u_n\} = h(\Delta) \qquad (2)$$

элементарных кодов, и сама функция h при этом не принципиальна. [8] Очевидно, что свойство префикса гарантирует однозначную декодируемость кода и эффективный алгоритм декодирования.

Префиксному коду (2) можно сопоставить корневое дерево, в котором каждой цепи от корня до концевой вершины соответствует элементарный код, и наоборот. Пусть, например, кодирующий алфавит $\Delta = \{b_1, b_2, b_3, b_4\}$, а множество элементарных кодов

$$C = \{\, b_1b_3, b_3, b_1b_1, b_1b_2, b_4b_2b_3, b_1b_4, b_4b_1, b_4b_2b_4 \,\}\,.$$

Здесь свойство префикса выполняется, и коду соответствует кодовое дерево, изображённое на рис. 1.

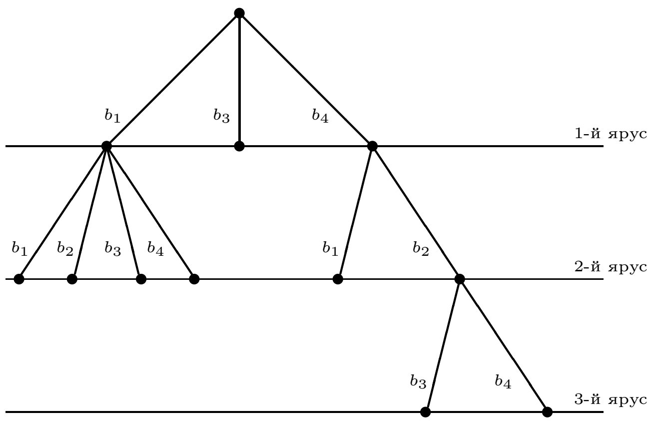

Рис. 1

При этом, согласно [1, стр. 147], префиксный код является полным, если порядок ветвления всех неконцевых вершин кодового дерева равен мощности кодирующего алфавита Δ.

[8] В (2), как и в некоторых местах далее, мы рассматриваем Δ не как алфавит (множество букв), а как множество *однобуквенных слов.*

Несложно доказывается, что в случае конечных алфавитов (который мы рассматриваем) понятия «полный код» и «максимальный код» всегда совпадают.

Произведём подсчёт количества различных максимальных префиксных кодов с ограничением t на длину слов над кодирующим алфавитом мощности q. Эта задача эквивалентна задаче подсчёта всех корневых деревьев с порядком ветвления не концевых вершин q и имеющим не более t ярусов.

Первый случай: $t = 1$. Не более 1 яруса (то есть ровно 1 ярус) имеет 1 кодовое дерево, соответствующее множеству всех однобуквенных элементарных кодов (2).

Второй случай: $t = 2$. Не более чем 2 яруса имеют 2^q деревьев. Из q концевых вершин первого яруса выбираем произвольное подмножество вершин, от которых пойдут пучки из q ребер во 2-й ярус. Количество подмножеств q-элементного множества равно 2^q, включая пустое подмножество, соответствующее 1-ярусному дереву.

Третий случай: $t = 3$. Сначала подсчитаем деревья, имеющие *ровно* 3 яруса; посчитаем q вспомогательных сумм – по количеству вершин для ветвления в первом ярусе. [9)]

$$C(q,1)\cdot C(q,1) + C(q,1)\cdot C(q,2) + \cdots + C(q,1)\cdot C(q,q)\ + \tag{3}$$
$$+\ C(q,2)\cdot C(2q,1) + C(q,2)\cdot C(2q,2) + \cdots + C(q,2)\cdot C(2q,2q)\ + \tag{4}$$
$$+\ \ldots\ +$$
$$+\ C(q,q)\cdot C(q^2,1) + C(q,q)\cdot C(q^2,2) + \cdots + C(q,q)\cdot C(q^2,q^2)\,. \tag{5}$$

Здесь:

- (3) соответствует ситуациям, в которых в первом ярусе выбрана 1 вершина для ветвления, а во втором – 1, 2, ..., q из q имеющихся;
- (4) соответствует ситуациям, в которых в первом ярусе выбраны 2 вершины для ветвления, а во втором – 1, 2, ..., $2q$ из $2q$ имеющихся;
- (5) соответствует ситуациям, в которых в первом ярусе выбраны q вершин для ветвления, а во втором – 1, 2, ..., q^2 из q^2 имеющихся.

Таким образом, общее число деревьев, имеющих *не более* чем 3 яруса – 2^q плюс сумма (3), (4) и (5). Заменяя 2^q на сумму $C(q,0) + C(q,1) + \cdots + C(q,q)$, получаем

$$\begin{gathered}
C(q,0) + C(q,1) + \cdots + C(q,q)\ + \\
+\ C(q,1)\cdot C(q,1) + C(q,1)\cdot C(q,2) + \cdots + C(q,1)\cdot C(q,q)\ + \\
+\ C(q,2)\cdot C(2q,1) + C(q,2)\cdot C(2q,2) + \cdots + C(q,2)\cdot C(2q,2q)\ + \\
+\ \ldots\ + \\
+\ C(q,q)\cdot C(q^2,1) + C(q,q)\cdot C(q^2,2) + \cdots + C(q,q)\cdot C(q^2,q^2)\,.
\end{gathered}$$

Далее, вынеся за скобки общие множители, перепишем последнее в виде

$$\begin{gathered}
C(q,0) + C(q,1)\cdot(1 + C(q,1) + C(q,2) + \cdots + C(q,q))\ + \\
+\ C(q,2)\cdot(1 + C(2q,1) + C(2q,2) + \cdots + C(2q,2q))\ + \\
+\ \ldots\ + \\
+C(q,q)\cdot(1 + C(q^2,1) + C(q^2,2) + \cdots + C(q^2,q^2))\,;
\end{gathered}$$

после замены суммы в скобках на степени двойки получаем

$$C(q,0) + C(q,1)\cdot 2^q + C(q,2)\cdot 2^{2q} + \cdots + C(q,q)\cdot 2^{q\cdot q}.$$

9) $C(n,m)$ – число сочетаний из n по m.

Итого мы получаем такое число максимальных префиксных кодов с длиной слов не более 3 в q-буквенном кодирующем алфавите:

$$\sum_{i=0}^{q} C(q,i)\cdot 2^{i\cdot q}.$$

Аналогичными рассуждениями мы получаем число максимальных префиксных кодов с длиной слов не более 4 в q-буквенном кодирующем алфавите:

$$1+C(q,1)\cdot\sum_{i=0}^{q} C(q,i)\cdot 2^{i\cdot q}+C(q,2)\cdot\sum_{i=0}^{2q} C(2q,i)\cdot 2^{i\cdot q}+$$
$$+\cdots+C(q,q)\cdot\sum_{i=0}^{q^2} C(q^2,i)\cdot 2^{i\cdot q}.$$

Продолжая аналогичным образом, мы по индукции доказываем следующую теорему.

Теорема 2. *Число максимальных префиксных кодов с длиной слов не более t ($t \geqslant 3$) в q-буквенном кодирующем алфавите задаётся формулой*

$$\sum_{i_1=0}^{q} C(q,i_1)\cdot\Big(\sum_{i_2=0}^{i_1\cdot q} C(i_1\cdot q,i_2)\cdot\Big(\sum_{i_3=0}^{i_2\cdot q} C(i_2\cdot q,i_3)\cdot$$
$$\cdot\ \ldots\ \cdot\Big(\sum_{i_{t-2}=0}^{i_{t-3}\cdot q} C(i_{t-3}\cdot q,i_{t-2})\cdot 2^{i_{t-2}\cdot q}\Big)\ldots\Big)\Big). \quad \square$$

(Параметры $i_1, i_2, \ldots i_{t-2}$ соответствуют количеству ветвящихся вершин в 1-м, 2-м, …, $t-2$-м ярусах соответственно.)

6. ЗАКЛЮЧЕНИЕ. ПРИМЕРЫ И НЕКОТОРЫЕ ГИПОТЕЗЫ

В данном разделе рассматриваются некоторые интересные примеры и задачи для дальнейшего решения.

Во-первых, рассмотрим некоторые примеры. В префиксном случае пары языков A и B, для которых выполнено условие (2), могут быть построены на основе сформулированного выше критерия эквивалентности $A \equiv B$. Кроме того очевидно, что также условие (2) выполняется для произвольных языков A и B, каждый из которых может быть записан в виде

$$\bigcup_{i\in I}\Sigma^i$$

для некоторого множества индексов $I \subseteq \mathbb{N}_0$.

Приведём менее очевидный пример – пример *непрефиксных* языков A и B над алфавитом $\Sigma=\{0,1\}$:

$$A=\Sigma\cup\{\,00,01,11\,\}\cup\Sigma^3, \quad B=\Sigma\cup\Sigma^2\,; \tag{1}$$

тогда

$$A\cdot B=B\cdot A=\bigcup_{2\leqslant i\leqslant 5}\Sigma^i\,.$$

Все примеры (эти и любые другие) могут быть изменены с помощью применения к языкам A и B одного и того же морфизма, причём не обязательно инъективного.

Далее сформулированы две гипотезы, для которых у авторов в настоящее время нет ни доказательств, ни контрпримеров.

Первая гипотеза.

Если (2), то выполнено равенство

$$\mathrm{pv}\,(A) \cdot \mathrm{pv}\,(B) = \mathrm{pv}\,(B) \cdot \mathrm{pv}\,(A)\,. \quad \Box \tag{2}$$

При этом отметим, что для произвольного языка

$$A \in \mathrm{mp}^{+}\,(\Sigma)$$

по определению множества mp выполнено условие

$$\mathrm{pv}\,(A) \in \mathrm{mp}\,(\Sigma)\,.$$

Поэтому если первая гипотеза верна (т.е. если равенство (2) действительно выполняется для любых коммутирующих языков A и B), то, согласно сформулированному в разделе 2. критерию эквивалентности $A \equiv B$,

$$\mathrm{pv}\,(A) = C^{k} \quad \text{и} \quad \mathrm{pv}\,(B) = C^{l}$$

для некоторых $C \subseteq \Sigma^{*}$ (префиксного множества) и чисел $k, l \in \mathbb{N}_0$.

На основе этих равенств формулируется

Вторая гипотеза.

При сделанных обозначениях можно выбрать множество C таким образом, что $A, B \subseteq C^{*}$. $\Box$

(Заметим, что, например, пары бесконечных языков

$$A = \Sigma^{2}\Sigma^{*} \quad \text{и} \quad B = \Sigma^{4}\Sigma^{*}$$

для *произвольного* алфавита Σ контрпримерами ко второй гипотезе *не* являются, поскольку мы можем выбрать не только $C = \Sigma^{2}$, но и $C = \Sigma$.)

Для анализа обеих сформулированных гипотез может оказаться полезным следующий факт: вообще говоря,

$$\mathrm{pv}\,(A) \cdot \mathrm{pv}\,(B) \neq \mathrm{pv}\,(A \cdot B)\,. \tag{3}$$

Для подтверждения (3) рассмотрим такой пример:

$$\Sigma = \{0,1\}, \; A = B = \{\,0, 01, 100, 101, 110, 111\,\}\,;$$

заметим, что здесь

$$A,\ B,\ AB \in \mathrm{mp}^{+}\,(\Sigma).$$

Неравенство (3) при этом можно проверить непосредственно. [10]

---

[10] А можно показать и почти без «вычислений», используя следующие соображения:

$$\mathrm{pv}\,(A) = \mathrm{pv}\,(B) = \{\,0, 100, 101, 110, 111\,\}\ ,$$

значит,

$$\mathrm{pv}\,(A) \cdot \mathrm{pv}\,(B) \not\ni 010\,,$$

в то время как $\mathrm{pv}\,(AB) \ni 010$.

Отметим ещё, что с задачей описания критериев равенства $A \cdot B = B \cdot A$, а также со многими из задач, рассматривавшихся в предыдущих разделах и в процитированных нами предыдущих статьях (прежде всего – [4], [5]), непосредственно связана задача «извлечения корня» из заданного языка: для заданного языка $A \subseteq \Sigma^*$ требуется найти максимально возможное $n \in \mathbb{N}$ и зависящий от n язык $B \subseteq \Sigma^*$, такие что $A = B^n$. До конца эта задача авторами ещё не исследована, поэтому приведём лишь один интересный пример, имеющий некоторую аналогию с примером (1): для произвольного алфавита Σ (в том числе – при $|\Sigma| = 1$ и $|\Sigma| = \omega$) из языка

$$\bigcup_{2 \leqslant i \leqslant 10} \Sigma^i$$

извлекается не только «очевидный» квадратный корень

$$\bigcup_{1 \leqslant i \leqslant 5} \Sigma^i ,$$

но и

$$\bigcup_{i \in \{ 1,2,4,5 \}} \Sigma^i .$$

Таким образом, в случае языков операция извлечения корня заданной степени не является (однозначной) функцией.

Сформулируем также в общем виде одну проблему, постановка и частичное решение которой для двоичного алфавита (т.е. $|\Sigma| = 2$) и значения $d = 3$ (см. определение ниже) приведено в [15, стр. 262].

В помехоустойчивом кодировании обычно используется равномерный код, то есть длины всех слов одинаковы. Расстоянием Хэмминга между словами одинаковой длины называется количество позиций, в которых различаются эти слова. Минимум из попарных расстояний различных кодовых слов называют минимальным расстоянием кода и обозначают $d(C)$ или просто d. (Пример: $C = \{010, 110, 101\}$; при этом $d = 1$, т.к. расстояние между словами 010 и 110 равно 1, между словами 010 и 101 равно 3, между словами 110 и 101 равно 2, а $\min\{1, 3, 2\} = 1$.) От этого значения зависит корректирующая способность кода: код с минимальным расстоянием d обнаруживает $d{-}1$ и исправляет $[(d{-}1)/2]$ ошибок.

Возникает такая задача: оценить максимальные размеры равномерных кодов (двоичных, троичных и т.д.) фиксированной длины n, имеющих минимальное расстояние d – т.е. любые два слова должны находиться на расстоянии d или больше. Отметим, что [15, стр. 262] приведено частное решение – для двоичного алфавита приведены максимальные размеры кодов, исправляющих 1 ошибку (т.е. $d = 3$). Результаты, приведённые в [15], таковы: для $n = 3$ ответ 2; для $n = 4$ ответ 2; для $n = 5$ ответ 4; для $n = 6$ ответ 8; для $n = 7$ ответ 16; для $n = 8$ ответ 20. Итак, на основе материала данной статьи, возможно, удастся отыскать решение для бо́льших значений $|\Sigma|$ и/или d.

СПИСОК ЛИТЕРАТУРЫ

[1] Лаллеман Ж. Полугруппы и комбинаторные приложения / Ж. Лаллеман. — М.: Мир, 1985. — 440 с.

[2] Melnikov B. The equality condition for infinite catenations of two sets of finite words / B. Melnikov // Int. J. of Found. of Comp. Sci. — 1993. — Vol. 4, № 3. — P. 267–274.

[3] Melnikov B. Some equivalence problems for free monoids and for subclasses of the CF-grammars class / B. Melnikov // World Sci. Publ. — 1995. — pp. 67–68.

[4] Мельников Б.Ф. Описание специальных подмоноидов глобального надмоноида свободного моноида / Б.Ф. Мельников // Известия высших учебных заведений. Математика. — 2004. — №. 3. — С. 46–56.

[5] Мельников Б.Ф. Некоторые следствия условия эквивалентности однозначных скобочных грамматик / Б.Ф. Мельников // Вестник Московского университета. Серия 15: Вычислительная математика и кибернетика. — 1991. — № 3. — С. 51–53.

[6] Дубасова О. Об одном расширении класса контекстно-свободных языков / О. Дубасова, Б. Мельников // Программирование. — 1995. — №. 6. — С. 46–58.

[7] Melnikov B. Some grammatical structures of programming languages as simple bracketed languages / B. Melnikov, E. Kashlakova // Informatica (Lithuanian Acad. Sci. Ed.). — 2000. — Vol. 11, №. 4. — P. 441–454.

[8] Гинзбург С. Математическая теория контекстно-свободных языков / С. Гинзбург. — М.: Мир, 1970. — 326 с.

[9] Станевичене Л. Об одном средстве исследования бесконтекстных языков / Л. Станевичене // Кибернетика. — 1989. — № 4. — С. 135–136.

[10] Stanevicienė L. D-graphs in context-free language theory / L. Stanevicienė // Informatica (Lithuanian Acad. Sci. Ed.). — 1997. — Vol. 8, №. 1. — С. 43–56.

[11] Мейтус В. Разрешимость проблемы эквивалентности детерминированных магазинных автоматов / В. Мейтус // Кибернетика и системный анализ. — 1992. — №. 5. С. 20–45.

[12] Sénizergues G. $L(A) = L(B)$? decidability results from complete formal systems / G. Sénizergues // Theor. Comput. Sci. — 2001. — Vol. 251, №. 1–2. P. 1–166.

[13] Вылиток А. О построении графа магазинного автомата / А. Вылиток // Вестник Московского университета. Серия 15: Вычислительная математика и кибернетика. — 1996. — № 3. — С. 68–73.

[14] Вылиток А. Об одном расширении класса конечных автоматов для задания контекстно-свободных языков / А. Вылиток, М. Зубова, Б. Мельников // Вестник Московского университета. Серия 15: Вычислительная математика и кибернетика. — 2013. — № 1. — С. 39–45.

[15] Лидл Р. Прикладная абстрактная алгебра / Р. Лидл, Г. Пильц. — Екатеринбург: Уральский университет, 1996. — 226 с.

[16] Вернер М. Основы кодирования / М. Вернер. — М.: Техносфера, 2006. — 288 с.

[17] Корабельщикова С. Аппроксимация полугруппы характеров гомоморфизмами в мультипликативную полугруппу конечного поля / С. Корабельщикова, И. Игнатьева // Вестник Северного (Арктического) федерального университета. Серия: Естественные науки. — 2011. — №. 2. — С. 107–110.

[18] Зяблицева Л. Некоторые специальные полугруппы и их гомоморфизмы / Л. Зяблицева, И. Попов, С. Корабельщикова. — Архангельск: ИПЦ САФУ им. М.В. Ломоносова, 2013. — 128 с.

[19] Корабельщикова С. Система обработки файлов в целях защиты при их частичном повреждении / С. Корабельщикова, А. Лудков // Свидетельство о государственной регистрации программ для ЭВМ. № 2010610352.

[20] Мельников Б. Подклассы класса контекстно-свободных языков / Б. Мельников. — М.: Изд-во МГУ, 1995. — 174 с.

[21] Алексеева А., Мельников Б. Итерации конечных и бесконечных языков и недетерминированные конечные автоматы / А. Алексеева, Б. Мельников // Вектор науки Тольяттинского государственного университета. — 2011. — № 3. — С. 30–33.

[22] Яблонский С. Введение в дискретную математику / С. Яблонский. — М.: Высш. шк., 2003. — 384 с.

[23] Зубова М.А. Об одном алгоритме построения универсального автомата Конвея / М.А. Зубова, Б.Ф. Мельников // Вестн. Воронеж. гос. ун-та. Сер. Физика, математика. — 2012. — № 1. — С. 135–137.

[24] Долгов В.Н. Построение универсального конечного автомата. I. От теории к прак-

тическим алгоритмам / В.Н. Долгов, Б.Ф. Мельников // Вестн. Воронеж. гос. ун-та. Сер. Физика, математика. — 2013. — № 2. — С. 173–181.

[25] Долгов В.Н. Построение универсального конечного автомата. II. Примеры работы алгоритмов / В.Н. Долгов, Б.Ф. Мельников // Вестн. Воронеж. гос. ун-та. Сер. Физика, математика. — 2014. — № 1. — С. 78–85.

REFERENCES

[1] Lallement G. Semigroups and combinatorial applications. [Lalleman Zh. Polugruppy i kombinatornye prilozheniya]. Moscow: Mir, 1985, 440 p.

[2] Melnikov B. The equality condition for infinite catenations of two sets of finite words. Int. J. of Found. of Comp. Sci., 1993, Vol. 4, № 3, pp. 267–274.

[3] Melnikov B. Some equivalence problems for free monoids and for subclasses of the CF-grammars class. World Sci. Publ., 1995, pp. 67–68.

[4] Melnikov B.F. Description of special submonoids global supermonoid free monoid. [Mel'nikov B.F. Opisanie special'nyx podmonoidov global'nogo nadmonoida svobodnogo monoida]. *Izvestiya vysshix uchebnyx zavedenij. Matematika — Russian Mathematics*, 2004, no. 3, pp. 46–56.

[5] Melnikov B.F. Some consequences of the conditions for the equivalence of unambiguous grammars bracketed. [Mel'nikov B.F. Nekotorye sledstviya usloviya e'kvivalentnosti odnoznachnyx skobochnyx grammatik]. *Vestnik Moskovskogo universiteta. Seriya 15: Vychislitel'naya matematika i kibernetika — Moscow University Computational Mathematics and Cybernetics*, 1991, no. 3, pp. 51–53.

[6] Dubasova O., Melnikov B. On an extension of the class of context-free languages. [Dubasova O., Mel'nikov B. Ob odnom rasshirenii klassa kontekstno-svobodnyx yazykov]. *Programmirovanie — Programming and Computer Software*, 1995, no. 6, pp. 46–58.

[7] Melnikov B., Kashlakova E. Some grammatical structures of programming languages as simple bracketed languages. Informatica (Lithuanian Acad. Sci. Ed.), 2000, Vol. 11, no. 4, pp. 441–454.

[8] Ginsburg S. The Mathematical Theory of Context Free Languages. [Ginzburg S. Matematicheskaya teoriya kontekstno-svobodnyx yazykov]. Moscow: Mir, 1970, 326 p.

[9] Stanevicien'e L. On a research facility without contextual languages. [Stanevichene L. Ob odnom sredstve issledovaniya beskontekstnyx yazykov]. *Kibernetika — Cybernetics*, 1989, no. 4, pp. 135–136.

[10] Staneviciené L. D-graphs in context-free language theory. Informatica (Lithuanian Acad. Sci. Ed.), 1997, Vol. 8, №. 1, pp. 43–56.

[11] Meitus V. Solvability of the equivalence problem for deterministic pushdown automaton. [Mejtus V. Razreshimost' problemy e'kvivalentnosti determinirovannyx magazinnyx avtomatov]. *Kibernetika i sistemnyj analiz — Cybernetics and Systems Analysis*, 1992, no. 5, pp. 20–45.

[12] Sénizergues G. $L(A) = L(B)$? decidability results from complete formal systems. Theor. Comput. Sci., 2001, Vol. 251, no. 1–2, pp. 1–166.

[13] Vylitok A. On the construction of the graph pushdown automaton. [Vylitok A. O postroenii grafa magazinnogo avtomata]. *Vestnik Moskovskogo universiteta. Seriya 15: Vychislitel'naya matematika i kibernetika — Moscow University Computational Mathematics and Cybernetics*, 1996, no. 3, pp. 68–73.

[14] Vylitok A., Zubova M., Melnikov B. On an extension of the class of finite automata to specify context-free languages. [Vylitok A., Zubova M., Mel'nikov B. Ob odnom rasshirenii klassa konechnyx avtomatov dlya zadaniya kontekstno-svobodnyx yazykov]. *Vestnik Moskovskogo universiteta. Seriya 15: Vychislitel'naya matematika i kibernetika — Moscow University Computational Mathematics and Cybernetics*, 2013, no. 1, pp. 39–45.

[15] Lidl R., Pilz G. Applied Abstract Algebra. [Lidl R., Pil'c G. Prikladnaya abstraktnaya algebra]. Ekaterinburg: Ural University, 1996, 226 p.

[16] Verner M. Basics of coding. [Verner M. Osnovy kodirovaniya]. Moscow: Technosphere, 2006, 288 p.

[17] Korabelshchikova S., Ignatiyeva I. Approximation of semi-group of characters with homomorphisms into the multiplicative semi-gpoup of a finite field. [Korabel'shhikova S., Ignat'eva I. Approksimaciya polugruppy xarakterov gomomorfizmami v mul'tiplikativnuyu polugruppu konechnogo polya]. *Vestnik Severnogo (Arkticheskogo) federal'nogo universiteta. Seriya: Estestvennye nauki — Bulletin of the North (Arctic) Federal University. Series: Natural Sciences*, 2011, no. 2, pp. 107–110.

[18] Zyablitseva L., Korabelshchikova S., Popov I. Some special semigroups and their homomorphisms. [Zyabliceva L., Popov I., Korabel'shhikova S. Nekotorye special'nye polugruppy i ix gomomorfizmy]. Arkhangelsk: Northern (Arctic) Federal University, 2013, 122 p.

[19] Korabelshchikova S., Ludkov A. Processing system files to protect them with partial damage. [Korabel'shhikova S., Ludkov A. Sistema obrabotki fajlov v celyax zashhity pri ix chastichnom povrezhdenii. Svidetel'stvo o gosudarstvennoj registracii programm dlya E'VM]. Certificate of state registration of computer programs, No. 2010610352.

[20] Melnikov B. Subclasses of the class context-free languages. [Mel'nikov B. Podklassy klassa kontekstno-svobodnyx yazykov]. Moscow: Moscow State University, 1995, 174 p.

[21] Alekseyeva A., Melnikov B. Iteration of finite and infinite languages and nondeterministic finite automata. [Alekseeva A., Mel'nikov B. Iteracii konechnyx i beskonechnyx yazykov i nedeterminirovannye konechnye avtomaty]. *Vektor nauki Tol'yattinskogo gosudarstvennogo universiteta — Vector science of the Togliatti State University*, 2011, no. 3, pp. 30–33.

[22] Yablonskiy S. Introduction in discrete mathematics. [Yablonskij S. Vvedenie v diskretnuyu matematiku]. Moscow: Higher School, 2002, 384 p.

[23] Zubova M.A., Melnikov B.F. On Algorithm of Constructing Conway's Universal Automaton. [Zubova M.A., Mel'nikov B.F. Ob odnom algoritme postroeniya universal'nogo avtomata Konveya]. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Fizika. Matematika — Proceedings of Voronezh State University. Series: Physics. Mathematics*, 2012, no. 1, pp. 135–137.

[24] Dolgov V.N., Melnikov B.F. Construction of Universal Finite Automaton. I. From Theory to the Practical Algorithm. [Dolgov V.N., Mel'nikov B.F. Postroenie universal'nogo konechnogo avtomata. I. Ot teorii k prakticheskim algoritmam]. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Fizika. Matematika — Proceedings of Voronezh State University. Series: Physics. Mathematics*, 2013, no. 2, pp. 173–181.

[25] Dolgov V.N., Melnikov B.F. Construction of Universal Finite Automaton. II Examples of Algorithms Functioning. [Dolgov V.N., Mel'nikov B.F. Postroenie universal'nogo konechnogo avtomata. II. Primery raboty algoritmov]. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Fizika. Matematika — Proceedings of Voronezh State University. Series: Physics. Mathematics*, 2014, no. 1, pp. 78–85.

Корабельщикова Светлана Юрьевна, доцент кафедры математического анализа, алгебры и геометрии института математики, информационных и космических технологий Северного (Арктического) федерального университета имени М. В. Ломоносова, кандидат физико-математических наук, доцент, Архангельск, Российская Федерация
E-mail: kmv@atnet.ru

Korabelshchikova Svetlana Yurievna, Associate Professor of the Department of Mathematical Analysis, Algebra and Geometry, Institute of Mathematics, Information and Space Technologies, Northern (Arctic) Federal University named after M. V. Lomonosov, Arkhangelsk, Russian Federation
E-mail: kmv@atnet.ru

Мельников Борис Феликсович, заведующий кафедрой прикладной математики и информатики Тольяттинского филиала Самарского государственного университета, доктор физико-математических наук, профессор, Тольятти, Российская Федерация
E-mail: bormel@rambler.ru

Melnikov Boris, Head of Department of Applied Mathematics and Informatics, Togliatti Branch of Samara State University, professor, Togliatti, Russian Federation
E-mail: bormel@rambler.ru