УДК 519

Е. В. Прибавкина

2-СЖИМАЮЩИЕ СЛОВА И ПРОБЛЕМА РЕКОНСТРУКЦИИ ПОСЛЕДОВАТЕЛЬНОСТИ*

1. Введение

Проблема реконструкции слова по множеству его факторов состоит в нахождении кратчайшего слова, имеющего в качестве факторов все элементы из данного множества. Эта задача возникает в области биоинформатики при анализе генетических последовательностей. Дело в том, что классическая технология секвенирования позволяет биологам в ходе одного эксперимента читать только короткие (300–500-буквенные) фрагменты ДНК. Чтобы получить полный геном, необходимо совместить должным образом эти короткие фрагменты.

В общем случае задача реконструкции последовательности является NP-трудной [1], поэтому в последнее время изучались разные ее вариации при некоторых дополнительных ограничениях на слово и/или множество его факторов [2–5]. Одной из таких модификаций, которая будет рассматриваться в данной работе, является секвенирование гибридизацией (SBH). Дадим математическую интерпретацию этой задачи: по данному множеству X слов длины k определить, существует ли слово w, содержащее все элементы множества X в качестве факторов и такое, что любой его фактор длины k является элементом множества X (такое слово w называется *накрывающим для множества* X). Для этой задачи известен полиномиальный от $|X|$ алгоритм [2], который в случае положительного ответа находит кратчайшее накрывающее для X слово.

С другой стороны, задача реконструкции слова по множеству его факторов может иметь другое применение. В комбинаторике слов часто случается, что свойства определенных факторов слова полностью определяют свойства самого слова [6–8]. Такие факторы мы будем называть *существенными*. Поэтому естественным вопросом является возможность реконструкции (кратчайшего) слова с данным свойством по множеству его существенных факторов.

---

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

Данная работа является попыткой применить эту идею: мы изучаем возможность реконструкции 2-*сжимающих* и 2-*синхронизирующих слов* по множеству их *внутренних отрезков*. Основной результат работы (предложения 1 и 2) показывает, что кратчайшие 2-сжимающие и 2-синхронизирующие слова являются накрывающими для некоторого специального множества своих факторов. Этот результат в некоторых случаях позволяет ускорить процесс проверки данного слова на 2-сжимаемость и 2-синхронизируемость. Кроме того, такой подход позволяет находить более короткие 2-сжимающие и 2-синхронизирующие слова по существующим примерам.

2. Внутренние отрезки

Зафиксируем конечный алфавит Σ. Детерминированный конечный автомат $\mathscr{A} = \langle Q, \Sigma, \delta\rangle$ определяется заданием конечного *множества состояний* Q и *функции переходов* $\delta : Q \times \Sigma \to Q$. Функция δ естественным образом продолжается на свободный моноид Σ^*; это продолжение мы также будем обозначать через δ. Таким образом, каждое слово $w \in \Sigma^*$ (в частности, каждая буква алфавита Σ) порождает преобразование $\delta(\_\!\_, w) : Q \to Q$ множества Q. Имея дело с конечными детерминированными автоматами, мы будем отождествлять слово w с этим преобразованием. Для каждого $v \in \Sigma^*$ и $q \in Q$ будем писать $q \,.\, v = \delta(q, v)$ и положим $Q \,.\, v = \{q \,.\, v \mid q \in Q\}$.

Пусть n – натуральное число. Автомат $\mathscr{A}$ называется n-*сжимаемым*, если существует слово $v \in \Sigma^+$ такое, что справедливо неравенство $|Q| - |Q \,.\, v| \geq n$ (при этом говорят, что слово v *сжимает автомат* $\mathscr{A}$ *на* n *состояний*). Слово $w \in \Sigma^*$ называется n-*сжимающим*, если оно сжимает на n состояний все n-сжимаемые автоматы с входным алфавитом Σ. Это понятие имеет различные приложения в универсальной алгебре, теории полугрупп преобразований, теории автоматов и биоинформатике (подробнее см. [9]).

Наряду с n-сжимающими словами рассматривается также понятие n-*синхронизирующих* слов. Напомним, что автомат $\mathscr{A} = \langle Q, \Sigma, \delta\rangle$ называется синхронизируемым, если существует *синхронизирующее* слово $v \in \Sigma^*$, переводящее все состояния в одно выделенное: $|Q \,.\, v| = 1$. Слово $w \in \Sigma^*$ называется n-*синхронизирующим*, если оно синхронизирует все синхронизируемые автоматы с $n+1$ состоянием и входным алфавитом Σ.

Для первого нетривиального случая $n = 2$ в [6] дана теоретико-групповая характеризация 2-сжимающих слов, а именно, по данному слову w авторы строят множество его факторов и показывают, что слово w является 2-сжимающим в том и только в том случае, когда это множество удовлетворяет определенным условиям в терминах свободных групп. В той же работе подобный критерий дан и для 2-синхронизирующих слов. В общем виде эта характеризация довольно громоздка, но здесь нам понадобятся лишь несколько относительно простых следствий (леммы 1–3).

Здесь мы исследуем возможность реконструкции 2-сжимающих и 2-синхронизирующих слов по множеству их факторов, которое играет важную роль в вышеупомянутом критерии. Это множество определяется следующим образом.

Выберем произвольную букву $a \in \Sigma$ и обозначим $\Pi = \Sigma \setminus \{a\}$ (мы считаем, что $|\Sigma| > 1$). Тогда любое слово $w \in \Sigma^+$ можно единственным образом представить в виде

$$w = a^{\alpha_0} p_1 a^{\alpha_1} p_2 \cdots p_m a^{\alpha_m}, \tag{1}$$

где $\alpha_0, \alpha_m \geq 0$; $\alpha_i > 0$, $i = 1, \ldots, m-1$; $p_i \in \Pi^+$, $i = 1, \ldots, m$; $m \geq 0$. Фактор p_i называется *внутренним отрезком по отношению к букве* a, если $\alpha_{i-1} \cdot \alpha_i \neq 0$, т. е. и слева, и справа от p_i есть буква a. Понятие внутреннего отрезка было введено в [6] в более общей форме, здесь мы даем это определение в том виде, в каком оно использовалось в [8]. Обозначим через S_a множество всех внутренних отрезков слова w по отношению к букве a и положим $S = \bigcup_{a \in \Sigma} S_a$. Отметим, что множества S_a для разных букв $a \in \Sigma$ могут иметь общие элементы, поскольку одно и то же слово из Π^+ может несколько раз появляться в качестве фактора в разложении (1) между разными буквами $a \in \Sigma$.

Лемма 1. *Пусть w – 2-сжимающее слово. Тогда для каждого выбора буквы $a \in \Sigma$ множество S_a порождает свободную группу $FG(\Pi)$ над алфавитом $\Pi = \Sigma \setminus \{a\}$.*

Это утверждение является следствием характеризации 2-сжимающих слов из [6]. Доказательство в явном виде см. в [8].

Следующее утверждение о свойствах множества внутренних отрезков 2-синхронизирующего слова следует из характеризации 2-синхронизирующих слов в [6]. Для полноты изложения мы приводим его с доказательством.

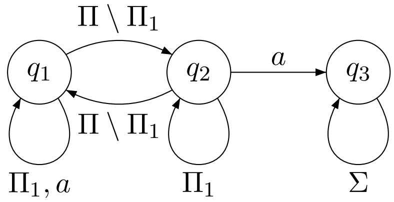
Лемма 2. *Пусть w – 2-синхронизирующее слово. Тогда для каждого выбора буквы $a \in \Sigma$ множество S_a не является языком ни над каким собственным подмножеством алфавита $\Pi = \Sigma \setminus \{a\}$, т. е. $S_a \not\subseteq \Pi_1^*$ ни для какого $\Pi_1 \subsetneqq \Pi$.*

Доказательство. Рассуждая от противного, предположим, что для некоторого выбора буквы $a \in \Sigma$ существует $\Pi_1 \subsetneqq \Pi$, что $S_a \subset \Pi_1^*$. Тогда рассмотрим автомат $\mathscr{A}_{S_a} = \langle Q, \Sigma, \delta \rangle$ с $Q = \{q_1, q_2, q_3\}$ и функцией переходов, заданной следующим образом:

$q_1 \,.\, x = q_1$ для любого $x \in \Pi_1 \cup \{a\}$; $\quad q_1 \,.\, y = q_2$ для любого $y \in \Pi \setminus \Pi_1$;

$q_2 \,.\, x = q_2$ для любого $x \in \Pi_1$; $\quad q_2 \,.\, y = q_1$ для любого $y \in \Pi \setminus \Pi_1$;

$q_2 \,.\, a = q_3$; $\quad q_3 \,.\, x = q_3$ для любого $x \in \Sigma$.



Автомат $\mathscr{A}_{S_a}$

Графически автомат $\mathscr{A}_{S_a}$ представлен на рисунке. Легко видеть, что этот автомат синхронизируем, например, словом aba, где b – произвольная буква из $\Pi \setminus \Pi_1$, однако слово w его не синхронизирует: $Q \,.\, w = Q \setminus \{q_2\}$. Получаем противоречие с тем, что слово w – 2-синхронизирующее.

Для каждого множества S_a рассмотрим множество $\widehat{S}_a = \{axa \mid x \in S_a\}$, его элементы мы будем называть *расширенными внутренними отрезками*. Множество всех расширенных отрезков обозначим через $\widehat{S} = \bigcup_{a\in\Sigma} \widehat{S}_a$. Отметим, что теперь никакие два множества вида $\widehat{S}_a$ не пересекаются.

Следующее утверждение является техническим следствием критерия из работы [6].

Лемма 3. *Пусть два слова v и w обладают одним и тем же множеством $\widehat{S}$ расширенных внутренних отрезков. Тогда*

(i) v *– 2-сжимающее* $\iff$ w *– 2-сжимающее;*

(ii) v *– 2-синхронизирующее* $\iff$ w *– 2-синхронизирующее.*

Пусть дано множество $\widehat{S}$ расширенных внутренних отрезков некоторого 2-сжимающего (2-синхронизирующего) слова. Обозначим через $\mathcal{CW}_S$ (соответственно $\mathcal{SW}_S$) множество всех кратчайших слов, расширенные внутренние отрезки которых совпадают с $\widehat{S}$. По лемме 3 каждое слово из множества $\mathcal{CW}_S$ ($\mathcal{SW}_S$) также является 2-сжимающим (2-синхронизирующим).

3. Построение множества, для которого 2-сжимающее (2-синхронизирующее) слово является накрывающим

По множеству $\widehat{S}$ расширенных внутренних отрезков слова построим множество $X_S \subset \Sigma^3$ как множество всех факторов длины 3 слов из множества $\widehat{S}$. Например, если $\widehat{S} = \{aba, aca, bacab, cabac\}$, то $X_S = \{aba, aca, bac, cab\}$.

Предложение 1. *Если $\widehat{S}$ – множество расширенных внутренних отрезков 2-сжимающего слова, то любое слово w_S из $\mathcal{CW}_S$ является накрывающим для множества X_S.*

Доказательство. По построению каждый элемент множества X_S является фактором слова w_S. Таким образом, нам нужно доказать, что любой фактор x слова w_S длины 3 принадлежит X_S.

1. Если $x = abc$, где $a, b, c \in \Sigma$, является префиксом слова w_S, то $a \neq b$ (иначе, если $x = aac$, то $w_S = aacu$ для некоторого $u \in \Sigma^+$ и слово $w' = acu$ будет иметь тот же набор $\widehat{S}$ расширенных внутренних отрезков, что и w_S, и иметь меньшую длину, что противоречит выбору слова w_S). Таким образом, префикс x может иметь вид aba или abc для различных букв a, b и c. В первом случае x принадлежит множеству $\widehat{S}_a$, а значит и X_S, во втором случае x также принадлежит множеству X_S, поскольку является фактором некоторого элемента множества $\widehat{S}_a$ (соответствующего внутреннему отрезку между первым и вторым вхождением буквы a в слово w_S). Аналогично рассматривается случай, когда x является суффиксом слова w.

2. Пусть теперь x является внутренним фактором слова w_S. Рассмотрим четыре возможных случая.

2.1. $x = aaa$ для некоторой буквы $a \in \Sigma$. Поскольку x является внутренним фактором, слева и справа от x в слове w_S есть буквы, например, b и c (заметим, что b и c отличны от a, иначе, как и в первом случае, мы бы имели слово, начинающееся или заканчивающееся степенью буквы a, что противоречило бы выбору слова w_S). Таким образом, $w_S = \cdots b \cdots aaa \cdots c \cdots$ и $a \neq b$, $a \neq c$. Если по обе стороны от x встречается одна и та же буква, например b, то $x \in X_S$ (как фактор элемента множества $\widehat{S}_b$). В противном случае пусть буква b встречается только слева от x в w_S, а буква c – только справа, тогда мы имеем $w_S = \cdots b \cdots b \cdots aaa \cdots c \cdots c \cdots$. Но тогда буква c не входит во внутренние отрезки слова w_S по отношению к букве b, следовательно, множество S_b не может порождать свободную группу $FG(\Sigma \backslash \{b\})$, что противоречит свойству внутренних отрезков 2-сжимающего слова (лемма 1).

2.2. $x = aab$, $a \neq b$. Как и в случае 2.1, слева от x есть буквы, отличные от a. Если слева от x есть буква b, то x является фактором некоторого элемента множества $\widehat{S}_b$, а значит $x \in X_S$. В противном случае пусть слева от x

есть буква $c \neq b$, тогда, рассуждая аналогично первому случаю, мы приходим к выводу, что справа от x тоже есть буква c, следовательно, x является фактором некоторого элемента множества $\widehat{S}_c$, т. е. $x \in X_S$. Аналогично рассматривается случай, когда $x = abb$ и $a \neq b$.

2.3. $x = aba$ и $a \neq b$. В этом случае $x \in \widehat{S}_a$ по определению, т. е. $x \in X_S$.

2.4. $x = abc$ для различных букв a, b, c. Как и в предыдущих случаях, если слева от x в w_S есть буква c или справа от x в w_S есть буква a, то $x \in X_S$. Если же слева от x нет буквы c, а справа нет буквы a, то внутренние отрезки относительно буквы a не содержат буквы c, а значит, множество S_a не может порождать свободную группу $FG(\Sigma \setminus a)$, вновь получаем противоречие с леммой 1.

Предложение 2. *Если $\widehat{S}$ – множество расширенных внутренних отрезков 2-синхронизирующего слова, то любое слово w_S из $\mathcal{SW}_S$ является накрывающим для множества X_S.*

Доказательство. Доказательство аналогично доказательству предыдущего предложения за исключением ситуаций, когда получается противоречие с леммой 1, т. е. когда в слове w_S некоторая буква a встречается только слева от некоторого $x \in X_S$, а некоторая другая буква b – только справа. Но в этом случае получается противоречие со свойством внутренних отрезков 2-синхронизирующих слов (леммой 2).

Замечание 1. *Отметим, что число 3 является максимальным возможным значением для k при сведении нашей задачи к SBH, поскольку 3 является минимальным возможным значением длины расширенных внутренних отрезков – отрезков вида aba.*

Пример 1. Рассмотрим слово

$$v_1 = aba \cdot c \cdot b^2c^2a^2c \cdot bab \cdot cac \cdot b \cdot cab$$

длины 21 над алфавитом $\{a, b, c\}$. Известно, что v_1 является одним из кратчайших 2-сжимающих слов над трехбуквенным алфавитом. Множество его расширенных внутренних отрезков:

$$\begin{aligned} \widehat{S} = \{&aba,\ abca,\ acba,\ acb^2c^2a,\ acbca, \\ &bab,\ bacb,\ bcab,\ \ bcacb,\ bc^2a^2cb, \\ &cac,\ ca^2c,\ cbabc,\ cbc,\ cb^2c\}, \end{aligned}$$

из которого мы получаем множество

$$X_S = \{aac,\ aba,\ abc,\ acb,\ bab,\ bac,\ bbc,\ bca,\\ bcc,\ caa,\ cab,\ cac,\ cba,\ cbb,\ cbc,\ cca\}.$$

Применяя алгоритм Shortest SBH Reconstruction из [2], мы получаем, что длина кратчайшего накрывающего слова для X_S равна 21. Более того, в списке всех таких слов, кроме v_1, присутствуют остальные три кратчайших 2-сжимающих слова с тем же множеством $\widehat{S}$ расширенных внутренних отрезков:

$$v_2 = aba \cdot c \cdot bab \cdot cac \cdot b^2c^2a^2c \cdot b \cdot cab,$$
$$v_3 = aba \cdot c \cdot b^2c^2a^2c \cdot b \cdot cac \cdot bab \cdot cab,$$
$$v_4 = aba \cdot c \cdot b \cdot cac \cdot b^2c^2a^2c \cdot bab \cdot cab.$$

4. Заключение

Мы применяем метод реконструкции последовательности к словам со свойством 2-сжимаемости (2-синхронизируемости) и к специальному множеству их факторов – расширенным внутренним отрезкам. В результате мы показываем, что для данного множества $\widehat{S}$ расширенных внутренних отрезков 2-сжимающего (2-синхронизирующего) слова кратчайшее слово w_S с тем же множеством $\widehat{S}$ является также 2-сжимающим (соответственно 2-синхронизирующим) и накрывающим для множества X_S всех факторов длины 3 слов из $\widehat{S}$.

Этот результат может иметь следующее применение. Критерий из [6] дает теоретико-групповой алгоритм распознавания 2-сжимающих слов, его улучшенную версию (под названием C2C) можно найти в [10]. Время его работы экспоненциально от длины входного слова. В работе [11] показано, что задача проверки слова на 2-сжимаемость является co-NP-полной, следовательно, маловероятно существование эффективного алгоритма для этой задачи, но представляет интерес любое улучшение существующего алгоритма.

В некоторых случаях наш подход может служить таким улучшением. Идея состоит в том, чтобы, до применения алгоритма C2C к данному слову, сначала разрезать его на множество факторов и применить реконструкцию. Таким образом, если нам повезет, мы получим слово короче исходного, сохраняющее свойство 2-сжимаемости. В этом случае далее мы будем применять алгоритм C2C уже к более короткому слову. Формально, пусть дано слово w. Сначала мы находим множество $\widehat{S}$ его расширенных внутренних отрезков, по которому строим множество X_S. Далее мы применяем алгоритм Shortest SBH

Reconstruction для нахождения кратчайшего накрывающего слова для множества X_S. Если такое слово не найдено, слово w не является 2-сжимающим. В случае когда среди всех кратчайших накрывающих слов для X_S есть слово v с тем же множеством $\widehat{S}$ расширенных внутренних отрезков, то ясно, что v не длиннее исходного слова и сохраняет свойство 2-сжимаемости. Однако если среди всех кратчайших накрывающих слов для X_S нет слова со множеством расширенных внутренних отрезков $\widehat{S}$, то наш подход не дает улучшения, поскольку в этом случае мы не можем сказать ничего о свойствах слова w. В этом случае мы должны искать накрывающее слово для X_S с дополнительным условием: оно должно содержать все слова из $\widehat{S}$ в качестве факторов. Эта задача исследовалась, например, в [5] и была решена в некоторых частных случаях, однако в общем виде она остается открытой.

Рассмотрим несколько примеров, иллюстрирующих наш подход.

Пример 2. Рассмотрим слово

$$w_1 = a^2 \cdot (ba)^2 cb \cdot (cac \cdot bab)^2 cac \cdot bc \cdot a^2c^2b^2 \cdot cac \cdot bc \cdot (ab)^2$$

длины 40. Множество его внутренних отрезков:

$$\begin{aligned}\widehat{S} = \{&aba,\ abca,\ acba,\ acbca,\ ac^2b^2ca,\\ &bab,\ bacb,\ bca^2c^2b,\ bcab,\ bcacb,\\ &cac,\ ca^2c,\ cbabc,\ cbc,\ cb^2c\},\end{aligned}$$

откуда мы получаем множество

$$\begin{aligned}X_S = \{&aac,\ aba,\ abc,\ acb,\ acc,\ bab,\ bac,\ bbc,\\ &bca,\ caa,\ cab,\ cac,\ cba,\ cbb,\ cbc,\ ccb\}.\end{aligned}$$

Применяя алгоритм Shortest SBH Reconstruction, мы получаем кратчайшие накрывающие слова для X_S длины 21, среди которых есть четыре слова с тем же множеством $\widehat{S}$ расширенных внутренних отрезков:

$$\begin{aligned}u_1 &= aba \cdot c \cdot bab \cdot ca^2c^2b^2 \cdot cac \cdot b \cdot cab,\\ u_2 &= aba \cdot c \cdot bab \cdot cac \cdot b \cdot ca^2c^2b^2 \cdot cab,\\ u_3 &= aba \cdot c \cdot b \cdot ca^2c^2b^2 \cdot cac \cdot bab \cdot cab,\\ u_4 &= aba \cdot c \cdot b \cdot cac \cdot bab \cdot ca^2c^2b^2 \cdot cab.\end{aligned}$$

Применяя C2C к любому из этих слов, получаем, что исходное слово w_1 является 2-сжимающим.

Пример 3. Рассмотрим одно из кратчайших 2-сжимающих слов над трехбуквенным алфавитом:

$$w_2 = abc \cdot cab \cdot cac \cdot bca \cdot abb \cdot cab \cdot abc.$$

Множество его расширенных внутренних отрезков:

$$\begin{aligned}\widehat{S} = \{&aba,\ ab^2ca,\ abca,\ abc^2a,\ acbca,\\ &bab,\ bca^2b,\ bcab,\ bcacb,\ bc^2ab,\\ &ca^2b^2c,\ cababc,\ cabc,\ cac,\ cbc\},\end{aligned}$$

откуда мы получаем множество

$$X_S = \{aab,\ aba,\ abb,\ abc,\ acb,\ bab,\ bbc,\ bca,\ bcc,\ caa,\ cab,\ cac,\ cbc,\ cca\}.$$

Применяя алгоритм Shortest SBH Reconstruction, мы получаем, что кратчайшее накрывающее слово для X_S имеет длину 16 (например, $u = cab \cdot ab^2 \cdot c^2a^2 \cdot bca \cdot cbc$). Множество всех таких слов не содержит слова с тем же множеством $\widehat{S}$ расширенных внутренних отрезков (иначе они были бы более короткими 2-сжимающими словами).

В случае 2-синхронизирующих слов существует очевидный алгоритм распознавания: взять все автоматы с тремя состояниями над фиксированным алфавитом (их конечное число), выделить среди них синхронизируемые (существует эффективный алгоритм проверки автомата на синхронизируемость) и проверить, синхронизирует ли данное слово все такие автоматы. Кроме того, известно (см. например [7]), что язык 2-синхронизирующих слов, в отличие от 2-сжимающих, является регулярным. Однако такая распознаваемость в принципе не ведет ни к каким практическим алгоритмам. Характеризация в [6], так же как и в случае 2-сжимающих слов, ведет к экспоненциальному от длины входного слова алгоритму распознавания 2-синхронизирующих слов. Следовательно, наш подход можно применять и в этом случае.

Кроме того, наш подход можно применять для построения коротких примеров 2-сжимающих и 2-синхронизирующих слов. В случае двухбуквенного алфавита все кратчайшие такие слова были найдены еще в самой первой работе по данной теме [12]. В [13] с помощью компьютерной реализации вышеупомянутых алгоритмов был найден полный список всех 2-сжимающих и 2-синхронизирующих слов над трехбуквенным алфавитом. Для алфавитов большей мощности существуют лишь отдельные примеры или серии примеров, (см. [6, 7, 14, 15]), построенные «руками» на основе свойств этих слов, поэтому они вряд ли являются кратчайшими. Применение реконструкции к этим примерам в некоторых случаях позволяет уменьшить их длину. Так, применяя наш алгоритм к 2-сжимающему слову над пятибуквенным алфавитом из работы [14], удалось получить более короткое 2-сжимающее слово. Применяя алгоритм к 2-синхронизирующим словам над четырех- и пятибуквенным алфавитами и к 2-сжимающему слову над четырехбуквенным алфавитом из той же работы, уменьшить примеры не удалось, однако удалось получить еще три 2-сжимающих слова той же длины над четырехбуквенным алфавитом.

1. Gallant J., Maier D., Storter J. On finding minimal length superstrings // J. Comput. System Sci. 1980. Vol. 20. P. 50–58.

2. Ananichev D. S. The problem of the shortest SBH reconstruction is in P. (В печати).

3. Fici G., Mignosi F., Restivo A. et. al. Word assembly through minimal forbidden factors // Theoret. Comput. Sci. 2006. Vol. 359. P. 214–230.

4. Mignosi F., Restivo A., Sciortino M. Forbidden factors and fragment assembly // Developments in Language theory [Lecture Notes in Computer Science; Vol. 2295]. B.; Heidelberg: Springer, 2002. P. 349–358.

5. Pevzner P. A., Tang H., Waterman M. S. An eulerian path approach to DNA fragment assembly problem // Proc. of the National Academy of Sciences of the USA. 2001. Vol. 98, № 17. P. 9748–9753.

6. Ananichev D. S., Cherubini A., Volkov M. V. Image reducing words and subgroups of free groups // Theoret. Comput. Sci. 2003. Vol. 307. P. 77–92.

7. Ananichev D. S., Volkov M. V. Collapsing words vs. synchronizing words // Developments in Language theory [Lecture Notes in Computer Science; Vol. 2295]. B.; Heidelberg: Springer, 2002. P. 166–174.

8. Pribavkina E. V. On some properties of the langusge of 2-collapsing words // Ibid [Lecture Notes in Computer Science; Vol. 3572]. B.; Heidelberg: Springer, 2005. P. 374–384.

9. Ananichev D. S., Petrov I. V., Volkov M. V. Collapsing words: a progress report // Ibid. P. 11–21.

10. Ananichev D. S., Cherubini A., Volkov M. V. An inverse automata algorithm for recognizing 2-collapsing words // Ibid [Lecture Notes in Computer Science; Vol. 2450]. B.; Heidelberg: Springer, 2003. P. 270–282.

11. Cherubini A., Kisielewicz A. Recognizing collapsing words is co-NP-complete: Preprint. 2005.

12. Sauer N., Stone M. G. Composing functions to reduce image size // Ars Combinatoria. 1991. Vol. 31. P. 171–176.

13. Ananichev D. S., Petrov I. V. Quest for short synchronizing words and short collapsing words // WORDS 2003. Proc. 4th Int. Conf. Univ. of Turku, Turku, 2003. P. 411–418.

14. Cherubini A., Kisielewicz A., Piochi B. A bound for the length of shortest 2-collapsing words // Proc. 6th Int. Conf. on Words. Technical report of Institut de Mathématiques de Luminy. France, Marseille, 2007. P. 90–99.

15. Margolis S. W., Pin J.-E., Volkov M. V. Words guaranteeing minimum image // Int. J. Foundations Comp. Sci. 2004. Vol. 15. P. 259–276.

Статья поступила 08.04.2008