Периодичность морфических слов*

И. В. МИТРОФАНОВ
Московский государственный университет
им. М. В. Ломоносова
e-mail: phortim@yandex.ru

УДК 510.53+519.16+512.53

Ключевые слова: комбинаторика слов, морфическая последовательность, HD0L-система, разрешимость.

Аннотация

В работе доказывается разрешимость задачи заключительной периодичности для морфических слов.

Abstract

I. V. Mitrofanov, Periodicity of morphic words, Fundamentalnaya i prikladnaya matematika, vol. 18 (2013), no. 4, pp. 107—119.

In this paper, we prove the decidability of the ultimate periodicity problem (the HD0L periodicity problem).

1. Введение

В работе доказывается разрешимость следующей задачи.

Дано: два алфавита A и B, подстановка $\varphi\colon A^* \to A^*$, продолжающаяся над буквой a, и морфизм $h\colon A^* \to B^*$.

Ответ: существуют ли такие конечные слова u и v, что $h(\varphi^\infty(a)) = uv^\infty$ (т. е. является ли сверхслово заключительно периодичным).

Далее мы будем называть эту задачу *проблемой заключительной периодичности*.

Теорема 1.1. *Проблема заключительной периодичности разрешима.*

Ранее результат был опубликован автором в препринте [10] с более сложным доказательством, основанным на совместной с А. Я. Беловым работе [9]. Другое доказательство в это же время независимо получил Ф. Дюран [3].

Проблема заключительной периодичности (в англоязычных публикациях «the HD0L ultimate periodicity problem») была открытым вопросом около 30 лет.

*Работа выполнена при частичной финансовой поддержке фонда Дмитрия Зимина «Династия».

Фундаментальная и прикладная математика, 2013, том 18, № 4, с. 107—119.

В 1986 году в [6, 13] проблема была решена для чисто морфических последовательностей, в [7] — для автоматных последовательностей (частный случай морфических).

В [8] даётся эквивалентная формулировка проблемы в терминах распознаваемых подмножеств натуральных чисел и абстрактных систем счисления. Описание распознаваемых подмножеств натуральных чисел в абстрактных системах счисления [14] вместе с теоремой 1.1 даёт алгоритм проверки того, представляется ли распознаваемое подмножество натуральных чисел в произвольной системе счисления в виде объединения конечного количества арифметических прогрессий.

Структура работы такова. В разделе 2 приведены основные определения. В разделе 3 задача решается для примитивных подстановочных систем, в разделе 4 из результата предыдущего раздела выводится решение общей задачи.

2. Слова, морфизмы, подстановки и морфические сверхслова

Алфавит — это произвольное конечное множество. Его элементы будут называться *буквами*. Отображение начального куска натурального ряда в алфавит A называется конечным словом (или просто словом) над A. Множество конечных слов вместе с пустым обозначается A^*. Пустое слово обозначают ε. *Бесконечное слово*, или *сверхслово*, — это отображение из $\mathbb{N}$ в алфавит A. Множество сверхслов обозначается A^ω.

Для конечного слова определена *длина* — количество букв в нём. Длина слова u также будет обозначаться $|u|$. Если слово u_1 конечно, то определена *конкатенация* слов u_1 и u_2 — слово u_1u_2, получающееся приписыванием второго к первому справа.

Слово v является *подсловом* слова u, если $u = v_1vv_2$ для некоторых слов v_1, v_2. Если v_1 или v_2 — пустое слово, v называется *началом* или соответственно *концом* слова u. На словах существует естественная структура частично упорядоченного множества: $u_1 \sqsubseteq u_2$, если u_1 является подсловом u_2. Будем писать $u_1 \sqsubseteq_k u_2$, если слово u_1 входит в u_2 не менее k раз.

Сверхслово W называется *рекуррентным*, если любое его подслово встречается в W бесконечно много раз, иначе говоря, $v \sqsubseteq W$ влечёт $v \sqsubseteq_\infty W$.

Сверхслово W называется *периодичным*, если $W = uuuuu\ldots$ для некоторого непустого u. Само слово u, равно как и его длина, называются *периодом* сверхслова W. Сверхслово называется *заключительно периодичным*, если оно представляется в виде конкатенации $W = uW'$, где u — конечное слово, а W' — периодичное сверхслово.

Предложение 2.1. *Если сверхслово заключительно периодичное и рекуррентное, то оно периодичное.*

Множество слов A^* над алфавитом A можно считать свободным моноидом с операцией конкатенацией и единицей — пустым словом. Отображение $\varphi\colon A^* \to B^*$ называется *морфизмом*, если оно сохраняет операцию моноида. Очевидно, морфизм достаточно задать на буквах алфавита A. Морфизм называется нестирающим, если образом никакой буквы не является пустое слово. Если $|\varphi(a_i)| = 1$ для любой буквы $a_i \in A$, то φ — *кодирование*. Если алфавиты A и B совпадают, то φ называется *подстановкой*.

Морфизм можно продолжить на бесконечное слово по правилу: если u — начало W, то $\varphi(u)$ — начало $\varphi(W)$.

Если φ — такая подстановка, что $\varphi(a_1) = a_1 v$ для некоторого слова v и для всех $k \in \mathbb{N}$ $\varphi^k(v) \neq \varepsilon$, то говорится, что φ *продолжается над* a_1, и бесконечное слово

$$\varphi^\infty(a_1) := a_1 v \varphi(v) \varphi^2(v) \varphi^3(v) \varphi^4(v) \ldots$$

называется *чисто морфическим* или *бесконечной неподвижной точкой* морфизма φ. Если задан морфизм $h\colon A^* \to C^*$, то сверхслово $h\big(\varphi^\infty(a_1)\big)$ называется *морфическим*.

Определение 2.2. Подстановка φ называется *примитивной*, если существует такое $k \in \mathbb{N}$, что для любой буквы $a_i \in A$ в образе $\varphi^k(a_i)$ содержатся все буквы из A. Если φ продолжается над a_1, то сверхслово $h\big(\varphi^\infty(a_1)\big)$ для произвольного h называется *примитивным морфическим словом*.

Определение 2.3. Матрица подстановки — это такая матрица, у которой в i-м столбце в j-й строке стоит число вхождений буквы a_j в $\varphi(a_i)$.

Подстановка φ является примитивной тогда и только тогда, когда некоторая степень её матрицы состоит из положительных чисел.

3. Проверка периодичности в примитивном случае

В этом разделе мы покажем разрешимость следующей задачи.

Дано: два конечных алфавита A и B, примитивная подстановка $\varphi\colon A^* \to A^*$, продолжающаяся над буквой $a_1 \in A$, и нестирающий морфизм $h\colon A^* \to B^*$.

Ответ: периодично ли слово $h\big(\varphi^\infty(a)\big) = u^\infty$ и, если да, слово u, являющееся наименьшим периодом.

В дальнейшем мы считаем, что все входные данные зафиксированы. Это значит, что величину, зависящую только от h и φ, мы будем рассматривать как константу и писать не $C(\varphi, h)$, а просто C.

Сформулируем два полезных свойства примитивной подстановочной системы [2].

Предложение 3.1. *Существуют положительные $K_1 < K_2$ и $\lambda > 1$, такие что для любых k и a_i выполнено*

$$K_1 \lambda^k < \big|h\big(\varphi^k(a_i)\big)\big| < K_2 \lambda^k.$$

Предложение 3.2. *Существует такое K, что если u_1 и u_2 — подслова слова $h\big(\varphi^\infty(a_1)\big)$, такие что $|u_2| > K|u_1|$, то u_1 является подсловом u_2.*

3.1. Схемы расположений подслов

Здесь мы определим *схему расположений подслов* — объект, с которым работает описываемый в дальнейшем алгоритм.

Сначала дадим неформальное описание. Хочется сказать, что вхождения слов $u_1 = abba$ и $u_2 = baab$ расположены в слове $u_3 = abbabaabaab$ так же, как вхождения слов $v_1 = aa$ и $v_2 = bb$ расположены внутри слова $v_3 = aabbb$. Слово u_1 является началом u_3, а слово u_2 имеет в u_3 два перекрывающихся вхождения: первое начинается там, где кончается единственное вхождение слова u_1, а второе является концом слова u_3. Та же самая фраза остаётся верной, если вместо u_1, u_2, u_3 написать соответственно v_1, v_2, v_3. Хочется сказать, что у $(\{u_3\}, \{u_1, u_2\})$ такая же схема вхождения, как у $(\{v_3\}, \{v_1, v_2\})$.

Определение 3.3. *Узлом* конечного слова u длины n будем называть одну из $n+1$ позиций: начало слова (*начальный* узел), конец слова (*конечный* узел) или один из $n-1$ промежутков между его буквами (*обычный* узел).

Каждая пара узлов определяет некоторое подслово в u, возможно пустое.

Определение 3.4. Пусть $U = (U_1, U_2, \dots, U_n)$ и $u = (u_1, u_2, \dots, u_m)$ — два упорядоченных множества конечных слов над одним и тем же алфавитом. Для каждого $U_i \in U$ определим множество *интересных* узлов: начальный, конечный, а также все те узлы, которые являются концами или началами каких-либо вхождений слов из u.

Назовём *схемой расположения подслов* для U и u следующую пару:

1) упорядоченное множество из n чисел $t_1, t_2, \dots, t_n$, где t_i — это число интересных узлов в слове U_i;
2) таблица размером $n \times m$, в клетках которой находятся множества упорядоченных пар чисел, полученные по следующему правилу: в клетке, стоящей на строке i в столбце j, парами чисел, не превосходящих t_i, описываются все вхождения слова u_j в U_i (каждая пара чисел задаёт начало и конец некоторого вхождения).

Такую схему будем обозначать $S(U, u)$.

Пример 3.5. Пусть $n = 3$, $m = 2$, $U_1 = abcabc$, $U_2 = bcabca$, $U_3 = bbbbbc$, $u_1 = abc$, $u_2 = bc$.

В слове U_1 пять интересных узлов: ${}_1a_2bc_3a_4bc_5$.

В слове U_2 также пять интересных узлов: ${}_1bc_2a_3bc_4a_5$.

В слове U_3 три интересных узла: ${}_1bbbb_2bc_3$.

Тогда $S\big((U_1, U_2, U_3), (u_1, u_2)\big)$ состоит из вектора $(5, 5, 3)$ и следующей таблицы:

(1,3), (3,5)	(2,3), (4,5)
(2,4)	(1,2), (3,4)
∅	(2,3)

.

3.2. Схемы вхождений подслов, связанные с итерациями подстановки

Определение 3.6. Упорядоченное множество подслов слова $\varphi^\infty(a_1)$, состоящих из одной или двух букв, назовём *порождающими словами* и будем обозначать G. Согласно, например, предложению 4.2, все слова из G алгоритмически находятся, порядок на G выбирается произвольный.

Введём обозначение

$$U^k := \{h(\varphi^k(g)) \mid g \in G\}.$$

Мы считаем, что элементы множества U^k упорядочены так же, как соответствующие элементы G.

Пример 3.7. Найдём $S(U^5, U^4)$ для подстановочной системы Фибоначчи.
$A = B = \{a, b\}$, $\varphi(a) = ab$, $\varphi(b) = a$, $h = \mathrm{Id}$.
$h(\varphi^\infty(a) = abaababaab\ldots$.
$G = (a, b, aa, ab, ba)$.
$U^4 = (abaababa, abaab, abaababaabaababa, abaababaabaab, abaababaababa)$.
U^5 с отмеченными интересными узлами:

$$({}_1abaab_2aba_3ab_4aab_5, {}_1abaab_2aba_3, {}_1abaab_2aba_3ab_4aab_5aba_6ab_7aba_8ab_9aab_{10},$$
$$ {}_1abaab_2aba_3ab_4aab_5aba_6ab_7aba_8, {}_1abaab_2aba_3ab_4aab_5aba_6ab_7aab_8).$$

Таким образом, $S(U^5, U^4)$ состоит из вектора $(5, 3, 10, 8, 8)$ и следующей таблицы:

(1,3)	(1,2), (2,4), (3,5)	∅	(1,5)	∅
(1,3)	(1,2)	∅	∅	∅
(1,3), (3,6), (5,8)	(1,2), (2,4), (3,5), (5,7), (7,9), (8,10)	(1,6)	(1,5), (5,10)	(3,8)
(1,3), (3,6), (5,8)	(1,2), (2,4), (3,5), (5,7)	(1,6)	(1,5)	(3,8)
(1,3), (3,6)	(1,2), (2,4), (3,5), (5,7), (6,8)	(1,6)	(1,5), (3,8)	∅

.

Теорема 3.8. *Можно алгоритмически (по подстановочной системе) найти такое число d, что из условий $k_1 > l_1 + d$, $k_2 > l_2 + d$ и $S(U^{k_1}, U^{l_1}) = S(U^{k_2}, U^{l_2})$ следует $S(U^{k_1+1}, U^{l_1+}1) = S(U^{k_2+1}, U^{l_2+1})$. Равенство схем вхождения подслов следует понимать как совпадение таблиц.*

Доказательство. Это утверждение можно интерпретировать следующим образом: $S(U^{k+1}, U^{l+1})$ можно однозначно определить по $S(U^k, U^l)$.

Предложение 3.9. *$S(U^k, U^{l+1})$ можно однозначно определить по $S(U^k, U^l)$.*

Доказательство. Пусть $\varphi(g_i) = a_{i_1}a_{i_2}\ldots a_{i_t}$. Тогда

$$U_i^{l+1} = h(\varphi^{l+1}(g_i)) = h(\varphi^l(a_{i_1}))h(\varphi^l(a_{i_2}))\ldots h(\varphi^l(a_{i_t})).$$

Каждое из $h(\varphi^l(a_i))$ принадлежит U^l, поэтому у нас есть проверяемый критерий: пара узлов (p, p') слова $u \in U^k$ задаёт некоторое вхождение слова U_i^{l+1} тогда и только тогда, когда существует последовательность узлов $p_0 = p$,

$p_1, \ldots, p_t = p'$, такая что (p_0, p_1) — вхождение слова $h\bigl(\varphi^l(a_{i_1})\bigr)$, (p_1, p_2) — вхождение слова $h\bigl(\varphi^l(a_{i_2})\bigr)$,…, (p_{t-1}, p_t) — вхождение слова $h\bigl(\varphi^l(a_{i_t})\bigr)$. $\square$

Если мы знаем, как именно внутри слова v расположены все вхождения слов A и B, то мы можем определить и все вхождения слова AB. Но если слово v не входит ни в A, ни в B, то в AB оно может как входить, так и не входить. Поэтому переход от $S(U^k, U^l)$ к $S(U^{k+1}, U^l)$ менее тривиальный.

Лемма 3.10. *Пусть про конечные слова $A, B, C, u_1, u_2, \ldots, u_t$ известно, что $|B| \geqslant 2\max_i(|u_i|)$, а также известна схема $S\bigl((AB, B, BC), (u_1 \ldots u_t)\bigr)$. Тогда можно найти $S\bigl(ABC, (u_1 \ldots u_t)\bigr)$.*

Доказательство. В слове ABC вхождение B, идущее после A и перед C, назовём *центральной* частью. *Центральными левыми* узлами назовём узлы, являющиеся левыми концами каких-либо вхождений слов из (u_i), целиком лежащих в центральной части. Аналогично введём определение *центральных правых* узлов. Объединение этих множеств назовём *центральными* узлами.

Пусть W — конечное слово и фиксировано какое-то вхождение слова W в ABC. Тогда числам в $S(W, (u_i))$ соответствуют узлы слова ABC, а парам чисел — подслова слова ABC. Среди вхождений B в ABC выберем центральную часть, также естественным образом фиксируются вхождения AB (начало) и BC (конец).

Центральным левым узлам взаимно-однозначно соответствуют числа i, встречающееся в таблице $S\bigl(B, (u_i)\bigr)$ в парах первыми членами, т. е. как $(i, \cdot)$. Так по схеме $S(B, (u_i))$ можно найти количество центральных левых узлов. Также центральным левым узлам соответствуют некоторые из чисел в таблице $S\bigl(AB, (u_i)\bigr)$, встречающиеся в парах первыми членами.

Пусть какому-то центральному левому узлу в $S\bigl(AB, (u_i)\bigr)$ соответствует число i. Тогда если в таблице $S\bigl(AB, (u_i)\bigr)$ встречается пара вида $(i', \cdot)$, где $i' > i$, то числу i' также соответствует некоторый центральный левый узел. Таким образом, по схемам вхождения подслов можно найти все числа, которым в $S\bigl(AB, (u_i)\bigr)$ соответствуют центральные левые узлы (поскольку известно общее количество таких чисел). Далее можно найти все числа, соответствующие в $S\bigl(AB, (u_i)\bigr)$ центральным правым узлам. Числа из $S\bigl(AB, (u_i)\bigr)$, которым не соответствуют центральные узлы, назовём *левыми периферийными*. Каждому левому периферийному числу соответствует интересный узел в ABC, исключение — число r, кодирующее конец слова AB, которое может не соответствовать интересному узлу в ABC.

Каждый узел слова ABC, соответствующий левому периферийному числу, являются концом или началом некоторого вхождения u_i, пересекающегося с начальным вхождением слова A. Аналогично можно выяснить, каким числам из $S\bigl(BC, (u_i)\bigr)$ соответствуют центральные узлы, и ввести определение *правых периферийных* чисел. Также по схеме выясняется, каким парам чисел, где одно число из $S\bigl(AB, (u_i)\bigr)$, а второе — из $S\bigl(BC, (u_i)\bigr)$, соответствует один и тот же центральный узел.

Так как $|B| \geqslant 2\max(|u_i|)$, то все узлы, соответствующие левым периферийным числам, расположены строго левее, чем узлы, соответствующие правым периферийным числам.

Так как $|B| \geqslant 2\max(|u_i|)$, то каждое вхождение слова из (u_i) в ABC соответствует хоть какой-то паре чисел из схемы вхождения подслов, а все интересные узлы в ABC либо центральные, либо соответствуют правым периферийным числам, либо соответствуют левым периферийным числам.

Таким образом, восстанавливается схема $ABC, (u_i)$. □

Пример 3.11. Пусть дана схема $S\big((AB, B, BC), (u_1)\big)$. Вектор: $(5, 4, 6)$, таблица:

(1,2), (2,3), (3,4)
(2,3)
(2,3), (3,4), (4,5)

.

Из второй строчки делаем вывод, что есть один центральный левый узел и один центральный правый узел. В первой строчке центральному левому узлу соответствует цифра 2, а центральному правому — цифра 3. В третьей строчке этим узлам соответствуют 1 и 2.

Значит, в ABC семь интересных узлов (два центральных, два соответствующих левым периферийным числам и три соответствующих правым периферийным числам).

Таблица будет выглядеть так:

(1,2), (2,3), (3,4), (4,5), (5,6)

.

Следствие 3.12. *Если известно, что любое из слов U^k хотя бы в два раза длиннее любого из слов U^l, то $S(U^{k+1}, U^l)$ однозначно определяется по $S(U^k, U^l)$.*

Доказательство. Нас интересует схема расположения подслов для слова $u \in U^{k+1}$ и множества U^l. Слово u представляется в виде конкатенации $u = u_1 u_2 \ldots u_t$, где все u_i принадлежат U^k и все $u_i u_{i+1}$ принадлежат U^k. По правилу, описанному в лемме 3.10, найдём все схемы вхождения подслов вида $S(u_i u_{i+1} u_{i+2}, U^l)$, потом $S(u_i u_{i+1} u_{i+2} u_{i+3}, U^l)$ и т. д. Действуя так, доберёмся до $S(u_1 u_2 \ldots u_t, U^l)$. □

Согласно предложению 3.1 можно найти (алгоритмически) такое d, что из условия $k > l + d$ следует, что

$$\min_{u \in U^k} |u| > 2 \max_{u \in U^l} |u|.$$

Пусть $k - l > d$ и $S(U^{k_1}, U^{l_1}) = S(U^{k_2}, U^{l_2})$. Из следствия 3.12 следует, что $S(U^{k_1+1}, U^{l_1}) = S(U^{k_2+1}, U^{l_2})$ и $S(U^{k_1+1}, U^{l_1+1}) = S(U^{k_2+1}, U^{l_2+1})$. □

3.3. Алгоритм

В дальнейшем полагаем, что d — это число d, возникающее в теореме 3.8. Оно может быть эффективно найдено как функция от (φ, h).

Размером схемы вхождения подслов будем называть наибольшее количество элементов в клетке соответствующей таблицы. Так, размер схемы из примера 3.7 равен 5.

Найдём такое число N, что $N > 2K_2K\lambda^d/K_1$, где K — константа равномерной рекуррентности, а K_1, K_2, λ — числа из предложения 3.1 (заметим, что точно эти числа мы найти не можем, но можем получить на них какую-то алгоритмическую оценку).

Алгоритм 3.13.

1. Выбрать произвольное k_0.
2. Последовательно для всех натуральных i строить $S(U^{k_0+i+d}, U^{k_0+i})$, пока не будут выполнены условия пункта 3 или 4.
3. Если какие-то две из построенных схем совпали, то слово $h\big(\varphi^\infty(a_1)\big)$ непериодично.
4. Если размер некоторой схемы $S(U^{k_0+i+d}, U^{k_0+i})$ больше, чем N, то $h\big(\varphi^\infty(a_1)\big)$ периодично и длина периода не превосходит $\max\limits_{u \in U^{k_0+i}} |u|$.

Доказательство. Несложно понять, что различных схем размера, не превосходящего N, может встретится лишь конечное число. Значит, либо условия пункта 3, либо условия пункта 4 на каком-то шаге окажутся выполнены.

Пусть выполнены условия пункта 3 и

$$S(U^{k_0+i_1+d}, U^{k_0+i_1}) = S(U^{k_0+i_2+d}, U^{k_0+i_2}).$$

Тогда согласно теореме 3.8 последовательность схем будет периодична с длиной периода $i_2 - i_1$, в частности, размеры всех схем в последовательности будут ограничены. Но если $|u_1| > 2|u_2|$, $u_1, u_2 \in (\varphi^\infty(a_1))$, $\big(\varphi^\infty(a_1)\big)$ периодично с длиной периода p, то у слова u_2 не менее чем $|u_2|/p$ вхождений в u_1. Выбирая $u_1 \in U^{i+d}$, $u_2 \in U^i$ для достаточно большого i, получаем противоречие с периодичностью.

Пусть выполнены условия пункта 4 и нашлись два слова $u_1 \in U^{i+d}$, $u_2 \in U^i$, такие что у u_2 не менее чем N вхождений в u_1. По принципу Дирихле какие-то два вхождения расположены друг относительно друга на расстоянии x, не превосходящем $|u_1|/N$. Согласно предложению 3.1

$$\frac{|u_1|}{|u_2|} \leqslant \frac{K_2\lambda^d}{K_1}.$$

Из выбора N следует, что $Kx < |u_2|$. Рассмотрим эти два вхождения слова u_2:

$$\overbrace{b_1 \dots b_x\, b_{x+1} b_{x+2} \dots b_{k'-1} b_{k'}}^{u_2} b_{k'+1} \dots b_{k'+x}.$$

$$b_1 \dots b_x\, \underbrace{b_{x+1} b_{x+2} \dots b_{k'-1} b_{k'} b_{k'+1} \dots b_{k'+x}}_{u_2}.$$

Несложно убедиться, что любые две буквы, находящиеся в слове u_2 на расстоянии x, одинаковы, т. е. любое подслово длины x слова u_2 является циклическим сдвигом слова $b_1b_2\ldots b_x$. Из определения константы равномерной рекуррентности следует, что любое подслово слова $hbigl(\varphi^\infty(a_1)\big)$ является подсловом слова u_2 и, следовательно, циклическим сдвигом слова $b_1b_2\ldots b_x$. Отсюда следует, что $h\big(\varphi^\infty(a_1)\big)$ периодично с длиной периода x. □

4. Сведение общего случая к примитивному

Сперва воспользуемся следующим классическим в комбинаторике слов результатом.

Теорема 4.1 [2, гл. 7]. *Если $f\colon A^* \to B^*$ и $g\colon A^* \to A^*$ — произвольные морфизмы и $f\big(g^\infty(a_1)\big)$ — бесконечное слово, то можно найти такие алфавит A', букву $a'_1 \in A'$, нестирающую подстановку φ, действующую на алфавите A и кодирование $\tau\colon A' \to B$, что $f\big(g^\infty(a_1)\big) = \tau\big(\varphi^\infty(a'_1)\big)$.*

Далее считаем φ нестирающей подстановкой, а h — кодированием.

Предложение 4.2. *Пусть φ — подстановка, действующая на алфавите A, h — морфизм из A^* в B^*, u — конечное слово из B^*. Тогда существует алгоритм, проверяющий, сколько раз u встречается в слове $W = h\big(\varphi^\infty(a_1)\big)$.*

Доказательство. Пусть $|u| = n$. Очевидно, можно считать, что все буквы алфавита A встретятся в $\varphi^\infty(a_1)$. Также можно считать, что все буквы алфавита встречаются в $\varphi(a_1)$ (иначе возьмём нужную степень морфизма). Если $a \in A$, то $\chi_n(a)$ — это число вхождений слова u в $h\big(\varphi^k(a)\big)$, $l_k(a)$ и $r_k(a)$ — это два слова длины n, являющиеся соответственно началом и концом $h\big(\varphi^k(a)\big)$. Если $\big|h\big(\varphi^k(a)\big)\big| < n$, то $l_k(a) = r_k(a) = h\big(\varphi^k(a)\big)$.

Обозначим Ω_k набор из $2|A|$ слов: $r_k(a_i)$ и $l_k(a_i)$ для всех $a_i \in A$. Числа $\chi_k(a_i)$ образуют вектор χ_k.

Если $\varphi(a_i) = a_{i_1}a_{i_2}\ldots a_{i_m}$, то

$$h\big(\varphi^{k+1}(a_i)\big) = h\big(\varphi^k(a_{i_1})\big)h\big(\varphi^k(a_{i_2})\big)\ldots h\big(\varphi^k(a_{i_m})\big).$$

Слово l_{k+1} получается, если взять первые n букв слова $l_k(a_{i_1})l_k(a_{i_2})\ldots l_k(a_{i_k})$, а слово r_{k+1} — если взять последние n букв слова $r_k(a_{i_1})r_k(a_{i_2})\ldots r_k(a_{i_k})$. Стало быть, набор Ω_{k+1} однозначно определяется по набору Ω_k. Так как различных наборов не больше чем $|B|^{2|A|n}$, то последовательность Ω_k зациклится.

Слово $h\big(\varphi^{k+1}(a_i)\big)$ делится на блоки $h\big(\varphi^k(a_{i_k})\big)$, и вхождения u в $h\big(\varphi^{k+1}(a_i)\big)$ бывают двух типов: те, которые целиком лежат в каком-либо блоке, и те, которые принадлежат хотя бы двум блокам. Количество первых равно $\chi_k(a_{i_1}) + \chi_k(a_{i_2}) + \ldots + \chi_k(a_{i_m})$. Это число равно i-й компоненте вектора $M\chi_k$, где M — это матрица подстановки φ.

Количество вхождений, принадлежащих хотя бы двум блокам, можно вычислить по Ω_k. Таким образом, $\chi_{k+1} = A\chi_k + f(\Omega_k)$.

Пусть последовательность наборов Ω_k периодична начиная с $k = k_0$ и длина периода равна T. Тогда получим

$$\chi_{k_0+(k+1)T} = M\Big(M\big(\ldots\big(M\chi_{k_0+kT} + f(\Omega_{k_0+kT})\big) + \ldots\big)\Big) + \\ + f(\Omega_{k_0+(k+1)T-2}) + f(\Omega_{k_0+(k+1)T-1}).$$

Раскроем все скобки: $\chi_{k_0+(k+1)T} = M^T\chi_{k_0+kT} + C$ для некоторого вектора C с неотрицательными компонентами. Это значит, что можно определить, стремится ли $\chi_k(a_1)$ к бесконечности при k, стремящемся к бесконечности. Если стремится, то u встречается в W бесконечно много раз. Иначе можно найти предел $\chi_k(a_1)$, который равен числу вхождений u в W. □

Следствие 4.3. *Можно определить, верно ли, что сверхслово заключительно периодично с данным периодом u, т. е. верно ли, что все слова длины $|u|$, которые бесконечно много раз встречаются в W, являются циклическими сдвигами u.*

Далее φ — нестирающий морфизм из A^* в A^*, h — кодирование из A в B, $W = h\big(\varphi^\infty(a_1)\big)$.

Слово $w \in A^*$ будем называть *φ-ограниченным*, если последовательность

$$w, \varphi(w), \varphi^2(w), \varphi^3(w), \ldots$$

периодична начиная с некоторого момента. В противном случае $|\varphi^n(w)| \to \infty$ при $n \to \infty$; слово w называется *φ-растущим*. Очевидно, слово является φ-ограниченным тогда и только тогда, когда оно состоит из φ-ограниченных букв.

Теорема 4.4. *Существует алгоритм, который определяет, конечно ли в $\varphi^\infty(a)$ число различных φ-ограниченных подслов. Если это число бесконечно, то можно проверить, является ли W периодическим. Если оно конечно, то все эти слова алгоритмически находятся.*

Доказательство. Прежде всего отметим, что все φ-возрастающие буквы алгоритмически находятся (см., например, [11]). Далее φ-растущие буквы будем писать как a_1, a_2 и т. д.

Построим ориентированный граф Q, на рёбрах которого будут записаны упорядоченные пары слов. Вершинами этого графа будут служить φ-растущие буквы из A, а также всевозможные упорядоченные пары φ-растущих букв. Введём фиктивную букву t, также к вершинам Q добавим всевозможные пары вида a_it, где a_i — φ-растущая буква.

Из вершины a_i в a_j идёт ребро, если $a_j \in a_i$. На таких рёбрах пара слов — $\{\varepsilon, \varepsilon\}$ (ε — пустое слово). Из вершины a_i в a_ja_k ведёт ребро со словами $\{\omega, \varepsilon\}$, если для некоторого φ-ограниченного слова ω слово $a_j\omega a_k$ является подсловом $\varphi(a_i)$ (из a_i в a_ja_k могут вести несколько рёбер). Из a_i и a_it ведёт по ребру с парой $\{\omega, \varepsilon\}$ в a_jt, если ω φ-ограниченное и $\varphi(a_i)$ оканчивается на $a_j\omega$.

Из a_ia_j ведёт ребро в a_ka_l с парой φ-ограниченных слов $\{\omega_1; \omega_2\}$, если $\varphi(a_1)$ кончается на $a_k\omega_1$, а $\varphi(a_2)$ начинается на ω_2a_l.

Предложение 4.5. *Пусть $k \in \mathbb{N}$. Рассмотрим какой-нибудь путь длины k по рёбрам графа Q, выходящий из a_1. Последовательность пар слов на рёбрах этого пути*

$$\{u_1, v_1\}, \{u_2, v_2\}, \ldots, \{u_k, v_k\}.$$

Тогда в $\varphi^k(a_1)$ есть φ-нерасширяемое слово

$$u_k\varphi(u_{k-1})\ldots\varphi^{k-1}(u_1)\varphi^{k-1}(v_1)\varphi^{k-2}(v_2)\ldots v_k.$$

При этом если путь оканчивается на a_ia_j, то есть вхождение этого слова, обрамлённое буквами a_i и a_j. Наоборот, любое φ-ограниченное подслово $\varphi^k(a_1)$ можно построить, получив по указанному правилу слово и взяв его подслово.

Это несложно показывается индукцией по k.

Первый случай: в любом ориентированном цикле графа Q, до которого можно добраться из a_1, на рёбрах цикла написаны пары пустых слов. Тогда в любом пути, выходящем из вершины a_1, число рёбер, на которых написаны непустые слова, не превосходит количества вершин в Q. Следовательно, число различных φ-ограниченных слов конечно, и по графу Q можно их все найти.

Второй случай: есть цикл, и в этом цикле не все слова пустые. Пусть, например, в цикле есть пары слов, в которых первое слово не пустое. Тогда для некоторой буквы a_i и непустого φ-ограниченного u, слово $\varphi(a_i)$ оканчивается на a_iu. Образ u при подстановке φ зациклится. Следовательно, для некоторого непустого слова U для любого $k \in \mathbb{N}$ слово $h(U^k)$ является подсловом W. Тогда сверхслово W заключительно периодично, если и только если оно заключительно периодично с периодом $h(U)$. А проверять, есть ли заключительная периодичность с данным периодом, мы умеем (см. следствие 4.3). □

Пусть I_φ — множество всех φ-растущих букв, B_φ — множество φ-ограниченных подслов сверхслова $\varphi^\infty(a_1)$ (включая пустое). Можно считать, что B_φ конечно и что мы знаем все слова в B_φ. Рассмотрим (конечный) алфавит C, состоящий из символов $[twt']$, где t и t' — буквы из I_φ, а w — слово из B_φ и слово twt' является подсловом $\varphi^\infty(a)$.

Определим морфизм $\psi\colon C^* \to C^*$ следующим образом:

$$\psi([twt']) = [t_1wt_2][t_2wt_3]\ldots[t_kw_kt_{k+1}],$$

где $\varphi(tw) = w_0t_1w_1t_2\ldots t_kw'_k$, слово $\varphi(t')$ начинается с w''_kt_{k+1} и $w_k = w'_kw''_k$ (слова w_i, w'_k и w''_k принадлежат B_φ).

Также определим $f\colon C^* \to A^*$ по правилу

$$f([twt']) = tw.$$

Предложение 4.6. *Все буквы алфавита C являются ψ-растущими.*

Доказательство. Заметим, что в $\psi^n([twt'])$ столько же букв, сколько в слове $\varphi^n(t)$ φ-растущих букв. Очевидно, в образе $\varphi(t)$ от произвольной буквы $t \in I_\varphi$ содержится хотя бы одна буква из I_φ. Более того, в слове $\varphi^n(t)$ для некоторого n содержатся хотя бы две буквы из I_φ, иначе $\varphi^n(t) = w_nt_{i_n}v_n$ (где w_n и v_n принадлежат B_φ) и $|\varphi^nt|$ ограниченно. □

Пусть $\varphi^\infty(a)$ имеет вид $a_1 w_1 a_2 \ldots$, где $a_1, a_2 \in I_\varphi$, $w_1 \in B_\varphi$. Тогда, как несложно убедиться, для любого n слово $\varphi^n(a)$ является началом слова $f\big(\psi^n([a_1 w_1 a_2])\big)$, следовательно, слова W и $(h \circ f)\big(\psi^\infty([a_1 w_1 a_2])\big)$ совпадают. Заметим, что морфизм $h' = h \circ f$ нестирающий.

Замечание 4.7. Конструкция морфизма ψ встречалась в работах [11, 12].

Букву $c_i \in C$ назовём *рекуррентной*, если для некоторого $k \in \mathbb{N}$ выполнено $c_i \sqsubseteq \psi^k(c_i)$. Для каждой рекуррентной буквы $c_i \in C$ существует такое число $k(c_i)$, что если $k(c_i) \mid n$, то $c_i \sqsubseteq \psi^n(c_i)$. Следовательно, существует такое $n \in \mathbb{N}$, что если c_i — произвольная рекуррентная буква алфавита C, то $c_i \sqsubseteq \psi^n([a_1 w_1 a_2])$. Положим $\rho = \psi^n$.

Рассмотрим ориентированный граф G_ρ, вершинами которого являются буквы алфавита C и из c_i ведёт стрелка в c_j тогда и только тогда, когда c_j содержится в $\rho(c_i)$.

Пусть D — сильносвязная компонента этого графа, до которой можно дойти по стрелкам из $[a_1 w a_2]$. Рассмотрим ограничение ρ на D^*. Все буквы из D являются рекуррентными. Следовательно, если $d \in D$, то $d \sqsubseteq \rho(d)$. Также для любого k выполнено $\rho^k(d) \sqsubseteq \rho^{k+1}(d)$. Следовательно, существует такое m, что для любых букв $d_1, d_2 \in D$ $d_2 \sqsubseteq \rho^m(d_1)$. Поэтому морфизм ρ в ограничении на D^* является примитивным.

Найдётся такая буква $d \in D$, что $\rho^l(d)$ начинается на d для некоторого l. Обозначим $\rho_2 = \rho^l$. Так как все буквы из D являются ρ_2-растущими, то $\rho_2^\infty(d)$ — бесконечное сверхслово, при этом все его конечные подслова являются подсловами $\varphi^\infty(a_1)$.

Пользуясь алгоритмом, построенным в разделе 3, определим периодичность сверхслова $h'\big(\psi^\infty(d)\big)$. Если сверхслово непериодично, то в W бесконечно много специальных справа подслов и само W не является заключительно периодичным.

Если же сверхслово $h'\big(\psi^\infty(d)\big)$ периодично и его период — слово u, то, если W заключительно периодично, его периодом является то же самое слово u. Согласно следствию 4.3 мы можем проверить, верно ли, что $h'\big(\psi^\infty(d)\big)$ заключительно периодично с периодом u. Таким образом, теорема 1.1 доказана.

Литература

[1] Мучник Ан. А., Притыкин Ю. Л., Семёнов А. Л. Последовательности, близкие к периодическим // Успехи мат. наук. — 2009. — Т. 64, № 5. — С. 21—96.

[2] Allouche J.-P., Shallit J. Automatic Sequences: Theory, Applications, Generalizations. — Cambridge: Cambridge Univ. Press, 2003.

[3] Durand F. Decidability of the HD0L ultimate periodicity problem. — arXiv:1111.3268v1.

[4] Durand F. HD0L ω-equivalence and periodicity problems in the primitive case (to the memory of G. Rauzy) // J. Uniform Distrib. Theory.

[5] Ehrenfeucht A., Rozenberg G. Repetition of subwords in DOL languages // J. Inform. Control. — 1983. — Vol. 59, no. 1-3. — P. 13—35.

[6] Harju T., Linna M. On the periodicity of morphisms on free monoids // Inform. Théor. Appl. — 1986. — Vol. 20, no. 1. — P. 47—54.

[7] Honkala J. A decision method for the recognizability of sets defined by number systems // Inform. Théor. Appl. — 1986. — Vol. 20, no. 4. — P. 395—403.

[8] Honkala J., Rigo M. Decidability questions related to abstract numeration systems // Discrete Math. — 2004. — Vol. 285, no. 1-3. — P. 329—333.

[9] Kanel-Belov A., Mitrofanov I. Periodicity of Rauzy scheme and substitutional systems. — arXiv:1107.0185 [math.DS].

[10] Mitrofanov I. A proof for the decidability of HD0L ultimate periodicity. — arXiv:1110.4780.

[11] Nicolas F., Pritykin Yu. On uniformly recurrent morphic sequences // Int. J. Foundations Comput. Sci. — 2009. — Vol. 20, no. 5. — P. 919—940.

[12] Pansiot J.-J. Complexité des facteurs des mots infinis engendrés par morphismes itérés // Proc. ICALP'84. — Berlin: Springer, 1984. — (Lect. Notes Comput. Sci.; Vol. 172). — P. 380—389.

[13] Pansiot J.-J. Decidability of periodicity for infinite words // Inform. Théor. Appl. — 1986. — Vol. 20, no. 1. — P. 43—46.

[14] Rigo M., Maes A. More on generalized automatic sequences // J. Automata, Languages and Combinatorics. — 2002. — Vol. 7, no. 3. — P. 351—376.