НОВЫЕ СРЕДСТВА
КИБЕРНЕТИКИ, ИНФОРМАТИКИ,
ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ
И СИСТЕМНОГО АНАЛИЗА

Ю.Г. СМЕТАНИН, М.В. УЛЬЯНОВ

УДК 519.16, 519.17

РЕКОНСТРУКЦИЯ СЛОВ ПО КОНЕЧНОМУ МУЛЬТИМНОЖЕСТВУ ПОДСЛОВ В ГИПОТЕЗЕ СДВИГА 1. II. РЕКОНСТРУКЦИЯ ПРИ НАЛИЧИИ ЗАПРЕЩЕННОГО СЛОВА[1]

Аннотация. Рассматривается расширение задачи реконструкции слов по заданному мультимножеству подслов, предположительно порожденных смещением окна фиксированной длины со сдвигом 1. Это связано с наличием дополнительных ограничений на допустимые решения. Изучен случай, когда эти ограничения определяются запрещенными словами. Получено решение задачи, основанное на поиске эйлеровых путей в мультиорграфе де Брейна с дополнительной операцией редукции ребер и применением специальных алгебраических операций умножения матриц смежности, определенных в первой части статьи.

Ключевые слова: реконструкция слов, запрещенное слово, эйлеровы пути, мультиорграф де Брейна, комбинаторика слов.

ВВЕДЕНИЕ

В первой части настоящей статьи [1] был предложен набор алгебраических операций над словами, в терминах которых сформулирована задача реконструкции слов над конечным алфавитом по подсловам этого слова. Задача реконструкции была сведена к задаче поиска эйлеровых путей (циклов) в мультиорграфе де Брейна. Предложен метод перечисления всех решений задачи реконструкции на основе возведения в степень матрицы смежности мультиорграфа де Брейна с использованием специально введенных операций.

В дополнение к приведенным в [1] результатам отметим, что из предложенного сведения задачи реконструкции без запретов к поиску эйлеровых путей в мультиорграфе де Брейна очевидно следует однозначность реконструкции в случае, если полустепень выхода каждой вершины не превосходит единицы. Из этого, в свою очередь, следует представленный A. Carpi и A. de Luca в [2] совершенно иными методами результат о достаточной длине подслов для однозначной реконструкции.

В настоящей статье рассматривается задача реконструкции по подсловам, когда для допустимых решений существуют дополнительные ограничения. Наибольший практический интерес представляет случай, когда реконструируемое слово не содержит заранее заданного набора запрещенных слов. Этот случай соответствует описанию последовательностей в символической динамике [3], где основным понятием является пространство сдвигов (shift space), определяемое наборами запрещенных слов. В других терминах рассматриваемую задачу можно считать задачей восстановления слов из языка, определяемого набором запрещенных слов.

Один из возможных вариантов решения задачи реконструкции на основе подслов фиксированной длины сдвига 1 при единственном запрещенном слове составляет предмет исследования настоящей работы, представляющей вторую

[1]Работа выполнена при поддержке РФФИ, грант № 13-07-00516.

Окончание. Начало в № 1, 2014.

часть статьи. Для решения задачи в этой постановке предлагается использовать методику, описанную в первой части. Для частного случая этой задачи разработан метод редукции мультиорграфа де Брейна, позволяющий проверять наличие решения с помощью поиска эйлеровых путей. Описываемый метод естественно распространяется на случай нескольких запрещенных слов, а также на задачи реконструкции пространств сдвигов конечного типа в символической динамике.

ТЕРМИНОЛОГИЯ И ОБОЗНАЧЕНИЯ

В тексте настоящей статьи используются как общие обозначения, связанные с реконструкцией слов, введенные в первой части статьи, так и обозначения, отражающие специфику данной части, состоящую в наличии запрещенных слов. Тем не менее в целях логической ясности изложения приведем все используемые обозначения полностью:

• $\Sigma = \{0, 1\}$ — бинарный алфавит, s — произвольный символ алфавита;

• $\Sigma^0 = \varnothing$ — пустое множество;

• Σ^k — k-я декартова степень множества Σ (k-элементный кортеж);

• $\Sigma^* = \bigcup\limits_{k=0}^{\infty} \Sigma^k$ — транзитивное замыкание Σ (множество всех возможных кортежей);

• c — произвольный кортеж (для кортежа c определим длину — число составляющих его элементов как мощность мультимножества, образованного из кортежа $|c| = n$, например $|(0,1,1,0)| = 4$. Длина пустого кортежа равна нулю);

• $R(c, i)$ — функция выбора элемента кортежа, определенная при $1 \le i \le |c|$ и возвращающая элемент с номером i из кортежа c; если $i > |c|$, то $R(c, i) = \varnothing$; если элементы кортежа принадлежат Σ, то функция возвращает символ алфавита, например $R((1,0,0,1),3) = 0$;

• w — слово (над алфавитом) — последовательность символов алфавита, при этом собственно символы алфавита есть слова по определению;

• λ — пустое (не содержащее символов) слово;

• знак + представляет операцию конкатенации (склейки) слов: $w_1 + w_2 = w_1 w_2$; результат операции + есть слово, представляющее последовательность символов слова w_1, за которой следует последовательность символов слова w_2, например $01 + 11 = 0111$;

• $D(\cdot)$: $D(c) = w$, где $c \in \Sigma^*$ — кортеж, w — слово; оператор $D(\cdot)$ — деструктор кортежа, представляющий оператор создания слова из кортежа путем конкатенации символов алфавита Σ, $D(c) = w = R(c,1) + R(c,2) + \ldots + R(c, |c|)$, например $D((1,1,0,1)) = 1101$;

• $L(\cdot)$: $L(C) = W$, где $C \subseteq \Sigma^*$ — множество кортежей, W — множество слов; $L(\cdot)$ — оператор создания множества слов, состоящих из символов алфавита Σ, который действует на множество кортежей посредством последовательного применения оператора $D(\cdot)$,

$$L(C) = W = \{w \mid \forall c \in C \ w = D(c)\},$$

например $L(\{(1,1,0,1), (0,1,1)\}) = \{1101, 011\}$;

• $w = s_1 s_2 \ldots s_k \in L(\Sigma^k)$ — произвольное слово из множества $L(\Sigma^k)$ над алфавитом Σ;

• $|w| = k = |L^{-1}(w)|$ — длина слова, определяемая как число элементов в кортеже, порождающем это слово;

• $L_k = L(\Sigma^k) = \{w \mid |w| = k\}$ — множество всех слов длины k над алфавитом Σ;

• $L^* = L(\Sigma^*)$ — полный язык над алфавитом Σ, т.е. множество всех возможных слов;

• $L \subset L^*$ — произвольный неполный язык над алфавитом Σ;

• если $w = s_1 s_2 \ldots s_n \in L(\Sigma^n)$, то при $k < n$ имеем $v = s_{i_1} s_{i_2} \ldots s_{i_k}$, $1 \le i_1$, $i_2 = i_1 + 1, \ldots, i_k = i_{k-1} + 1 \le n$, — подслово длины k слова w;

• $Q(w, i, l)$ — оператор выделения подслова длины l в слове w, начиная с символа в позиции i; если $|w| = n$, то оператор определен при $i + l - 1 \le n$:

$$Q(s_1 s_2 \ldots s_n, i, l) = u = s_i s_{i+1} \ldots s_{i+l-1};$$

• $P(w) = Q(w, 1, k-1) = s_1 s_2 \ldots s_{k-1} \in L(\Sigma^{k-1})$ — полный префикс длины $|w| - 1$ слова w, $|w| = k \ge 2$;

• $S(w) = Q(w, 2, k-1) = s_2 \ldots s_k \in L(\Sigma^{k-1})$ — полный суффикс длины $|w| - 1$ слова w, $|w| = k \ge 2$;

• $Sn(w) = Q(w, k, 1) = s_k \in L(\Sigma^1)$ — суффикс слова w длины 1 — символ алфавита, $|w| = k \ge 2$;

• $V(L_k, m)$ — оператор выборки: его результат представляет произвольное подмножество (возможно, с повторениями) из m слов множества L_k:

$$V(L_k, m) = \{ v_i \mid i = \overline{1, m}; v_i = s_1^{(i)} s_2^{(i)} \ldots s_k^{(i)} \in L_k \};$$

заметим, что ввиду особенностей задачи допускаем рассмотрение $V(L_k, m)$ как мультимножества с кратностями элементов; в этом случае m — сумма кратностей элементов;

• $SH1(w, k)$ — оператор сдвига 1; оператор, определенный при $|w| > k$, порождает множество подслов длины k мощности $|w| - k + 1$, выполняя сдвиг на единицу окна длины k по слову w, начиная с крайней левой позиции слова w:

$$SH1(w, k) = \{ u_j \mid j = 1, |w| - k + 1;\ u_j = Q(w, j, k) \};$$

для оператора $SH1(w, k)$ допускаем создание мультимножества, например

$$SH1(1101010, 4) = \{1101, 1010, 0101, 1010\} = \{1101, 1010^{(2)}, 0101\};$$

• ω — запрещенное слово; $w \setminus \omega$ — операция «коллапса» — при очевидном условии $|w| \ge |\omega|$ результатом операции является или слово w, если оно не содержит запрещенного подслова ω, или λ, если слово w содержит ω:

$$w \setminus \omega = \begin{cases} w & \forall j = 1, |w| - |\omega| + 1: \omega \ne Q(w, j, |\omega|), \\ \lambda, & \exists j: \ \omega = Q(w, j, |\omega|); \end{cases}$$

• $\Omega(L, \omega)$ — оператор редукции языка L по запрету ω.

Допускаем, что язык L может быть мультимножеством, оператор $\Omega(L, \omega)$, действуя на язык L, оставляет только те слова из языка L, которые не содержат запрещенного слова ω:

$$\Omega(L, \omega) = \{ u \mid u = w \setminus \omega \ \ \forall w \in L \},$$

обозначим $L \setminus \omega$ язык L, редуцированный по запрету ω; таким образом, $L \setminus \omega = \Omega(L, \omega)$.

ПОСТАНОВКА ЗАДАЧИ РЕКОНСТРУКЦИИ ПРИ НАЛИЧИИ ЗАПРЕЩЕННОГО СЛОВА

Как и в [1], считаем заданными длину подслова k, число подслов m и исходное мультимножество слов $V(L_k, m)$ над алфавитом $\Sigma = \{0, 1\}$, рассматриваемое как базис реконструкции, т.е. как мультимножество подслов сдвига 1 относительно некоторого неизвестного слова w. Далее, в постановке 2, задано единственное запрещенное слово ω для реконструируемых слов w. Отметим, что формулировка постановки 2 тесно связана с постановкой 1.

Постановка 1. Задача реконструкции без запретов.

Содержательно. В условиях гипотезы сдвига 1 относительно мультимножества $V(L_k, m)$ возможно ли выполнить реконструкцию слова w в принципе, и если эта задача имеет решение, то какой вид имеют возможные реконструкции?

Формально. Введем в рассмотрение множество

$$W = \{w \,\big|\, |w| = m + k - 1,\ V(L_k, m) = SH1(w, k)\}\,; \tag{1}$$

при этом выражение (1) понимется как равенство мультимножеств (равны как элементы, так и их кратности). Тогда если $|W| = 0$, то решения нет (реконструкция слов невозможна); если $|W| \geq 1$, то решение есть (реконструкция возможна).

В дальнейшем возможные решения задачи в постановке 1 будем называть простыми решениями задачи реконструкции.

Постановка 2. Задача реконструкции с запретом.

Содержательно. В условиях гипотезы сдвига 1 относительно мультимножества $V(L_k, m)$ возможна ли в принципе реконструкция такого слова или слов w, которые не содержат слова ω в качестве подслова, и какой вид имеют эти реконструируемые слова?

Формально. Введем в рассмотрение аналогично постановке 1 множество $W \setminus \omega = \Omega(W, \omega)$, где W определено в соответствии с (1). Тогда, рассматривая совместно множества W и $W \setminus \omega$, выделим следующие случаи.

Случай 1. $|W| = 0$ — множество простых решений задачи реконструкции пусто, и очевидно, что $|W \setminus \omega| = 0$, тогда множество $V(L_k, m)$ в принципе не является реконструирующим.

Случай 2. $|W| \geq 1$ и $|W \setminus \omega| = 0$ — решения в постановке 2 нет, все простые реконструкции содержат запрещенное слово. Таким образом, реконструкция на базе $V(L_k, m)$ без запрещенного слова ω невозможна.

Случай 3. $|W| \geq 1$ и $|W| = |W \setminus \omega|$ — решение в постановке 2 существует, и все простые реконструкции не содержат запрещенного слова; тогда мультимножество $V(L_k, m)$ доставляет полное решение задачи реконструкции в постановке 2, которое совпадает с простым решением в постановке 1. Заметим, что этот случай эквивалентен реконструкции из пространства сдвигов, определяемого запрещенным словом.

Случай 4. $|W| \geq 1$ и $|W| > |W \setminus \omega|$ — решение в постановке 2 существует, но есть такие простые реконструкции, которые содержат запрещенное слово; тогда мультимножество $V(L_k, m)$ доставляет частичное решение задачи реконструкции в постановке 2. В этом случае представляет интерес получение как точного числа решений $M_\omega = |W \setminus \omega|$, так и самих решений задачи — множества слов $W \setminus \omega$.

ОСОБЕННОСТИ ПОСТАНОВКИ 2 И ТРИВИАЛЬНЫЕ СЛУЧАИ

Особенности постановки 2 связаны не только с наличием самого запрещенного слова ω, но и с соотношениями между его длиной и длинами реконструирующих и реконструируемых слов. В целях изложения этих особенностей обозначим n длину запрещенного слова ω и отметим, что длины всех подслов в мультимножестве $V(L_k, m)$ равны k, а длина реконструируемых слов $w \in W$, если $V(L_k, m)$ является реконструирующим подмножеством, равна $r = |w| = m + k - 1$.

В зависимости от соотношений между указанными длинами возникают варианты, приводящие к различным случаям решения в постановке 2. Далее под тривиальными случаями будем понимать такие решения, которые выполняются с незначительными вычислительными затратами, т.е. с малой трудоемкостью. При этом полное исследование множества W не считаем тривиальным случаем, так как само получение множества реконструируемых слов связано с перечислением всех эйлеровых путей на основе специального символьного возведения в степень матрицы стоимостей, подробно описанного в [1].

Рассмотрим в развернутом виде тривиальный случай и варианты, возникающие в постановке 2.

Пусть $|W| = 0$ — решения задачи в постановке 1 не существует и, очевидно, нет решений в постановке 2. Этот случай элементарно идентифицируется проверкой существования эйлерова пути (цикла), полученном в постановке I мультиорграфа де Брейна [4]. Наличие эйлерова цикла (или пути) устанавливается через подсчет полустепеней захода и выхода вершин мультиорграфа ([5], с. 145) и требует не более чем однократного прохода по матрице смежности мультиорграфа де Брейна. Сложность такой проверки оценивается как $O(m^2)$.

Далее рассмотрим варианты, возникающие при наличии простой реконструкции, т.е. наличии эйлерова пути в мультиорграфе де Брейна, обеспечивающего выполнение условия $|W| \geq 1$ (случаи 2-4). Эти варианты идентифицируются следующими соотношениями между значениями длин n, k и r.

Вариант А. Длина n запрещенного слова больше длины r реконструируемого слова $(n > r)$. Тогда получаем вариант случая 3: $|W| \geq 1$ и $|W| = |W \setminus \omega|$ — все возможные реконструкции не содержат запрещенного слова ω, и решение для постановки 2 доставляется простой реконструкцией. Вариант идентифицируется за $O(1)$ операций.

Вариант Б. Длина n запрещенного слова меньше длины k слов $(n < k)$ из исходного мультимножества $V(L_k, m)$. Для исследования этого варианта рассмотрим множество из подслов длины n, полученных сдвигом 1 по всем словам из $V(L_k, m)$:

$$U = \bigcup_{i=1}^{m} SH1(v_i, n), v_i \in V(L_k, m).$$

Тогда если $\omega \in U$, то имеет место случай 2: $|W| \geq 1$, $|W \setminus \omega| = 0$ — решения в постановке 2 нет, все простые реконструкции содержат запрещенное слово, поскольку оно является подсловом одного из слов исходного множества $V(L_k, m)$; если $\omega \notin U$, то $|W| = |W \setminus \omega|$ и получаем вариант случая 3: $|W| \geq 1$, $|W| = |W \setminus \omega|$ — все возможные простые реконструкции не содержат запрещенного слова и множество W доставляет решение данной задачи в постановке 2.

Сложность проверки варианта Б определяется сложностью построения множества U, которая оценивается как $O(kmn - mn^2 + mn)$ операций над символами, поскольку сдвиг необходимо выполнить для всех m слов из $V(L_k, m)$ и для каждого из них получить $k - n + 1$ подслов сдвига 1, имеющих длину n.

Вариант В. Длина n запрещенного слова равна длине k подслов $(n = k)$ из $V(L_k, m)$; тогда проверяем, является ли ω элементом мультимножества $V(L_k, m)$. Если $\omega \in V(L_k, m)$, то имеет место случай 2: $|W| \geq 1$, $|W \setminus \omega| = 0$ — решения в постановке 2 нет, все простые реконструкции содержат запрещенное слово, поскольку оно является одним из слов исходного мультимножества $V(L_k, m)$. Если $\omega \notin V(L_k, m)$, то $|W| = |W \setminus \omega|$ и получаем вариант случая 3: $|W| \geq 1$ и $|W| = |W \setminus \omega|$ — все возможные реконструкции не содержат запрещенного слова и множество W определяет решение для постановки 2.

Трудоемкость проверки варианта В определяется сложностью проверки вхождения ω в мультимножество $V(L_k, m)$, т.е. составляет $O(mn)$ символьных операций.

РЕШЕНИЕ ПОСТАНОВКИ 2 ПЕРЕЧИСЛЕНИЕМ ВСЕХ ЭЙЛЕРОВЫХ ПУТЕЙ

Соотношение длин $k < n \leq r$ приводит к нетривиальному варианту постановки 2. Очевидный путь исследования для этого варианта — решение постановки 1 с последующей проверкой вхождения запрещенного слова в множество слов реконструкции. Фактически это переборное решение для постановки 2 на основе решения в постановке 1 включает три этапа.

1. Решение постановки 1 путем перечисления всех эйлеровых путей на основе специального символьного возведения в степень матрицы смежности, описанного в первой части статьи. В результате получаем множество слов W, являющихся простыми реконструкциями без запретов.

2. Построение множества слов, не содержащих запрещенного слова, путем применения оператора Ω (оператора редукции языка) с запретом ω к мно-

жеству W, т.е. проверки вхождения запрещенного слова в множество реконструкции. В результате получаем множество реконструкций без запрещенного слова: $W \setminus \omega = \Omega(W, \omega)$.

3. Идентификация всех возможных четырех случаев в постановке 2 задачи реконструкции при наличии запрещенного слова путем сравнения мощности множеств W и $W \setminus \omega$.

В рамках такого решения выделим вариант, когда $n = r$, т.е. длина n запрещенного слова равна длине r реконструируемого слова, и проверяем, является ли ω элементом множества простых реконструкций W. При этом если $\omega \in W$, то имеет место случай 4: $|W| \geq 1$ и $|W| > |W \setminus \omega|$ — мультимножество $V(L_k, m)$ доставляет частичное решение задачи реконструкции в постановке 2. Особенность данного варианта состоит в том, что в множестве W есть только одно запрещенное слово ω, и следовательно $M_\omega = |W| - 1$. Если $\omega \notin W$, то $|W| = |W \setminus \omega|$ и получаем вариант случая 3: $|W| \geq 1$ и $|W| = |W \setminus \omega|$ — все возможные реконструкции не содержат запрещенного слова и множество W определяет решение для постановки 2.

При определенных условиях решение для постановки 2 может быть получено с меньшей трудоемкостью, чем через прямое решение задачи реконструкции без запретов. Вариант такого частичного решения, основанный на специальной процедуре редукции мультиорграфа де Брейна, представлен ниже.

РЕШЕНИЕ В ЧАСТНЫХ СЛУЧАЯХ НА ОСНОВЕ РЕДУКЦИИ ГРАФА ДЕ БРЕЙНА

Рассмотрим нетривиальный вариант $|W| \geq 1$ при соотношении длин $k < n \leq r$ и введем специальную процедуру редукции мультиорграфа де Брейна $G = (D, H)$, построенного на основе мультимножества $V(L_k, m)$. Отметим, что дуги h_i в G описываются пятеркой $h_i = (d_j, d_l, e_i, r_i, vp_i)$, где d_j — начальная вершина дуги, d_l — конечная вершина дуги, e_i — символическое имя дуги, r_i — кратность, vp_i — значение слова из $V(L_k, m)$.

Процедура редукции включает следующие два этапа.

Этап 1. *Поиск запрещенного слова в мультиорграфе де Брейна.*

1. Строим упорядоченное мультимножество Q подслов длины k из запрещенного слова посредством оператора сдвига 1:

$$Q = SH1(\omega, k) = \{q_i \mid i = 1, \ n - k + 1\}.$$

2. Полагаем $i = 1$ и присваиваем кортежу t значение пусто.

3. Ищем в $G = (D, H)$ дугу h_j, для которой $R(h_j, 5) = q_i$, т.е. дугу, значение которой совпадает с i-м сдвигом по запрещенному слову. Если такой дуги нет, то имеем вариант случая 3: $|W| \geq 1$ и $|W| = |W \setminus \omega|$, и решаем задачу в постановке 1. В этом случае этап 1 закончен. Если такая дуга есть, то добавляем в кортеж t элемент h_j: $t = t \times h_j$ и продолжаем этап 1.

4. Увеличиваем значение $i = i + 1$ и переходим к шагу 3 для всех $i \leq n - k + 1$.

5. Путь по дугам, описываемый кортежем t, порождает запрещенное слово.

Этап 2. *Собственно редукция мультиорграфа* $G = (D, H)$.

По мультиорграфу $G = (D, H)$ строим редуцированный мультиорграф $G' = (D, H')$ следующим образом.

1. По всем элементам кортежа t, т.е. по всем $h_i, i = 1, n - k + 1,$ (нумерация в порядке построения кортежа) выполняем редукцию ребер:

$$h_i = (d_j, d_l, e_i, r_i, vp_i) \to h'_i = (d_j, d_l, e_i, r_i - 1, vp_i);$$

если $r_i - 1 = 0$, то считаем, что дуга h_i удалена из G'.

2. Добавляем в граф $G' = (D, H')$ новую дугу, имеющую значение запрещенного слова и соединяющую начальную и конечную вершины найденного пути:

$$H' = H' \cup h^*, \ h^* = (R(R(t, 1), 1), R(R(t, n - k + 1), 2), e^*, 1, \omega).$$

Конец процедуры.

Заметим, что сама процедура редукции имеет малую трудоемкость, поскольку на этапе 1 происходит просмотр всех m дуг в $G=(D,H)$ для поиска подслова q_1 длины n и не более двух дуг на всех последующих $n-k$ шагах, это дает оценку сложности $O(mn+2n^2-2kn)$ для этапа 1, а этап 2, что очевидно, не увеличивает эту сложность.

Исследование редуцированного мультиорграфа $G'=(D,H')$ позволяет получить некоторую информацию о решениях постановки 2. Отметим, что найденный в процедуре редукции путь, отражающий запрещенное слово, еще не является гарантией того, что эйлеров путь (цикл) проходит именно в такой последовательности по этим дугам. Поэтому необходимо исследование наличия эйлерового пути (цикла) в $G'=(D,H')$.

Наличие в $G'=(D,H')$ эйлерова пути (эйлерова цикла) определяет существование решения задачи реконструкции, включающего запрещенное слово. Поскольку при редукции для всех вершин сохраняется соотношение между полустепенями захода и выхода вершин, то исходный эйлеров мультиорграф $G=(D,H)$, преобразованный процедурой редукции в $G'=(D,H')$, либо остается графом с эйлеровым путем (циклом) и, возможно, содержит изолированные вершины, либо становится несвязным графом с несколькими компонентами, каждая из которых является эйлеровым графом.

В первом случае задача реконструкции имеет решения, включающие запрещенное слово, при этом неизвестно — есть ли решения, не содержащие запрета; для ответа на этот вопрос необходимо полное исследование через решение в постановке 1.

Во втором случае задача реконструкции не имеет решений, содержащих запрещенное слово: $|W|\geq 1$ и $|W|=|W\setminus\omega|$. Особо отметим, что этот случай невозможен, если кратность всех дуг в пути, соответствующем запрещенному слову в исходном мультиорграфе $G=(D,H)$, больше или равна двум. При этом условии нет необходимости проверять наличие эйлерова пути (цикла) в редуцированном графе, поскольку это очевидно.

Таким образом, предложенная процедура редукции позволяет с малой трудоемкостью идентифицировать частный случай отсутствия решений, содержащих запрещенное слово, что представляет, в частности, интерес относительно символической динамике [3] и ряде приложений [6–10].

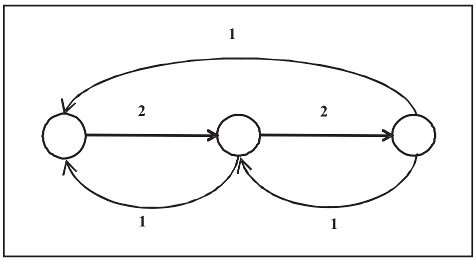

Рис. 1. Исходный мультиорграф де Брейна — дуги запрещенного слова имеют кратность два

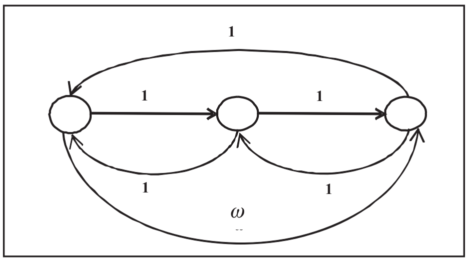

Рис. 2. Редуцированный мультиорграф де Брейна — наличие эйлерового цикла приводит к решениям, включающим запрещенное слово

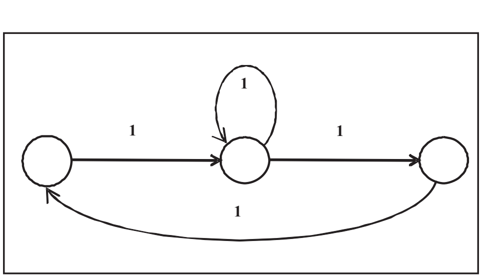

Рис. 3. Исходный мультиорграф де Брейна — дуги запрещенного слова имеют кратность один

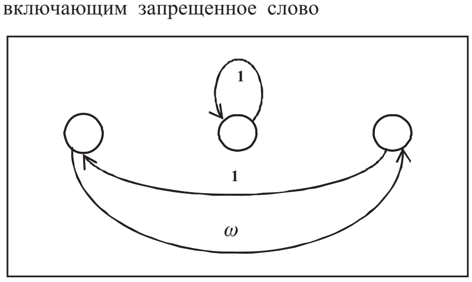

Рис. 4. Редуцированный мультиорграф де Брейна не имеет эйлеровых путей (циклов) — нет реконструкций, включающих запрещенное слово

Проиллюстрируем процедуру редукции, рассматривая только структуру графов с указанием кратности дуг. Жирным выделены дуги, представляющие запрещенное слово. Цифры над дугами соответствуют кратности вхождения подслов в слово. На рис. 1 и 2 приведен случай, когда в мультиорграфе $G' = (D, H')$ после редукции сохраняется эйлеров цикл и, следовательно, есть реконструкции, содержащие запрещенное слово. На рис. 3 и 4 приводится случай, когда в редуцированном графе пропадают эйлеровы пути (циклы) и граф становится несвязным — в решении постановки 2 нет реконструкций с запрещенным словом.

ЗАКЛЮЧЕНИЕ

В настоящей статье рассмотрена постановка 2 задачи реконструкции по заданному мультимножеству подслов, предположительно порожденных смещением окна фиксированной длины со сдвигом 1 по некоторому слову. Особенность этой постановки состоит в наличии запрещенного слова в реконструкции. Принятый в первой части операторный подход расширен на описание операций с запрещенными словами. Рассмотрены четыре возможных случая решения задачи реконструкции в постановке 2, специально выделены и исследованы тривиальные случаи. Предложен вариант полного решения задачи, основанный на решении постановки 1 путем перечисления всех эйлеровых путей (циклов) в мультиорграфе де Брейна с применением специальных алгебраических операций умножения матриц смежности, определенных в первой части статьи. Дополнительно предложен вариант частичного решения, основанный на введенной авторами процедуре редукции и последующем исследовании редуцированного мультиорграфа де Брейна. Рассмотренный метод позволяет эффективно с учетом трудоемкости идентифицировать случай отсутствия запрещенного слова в множестве слов реконструкции. Этот случай представляет определенный интерес, поскольку основным понятием является пространство сдвигов в символической динамике. В этом аспекте отметим, что для задачи проверки, определяет ли данное мультимножество подслов пространство сдвигов, заданное определенными запретами, предложенная процедура редукции и последующего исследования мультиорграфа доставляет полное решение.

СПИСОК ЛИТЕРАТУРЫ

1. Сметанин Ю.Г., Ульянов М.В. Реконструкция слов по конечному мультимножеству подслов в гипотезе сдвига 1. I. Реконструкция без запретов // Кибернетика и системный анализ. — 2014. — № 1. — С. 168–177.
2. Carpi A., de Luca A. Words and special factors // Theoretical Computer Sci. — 2001. — **259**. — P. 145–182.
3. Lind D., Marcus B. An introduction to symbolic dynamics and coding. — Cambridge, UK: Cambridge Univ. Press, 1995. — 495 p.
4. de Bruijn N.G. A combinatorial problem // Nederl. Akad. Wetensch. Proc. 49. — P. 758–764; Indagoniones Math. — 1946. — **8**. — P. 461–467.
5. Шапорев С.Д. Дискретная математика. Курс лекций. — СПб.: БХВ-Петербург, 2006. — 400 с.
6. Lothaire M. Combinatorics of words. Encyclopedia of mathematics and its applications. — Reading, Mass.: Addison-Wesley Publ. Co. — 1983. — **17**. — 228 p. — http://www-igm.univ-mlv.fr/~berstel/Lothaire/.
7. Lothaire M. Algebraic combinatorics on words. — Cambridge, UK: Cambridge Univ. press, 2002. — 455 p. — http://www-igm.univ-mlv.fr/~berstel/Lothaire/.
8. Lothaire M. Applied combinatorics on words, 2005. — 610 c.
9. Dreyer W., Kotz Dittrich A., Schmidt D. Research perspectives for time series management systems // SIGMOD Record. — 1994. — **23**, N 1. — P. 10–15.
10. Гасфилд Д. Строки, деревья и последовательности в алгоритмах: Информатика и вычислительная биология / Пер. с англ. — СПб.: Невский диалект; БХВ–Петербург, 2003. — 654 с.

Поступила 26.09.2014