ЧЕБЫШЕВСКИЙ СБОРНИК
Том 22. Выпуск 1.

УДК 519.115.8 DOI 10.22405/2226-8383-2021-22-1-57-66

О возможности восстановления периодического слова по подсловам фиксированной длины[1]

В. А. Алексеев, Ю. Г. Сметанин

Василий Антонович Алексеев — ассистент кафедры информатики и вычислительной математики, ассистент кафедры высшей математики, Московский физико-технический институт (национальный исследовательский университет) (г. Москва).
e-mail: wasya.alekseev@gmail.com

Юрий Геннадьевич Сметанин — доктор физико-математических наук, главный научный сотрудник, Федеральный исследовательский центр «Информатика и управление» Российской академии наук (г. Москва).
e-mail: ysmetanin@rambler.ru

Аннотация

Рассмотрена задача реконструкции слов из конечного алфавита по частичной информации при дополнительных ограничениях на допустимые слова. А именно, ставится задача о восстановлении периодического слова по мультимножеству его подслов одной длины. Для некоторых видов частичной информации и ограничений получены условия однозначной реконструкции. Показано, что периодическое слово с периодом p однозначно определяется мультимножеством его подпоследовательностей длины $k \geq \left\lfloor \frac{16}{7}\sqrt{p} \right\rfloor + 5$. Для слова, состоящего из непериодического префикса длины q и периодического суффикса с периодом p, повторяющегося l раз, получена аналогичная оценка $k \geq \left\lfloor \frac{16}{7}\sqrt{P} \right\rfloor + 5$ при условии $l \geq q^{\left\lfloor \frac{16}{7}\sqrt{P} \right\rfloor + 5}$, где $P = \max(p, q)$.

Ключевые слова: реконструкция слова, мультимножество подслов, подслова фиксированной ванной длины, периодическое слово.

Библиография: 14 названий.

Для цитирования:
В. А. Алексеев, Ю. Г. Сметанин. О возможности восстановления периодического слова по подсловам фиксированной длины //Чебышевский сборник, 2021, т. 22, вып. 1, с. 57–66.

[1]Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта №19-07-00150.

CHEBYSHEVSKII SBORNIK
Vol. 22. No. 1.

UDC 519.115.8 DOI 10.22405/2226-8383-2021-22-1-57-66

On the possibility of a periodic word reconstruction from the subwords of fixed length

V. A. Alekseev, Y. G. Smetanin

Vasily Antonovich Alekseev — assistant at the department of informatics and computational mathematics, assistant at the department of higher mathematics, Moscow Institute of Physics and Technology (MIPT) (Moscow).
e-mail: wasya.alekseev@gmail.com
Yuri Gennadievich Smetanin — doctor of physical and mathematical sciences, chief researcher, Federal Research Center «Computer Science and Control», Russian Academy of Sciences (Moscow).
e-mail: ysmetanin@rambler.ru

Abstract

The problem being considered is the reconstruction of periodic words from a finite alphabet using multiset of fixed length subwords. This is a special case of a more general problem of reconstruction with incomplete information and under restrictions on the words in question. For some constraints on the multiset of subwords, conditions for possibility of reconstruction are obtained. It is shown that a periodic word with period p is uniquely determined by the multiset of its subwords of length $k \geq \left\lfloor \frac{16}{7}\sqrt{p} \right\rfloor + 5$. For a word consisting of a non-periodic prefix of length q and a periodic suffix with period p, repeated l times, a similar estimate is obtained: $k \geq \left\lfloor \frac{16}{7}\sqrt{P} \right\rfloor + 5$, provided $l \geq q^{\left\lfloor \frac{16}{7}\sqrt{P} \right\rfloor + 5}$ where $P = \max(p, q)$.

Keywords: word reconstruction, multiset of subwords, subwords of fixed length, periodic word.

Bibliography: 14 titles.

For citation:
V. A. Alekseev, Y. G. Smetanin. 2021, “On the possibility of a periodic word reconstruction from the subwords of fixed length”, *Chebyshevskii sbornik*, vol. 22, no. 1, pp. 57–66.

1. Введение

Символьное кодирование — способ представления информации, при котором данные кодируются в виде слов над определённым алфавитом [13]. Такой способ представления информации используется, например, при передаче данных по каналам связи [11], при работе с временными рядами [14], при анализе генома в биоинформатике [5]. Часто закодированные данные доступны не полностью, а в виде фрагментов. Это может происходить как из-за внешних факторов, например при потере или искажении данных при передаче [12], так и просто в рамках постановки конкретной задачи [3]. Одной из математических абстракций упомянутой проблемы неполной информации является постановка о восстановлении слова по множеству его подслов [2, 9].

В работах [6, 7] получены оценки для длины подслова, достаточной, чтобы данное слово можно было восстановить по мультимножеству всех его подслов. В данной же работе рассматривается частный случай периодического слова, такого что оно состоит из повторяющейся многократно последовательности символов. На основе доказанного другими авторами для случая произвольного слова, выводятся улучшенные оценки для периодического слова. Также рассматривается случай слова, состоящего из непериодического префикса и периодического суффикса, причём длина префикса меньше длины суффикса.

2. Постановка задачи

Греческими буквами обозначаются слова, малыми латинскими — символы алфавита

- E_2^n — множество слов длины n из алфавита $\{0,1\}$
- для слова $\alpha = (a_1a_2\dots a_n) \in E_2^n$ символом $|\alpha|$ обозначается сумма его элементов: $|\alpha| = a_1 + a_2 + \ldots + a_n$
- λ — пустое слово, длина которого равна нулю

Определение 1. *Для заданного слова $\alpha = (a_1a_2\dots a_n) \in E_2^n$ и заданного опорного вектора $v = (v_1v_2\dots v_n)$, $v_i \in \{0,1\}, i = 1\dots n$ операция фрагментирования $\langle\alpha \cdot v\rangle$ строит слово длины $|v|$ по следующему правилу:*

$$\langle\alpha \cdot v\rangle = \begin{cases} a_i, & v_i = 1 \\ \lambda, & v_i = 0 \end{cases}$$

В общем виде рассматриваемая задача формулируется следующим образом: заданы

- набор опорных векторов $V = \{v^1, v^2, \dots, v^N\}, v_i \in E_2^n$, $|v^i| = k, i = 1\dots N$
- и множество слов $X = \{x^1, x^2, \dots, x^N\}, x^i \in E_2^k, i = 1\dots N$

Требуется проверить, является ли набор векторов X набором подпоследовательностей (фрагментов) фиксированной длины некоторого неизвестного слова $\alpha \in E_2^n$, построенных с помощью операции фрагментирования векторами из V, и найти все возможные решения.

Для возникающей задачи известно её сведение к проверке единственности решения диофантовых уравнений определенного вида [1]. Показано, что задача о существовании и единственности решения является NP-полной. При этом все оценки, получаемые для случая двоичного алфавита, остаются верными и для произвольного алфавита [2], так как в случае алфавита $\{0, 1, \dots, k\}$ задача сводится к набору из k задач в двоичном алфавите с помощью отображений:

$$\begin{cases} \phi_i \colon x \to \begin{cases} 1, & x = i \\ 0, & x \neq i \end{cases} \\ i = 0, 1, \dots, k \end{cases}$$

Поэтому далее будут рассмотрены только двоичные случаи.

В случае полного мультимножества фрагментов ($V = E_2^n$) установлено [10], что для однозначного восстановления необходима длина $k \geq \exp\Big(\Omega\Big(\sqrt{\log n}\Big)\Big)$. Для того же случая полного мультимножества фрагментов известны следующие оценки на достаточную длину:

- $k \geq \left\lfloor \frac{n}{2} \right\rfloor$ — показано в [2]
- $k \geq \big(1 + o(1)\big)\sqrt{n \log n}$ — показано в [7]
- $k \geq \left\lfloor \frac{16}{7}\sqrt{n} \right\rfloor + 5$ — показано в [6]

Было изучено несколько частных случаев задачи. Так, для слов, состоящих из l серий, достаточно фрагментов длины l [4].

В данной же работе доказывается, что для однозначного восстановления периодического слова длины n с периодом p по мультимножеству всех его подпоследовательностей длины k достаточно выполнения условия $k \leq \big(1 + o(1)\big)\sqrt{p \log p}$. При этом утверждение остаётся верным, если у слова есть маленький непериодический префикс.

Перейдём к задаче реконструкции при $V = E_2^n$, но с дополнительными ограничениями на допустимые решения.

3. Результаты

Как уже отмечалось, объект дальнейшего анализа — двоичные слова $\alpha = (a_1 a_2 \ldots a_n)$, $a_i \in \{0, 1\}$. Более того, нас будут интересовать *периодические* слова.

Определение 2. *Слово* $\alpha = (a_1 a_2 \ldots a_n), a_i \in \{0, 1\}$ *называется периодическим с периодом* p*, если*

$$\alpha = (a_1 a_2 \ldots a_p)^l, a_i \in \{0, 1\}$$

где $x^s = \underbrace{xx \ldots x}_{s}$.

В [7] получена оценка наименьшей длины фрагментов $k = f^*(n)$, при которой гарантирована однозначность восстановления слова:

$$f^*(n) \leq (1 + o(1))\sqrt{n \log n}$$

что позволит нам доказать следующую теорему.

Теорема 1. *Для однозначного восстановления периодического слова длины* n *с периодом* p *по мультимножеству всех его подпоследовательностей длины* k *достаточно выполнения условия*

$$k \geq (1 + o(1))\sqrt{p \log p}$$

при этом наименьшая длина подпоследовательностей $f^*(n)$*, обеспечивающая однозначное восстановление слова*

$$f^*(n) \leq (1 + o(1))\sqrt{p \log p}$$

Начнём с леммы, которая непосредственно следует из системы уравнений в [1].

Лемма 1. *Для любого слова по набору его фрагментов вида* $x^j 1$ *однозначно определяется набор его моментов вида* $\sum\limits_{r=1}^{n} a_r r^j$.

Перед доказательством леммы введём определение

Определение 3. *За* $N_\beta(\alpha)$ *обозначим число фрагментов, равных* β*, в слове* α.

В [1] показано, что по $N_\beta(\alpha)$ для всех двоичных слов β длины k однозначно восстанавливаются числа фрагментов всех длин меньше k.

Переходим к доказательству леммы.

Доказательство.

$$N_1(\alpha) = \sum_{r=1}^{n} a_r$$

$$N_{x1}(\alpha) = \sum_{r=1}^{n}(r-1)a_r = \sum_{r=1}^{n} ra_r - \underbrace{N_1(\alpha)}_{f_1\left(N_1(\alpha)\right)}$$

$$N_{x^21}(\alpha) = \sum_{r=1}^{n}\binom{r-1}{2}a_r = \frac{1}{2!}\sum_{r=1}^{n} r^2 a_r - \underbrace{\frac{1}{2!}\left(3N_{x1}(\alpha) + N_1(\alpha)\right)}_{f_2\left(N_1(\alpha), N_{x1}(\alpha)\right)}$$

$\ldots$

$$N_{x^{k-1}1}(\alpha) = \sum_{r=1}^{n}\binom{r-1}{k-1}a_r = \frac{1}{(k-1)!}\sum_{r=1}^{n} r^{k-1}a_r - f_{k-1}\Big(N_1(\alpha), N_{x1}(\alpha), \ldots, N_{x^{k-2}1}(\alpha)\Big)$$

Функции $f_2, f_3, \ldots, f_{k-1}$ вычисляются на основе комбинаторных соотношений. Например, найдём выражение функции $f_p(N_1(\alpha), \ldots, N_{x^p1}(\alpha))$, $2 \le p \le k-1$ через полученные на предыдущих шагах $f_1, \ldots, f_{p-1}$ (при этом можно положить $f_0 \equiv 0$):

$$\begin{aligned}
N_{x^p1}(\alpha) &= \sum_{r=1}^{n}\binom{r-1}{p}a_r = \sum_{r=1}^{n}\frac{(r-1)!}{p!(r-1-p)!}a_r = \sum_{r=1}^{n}\frac{(r-1)\cdot(r-2)\cdot\ldots\cdot(r-p)}{p!}a_r \\
&= \sum_{r=1}^{n}\frac{1}{p!}a_r\left\{r^p + r^{p-1}\sum_{i_1=1}^{p}(-i_1) + r^{p-2}\sum_{\substack{i_1,i_2\in\{1,2,\ldots,p\}\\ i_1\neq i_2}}(-i_1)(-i_2) + \ldots + r^0\prod_{i_p=1}^{p}(-i_p)\right\} \\
&= \frac{1}{p!}\sum_{r=1}^{n} r^p a_r + \frac{1}{p!}\sum_{l=1}^{p}\sum_{r=1}^{n} r^{p-l}a_r \sum_{\substack{S\subseteq\{1,\ldots,p\}\\ |S|=l}}\prod_{i\in S}(-i) \\
&= \frac{1}{p!}\sum_{r=1}^{n} r^p a_r + \frac{1}{p!}\sum_{l=1}^{p}\Big(N_{x^{p-l}1}(\alpha) + f_{p-l}\big(N_1(\alpha), \ldots, N_{x^{p-l-1}1}(\alpha)\big)\Big)\sum_{\substack{S\subseteq\{1,\ldots,p\}\\ |S|=l}}\prod_{i\in S}(-i) \\
&= \frac{1}{p!}\sum_{r=1}^{n} r^p a_r + f_p(N_1(\alpha), \ldots, N_{x^{p-1}1}(\alpha))
\end{aligned}$$

Так как, по предположению, все $f_1, \ldots, f_{p-1}$ известны, то известна и $f_p(N_1(\alpha), \ldots, N_{x^p1}(\alpha))$. □

И возвращаемся к доказательству теоремы (1).

Доказательство. Итак, доказательство базируется на доказанном в [7] для получения достаточной для однозначной реконструкции длины фрагменты в случае произвольных слов. Оценка величины $f^*(n)$ в [7] получена на основе анализа системы уравнений

$$\begin{cases}\sum_{r=1}^{n} a_r r^j = s_j(\alpha) \\ 0 \le j \le k-1\end{cases} \tag{1}$$

для которой при $f^*(n) \le (1+o(1))\sqrt{n\log n}$ решение единственно. Пусть слово α состоит из

$l = \frac{n}{p}$ периодов: $\alpha = (a_1 a_2 \dots a_p)^l$. Тогда нулевое уравнение $\sum\limits_{r=1}^{n} a_r = s_0(\alpha)$ можно переписать в виде $\sum\limits_{r=1}^{p} a_r = \frac{s_0(\alpha)}{l}$.

И для произвольного индекса m от 1 до $k-1$ включительно:

$$\begin{aligned}
\sum_{r=1}^{n} a_r r^m &= \sum_{r=1}^{p} a_r r^m + \sum_{r=p+1}^{2p} a_r r^m + \ldots + \sum_{r=(l-1)p+1}^{lp} a_r r^m \\
&= \sum_{r=1}^{p} a_r r^m + \sum_{r=1}^{p} a_r (p+r)^m + \ldots + \sum_{r=1}^{p} a_r \big((l-1)p + r\big)^m \\
&= \sum_{r=1}^{p} a_r r^m + \sum_{r=1}^{p}\sum_{j=0}^{m} \binom{m}{j} p^j r^{m-j} a_r + \ldots + \sum_{r=1}^{p}\sum_{j=0}^{m} \binom{m}{j} \big((l-1)p\big)^j r^{m-j} a_r \\
&= l \cdot \sum_{r=1}^{p} a_r r^m + \sum_{j=1}^{m}\sum_{r=1}^{p} \binom{m}{j} \big((l-1)p\big)^j r^{m-j} a_r \\
&= f_m \left(\sum_{r=1}^{p} a_r r^m, \ldots, \sum_{r=1}^{p} a_r \right)
\end{aligned}$$

где f_m — линейная функция.

И далее доказательство теоремы в точности повторяет доказательство из [7].

□

Сформулируем более сильное утверждение.

Теорема 2. *Для однозначного восстановления периодического слова длины n с периодом p по мультимножеству всех его подпоследовательностей длины k достаточно выполнения условия*

$$k \geq \left\lfloor \frac{16}{7}\sqrt{p} \right\rfloor + 5$$

при этом наименьшая длина подпоследовательностей $f^*(n)$, обеспечивающая однозначное восстановление слова

$$f^*(n) \leq \left\lfloor \frac{16}{7}\sqrt{p} \right\rfloor + 5$$

Доказательство. Доказательство основано на [6], где авторы доказывают возможность однозначной реконструкции слова по его подсловам не из анализа системы уравнений (1), как в [7], но из анализа похожей системы

$$\begin{cases} \sum\limits_{r=1}^{n} a_r r^j = \sum\limits_{r=1}^{n} b_r r^j \\ 0 \leq j \leq k-1 \end{cases} \tag{2}$$

выписанной для двух слов $\alpha = a_1 \dots a_n$ и $\beta = b_1 \dots b_n$. Доказывается, что слова α и β имеют одинаковые мультимножества подслов длины k тогда и только тогда, когда система (2) имеет нетривиальное решение $(a_1, \dots, a_n, b_1, \dots, b_1)$. Таким образом, доказательство сводится к поиску условия, при котором система диофантовых уравнений (2) имеет только тривиальное решение. Авторы [6] ссылаются на [8], где получен результат

$$k \geq \left\lfloor \frac{16}{7}\sqrt{n} \right\rfloor + 5$$

Для случая же периодических слов ранее в данной работе при доказательстве теоремы (1) показано, как систему (2) можно переписать в переменных только $a_1, \dots, a_p, b_1, \dots, b_p$, где p — период слова. Таким образом, оценка на достаточную для однозначного восстановления исходного слова длину подпоследовательности получается равной

$$k \geq \left\lfloor \frac{16}{7}\sqrt{p} \right\rfloor + 5$$

□

Реконструкция слова остаётся возможной и при наличии у него *малого непериодического префикса*. А именно, пусть слово периодическое, начиная с некоторого индекса:

$$\alpha = a_1 a_2 \dots a_q (a_{q+1} a_{q+2} \dots a_{q+p})^l \tag{3}$$

Введём обозначение $P \equiv \max(p, q)$. Тогда можно сформулировать следующие теоремы.

Теорема 3. *При $l \geq q^{\sqrt{P \log P}}$ для однозначного восстановления слова (3) достаточно $k \geq (1 + o(1))\sqrt{P \log P}$*

Теорема 4. *При $l \geq q^{\left\lfloor \frac{16}{7}\sqrt{P} \right\rfloor + 5}$ для однозначного восстановления слова (3) достаточно $k \geq \left\lfloor \frac{16}{7}\sqrt{P} \right\rfloor + 5$*

Обе сформулированные теоремы доказываются схоже: сведением к раннее доказанным теоремам для полностью периодического слова (1) и (2). Способ сведения одинаков, поэтому ограничимся доказательством только теоремы (4).

Доказательство.

$$\sum_{r=1}^{n} a_r = a_a + \dots a_q + l \cdot \sum_{r=q+1}^{q+p} a_r = s_0(\alpha)$$

При $q < l$ верно $\frac{a_1 + \dots a_q}{l} < 1$, поэтому

$$\begin{cases} \displaystyle\sum_{r=q+1}^{q+p} a_r = \left\lfloor \frac{s_0(\alpha)}{l} \right\rfloor \\ \displaystyle\sum_{r=1}^{q} a_r = \left\{ \frac{s_0(\alpha)}{l} \right\} \cdot l \end{cases}$$

Рассмотрим случай $s_1(\alpha)$:

$$\begin{aligned} \sum_{r=1}^{q} a_r r + \sum_{r=q+1}^{n} a_r r &= \sum_{r=1}^{q} a_r r + a_{q+1} \sum_{j=q+1}^{q+p(l-1)+1} j + \dots + a_{q+p} \sum_{j=q+p}^{q+p(l-1)+p} j \\ &= \sum_{r=1}^{q} a_r r + \sum_{i=1}^{p} a_{q+i} \left(l(q+i) + \frac{p(l-1)}{2} \cdot l \right) \\ &= \sum_{r=1}^{q} a_r r + \left(ql + \frac{p(l-1)l}{2} \right) \sum_{r=q+1}^{q+p} a_r + l \sum_{r=q+1}^{q+p} a_r r \\ &= \sum_{r=1}^{q} a_r r + l \cdot \sum_{r=q+1}^{q+p} a_r r + f_1(s_0) \end{aligned}$$

При $\frac{r(r+1)}{2} < l$ уравнения для непериодического префикса и периодического хвоста разделяются аналогично.

Перейдём к случаю произвольного $m \in [2, k-1] \cap \mathbb{N}$:

$$\begin{aligned}\sum_{r=1}^{q} a_r r^m + \sum_{r=q+1}^{n} a_r r^m &= \sum_{r=1}^{q} a_r r^m + \sum_{r=q+1}^{q+p} a_r \big((q+r)^m + \ldots + (q+p(l-1)+r)^m\big)\\ &= \sum_{r=1}^{q} a_r r^m + \sum_{r=q+1}^{q+p} a_r \sum_{h=1}^{l}\sum_{j=0}^{m}\binom{m}{j}(q+r)^j\big(p(l-h)\big)^{m-j}\Big)\\ &= \sum_{r=1}^{q} a_r r^m + \sum_{r=q+1}^{q+p} a_r \sum_{h=1}^{l}\sum_{j=0}^{m}\sum_{s=0}^{j}\binom{m}{j}\binom{j}{s} q^s r^{j-s}\big(p(l-h)\big)^{m-j}\Big)\end{aligned}$$

Нас интересует показатель степени при r во втором слагаемом равный $j - s = m$. Откуда получаем $j = m$ и $s = 0$. При меньших показателях степени будем получать члены, уже найденные из предыдущих уравнений. Таким образом,

$$\sum_{r=1}^{q} a_r r^m + \sum_{r=q+1}^{n} a_r r^m = \sum_{r=1}^{q} a_r r^m + l \cdot \sum_{r=q+1}^{q+p} a_r r^m + f_m(s_{m-1}, \ldots, s_0)$$

где f_m — линейная функция.

Так как $\sum\limits_{k=1}^{n} k^p \to \frac{n^{p+1}}{p+1}$, то при $\frac{q^k}{k} < l$ разделяются все уравнения и получается две системы. Поэтому, если ввести обозначение $P \equiv \max(p, q)$, из доказанного ранее для случая периодического слова, при

$$l \geq q^{\left\lfloor \frac{16}{7}\sqrt{P}\right\rfloor + 5}$$

размер подпоследовательностей ограничивается

$$k \geq \left\lfloor \frac{16}{7}\sqrt{P}\right\rfloor + 5$$

□

4. Заключение

В работе получены оценки на длину подслова, достаточную для однозначного восстановления периодического слова по мультимножеству его подслов фиксированной длины. Рассмотрен случай слова, состоящего из непериодического префикса и периодического суффикса. Для таких слов также доказана возможность однозначного восстановления, но при достаточной длине суффикса.

СПИСОК ЦИТИРОВАННОЙ ЛИТЕРАТУРЫ

1. Леонтьев В. К., Сметанин Ю. Г. О восстановлении векторов по набору их фрагментов //Доклады Академии наук. – Российская академия наук, 1988. – Т. 302. – №. 6. – С. 1319-1322.

2. Manvel B. et al. Reconstruction of sequences //Discrete Mathematics. – 1991. – Т. 94. – №. 3. – С. 209-219.

3. Skiena S. S., Sundaram G. Reconstructing strings from substrings //Journal of Computational Biology. – 1995. – Т. 2. – №. 2. – С. 333-353.

4. Леонтьев В. К. Задачи восстановления слов по фрагментам и их приложения //Дискретный анализ и исследование операций. – 1995. – Т. 2. – №. 2. – С. 26-48.

5. Gusfield D. Algorithms on stings, trees, and sequences: Computer science and computational biology //Acm Sigact News. – 1997. – Т. 28. – №. 4. – С. 41-60.

6. Krasikov I., Roditty Y. On a reconstruction problem for sequences //journal of combinatorial theory, Series A. – 1997. – Т. 77. – №. 2. – С. 344-348.

7. Scott A. D. Reconstructing sequences //Discrete Mathematics. – 1997. – Т. 175. – №. 1-3. – С. 231-238.

8. Borwein P., Erdélyi T., Kós G. Littlewood-type problems on [0, 1] //Proceedings of the London Mathematical Society. – 1999. – Т. 79. – №. 1. – С. 22-46.

9. Levenshtein V. I. Efficient reconstruction of sequences from their subsequences or supersequences //Journal of Combinatorial Theory, Series A. – 2001. – Т. 93. – №. 2. – С. 310-332.

10. Dudık M., Schulman L. J. Reconstruction from subsequences //Journal of Combinatorial Theory, Series A. – 2003. – Т. 103. – №. 2. – С. 337-348.

11. Batu T. et al. Reconstructing strings from random traces //Departmental Papers (CIS). – 2004. – С. 173.

12. Kannan S., McGregor A. More on reconstructing strings from random traces: insertions and deletions //Proceedings. International Symposium on Information Theory, 2005. ISIT 2005. – IEEE, 2005. – С. 297-301.

13. Lin J. et al. Experiencing SAX: a novel symbolic representation of time series //Data Mining and knowledge discovery. – 2007. – Т. 15. – №. 2. – С. 107-144.

14. Smetanin Y., Ul'yanov M. Determining the characteristics of kolmogorov complexity of time series: an approach based on symbolic descriptions. – 2013.

REFERENCES

1. Leontiev, V. K., & Smetanin, Y. G. 1988, "On the reconstruction of vectors from a set of their fragments", *Doklady Academii nauk*, vol. 302, no. 6, pp. 1319-1322.

2. Manvel, B., Meyerowitz, A., Schwenk, A., Smith, K., & Stockmeyer, P. 1991, "Reconstruction of sequences", *Discrete Mathematics*, 94(3), 209-219.

3. Skiena, S. S., & Sundaram, G. 1995, "Reconstructing strings from substrings", *Journal of Computational Biology*, 2(2), 333-353.

4. Leontiev, V. K. 1995, "Words reconstruction by fragments problems and their applications", *Diskretniy analiz i issledovanie operaciy*, 2(2), 26-48.

5. Gusfield, D. 1997, "Algorithms on stings, trees, and sequences: Computer science and computational biology", *Acm Sigact News*, 28(4), 41-60.

6. Krasikov, I., & Roditty, Y. 1997, "On a reconstruction problem for sequences", *Journal of Combinatorial Theory*, Series A, 77(2), 344-348.

7. Scott, A. D. 1997, "Reconstructing sequences", *Discrete Mathematics*, 175(1-3), 231-238.

8. Borwein, P., Erdélyi, T., & Kós, G. 1999, “Littlewood-type problems on [0, 1]”, *Proceedings of the London Mathematical Society*, 79(1), 22-46.

9. Levenshtein, V. I. 2001, “Efficient reconstruction of sequences from their subsequences or supersequences”, *Journal of Combinatorial Theory*, Series A, 93(2), 310-332.

10. Dudık, M., & Schulman, L. J. 2003, “Reconstruction from subsequences”, *Journal of Combinatorial Theory*, Series A, 103(2), 337-348.

11. Batu, T., Kannan, S., Khanna, S., & McGregor, A. 2004, “Reconstructing strings from random traces”, *Departmental Papers (CIS)*, 173.

12. Kannan, S., & McGregor, A. 2005, “More on reconstructing strings from random traces: insertions and deletions”, *Proceedings. International Symposium on Information Theory, 2005 (ISIT 2005)*, pp. 297-301, IEEE.

13. Lin, J., Keogh, E., Wei, L., & Lonardi, S. 2007, “Experiencing SAX: a novel symbolic representation of time series”, *Data Mining and knowledge discovery*, 15(2), 107-144.

14. Smetanin, Y. G., & Ul’yanov, M. V. 2013, “Determining the characteristics of kolmogorov complexity of time series: an approach based on symbolic descriptions”.

Получено 12.12.2020 г.
Принято в печать 21.02.2021 г.