

# Множество максимальных префиксных кодов и соответствующая ему полугруппа

А. А. Мельникова

**Аннотация**—Одним из самых многочисленных классов алгебры являются полугруппы – по-видимому, в связи с тем, что единственная аксиома, которая должна выполняться для полугруппы, – это аксиома ассоциативности заданной на ней операции. В данной работе показано, что множество максимальных префиксных кодов является множеством образующих соответствующей им полугруппы относительно заданной на ней операции конкатенации. Важным вопросом остается представление о структуре полугруппы, о чем также говорится в статье.

В статье рассмотрены вопросы морфизмов максимальных префиксных кодов. На основе представления максимальных префиксных кодов как последовательности слов в алфавитном порядке (если алфавит упорядочен), формулируется предложение о возможности поставить в соответствие каждому такому коду последовательность длин его слов. Обсуждаются некоторые условия, при которых последовательности длин будет соответствовать максимальный префиксный код.

Рассматривается вариант сжатия таких кодов с использованием длин слов, описывается алгоритм. Как известно, кодирование, сжатие информации имеет большое значение на практике. Приводится пример выполнения алгоритма и рисунок к нему. Также говорится о возможных направлениях продолжения развития тематики статьи, возможностях исследования максимальных префиксных кодов, их свойств, структуры, инвариантов, возможностей сжатия и однозначности декодирования, восстановления исходного кода.

Приведены примеры выполнения определений и свойств для максимальных префиксных кодов.

**Ключевые слова**—регулярные языки, коды, максимальные префиксные коды, полугруппы, морфизмы, инверсные морфизмы.

## I. ВВЕДЕНИЕ

Одним из важных вопросов теории полугрупп является структура этих формализмов. Понятие полугруппы и соответствующий термин возникли в начале XX века, в конце 1920-х годов начались систематические исследования полугрупп (см., например, [1]). В 1960-е годы теория полугрупп стала активно развивающейся областью алгебры, использующей различные методы, подходы и решающей разнообразные задачи, тесно связанной со многими областями математики, как с собственно алгебраическими, в первую очередь, с теорией групп

и теорией колец, так и другими, например, алгебраической теорией автоматов (полугруппы автоматов), регулярных языков.

В те же годы появились и первые монографии, целиком посвященные алгебраической теории полугрупп. При работе с данными, в частности – со сжатием информации при её передаче, большое практическое значение имеет теория кодирования, также рассматриваемая в настоящей статье. Как указывается в [2], под кодированием в широком смысле понимается переход от одного способа задания информации к другому, допускающий восстановление исходной информации. Такой метод широко распространен в математике и ее приложениях. Возникновение теории кодирования как самостоятельного раздела математики в 40-х годах XX века, изучающего кодирование в различных его аспектах, было с появлением работ К. Шеннона. В математической теории кодирования изучаются методы кодирования, связанные с основными математическими моделями, которые отражают существенные, основные черты реальных информационных систем. В данной работе теория полугрупп применяется к конкретным разделам теории кодирования, для множества максимальных префиксных кодов (МПК), которые, в отличие от кодов фиксированной длины, более эффективно использовать в силу того, что они занимают меньше памяти для хранения и передачи информации.

Данную статью можно считать продолжением тематики статьи [3], в которой рассматривались некоторые алгоритмы работы с языками, используя расширенные максимальные префиксные коды и морфизмы над этими кодами.

В первой части статьи (раздел II) рассмотрены предварительные сведения, на основе которых в разделе III делается вывод о том, что множество максимальных префиксных кодов является множеством образующих соответствующей им полугруппы. Далее приведены примеры, поясняющие и иллюстрирующие рассмотренные вопросы. На основе того, что максимальные префиксные коды можно представить как последовательности слов в алфавитном порядке (если алфавит упорядочен), формулируется предложение о возможности каждому МПК поставить в соответствие последовательность длин его слов. Обсуждаются некоторые условия, при которых последовательности длин (натуральных чисел) будет соответствовать МПК. В IV разделе рассматривается вариант сжатия максимальных префиксных кодов с использованием длин слов, описывается алгоритм такого сжатия. Как уже отмечалось ранее, кодирование, сжатие информации имеет большое значение на практике. Приводится пример выполнения алго-

Статья получена 10 июля 2023.

Александра Александровна Мельникова, Димитровградский инженерно-технологический институт – филиал Национального исследовательского ядерного университета «МИФИ» (e-mail: super-avahi@yandex.ru).

ритма и рисунок к нему. В последнем разделе статьи говорится о возможных направлениях продолжения развития тематики статьи, возможностях исследования максимальных префиксных кодов, их свойств, структуры, инвариантов, возможностей сжатия.

## II. ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ

В данной работе будут использованы следующие определения, взятые из [3,4].

**Определение.** Непустой язык  $C$  в алфавите  $\Sigma$  называется кодом, если для любых слов  $x_{i1}, x_{i2}, \dots, x_{im}$  и  $x_{j1}, \dots, x_{jn}$  из  $C$ , таких, что

$$x_{i1} x_{i2} \dots x_{im} = x_{j1} \dots x_{jn}, \quad (1)$$

имеет место  $x_{i1} = x_{j1}$ .

Из данного определения следует, что если  $C$  является кодом, то из (1) получаем  $m = n$  и  $x_{it} = x_{jt}$  при  $t = 1, \dots, m$ . Поэтому каждое слово в  $C^*$  единственным образом декодируется как произведение слов из  $C$ . Очевидно также, что пустое слово не может входить в код, кроме того, непустое подмножество кода также является кодом. Будем считать, что алфавит  $\Sigma$  содержит не менее двух символов, так как однобуквенный алфавит тривиален с точки зрения кодирования. Приведем два довольно простых характеристических свойства кодов (см. [4]).

**Лемма 1.** Непустой язык  $C$  над алфавитом  $\Sigma$  является кодом тогда и только тогда, когда существует множество символов  $\Sigma^1$  и биекция  $\Sigma^1$  на  $C$ , которую можно продолжить до инъективного морфизма  $\Sigma^{1*}$  в  $\Sigma^*$ .

Следующий результат известен под названием критерия Шютценберже. Будем говорить, что язык  $L$  катенативно-независим, если никакое слово  $w$  из  $L$  нельзя представить как произведение

$$w = w_1 \dots w_n, \text{ где } n \geq 2 \text{ и каждое } w_i \in L, i = 1, 2, \dots, n.$$

**Лемма 2.** Катенативно-независимый непустой язык  $C$  над  $\Sigma$  является кодом тогда и только тогда, когда для любого слова  $w$  в алфавите  $\Sigma$  из того, что оба языка  $C^*w$  и  $wC^*$  пересекаются с  $C^*$ , следует, что  $w \in C^*$ .

Как указывается в [4], леммы 1 и 2 сами по себе не дают эффективного метода, позволяющего определить, является ли данный язык кодом, однако такие алгоритмы (для конечных языков) существуют. В сущности, результаты исследования приводят к следующему алгоритму: надо проверить, пусто ли множество  $x_1 C^* \cap x_2 C^*$  для каждой пары  $(x_1, x_2)$  различных элементов из  $C$ , т. е.  $x_1 \neq x_2$ . Старейшим является алгоритм, вытекающий из следующей теоремы и в первоначальном варианте принадлежащий Сардинасу и Паттерсону (см. [4]).

**Теорема 3.** Пусть  $C$  – непустой язык над  $\Sigma$ . Определим индуктивно языки  $C_0, C_1, C_2, \dots$  над  $\Sigma$ , полагая

$$C_0 = C,$$

$$C_{i+1} = \{w \in \Sigma^+ \mid (\exists x \in C) (\exists y \in C_i) uw = x \text{ или } xw = y\}.$$

Язык  $C$  является кодом тогда и только тогда, когда  $C_i \cap C = \emptyset$  для каждого  $i \geq 1$ .

Заметим, что в случае конечного языка  $C$  длина каждого слова в каждом  $C_i$  не превосходит длину самого длинного слова в  $C$ . Следовательно, существует лишь конечное число различных языков  $C_i$  и мы действительно получаем алгоритм, определяющий, является ли множество  $C$  кодом.

Для произвольного регулярного языка  $A$  рассмотрим

алфавит

$$\Delta_A = \{d_i \mid i \in I\},$$

где  $I$  – некоторое множество индексов, причем такой же мощности, как и мощность  $A$ . Записью  $h_A$  обозначим некоторый морфизм

$$h_A : \Delta_A \rightarrow \Sigma^*,$$

для которого

$$\bigcup_{i \in I} h_A(d_i) = A,$$

причем если  $u \neq v$ , то  $h_A(u) \neq h_A(v)$  (последнее условие важно только в случае бесконечных языков  $A$  и  $\Delta_A$ ). В этих последних обозначениях будем иногда, если понятно, какое множество  $A$  подразумевается, опускать нижний индекс и писать  $\Delta$ .

Рассмотрим произвольный алфавит  $\Delta$  и следующее рекуррентное определение множества максимальных префиксных кодов над  $\Delta$  в соответствии с терминологией теории кодирования (см. [4], гл. 4 и др.). Обозначим через  $\text{mp}(\Delta)$  определяемое множество языков.

- Считаем, что  $\Delta \in \text{mp}(\Delta)$ .
- Для произвольных  $C \in \text{mp}(\Delta)$  и  $B \subseteq C$  полагаем  $(C \setminus B) \cup B \Delta \in \text{mp}(\Delta)$ .

Заметим, что если алфавит  $\Delta$ , конечен то выполнено следующее.

Во-первых, приведённое определение действительно задаёт коды, являющиеся максимальными и префиксными, что легко доказывается индукцией по длине максимального слова в языке, поскольку, согласно определению, для произвольного

$$A \in \text{mp}(\Delta)$$

выполнены условия

$$\text{Pr}(A) \text{ и } \text{ci}(A) = 1 \quad (2).$$

Первое из условий (2) означает, что язык является префиксным, а второе означает, что кодовый индикатор языка равен 1. Максимальные коды можно охарактеризовать посредством некоторых числовых величин, связанных с языками. *Кодовым индикатором слова  $w \in A$  – язык над алфавитом  $\Delta$ , состоящим из  $n \geq 2$  букв, называется величина  $\text{ci}(w) = n^{-|\text{pr}(w)|}$ .* В предположении, что  $\Delta$  – минимальный алфавит языка  $A$ , т. е.  $A$  не является языком ни над каким собственным подмножеством  $\Delta$ , *кодовый индикатор  $\text{ci}(A)$  языка  $A$ , определяется как сумма кодовых индикаторов всех слов в  $A$ .* Таким образом, если язык  $A$  конечен, то  $\text{ci}(A)$  – положительное рациональное число. Если же язык  $A$  бесконечен, то  $\text{ci}(A)$  либо является положительным вещественным числом (если соответствующий ряд сходится), либо равно бесконечности. Конечный код  $A$  максимален тогда и только тогда, когда  $\text{ci}(A) = 1$  (см. [4]).

Во-вторых, в определении достаточно допускать только случай  $|B| = 1$ .

Если множество  $D$  таково, что для некоторого

$$C \in \text{mp}(\Delta)$$

выполнено включение  $C \subseteq D$ , то будем писать

$$D \in \text{mp}^+(\Delta).$$

Если для некоторого языка  $A \in \Sigma$  выполнено условие

$$C \in \text{mp}(\Delta_A) \text{ ( или } D \in \text{mp}^+(\Delta_A) ),$$

то для языков

$$C' = h_A(C) \text{ (соответственно, } D' = h_A(D) ),$$

будем писать

$$C' \in \text{mp}(A) \text{ (соответственно } D \in \text{mp}^+(A) )$$

Следующее определение дано в соответствии с [5].

**Определение** Полугруппой  $(S, \cdot)$  называется непустое множество  $S$  вместе с бинарной операцией  $\cdot$ , удовлетворяющей ассоциативному закону:

$$x \cdot (y \cdot z) = (x \cdot y) \cdot z$$

для любых  $x, y, z \in S$ .

Для произвольного непустого подмножества  $A$  полугруппы  $S$  существует наименьшая подполугруппа в  $S$ , содержащая  $A$ . Она состоит из всех конечных произведений элементов из  $A$  и обозначается через  $\langle A \rangle$  (см. [5]). Любое подмножество  $A$  из  $S$ , для которого  $\langle A \rangle = S$ , называется множеством образующих полугруппы  $S$ ; такое множество всегда существует: можно взять, например,  $A = S$ . Множество всех подполугрупп из  $S$  частично упорядочено по включению, и любые две подполугруппы  $T_1, T_2$  имеют наибольшую нижнюю границу  $\inf(T_1, T_2) = T_1 \cap T_2$  и наименьшую верхнюю границу  $\sup(T_1, T_2) = \langle T_1 \cup T_2 \rangle$ . Таким образом, подполугруппы произвольной полугруппы образуют решетку (см. стр.17, [5], а также [6]).

Проблемы решеток и полурешеток широко рассматриваются в настоящее время, о них идет речь, в частности, в статьях [7], [8-10].

### III. СТРУКТУРА ПОЛУГРУПП, СВЯЗАННЫХ С МНОЖЕСТВОМ МПК

В этом разделе применим определения и свойства из предыдущего раздела к множеству максимальных префиксных кодов.

**Предложение 1.** Рассмотрим непустой язык  $\text{mp}(\Delta)$  над алфавитом  $\Delta$ , являющийся множеством максимальных префиксных кодов (МПК) и на нем операцию конкатенации  $\circ$ , алгебра  $(\langle \text{mp}(\Delta) \rangle, \circ)$  образует полугруппу (свободную полугруппу), множеством образующих является  $\text{mp}(\Delta)$ .

**Доказательство.** Действительно, аксиома ассоциативности операции выполняется: для любых  $x, y, z \in \langle \text{mp}(\Delta) \rangle$ ,

$$x \circ (y \circ z) = (x \circ y) \circ z$$

по определению конкатенации. Т. к. каждое слово в  $\langle \text{mp}(\Delta) \rangle$ , единственным образом декодируется как произведение слов из  $\text{mp}(\Delta)$ , то и для левой и правой частей последнего равенства декодирование будет одинаковым в силу того, что ни один код не является префиксом другого.

Обозначим, как обычно, через  $\mathcal{P}(\text{mp}(\Delta))$  множество подмножеств множества  $\text{mp}(\Delta)$ . Для любого  $B \in \mathcal{P}(\text{mp}(\Delta))$ ,  $B$  является множеством образующих для подполугруппы  $\langle B \rangle$  из полугруппы  $\langle \text{mp}(\Delta) \rangle$ , а все такие подполугруппы полугруппы  $\langle \text{mp}(\Delta) \rangle$  образуют решетку.

В качестве примера рассмотрим язык, являющийся максимальным префиксным кодом

$$B = \{ aa, abaa, ababa, ababb, abb, ba, bb \}$$

для  $\Sigma = \{ a, b \}$ , причем слова  $B$  расположены в алфавитном порядке. Непосредственно можно убедиться в том, что он действительно префиксный, т. е.  $\text{Pr}(B)$  выполняется. Кроме того, кодовый индикатор  $\text{ci}(B)$  языка  $B$  равен 1 (в данном случае  $n = 2$ , раз алфавит состоит из двух букв):  $\text{ci}(B) = 3 \cdot (1/2^2) + (1/2^3) + (1/2^4) + 2 \cdot (1/2^5) = 1$ .

Пусть имеется некоторый морфизм  $h_A: \Delta_A \rightarrow \Sigma^*$ . Возникает вопрос, каким же был исходный язык  $A$  и

соответствующий морфизм, а, значит, и инверсный морфизм  $h_A^{-1}(B)$ ? Очевидный морфизм, который можно задать, такой:  $h_A(0) = a, h_A(1) = b$  для соответствующего МПК

$$A_1 = \{ 00, 0100, 01010; 01011; 011; 10; 11 \}$$

над  $\Delta = \{ 0, 1 \}$ . Кроме него можно рассматривать и другие морфизмы, например, занумеровав все слова множества  $B$  цифрами от 0 до 6:

$$A_2 = \{ 0, 1, 2, 3, 4, 5, 6 \}$$

над  $\Delta = \{ 0, 1, 2, 3, 4, 5, 6 \}$ , где  $h_A(0) = aa, h_A(1) = abaa, \dots, h_A(6) = bb$ . В связи с этим появляется вопрос и об их оптимальности по разным критериям, таким, минимальное количество символов алфавита и т. п. Второй вариант задания вряд ли можно назвать оптимальным.

Представление максимальных префиксных кодов как последовательности слов в алфавитном порядке (если алфавит упорядочен) дает возможность каждому МПК поставить в соответствие последовательность длин его слов, поэтому очевидно следующее предложение.

**Предложение 2.** Каждому максимальному префиксному коду, слова которого упорядочены в алфавитном порядке, соответствует последовательность длин этих слов. Но не каждой конечной последовательности натуральных чисел (задающих длины слов) соответствует МПК.

Для нашего примера, МПК соответствует последовательность длин (по алфавиту)  $\{ 2, 4, 5, 5, 3, 2, 2 \}$ . Эти же длины использовались для подсчета величины  $\text{ci}(B)$ , которую можно назвать инвариантом максимальной кода, а условие  $\text{ci}(B) = 1$  – одно из условий существования для последовательности МПК. Убедиться в том, что не всякой конечной последовательности соответствует максимальный префиксный код, можно, если, например, рассмотреть последовательность, в которой цифр 1 больше, чем число букв алфавита. Это последовательность  $\{ 1, 1, 1 \}$  для алфавита из двух букв  $\{ a, b \}$ . Поэтому одно из условий того, что последовательности длин соответствует МПК – то, что число единиц (слов из одного символа) не превосходит числа элементов алфавита  $n$ . Конечно, число слов длины  $t$  не должно превосходить числа перестановок с повторениями из  $t$  элементов. Если среди  $t$  элементов (букв) слова есть  $n_1$  элементов одного вида,  $n_2$  элементов второго вида, ...,  $n_k$  элементов  $k$ -го вида, то число перестановок с повторениями определяется формулой:

$$P_t(n_1, n_2, \dots, n_k) = \frac{t!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!},$$

где  $n_1 + n_2 + \dots + n_k = t$ , при этом  $k \leq n$ .

Если  $n = 2$ , т.е. в алфавите 2 буквы  $\{ a, b \}$ , то число слов длины  $t$  не должно превосходить числа перестановок с повторениями из  $t$  элементов, где буква «a» повторяется  $n_1$  раз, а буква «b»  $n_2 = t - n_1$  раз:

$$P_t(n_1, n_2) = \frac{t!}{n_1! \cdot (t - n_1)!} = C_t^{n_1}.$$

Для нашего примера при  $t = 5$ , буква «a» если повторяется  $n_1 = 3$  раза, а буква «b»  $n_2 = 5 - 3 = 2$  раза, то

$$P_5(3, 2) = \frac{5!}{3! \cdot 2!} = C_5^3 = 10.$$

В действительности же в  $B$  такое слово только одно, и вообще, нет слов с одинаковым числом вхождений каждой буквы.

Таким образом, можно считать, что последовательность – это неполный инвариант. Возникает задача описания полного множества инвариантов, а также, в том числе – минимального такого множества. Возможно, есть и другие условия, которые должны выполняться, чтобы последовательность задавала максимальный префиксный код. Это может быть темой дальнейших исследований.

#### IV. ВОЗМОЖНОСТИ СЖАТИЯ ЭЛЕМЕНТОВ

##### МНОЖЕСТВА МАКСИМАЛЬНЫХ ПРЕФИКСНЫХ КОДОВ

В этой части статьи рассматривается возможность (вариант) сжатия максимальных префиксных кодов с использованием длин слов, а сжатие информации имеет большое практическое значение.

Для сжатия МПК можно записать последовательность изменения длин каждого слова над предыдущим. Опишем этот процесс на разобранный выше примере последовательности слов (расположенных в алфавитном порядке) МПК *B*. Возможно «закодировать» МПК с помощью следующего алгоритма.

##### Алгоритм.

• *Шаг 1.* Сначала записываем длину 1-го слова, в примере это 2 – длина (*aa*).

• *Шаг 2.* Затем записываем, на сколько букв следующее слово изменяется по сравнению с предыдущим. В примере – это снова 2, увеличение длины 2-го слова (*abaa*) по сравнению с 1-м.

• *Шаг 3.* Повторяем шаг 2, пока не дойдем до последнего слова последовательности (тогда переходим к шагу 4), причем, если последующее слово – это слово минимально возможной (здесь) длины, то ничего не записываем в новую последовательность (сжатие исходной).

В примере: далее получаем 1 – увеличение длины для 3-го слова (*ababa*);

затем слово *abb* – но это слово минимально возможной (здесь) длины поэтому ничего не пишем;

далее минимально возможное слово было бы *b* но реально это *ba* поэтому пишем увеличение на 1.

• *Шаг 4.* После обработки последнего слова МПК получаем сжатую последовательность (что и требовалось).

В примере: получаем сжатую последовательность

{2, 2, 1, 1 }.

*Конец описания алгоритма.*

Подчеркнем, что алгоритм корректен в силу его построения, алфавитного порядка слов, исходная последовательность однозначно восстанавливается (раскодируется) по сжатой последовательности (см. рисунок 1 к алгоритму).

$\begin{array}{l} \overline{aa} \quad 2 \\ abaa \quad 1 \\ ababa \\ ababb \\ abb \quad - \text{слово мин. возможной длины} \\ ba \quad 1 \\ bb \end{array}$

**Рисунок 1.** Пример построения к алгоритму

Очевидно, что длина получившейся после выполнения алгоритма сжатой последовательности меньше дли-

ны исходной, а значит, занимаемая ей память меньше по объему.

#### V. ЗАКЛЮЧЕНИЕ

В дальнейшем возможно продолжить развивать тематику статьи, исследование максимальных префиксных кодов, их свойств, структуры, инвариантов, возможностей сжатия, в частности, показать, что, возможно, в МПК не содержится слов с одинаковым количеством вхождений одних и тех же букв, рассмотреть, какие ограничения накладываются на последовательность длин слов. Для этих исследований подойдут, например, некоторые статистические эксперименты, в которых каким-либо способом генерируются последовательности МПК с разным числом букв (скажем, от 2 до 10), разной минимальной и разной максимальной длиной. Будет возможность сравнить память, занимаемую исходной и сжатой каким-либо способом последовательностями. Некоторые интересные направления также задаются в статьях [3,7,11].

#### БИБЛИОГРАФИЯ

- [1] Общая алгебра. Том 2 / Артамонов В.А., Салий В.Н., Скорняков Л.А. и др. М.: Наука, Физматлит, 1991.
- [2] Марков А.А., Введение в теорию кодирования // М.: Наука, Гл. ред. физ.-мат. лит., 1982. – 192с.
- [3] Мельников Б.Ф., Мельникова А.А. Полиномиальный алгоритм проверки выполнения условия морфического образа расширенного максимального префиксного кода // International Journal of Open Information Technologies. 2022. Т. 10. № 12. С. 1–9.
- [4] Саломая А. Жемчужины теории формальных языков // М.: Мир, 1986.
- [5] Лаллеман Ж. Полугруппы и комбинаторные приложения, М.: Мир, 1985.
- [6] Биркгоф Г., Барти Т. Современная прикладная алгебра // М.: Мир, 1976.
- [7] Зяблищева Л.В., Корабельщикова С.Ю., Абрамян М.Э. Графы полурешеток, матричные представления полугрупп идемпотентов и полурешетки автомата Ватерлоо // International Journal of Open Information Technologies. 2023. Т. 11. № 7. С. 69–76.
- [8] Мельников Б.Ф. Полурешетки подмножеств потенциальных корней в задачах теории формальных языков. Часть I. Извлечение корня из языка // International Journal of Open Information Technologies. 2022. Т. 10. № 4. С. 1–9.
- [9] Мельников Б.Ф. Полурешетки подмножеств потенциальных корней в задачах теории формальных языков. Часть II. Построение инверсного морфизма // International Journal of Open Information Technologies. 2022. Т. 10. № 5. С. 1–8.
- [10] Мельников Б.Ф. Полурешетки подмножеств потенциальных корней в задачах теории формальных языков. Часть III. Условие существования решетки // International Journal of Open Information Technologies. 2022. Т. 10. № 7. С. 1–9.
- [11] Данг В.В., Додонова Н.Л., Корабельщикова С.Ю., Мельников Б.Ф. SH-слабая двойственность полугрупп и минимальная полугруппа SH-аппроксимации // Известия высших учебных заведений. Поволжский регион. Физико-математические науки. 2019. № 1 (49). С. 29–39.

Александра Александровна Мельникова,  
доцент Димитровградского инженерно-технологического института – филиала Национального исследовательского ядерного университета «МИФИ» (<https://diti-mephi.ru/>),  
email: super-avahi@yandex.ru,  
mathnet.ru: personid=148963,  
elibrary.ru: authorid=143351,  
scopus.com: authorId=6603567251,  
ORCID: orcidID=0000-0002-1658-6857.

# The set of maximal prefix codes and the corresponding semigroup

Aleksandra Melnikova

**Abstract**—One of the most numerous classes of algebra are semigroups - apparently, due to the fact that the only axiom that must be fulfilled for a semigroup is the axiom of associativity of the operation given on it. In this paper, it is shown that the set of maximal prefix codes is the set of forming the corresponding semigroup with respect to the concatenation operation specified on it. An important issue remains the idea of the structure of the semigroup, which is also discussed in the article.

The article deals with the issues of morphisms of maximal prefix codes. Based on the representation of the maximum prefix codes as a sequence of words in alphabetical order (if the alphabet is ordered), a proposal is formulated on the possibility of putting a sequence of the lengths of its words in accordance with each such code. Some conditions are discussed under which a maximum prefix code will correspond to a sequence of lengths.

The variant of compression of such codes with the use of word lengths is considered, the algorithm is described. As you know, encoding, compression of information is of great importance in practice. An example of the algorithm execution and a drawing for it are given. It also talks about possible directions for the continuation of the development of the subject of the article, the possibilities of studying maximum prefix codes, their properties, structure, invariants, compression capabilities and unambiguity of decoding, restoration of the source code.

Examples of definitions and properties for maximum prefix codes are given.

**Keywords**—regular languages, codes, maximal prefix codes, semigroups, morphisms, inverse morphisms.

## REFERENCES

- [1] Abstract algebra. Vol. 2 / Artamonov V.A., Saliy V.N., Skorniyakov L.A. et al. Moscow: Nauka Publ., Phizmatlit, 1991 (in Russian).
- [2] Markov A.A., Introduction to coding theory // M.: Nauka, Gl. ed. phys.-mat. lit., 1982. – 192 p. (in Russian).
- [3] Melnikov B., Melnikova A. A polynomial algorithm for checking the fulfillment of the condition of the morphic image of the extended maximal prefix code // International Journal of Open Information Technologies. – 2022. – Vol. 10. No. 12. – P. 1–9 (in Russian).
- [4] Salomaa A. Pearls of the theory of formal languages // Moscow: Mir, 1986 (in Russian).
- [5] Lallemand G. Semigroups and Combinatorial Applications. – NJ, Wiley & Sons, Inc. – 1979. – 376 p. (in Russian).
- [6] Birkhoff G., Barty T. Modern applied Algebra // Moscow: Mir, 1976. – 400 p. (in Russian).
- [7] Zyabliteva L.V., Korabelshchikova S.Yu., Abramyan M. E. Semilattice graphs, matrix representations of idempotent semigroups and Waterloo automaton semilattices // International Journal of Open Information Technologies. 2023. Vol. 11. No 7. P. 69–76 (in Russian).
- [8] Melnikov B.F. Semi-lattices of the subsets of potential roots in the problems of the formal languages theory. Part I. Extracting the root from the language // International Journal of Open Information Technologies. 2022. Vol. 10. No. 4. P. 1–9 (in Russian).
- [9] Melnikov B.F. Semi-lattices of the subsets of potential roots in the problems of the formal languages theory. Part II. Constructing an inverse morphism // International Journal of Open Information Technologies. 2022. Vol. 10. No 5. P. 1–8 (in Russian).
- [10] Melnikov B.F. Semi-lattices of the subsets of potential roots in the problems of the formal languages theory. Part III. The condition for

the existence of a lattice // International Journal of Open Information Technologies. 2022. Vol. 10. No. 7. P. 1–9 (in Russian).

- [11] Dang V.V., Dodonova N.L., Korabelshchikova S.Yu., Melnikov B.F. SH-weak duality of semigroups and minimum semigroup of SH-approximation // University proceedings. Volga region. Physical and mathematical sciences. 2019. No. 1 (49). P. 29–39 (in Russian).

Aleksandra Aleksandrovna MELNIKOVA,  
Associated Professor of Dimitrovgrad Engineering and  
Technology Institute – Branch of National Research Nuclear  
University “MEPhI” (<https://diti-mephi.ru/>),  
email: super-avahi@yandex.ru,  
mathnet.ru: personid=148963,  
elibrary.ru: authorid=143351,  
scopus.com: authorId=6603567251,  
ORCID: orcidID=0000-0002-1658-6857.