УДК 519.16

Ю. В. Гамзова

ЛОКАЛЬНО ПЕРИОДИЧЕСКИЕ БЕСКОНЕЧНЫЕ ЧАСТИЧНЫЕ СЛОВА

1. Введение

Частичное слово представляет собой комбинаторный объект, получаемый естественным обобщением понятия обычного конечного слова. На слово, т. е. конечную последовательность элементов фиксированного множества (алфавита), можно смотреть как на функцию из множества $\{1,\ldots,n\}$ в алфавит. Тогда частичное слово длины n над алфавитом A может быть определено (см. [1]) как *частичная* функция

$$W\colon \{1,\ldots,n\} \to A.$$

Частичное Z-слово определим как частичную функцию $W\colon \mathbb{Z} \to A$.

Комбинаторика частичных слов находится в начальной стадии своего развития, и большинство постановок задач являются прямыми обобщениями задач комбинаторики «обычных» слов.

В данной работе исследуется фундаментальное свойство периодических слов – свойство взаимодействия периодов.

Дадим необходимые определения для обычных слов. *Длина* слова W есть количество символов в этом слове (обозначается через $|W|$). Порядковый номер символа в слове называется его *позицией*. Натуральное число p, не превосходящее $|W|$, называется *периодом* слова W, если

$$W(i) = W(i+p), \quad i = 1,\ldots,|W|-p.$$

В частности, любое слово имеет тривиальный период, равный его длине. Далее мы будем рассматривать слова, имеющие по крайней мере два различных периода. Свойство взаимодействия периодов заключается в том, что достаточно длинное слово с периодами p и q имеет также и «производный» период

НОД(p, q). Точная формулировка для обычных слов дается в теореме Файна–Вильфа, которую мы приведем здесь в удобных для нас обозначениях.

Теорема 1.1 [2]. *Пусть p и q – натуральные числа, тогда каждое слово длины не менее чем $p + q -$ НОД(p, q) с периодами p и q имеет период* НОД(p, q). *Указанная оценка является точной.*

Перейдем к частичным словам. Частичное слово над алфавитом A будем представлять как обычное слово над расширенным алфавитом $A_\diamond = A \cup \{\diamond\}$, подставляя символ $\diamond$ в неопределенные позиции. Символ $\diamond$, играющий роль, отличную от роли других алфавитных символов, будем называть *джокером*.

В [1] рассмотрены две естественные возможности обобщения свойства периодичности на частичные слова. Приведем соответствующие определения.

Частичное слово W имеет *период* p, если для всех позиций, не занятых джокерами, выполняется условие

$$i \equiv j (\mathrm{mod} p) \Rightarrow W(i) = W(j).$$

Например, слово $W = ab\diamond a\diamond cabc$ имеет период 3.

Частичное слово W имеет *локальный период* p, если для всех позиций, не занятых джокерами, выполняется условие

$$i = j + p \Rightarrow W(i) = W(j).$$

Для обычных слов наличие локального периода p всегда влечет наличие периода p, но это свойство не выполняется для частичных слов. Например, частичное слово $W = abc\diamond bcd$ имеет локальный период 3, но не имеет периода 3. Заметим, что частичное слово имеет период p тогда и только тогда, когда в нем все джокеры можно заменить буквами так, что полученное обычное слово будет иметь период p.

Назовем *плотностью* джокеров в конечном частичном слове отношение количества джокеров к длине слова. Плотность джокеров в бесконечном частичном слове можно определить как верхний предел плотности джокеров в его подсловах длины n при $n \to \infty$. В случае частичных слов выполнение свойства взаимодействия периодов зависит не только от длины слова, но и от плотности джокеров и их расположения в слове. В работе [3] приведена оценка плотности джокеров такая, что свойство взаимодействия периодов будет выполняться для любого слова, в котором плотность джокеров меньше указанной. Для бесконечных частичных слов этот результат может быть переформулирован следующим образом.

Теорема 1.2 [3]. *Пусть бесконечное частичное слово U имеет периоды p, q. Тогда если плотность джокеров в U меньше $(p+q-2)/(p\,q)$, то U также имеет период* НОД(p,q).

Кроме того, в [4] приведен алгоритм проверки наличия свойства взаимодействия периодов для частичных слов, в которых плотность джокеров превышает указанную в теореме 1.2.

Свойство взаимодействия локальных периодов частичных слов пока до конца не изучено. В [1] был доказан аналог теоремы Файна–Вильфа для локально периодических частичных слов с одним джокером.

Теорема 1.3 [1]. *Пусть p и q – натуральные числа. Тогда любое частичное слово длины не менее чем $p+q$ с одним джокером и локальными периодами p и q имеет период* НОД(p,q). *Указанная оценка является точной.*

Однако если частичное слово содержит более одного джокера, то свойство взаимодействия локальных периодов может вообще не выполняться. Например, частичное слово U с периодами p и q, содержащее два джокера в позициях $q+1$ и $p+1$, может содержать в первой позиции любую букву. Приведенный пример показывает, что в частичном слове могут существовать некоторые локальные структуры, которые независимо от длины слова приводят к невыполнению свойства взаимодействия локальных периодов. Будем называть *специальными* конечные подпоследовательности, наличие которых приводит к невыполнению свойства взаимодействия локальных периодов во всем слове. Таким образом, для локально периодических слов более чем с одним джокером можно доказывать лишь «условные» теоремы. Такие результаты приведены в статьях [5, 6].

Теорема 1.4 [6]. *Пусть частичное слово U имеет локальные периоды p, q и не содержит специальных подпоследовательностей. Тогда если плотность джокеров в слове U меньше $2/(p+q)$, то U также имеет период* НОД(p,q).

Исследование специальных подпоследовательностей пока не производилось и представляет большой интерес. Кроме того, в бесконечном частичном слове вероятность наличия специальных подпоследовательностей равна единице (если джокеры распределены равномерно, т. е. джокер с равной вероятностью может находиться в любой позиции). Данная работа посвящена исследованию свойств специальных подпоследовательностей и созданию алгоритма для их поиска.

Как отмечалось в разных работах (см. [1, 7]), утверждения о взаимодействии произвольных периодов p и q тривиально вытекают из аналогичных утверждений для взаимно простых периодов ввиду следующего наблюдения.

Наблюдение. В случае, когда НОД$(p,q) = d > 1$, можно заменить (частичное) слово U набором, состоящим из d (частичных) слов $U_1, \ldots, U_d$, где $U_i = U(i)U(d+i)U(2d+i)\ldots$. Каждое из этих слов будет иметь взаимно простые периоды p/d, q/d. При этом слово U будет иметь период d тогда и только тогда, когда все U_i имеют период 1.

Таким образом, достаточно исследовать свойство взаимодействия для взаимно простых периодов. В этом случае свойство взаимодействия периодов заключается в том, что в достаточно длинном слове с периодами p и q все буквы должны быть одинаковыми. В дальнейшем мы будем рассматривать только взаимно простые периоды. Зафиксируем взаимно простые локальные периоды q и p, $2 \leq q < p$.

Формализуем нашу задачу. Для данного частичного слова W рассмотрим множество частичных слов с теми же периодами и той же областью определения. Среди этих слов выберем слово U с максимальным количеством различных букв. Количество различных букв в U назовем *размерностью* слова W и обозначим через $r(W)$. Последовательности позиций

$$D_i = \{j \in D(W) \mid U(j) = a\}, \quad i = 1, \ldots, r(W),$$

где a – некоторая буква, будем называть *доменами* W. Из определений следует, что количество различных букв в частичном слове не превышает его размерность, а две позиции могут содержать различные буквы, только если эти позиции принадлежат разным доменам. Таким образом, выполнение свойства взаимодействия периодов эквивалентно тому, что размерность слова равна единице. Наша задача заключается в исследовании конечных подпоследовательностей таких, что размерность любого слова, содержащего данную последовательность, больше единицы.

Объектом нашего изучения будут бесконечные частичные слова. Полученные результаты легко обобщить для конечных слов, если представить конечное слово как бесконечное слово, содержащее джокеры в позициях, не принадлежащих исходному слову. Таким образом, в дальнейшем под *словом* будем понимать бесконечное частичное слово.

В следующем разделе описывается используемая техника. Третий раздел посвящен описанию свойств специальных конечных подпоследовательностей. В четвертом разделе эти свойства используются для получения необходимых и достаточных условий наличия в слове специальных подпоследовательностей. Также в этом разделе приводится описание полиномиального алгоритма с конечной памятью для проверки этих условий. Последний раздел посвящен описанию прочих свойств специальных подпоследовательностей и их связи с различными вариантами обобщенного свойства взаимодействия периодов.

2. Техника: бланк

При исследовании частичных слов нас интересует только размерность слова. Из определения понятно, что размерность слова связана только с расположением джокеров и не связана с буквами слова. Поэтому для удобства будем представлять частичное слово как обычное слово над алфавитом $\{0,1\}$, в котором в позициях джокеров стоят нули, а в позициях букв – единицы.

Частичному слову с периодами p, q можно сопоставить *граф периодичности* – граф, вершины которого соответствуют позициям букв данного частичного слова, а ребра соединяют вершины u, v такие, что разность номеров соответствующих позиций частичного слова составляет p или q. Легко заметить, что если в графе периодичности между двумя вершинами существует маршрут, то в исходном слове в соответствующих позициях должны находиться одинаковые буквы. Таким образом, множество доменов частичного слова совпадает с множеством компонент связности данного графа.

Граф периодичности можно изобразить на бесконечном цилиндре двумя системами спиралей. Для удобства этот граф изобразим на плоскости. Для этого «разрежем» цилиндр вдоль. Получим бесконечную в одну сторону таблицу ширины p. Теперь «восстановим» разорванные связи: добавим слева и справа еще по одной такой же таблице, слева – со сдвигом вниз на q строк, справа – со сдвигом вверх на q строк. Продолжая этот процесс, мы получим бесконечную двумерную таблицу, в которой в ячейке (x, y) находится 0, если $W(xp+yq) = 0$, и 1 – в противном случае. Такую таблицу мы будем называть *бланком.*

Выберем в частичном слове некоторую начальную позицию (для удобства будем считать, что это нулевая позиция). Рассмотрим два способа разделения позиций слова на части, начиная с этой начальной позиции.

Будем называть *блоком* прямоугольную часть бланка размером $p\times q$. Блок содержит множество позиций частичного слова

$$\{n \mid n = apq+xp+yq,\ a \in \mathbb{Z},\ x \in \{0,\dots,q-1\},\ y \in \{0,\dots,p-1\}\}\,.$$

Отрезком будем называть множество позиций подслова длины p, т. е. множество позиций $\{ap,\dots,ap+p-1,\ a \in \mathbb{Z}\}$. Каждый блок содержит полностью только один отрезок.

Заметим, что по построению бланка одной позиции частичного слова соответствует бесконечно много позиций бланка. Рассмотрим позицию x. Существует единственное разложение $x = apq+bp+cq$ такое, что $b \in \{0,\dots,q-1\}$, $c \in \{0,\dots,p-1\}$. Определим $a(x) = a$, $b(x) = b$, $c(x) = c$. Тогда множество ячеек бланка, соответствующих данной позиции x, можно определить следующим образом: $((a(x)-n)q + b(x), c(x)+np)$, $n \in \mathbb{Z}$.

Пример 2.1. На рис. 1 представлена часть бланка бесконечного частичного слова с периодами $5, 7$. Выделенные позиции соответствуют одной и той же позиции исходного слова.

0	0	1	0	0	1	1	0	0	1	1	1	1	0	0	0	1	1	1	1	1
0	1	0	1	1	0	1	0	1	0	0	0	1	1	1	1	**1**	0	0	1	0
0	1	1	1	0	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1
1	0	1	0	1	1	1	0	0	1	1	0	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1
0	1	0	0	0	1	1	1	1	**1**	0	0	1	0	1	1	1	1	0	1	1
0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	1
0	0	1	1	0	1	0	1	1	1	1	0	1	1	0	0	1	1	1	0	1
0	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1
0	0	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	0	0	0	1
1	1	**1**	0	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1
1	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	1	1
1	1	1	1	0	1	1	0	0	1	1	1	0	1	0	0	1	1	1	1	0
1	1	1	1	1	0	0	1	1	1	1	1	1	1	0	0	0	1	1	1	0

Рис. 1

3. Разрезы

Пусть J – некоторое множество позиций. Обозначим через $P(J)$ частичное слово W такое, что $W(x) = 0$ тогда и только тогда, когда $x \in J$. Будем называть *разрезом* слова W множество позиций джокеров T такое, что слово $P(T)$ содержит 2 домена. Для удобства будем считать, что разрез является минимальным множеством с таким свойством, т. е. не может содержать в качестве подмножества другой разрез. Легко заметить, что разрезы – это вершинные разрезы графа периодичности обычного Z-слова с периодами p, q.

В данном разделе мы рассмотрим свойства разрезов. Эти свойства будут использованы в разделе 4 для конструирования алгоритма поиска конечных разрезов.

Сначала введем необходимые определения. На множестве позиций частичного слова определим следующие отношения:

отношение ρ («иметь одинаковые значения в бланке»):

$$x \,\rho\, y \Leftrightarrow W(x) = W(y);$$

отношение ϕ («быть соседом»):

$$x\ \phi\ y \Leftrightarrow (x - y = ap + bq,\ a,\, b \in \{-1, 0, 1\});$$

отношение ψ («быть близким соседом»):

$$x\ \psi\ y \Leftrightarrow (|x - y| = q) \vee (|x - y| = p).$$

Таким образом, $\psi \subset \phi$. Отношения ϕ и ψ симметричны. Рассмотрим также отношения

$$\phi_1 = \phi \cap \rho \ \text{ и } \ \psi_1 = \psi \cap \rho.$$

Эти отношения тоже симметричны. Рефлексивно-транзитивные замыкания этих отношений будут являться отношениями эквивалентности. Назовем отношение ϕ_1^* *отношением связанности*, а отношение ψ_1^* – *отношением сильной связанности.*

Путь – это последовательность позиций слова $x_1, \ldots,\ x_n$ такая, что для $i = 1, \ldots, n-1$ выполняется $x_i\ \psi\ x_{i+1}$. *Цепочка* – это последовательность позиций слова $x_1, \ldots,\ x_n$ такая, что $x_i\ \phi\ x_{i+1},\ i = 1, \ldots, n-1$. Будем считать, что путь и цепочка не должны содержать повторов позиций (кроме, возможно, первой и последней позиций). Цепочку будем называть *замкнутой*, если $x_1 = x_n$. Также будем считать, что замкнутая цепочка является минимальной, т. е. не содержит замкнутых подцепочек.

Предложение 3.1. *Для любой позиции x разреза T в слове $P(T)$ два близких соседа позиции x принадлежат различным доменам.*

Доказательство. По определению разреза при замене в любой из позиций слова $P(T)$ джокера на букву получившееся слово содержит один домен. Таким образом, при замене в позиции x джокера на букву в слове появляется путь, соединяющий позиции из разных доменов, проходящий через x. Это означает, что два близких соседа x принадлежат различным доменам.

Предложение 3.2. *Пусть T – разрез. Для любой позиции $x \in T$ можно указать не менее двух позиций джокеров $y,\ z \in T$, являющихся соседями x.*

Доказательство. Рассмотрим соседей позиции x. Если среди соседей x присутствует менее двух позиций джокеров, то все соседи x сильно связаны, что противоречит предложению 3.1.

Предложение 3.3. *Любая замкнутая цепочка позиций джокеров является разрезом.*

Доказательство. Пусть T – замкнутая цепочка позиций джокеров. В бланке слова $P(T)$ замкнутая цепочка ограничивает некоторую непрерывную область. Рассмотрим произвольную позицию из этой области. Любая цепочка, соединяющая данную позицию с позицией, не принадлежащей данной области, проходит через некоторую позицию цепочки. Это означает, что данные позиции принадлежат разным доменам, т. е. размерность $P(T)$ больше единицы. Следовательно, T содержит разрез. Кроме того, при замене в одной из позиций джокера на букву все позиции букв в $P(T)$ окажутся сильно связаны. Отсюда вытекает, что замкнутая цепочка является минимальным множеством с таким свойством, т. е. разрезом.

Предложение 3.4. *Любой бесконечный разрез является объединением бесконечных цепочек позиций джокеров. Любой конечный разрез является замкнутой цепочкой позиций джокеров.*

Доказательство. Согласно предложению 3.2, любой разрез представляет собой объединение бесконечных и конечных замкнутых цепочек позиций джокеров. Согласно предложению 3.3, разрез не может содержать как собственное подмножество замкнутую цепочку джокеров. Таким образом, бесконечный разрез может представлять собой только объединение бесконечных цепочек позиций джокеров, а конечный – замкнутую цепочку.

Пусть размерность слова W больше 1. Будем говорить, что множество позиций джокеров T *изолирует* домен (или объединение доменов) I, если в слове $P(T)$ домен, содержащий позиции из I, является объединением I и некоторого множества позиций джокеров из W. Будем называть T *изолирующим множеством* домена (или объединения доменов) I. Из определения следует, что для уменьшения размерности слова необходимо и достаточно заменить джокер на букву в одной из позиций каждого изолирующего множества некоторого домена. Таким образом, размерность слова определяется изолирующими множествами.

Легко понять, что любое изолирующее множество является разрезом, а любой разрез изолирует домен или объединение доменов. Таким образом, исследование множества доменов частичного слова можно свести к исследованию множества разрезов.

Поскольку нас интересуют конечные структуры, приводящие к наличию нескольких доменов в любом слове, содержащем данную структуру, в дальнейшем объектом нашего исследования будут конечные разрезы. В следующем разделе приведен алгоритм, позволяющий находить в частичном слове конечные разрезы. Бесконечные разрезы рассматривались в [6], теорема 1.4 задает минимальную плотность джокеров в слове с бесконечным разрезом.

4. Алгоритм поиска конечных разрезов

В данном разделе представлен алгоритм поиска конечных разрезов, время работы которого линейно зависит от длины слова, а объем используемой памяти постоянен (если считать периоды p, q параметрами).

Определим порядок на множестве позиций как лексикографический порядок троек $(a(x), b(x), c(x))$, т. е.

$$x \prec y \Leftrightarrow (a(x) < a(y)) \vee ((a(x) = a(y)) \wedge (b(x) < b(y))) \vee \\ \vee ((a(x) = a(y)) \wedge (b(x) = b(y)) \wedge (c(x) < c(y))).$$

Наибольшую (в смысле порядка $\prec$) позицию разреза будем называть *замыкающей.*

Для каждой позиции x определим *метку* следующим образом:

$$Label(x) = \begin{cases} \min_{\prec}\{y : \ y \ \phi_1 \ x_1 \ \phi_1 \ x_2 \ \phi_1 \ \ldots \ \phi_1 \ x_n \ \phi_1 \ x, \\ \qquad\qquad y, x_1, \ldots, x_n \prec x\}, & W(x) = 0 \\ \infty, & W(x) = 1. \end{cases}$$

По предложению 3.2 все позиции разреза связаны между собой, поэтому все позиции разреза будут иметь одинаковую метку.

Предложение 4.1. *Позиция x, замыкающая разрез, имеет двух соседей y, z с одинаковой конечной меткой, при этом $y \prec x$, $z \prec x$.*

Доказательство. Согласно предложению 3.4, конечный разрез является замкнутой цепочкой джокеров. Таким образом, у позиции x есть два соседа y, z, принадлежащих разрезу, причем все остальные позиции разреза представляют собой цепочку, соединяющую y и z. Поскольку x является наибольшей позицией разреза, все остальные позиции разреза имеют меньшие номера. Это означает, что y и z имеют одинаковые конечные метки.

Один и тот же домен могут изолировать несколько разрезов. Из всех разрезов, изолирующих данный домен, выберем разрез с минимальной замыкающей позицией. Такой разрез будем называть *границей* домена.

Предложение 4.2. *Позиция, замыкающая границу домена, является близким соседом некоторой меньшей позиции буквы.*

Доказательство. Пусть позиция x замыкает границу домена I. Положим $J = \{y \mid W(y) = 0 \vee y \prec x\}$. Поскольку x – минимальная из позиций, замыкающих I, в слове $P(J)$ нет ни одного разреза, изолирующего I, а значит, существует путь из домена, содержащего I, в другой домен. При замене в позиции x буквы на джокер появляется разрез, изолирующий I, а значит, не менее двух близких соседей позиции x в слове $P(J \cup \{x\})$ содержат буквы и принадлежат разным доменам. Пусть это позиции y, z. Предположим, что $x \prec y, x \prec z$. В слове $P(J \cup \{x\})$ позиция любого джокера не превышает x, а значит, позиции y и z сильно связаны, т. е. принадлежат одному домену, что приводит к противоречию. Таким образом, в слове $P(J \cup \{x\})$ один из близких соседов позиции x с номером $y \prec x$ содержит букву. По построению слова $P(J \cup \{x\})$, в исходном слове в позиции y также содержится буква.

Пусть позиция x замыкает границу домена. Перенумеруем соседей x следующим образом (закрашенная клетка в центре соответствует позиции x):

3	4	5
2		6
1	8	7

Теорема 4.1. *Для того чтобы позиция замыкала границу, необходимо и достаточно, чтобы для соседей данной позиции выполнялось одно из следующих условий:*

1. $c(x) > 0$, *в позиции* 4 *находится буква, позиция* 5 *и одна из позиций* 2, 3 *имеют одинаковую конечную метку;*

2. $c(x) = 0$:

   а) *в позиции* 2 *находится буква, позиция* 1 *и одна из позиций* 3, 4, 5 *имеют одинаковую конечную метку;*

   б) *в позиции* 4 *находится буква, позиция* 5 *и одна из позиций* 1, 2, 3 *имеют одинаковую конечную метку.*

Доказательство. Необходимость. Пусть позиция x замыкает границу. Согласно предложению 4.1, среди соседей x с меньшим номером имеются две позиции джокеров с одинаковой конечной меткой. Согласно предложению 4.2, среди близких соседей x с меньшим номером имеется позиция буквы. Рассмотрим возможные случаи.

1. $c(x) > 0$. В этом случае все рассматриваемые позиции принадлежат одному блоку. Соседи, меньшие x, – это позиции 2, 3, 4, 5; близкие соседи, меньшие x, – позиции 2, 4:

   а) предположим, что буква находится в позиции 4:

- пусть в позиции 2 находится буква, тогда позиции джокеров с одинаковой меткой – это позиции 3, 5;
- пусть в позиции 2 находится джокер, тогда позиция 5 должна принадлежать разрезу (в противном случае в слове $P(J \cup \{x\})$ все близкие соседи позиции x сильно связаны, что противоречит наличию в этом слове разреза). В этом случае позиции джокеров с одинаковой меткой – это позиция 5 и одна из позиций 2, 3;

б) предположим, что в позиции 4 находится джокер, тогда буква находится в позиции 2. В слове $P(J \cup \{x\})$ все близкие соседи позиции x сильно связаны, что противоречит наличию в этом слове разреза.

Итак, в этом случае в позиции 4 обязательно должна присутствовать буква, а позиции джокеров с одинаковой меткой – это позиция 5 и одна из позиций 2, 3.

2. $c(x) = 0$. В этом случае клетки 1, 2, 3 принадлежат одному блоку, а все остальные – следующему блоку. Тогда соседи, меньшие x, – это позиции 1, 2, 3, 4, 5, близкие соседи, меньшие x, – позиции 2, 4:

   а) предположим, что буква находится в позиции 2, тогда (по соображениям, аналогичным рассмотренным выше) позиции джокеров с одинаковой меткой – это позиция 1 и одна из позиций 3, 4, 5;

   б) предположим, что буква находится в позиции 4, тогда позиции джокеров с одинаковой меткой – это позиция 5 и одна из позиций 1, 2, 3.

Достаточность. Пусть для позиции x выполняется одно из указанных в теореме условий. Так как два меньших соседа x имеют одинаковую конечную метку, то существует цепочка, соединяющая эти позиции и не проходящая через x. Добавление к этой цепочке позиции x делает цепочку замкнутой. Так как близким соседом x является позиция буквы, позиция x замыкает границу.

Теорема 4.2. *Алгоритм, определяющий наличие границ, имеет линейную временну́ю сложность. Используемый объем памяти линейно зависит от периодов p и q.*

Доказательство. В качестве доказательства опишем реализацию алгоритма. Позиции слова обрабатываются в порядке $\prec$. Для каждой позиции определяется ее метка. Для этого достаточно знать метки соседей данной позиции с меньшими номерами. Эти позиции уже просмотрены, а значит, их метки уже определены. Для текущей позиции метка устанавливается равной минимальной из меток ее соседей с меньшими номерами. Кроме того, для данной позиции проверяется выполнение условий, указанных в теореме 4.1. Если одно из данных условий выполняется, то позиция объявляется замыкающей границу. Это означает, что в данном частичном слове имеется граница.

Каждая позиция просматривается только один раз, и при просмотре данной позиции выполняется постоянное количество операций. Таким образом, время работы алгоритма линейно зависит от длины просматриваемого отрезка слова.

Для хранения меток достаточно использовать массив размера p, содержащий сведения о метках предыдущей строки, и массив размера q, содержащий сведения о метках правой колонки предыдущего блока, так как по свойствам бланка позиции этой колонки являются левыми соседями позиций из левой колонки текущего блока. Таким образом, используемый объем памяти линейно зависит от периодов p, q.

Начальные данные в массивах зависят от условий поиска. Если анализируемая последовательность рассматривается как конечное слово, то в массивы записываются нули (напомним, что конечное частичное слово можно рассматривать как бесконечное с джокерами в остальных позициях). В противном случае в качестве начальных данных записывается ∞.

Алгоритм легко модифицировать для поиска всех разрезов. Для этого достаточно вместо условий, указанных в теореме 4.1, проверять условие из предложения 4.1. Результатом работы описанного алгоритма является позиция, замыкающая границу (или произвольный разрез). Кроме того, дополнительно можно по меткам границы восстановить всю границу (или весь разрез), а по позиции буквы, являющейся близким соседом позиции, замыкающей границу, можно восстановить изолируемый границей домен.

5. Свойства конечных разрезов

Будем называть *Z-доменом* (ω-доменом, конечным доменом) домен, образующий Z-слово (соответственно, ω-слово, конечное слово).

В этом разделе мы рассмотрим 2 класса конечных разрезов (преграды и ловушки) и исследуем их свойства. Будем называть *преградой* минимальную по количеству позиций конечную последовательность позиций джокеров T такую, что в слове $P(T)$ существуют позиции x', y' такие, что для любого $x < x'$ и любого $y > y'$ позиции x и y принадлежат разным доменам. Легко заметить, что преграда изолирует ω-домен или объединение ω-доменов. Все остальные конечные разрезы будем называть *ловушками.*

Поскольку конечный разрез $T = x_1, \dots, x_n$ представляет собой замкнутую цепочку позиций джокеров,

$$x_i - x_{i+1} = \delta_p(i)p + \delta_q(i)q, \quad \delta_p(i),\, \delta_q(i) \in \{-1, 0, 1\}, \quad i = 1, \dots, n-1.$$

Назовем *длиной* разреза T величину $\ell(T) = \left|\sum\limits_{i=1}^{n-1} \delta_p(i)\right|$, *шириной* разреза T – величину $w(T) = \left|\sum\limits_{i=1}^{n-1} \delta_q(i)\right|$.

Предложение 5.1. *Для конечного разреза T либо $\ell(T) = 0$ и $w(T) = 0$, либо $\ell(T) = q$ и $w(T) = p$.*

Доказательство. В бланке длина и ширина цепочки – это координаты вектора, соединяющего начало цепочки и ее конец. Поскольку цепочка замкнутая, то $x_1 = x_n$. В бланке для ячеек, соответствующих одинаковым позициям, координаты вектора, соединяющего такие ячейки, равны (mp, mq), $m \in \mathbb{N} \cup \{0\}$. Однако если $m > 1$, то цепочка в бланке пересечется с аналогичной цепочкой, соединяющей другие блоки, а значит, существует цепочка с меньшим количеством позиций, что противоречит минимальности разреза. Таким образом, $m = 0$ или $m = 1$.

5.1. Преграды

Простейшим примером преграды является множество из p последовательных позиций джокеров.

Предложение 5.2. *Конечный разрез T является преградой тогда и только тогда, когда $\ell(T) = p$, $w(T) = q$.*

Доказательство. Необходимость. Пусть $T = z_1, \dots, z_n$ – преграда. Тогда по определению преграды существуют позиции x', y' такие, что для любого $x < x'$ и любого $y > y'$ позиции x и y принадлежат разным доменам. Рассмотрим позиции $x_i = x' - i$, $y_i = y' + i$, $i = 1, \dots, p$. По построению

$$\{b(x_1), \dots, b(x_p)\} = \{b(y_1), \dots, b(y_p)\} = \{0, \dots, p-1\},$$

$$\{c(x_1), \dots, c(x_p)\} = \{c(y_1), \dots, c(y_p)\} = \{0, \dots, q-1\}.$$

Для любых $i, j = 1, \dots, p$ позиции x_i и y_j должны принадлежать разным доменам, а значит, на любом пути из x_i в y_j должна встретиться позиция из преграды. Поэтому выполняется условие

$$\{b(z_1), \dots, b(z_n)\} = \{0, \dots, p-1\},$$

$$\{c(z_1), \dots, c(z_n)\} = \{0, \dots, q-1\}.$$

Это означает, что $\ell(T) = p$, $w(T) = q$.

Достаточность. Допустим, что разрез T изолирует домен I и $\ell(T) = p$, $w(T) = q$. Это означает, что в слове $P(T)$ домену, содержащему I, принадлежат позиции подслова длины p. Возьмем

$$x' = \min\{x \mid x \in T\} - p, \quad y' = \max\{x \mid x \in T\} + p.$$

Тогда любые позиции $x < x'$ и $y > y'$ принадлежат разным доменам.

Пример 5.1. Продолжим рассмотрение бланка из примера 2.1. На рис. 2 выделены позиции, составляющие преграду.

0	0	1	0	0	1	1	0	0	1	1	1	1	0	0	0	1	1	1	1	1
0	1	0	1	1	0	1	0	1	0	0	0	1	1	1	1	1	0	0	1	0
0	1	1	1	0	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1
1	0	1	0	1	1	1	0	0	1	1	0	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	**0**	**0**
0	0	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	**0**	1	1
0	1	0	0	0	1	1	1	1	1	0	0	1	0	1	1	1	1	**0**	1	1
0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	**0**	**0**	1	1	1
0	0	1	1	0	1	0	1	1	1	1	0	1	1	**0**	**0**	1	1	1	0	1
0	1	0	1	1	1	1	1	1	1	1	1	**0**	**0**	1	1	1	1	1	1	1
0	0	1	1	1	1	1	1	1	1	1	**0**	1	1	0	0	1	0	0	0	1
1	1	1	0	0	1	0	1	1	1	1	**0**	1	1	1	1	1	1	1	1	1
1	0	1	1	1	1	1	1	1	**0**	**0**	1	1	1	1	1	1	0	1	1	1
1	1	1	1	0	1	1	**0**	**0**	1	1	1	0	1	0	0	1	1	1	1	0
1	1	1	1	1	**0**	**0**	1	1	1	1	1	1	1	0	0	0	1	1	1	0

Рис. 2

5.2. Ловушки

Из предложений 5.1, 5.2 немедленно следует, что конечная граница T является ловушкой тогда и только тогда, когда $\ell(T) = 0$, $w(T) = 0$.

Изолированный ловушкой домен I не может содержать две одинаковых позиции из разных блоков (иначе ловушка содержала бы преграду), это накладывает ограничения на домен I. Укажем два из них.

1. Домен I не может содержать p последовательных позиций. Таким образом, если домен изолирован ловушкой, то в $\lceil p/2 \rceil$-окрестности любой его позиции найдется позиция, не принадлежащая данному домену (т. е. с помощью скачка ограниченной длины можно «выбраться» из ловушки).

2. Напомним, что множество ячеек бланка, соответствующих данной позиции x, можно определить как $((a(x)-n)q+b(x), c(x)+np)$, $n \in \mathbb{Z}$. Для некоторой позиции x обозначим $\mathrm{row}(x)=a(x)q+b(x)$, $\mathrm{col}(x)=c(x)$ (это номер строки и столбца одной из ячеек, соответствующих данной позиции). Для любой позиции x из домена I джокер должен содержаться в позициях x_1, x_2 таких, что

$$
\begin{aligned}
&\big((\mathrm{col}(x){=}\mathrm{col}(x_1){=}\mathrm{col}(x_2))\wedge\\
&\qquad\qquad\wedge(\mathrm{row}(x_1){<}\mathrm{row}(x){<}\mathrm{row}(x_2))\wedge(\mathrm{row}(x_2){-}\mathrm{row}(x_1){\le}q)\big)\vee\\
&\vee\big((\mathrm{row}(x){=}\mathrm{row}(x_1){=}\mathrm{row}(x_2))\wedge\\
&\qquad\qquad\wedge(\mathrm{col}(x_1){<}\mathrm{col}(x){<}\mathrm{col}(x_2))\wedge(\mathrm{col}(x_2){-}\mathrm{col}(x_1){\le}p)\big).
\end{aligned}
$$

Если джокер не встречается ни в одной из указанных позиций, то домену I принадлежат две одинаковые позиции из разных блоков. Таким образом, осмотрев $2(p+q)$ указанных позиций, удовлетворяющих приведенным условиям, можно определить, что данная позиция не принадлежит домену, изолированному ловушкой. Это достаточное условие отсутствия ловушки, но не необходимое.

Пример 5.2. Продолжим рассмотрение бланка из примера 2.1. На рис. 3 выделены позиции, составляющие ловушку.

0	**0**	1	**0**	**0**	1	1	0	0	1	1	1	1	0	0	0	1	1	1	1	1
0	1	**0**	1	1	**0**	1	0	1	0	0	0	1	1	1	1	1	0	0	1	0
0	1	1	1	**0**	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1
1	**0**	1	**0**	1	1	1	0	0	1	1	0	1	0	1	1	1	1	0	1	1
1	1	**0**	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1
0	1	0	0	0	1	1	1	1	1	0	0	1	0	1	1	1	1	0	1	1
0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	1
0	0	1	1	0	1	0	1	1	1	1	0	1	1	0	0	1	1	1	0	1
0	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1
0	0	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	0	0	0	1
1	1	1	0	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1
1	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	1	1
1	1	1	1	0	1	1	0	0	1	1	1	0	1	0	0	1	1	1	1	0
1	1	1	1	1	0	0	1	1	1	1	1	1	1	0	0	0	1	1	1	0

Рис. 3

Итак, мы видели, что форма ловушки не может быть произвольной. Однако домен, изолированный ловушкой, хотя и конечен, но может содержать произвольное количество позиций (т. е. это количество нельзя оценить с помощью периодов p, q).

Пример 5.3. На рис. 4 выделены позиции, составляющие ловушку.

0	**0**	1	**0**	**0**	1	1	0	0	1	1	0	1	0	0	0	1	1	1	1	1
0	1	**0**	1	1	**0**	1	0	1	0	0	0	1	0	1	1	1	0	0	1	0
0	1	1	1	1	1	**0**	0	0	1	0	1	1	1	0	0	1	1	1	0	1
1	**0**	1	**0**	**0**	1	1	**0**	0	1	1	0	1	1	1	1	0	1	0	1	0
1	1	**0**	1	**0**	1	**0**	0	1	0	1	1	0	0	1	1	1	0	1	0	0
0	0	1	1	**0**	1	**0**	0	0	1	1	1	1	1	0	1	1	1	0	1	1
0	1	0	0	**0**	1	**0**	1	1	1	0	0	1	0	1	1	1	0	0	1	0
0	0	1	**0**	1	1	1	**0**	**0**	1	1	1	0	1	1	1	0	0	1	1	1
0	0	1	1	**0**	1	1	1	1	**0**	1	0	1	0	1	1	0	1	1	0	1
0	1	0	1	1	**0**	**0**	1	1	1	**0**	1	0	0	0	0	1	1	1	1	1
0	0	1	1	1	1	1	**0**	1	1	1	**0**	1	1	0	0	1	0	0	0	1
1	1	1	0	0	1	**0**	1	1	1	1	**0**	1	1	1	1	1	1	1	1	1
0	0	1	1	1	**0**	1	1	1	0	**0**	1	1	1	1	1	1	0	1	1	1
1	1	0	1	0	1	**0**	1	1	**0**	1	1	0	1	0	0	1	1	1	1	0
1	1	1	0	1	0	0	**0**	**0**	1	1	1	1	1	0	0	0	1	1	1	0

Рис. 4

5.3. Обобщение свойства взаимодействия периодов

Приведенную выше стандартную формулировку свойства взаимодействия периодов можно обобщить. Будем считать, что подпоследовательность позиций букв представляет собой Z-слово. Свойство взаимодействия периодов состоит в выполнении некоторых условий для множества доменов частичного слова. В зависимости от используемых условий для множества доменов можно получить различные варианты обобщенного свойства взаимодействия периодов. Приведем несколько примеров таких задач и их формулировку в терминах преград и ловушек.

1. Размерность частичного слова равна 1, т. е. слово содержит только один домен (стандартная формулировка свойства взаимодействия периодов) тогда и только тогда, когда слово не содержит преград и ловушек и плотность джокеров соответствует указанной в теореме 1.4.

2. В частичном Z-слове любой домен является Z-доменом тогда и только тогда, когда слово не содержит преград и ловушек.

3. В частичном Z-слове есть хотя бы один Z-домен тогда и только тогда, когда слово не содержит преград.

Литература

1. Berstel J., Boasson L. Partial words and a theorem of Fine and Wilf // Theor. Comp. Sci. 1999. № 218. P. 135–141.
2. Fine N. J., Wilf H. S. Uniqueness theorem for periodic functions // Proc. Amer. Math. Soc. 1965. № 16. P. 109–114.
3. Шур А. М., Гамзова Ю. В. Частичные слова и свойство взаимодействия периодов // Изв. РАН. Сер. математ. 2004. № 68. С. 199–222.
4. Гамзова Ю. В. Статистические закономерности взаимодействия периодов частичных слов // Дискретный анализ и исследование операций. Сер. 1. 2004. № 11. С. 20–35.
5. Blanchet-Sadri F., Hegstrom R. A. Partial words and a theorem of Fine and Wilf revisited // Theor. Comp. Sci. 2002. № 270. P. 401–419.
6. Blanchet-Sadri F. Periodicity on partial words // Comput. Math. Appl. 2004. № 47. P. 71–82.
7. Choffrut C., Karhumäki J. Combinatorics on words // Handbook of Formal Languages. Berlin, 1997. Vol. 1. P. 329–438.