

**ЯЗЫКИ С КОНЕЧНЫМ АНТИСЛОВАРЕМ:
ИНДЕКС РОСТА И СВОЙСТВА АВТОМАТОВ***

Комбинаторная сложность является наиболее естественной «подсчитывающей» функцией, сопоставляемой формальному языку. Для языка L над конечным алфавитом Σ она определяется равенством $C_L(n) = |L \cap \Sigma^n|$. Изучению комбинаторной сложности для различных языков посвящен достаточно большой массив работ; впервые эта функция изучалась еще Морсом и Хэдлундом [1], а систематические исследования начались с цикла работ Эренфойхта и Розенберга, начатого работой [2]. При этом к настоящему времени проблематика далеко не исчерпана.

Характеризация различных множеств на языке запрещенных объектов часто встречается в алгебре (например, для многообразий алгебраических структур) и дискретной математике (например, характеристика графов запрещенными подграфами). Не менее популярен такой способ характеристики формальных языков и в самом простом виде (список запрещенных в языке слов, или *антисловарь*), и в более сложном (например, список запрещенных шаблонов). Этим способом можно описать любой *факториальный* (т. е. замкнутый относительно взятия подслов) язык.

Оценить комбинаторную сложность любого факториального языка можно при помощи языков с конечными антисловарями (КАС-языков). КАС-языки являются регулярными, и для них существует алгоритм вычисления комбинаторной сложности. Впервые подобный метод оценки для формальных языков применил Бранденбург [3], хотя аналогичные идеи для оценки функций роста алгебр применялись и раньше (см., например, [4]).

В связи с подобной «универсальностью» класса КАС-языков естественным является вопрос описания комбинаторной сложности для самих этих языков. КАС-языки образуют собственный подкласс класса регулярных языков, для которых общее описание комбинаторной сложности недавно завершено [5]. Однако априори совершенно не ясно, какие из функций комбинаторной сложности регулярных языков могут быть реализованы внутри данного подкласса. Комбинаторная сложность регулярного языка определяется структурой распознающего его детерминированного конечного автомата. Поскольку ограничения на вид конечных автоматов, распознающих КАС-языки,

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

весьма серьезны, естественно ожидать, что функции, могущие являться комбинаторной сложностью КАС-языка, значительно менее разнообразны, чем такие функции для произвольных регулярных языков.

В работе [6], однако, показано, что комбинаторной сложностью КАС-языка над любым неодноэлементным алфавитом может быть полином любой степени, как и для произвольных регулярных языков. В данной работе мы начинаем изучение более сложного случая – когда комбинаторная сложность экспоненциальна.

Мы вводим класс детерминированных конечных автоматов, которые разумно считать «стандартными» для КАС-языков, устанавливаем ряд графовых и комбинаторных свойств таких автоматов и выделяем С-граф – основную часть такого автомата, ответственную за порядок роста комбинаторной сложности. Далее, мы описываем процедуру редукции конечного антисловаря, сохраняющую порядок роста комбинаторной сложности, и более глубоко изучаем свойства автоматов, распознающих языки с редуцированным антисловарем.

Наконец, на основании полученных свойств мы получаем полный список С-графов не более чем с четырьмя вершинами для языков в двухбуквенном алфавите. Список состоит из 25 графов, имеющих 14 различных индексов экспоненциального роста комбинаторной сложности.

1. Предварительные сведения

Мы рассматриваем конечные слова над конечным неодноэлементным алфавитом Σ . Длина слова W обозначается через $|W|$. Символ λ обозначает пустое слово. Слово U является *подсловом* (соответственно *префиксом*, *суффиксом*) слова W , если W можно представить в виде PUQ (соответственно UQ , PU) для некоторых (возможно, пустых) слов P и Q . Подслово (префикс, суффикс) W называется *собственным*, если оно не совпадает с самим W . Слово $W = a_1 \dots a_n$ имеет *период* $p > 0$, если $a_i = a_{i+p}$ для всех $i = 1, \dots, |W| - p$. Как обычно, U^n означает конкатенацию n копий слова U , Σ^n – множество всех n -буквенных слов, а Σ^* – множество всех слов над Σ . Язык $L \subseteq \Sigma^*$ называется *факториальным*, если он содержит все подслова входящих в него слов, и *антифакториальным*, если ни одно слово из L не является подсловом другого слова из L .

Слова U и W называются *сопряженными*, если $U = PQ$, $W = QP$ для некоторых слов P и Q . Сопряженность является отношением эквивалентности на Σ^* . Класс сопряженности можно отождествить с *циклическим словом*, т. е. со словом, в котором первая и последняя буквы считаются соседними и начальная позиция не определена; все слова из класса получаются из цик-

лического слова указанием начальной позиции. Слово W называется *примитивным*, если $W = U^n$ влечет $n = 1$. Поскольку периоды сопряженных слов совпадают, можно говорить и о *примитивных циклических* словах.

Мы рассматриваем *детерминированные конечные автоматы* (ДКА) с *частичной* функцией переходов. Такой автомат полностью описывается орграфом, вершинами которого являются состояния, а (помеченными) дугами – переходы. Таким образом, каждый путь в орграфе помечен словом, а каждый цикл – циклическим словом. Если мы рассматриваем метку цикла как обычное слово, мы указываем вершину цикла, с которой начинается чтение. Любой путь из начальной вершины в терминальную называется *распознающим*. ДКА называется *плотным*, если любая его вершина принадлежит некоторому распознающему пути. ДКА называется *древовидным*, если он является деревом с корнем в качестве начальной вершины и всеми листьями в качестве терминальных вершин.

Комбинаторная сложность языка $L \subseteq \Sigma^*$ – это функция, определяемая равенством $C_L(n) = |L \cap \Sigma^n|$. Комбинаторная сложность факториального языка либо ограничена константой, либо строго возрастает [7]. Как обычно, мы называем функцию *полиномиальной*, если она ограничена сверху некоторым полиномом, и *экспоненциальной*, если она ограничена снизу некоторой экспонентой с основанием, большим 1. *Индексом роста* языка L называется верхний предел $\alpha = \overline{\lim}_{n \rightarrow \infty} \sqrt[n]{C_L(n)}$. Основные свойства индекса роста для факториальных языков описаны в следующем утверждении:

Предложение 1.1. *Индекс роста факториального языка $L \in \Sigma^*$ определяется как $\alpha(L) = \lim_{n \rightarrow \infty} \sqrt[n]{C_L(n)} = \inf_{n \in \mathbb{N}} \sqrt[n]{C_L(n)}$. При этом*

- 1) $\alpha(L) = 0 \iff L$ конечен;
- 2) $1 < \alpha(L) \leq |\Sigma| \iff C_L$ экспоненциальна;
- 3) $\alpha(L) = 1$ в остальных случаях.

Для *регулярных* факториальных языков верно следующее утверждение.

Предложение 1.2 ([8]). *Пусть факториальный язык L распознается плотным ДКА A . Тогда функция $C_L(n)$ экспоненциальна, если в A найдутся два цикла с общей вершиной, и полиномиальна в противном случае.*

Слово W *запрещено* в языке L , если оно не является подсловом никакого элемента L . Запрещенное слово *минимально*, если никакое его собственное подслово не запрещено. Множество всех минимальных запрещенных в L слов называется *антисловарем* L . Антисловарь всегда является антифакториальным языком. Нетрудно видеть, что факториальный язык $L \subseteq \Sigma$ и его антисловарь M связаны соотношениями

$$L = \Sigma^* \setminus \Sigma^* M \Sigma^*, \quad M = \Sigma L \cap L \Sigma \cap (\Sigma^* \setminus L).$$

Таким образом, языки L и M однозначно определяют друг друга, причем L регулярен тогда и только тогда, когда регулярен M . В частности, интересные нас языки с *конечным* антисловарем (КАС-языки) образуют собственный подкласс класса факториальных регулярных языков. В [9] был предложен общий алгоритм вычисления индекса роста КАС-языков, основанный на идеях [4].

Алгоритм 1.

Вход: антисловарь M над алфавитом Σ .

Выход: индекс роста α языка L с антисловарем M .

Шаг 1. Положить $\ell = \max_{W \in M} |W|$ и вычислить множество $L \cap \Sigma^{\ell-1} = \{U_1, \dots, U_s\}$.

Шаг 2. Вычислить матрицу

$$A[i, j] = \begin{cases} 1, & \text{если } U_i c = d U_j \text{ для некоторых } c, d \in \Sigma; \\ 0 & \text{иначе.} \end{cases}$$

Шаг 3. В качестве α взять наибольшее собственное значение матрицы A .

Построенная на шаге 2 матрица A является матрицей смежности некоторого орграфа, а именно *графа де Брёйна* (о таких графах см. [10]). Наибольшее собственное значение матрицы смежности графа называется *индексом* графа и отвечает за ряд важных свойств, в первую очередь – определяет количество путей заданной длины в графе (см. [11]).

Алгоритм 1 можно сделать более эффективным, так как следующая теорема позволяет использовать матрицы смежности значительно меньших размеров, а приводимый ниже алгоритм 2 – строить распознающие автоматы с такими матрицами смежности. Теорема 1.1 выражает хорошо известную связь комбинаторной сложности регулярного языка с индексом ДКА в наиболее общей форме. Ее доказательство приведено в [12].

Теорема 1.1. *Пусть язык L распознается плотным ДКА \mathcal{A} . Тогда индекс роста языка L совпадает с индексом \mathcal{A} .*

2. Свойства КАС-автоматов

Приводимый ниже алгоритм из [13] строит по конечному антисловарю плотный ДКА, распознающий факториальный язык с этим антисловарем. Такой ДКА, называемый ниже КАС-автоматом, можно сопоставить любому КАС-языку.

Алгоритм 2.

Вход: антисловарь M над алфавитом Σ .

Выход: ДКА \mathcal{A} , распознающий факториальный язык L с анτισловарем M .

Шаг 1. Построить древовидный автомат \mathcal{T} , распознающий M . (M , очевидно, является префиксным кодом, а \mathcal{T} – дерево этого кода.)

Шаг 2. Каждой вершине из \mathcal{T} сопоставить слово, помечающее путь от корня до этой вершины. (Таким образом, множество вершин \mathcal{T} – это множество префиксов слов из M .)

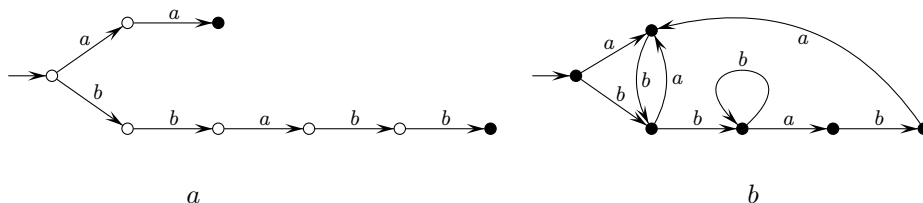
Шаг 3. Добавить к \mathcal{T} все возможные дуги по следующим правилам: дуга (U, W) , помеченная буквой a , добавляется, если

- U – нетерминальная вершина, и
- U не имеет исходящей дуги с меткой a , и
- W – длиннейший из суффиксов слова Ua , являющийся вершиной в \mathcal{T} .

(Такие дуги называются обратными, а дуги из \mathcal{T} – прямыми.)

Шаг 4. Удалить все терминальные вершины и сделать все оставшиеся вершины терминальными, получая автомат \mathcal{A} .

Пример 1. Автоматы \mathcal{T} и \mathcal{A} для анτισловаря $M = \{a^2, bbabb\}$ приведены на рисунках *а*, *б*. По матрице смежности автомата \mathcal{A} можно вычислить ее характеристический многочлен, равный $x^6 - x^5 - x^4 + x^3 - x$, и получить индекс роста $r(L) \sim 1,5551$.



Ниже обозначения \mathcal{T} и \mathcal{A} всегда используются для обозначения дерева и автомата, построенных по алгоритму 2. Ряд полезных свойств автомата \mathcal{A} приведен в следующей лемме.

- Лемма 2.1.** 1) \mathcal{A} является плотным ДКА.
 2) Множество вершин из \mathcal{A} совпадает с множеством всех собственных префиксов слов из M .
 3) Все пути из начальной вершины \mathcal{A} являются распознающими.
 4) Если U – вершина в \mathcal{A} , a – буква и слово Ua является запрещенным, то из U не выходит дуга с меткой a .
 5) Метка любого распознающего пути, заканчивающегося в вершине U , имеет суффикс U .

Доказательство. 1) Автомат \mathcal{A} является детерминированным по построению и плотным, поскольку все его вершины терминальные и достижимы из начальной вершины по дугам из \mathcal{T} .

2) Множество вершин из \mathcal{T} – это множество всех префиксов слов из M . Терминальные вершины в \mathcal{T} – это в точности листья, а они помечены словами из M . Когда мы удаляем эти вершины на шаге 4, остаются только собственные префиксы.

3) Очевидно, так как все вершины из \mathcal{A} являются терминальными.

4) В автомате, построенном на шаге 3, заведомо существует дуга из U с меткой a . Покажем, что эта дуга ведет в терминальную вершину, а значит, будет удалена на шаге 4. Если эта дуга прямая, то она ведет в вершину Ua из \mathcal{T} . Следовательно, Ua является префиксом слова из антисловаря. Так как по условию слово Ua запрещенное, имеем $Ua \in M$, т. е. Ua – терминальная вершина. Пусть теперь рассматриваемая дуга является обратной. Слово U не запрещено ввиду пункта 2). Значит, у запрещенного слова Ua есть запрещенный суффикс; пусть W – кратчайший из таких суффиксов. Тогда $W \in M$, и в то же время W – длиннейший из суффиксов Ua , являющийся вершиной в \mathcal{T} . Таким образом, рассматриваемая дуга ведет в W и W – терминальная вершина. Утверждение доказано.

5) Докажем индукцией по длине распознающего пути. База индукции: путь состоит из единственной дуги. Пусть a – метка этой дуги. Дуга с меткой a ведет из начальной вершины λ либо в вершину a , либо обратно в λ ; в обоих случаях требуемое условие выполнено. Шаг индукции: пусть доказываемое свойство выполнено для всех распознающих путей длины не более n . Выберем путь длины $n+1$. Пусть Wa – его метка, и последняя буква a помечает дугу (U_1, U_2) . Тогда $U_2 = U_1c$, если последняя дуга прямая, и U_2 является суффиксом U_1c , если она обратная. В обоих случаях U_2 – суффикс слова Wc , так как U_1 – суффикс W по предположению индукции. Шаг индукции выполнен.

Утверждение 1 леммы 2.1 позволяет применять к КАС-автоматам теорему 1.1 и тем самым сделать алгоритм 1 более эффективным. Кроме того, КАС-автоматы естественно считать «стандартными» автоматами для изучения КАС-языков: они достаточно компактны (хотя не обязательно минимальны по размеру), их построение производится быстро и, кроме того, в них сохраняется информация об антисловаре. А именно, имеется биекция между множеством вершин КАС-автомата и множеством собственных префиксов слов из антисловаря такая, что выполняется пункт 5 леммы 2.1. Заметим, что отсюда, в частности, следует, что все входящие в вершину дуги имеют равные метки.

Замечание 2.1. Алгоритм 2 можно применять не только к антифакториальным множествам слов. Если H – произвольный конечный префиксный код, алгоритм строит автомат, распознающий все слова, не имеющие

подслов из H . В самом деле, H распознается древовидным автоматом; все последующие шаги алгоритма и дальнейшие выводы остаются в силе.

Сейчас мы докажем несколько графовых свойств КАС-автоматов.

Лемма 2.2. Пусть U – вершина из \mathcal{T} . Если в \mathcal{T} есть вершина cU для любой буквы c , то вершина U не лежит ни в одном цикле автомата \mathcal{A} .

Доказательство. Любой цикл в \mathcal{A} содержит хотя бы одну обратную дугу, так как все прямые дуги принадлежат дереву. Значит, чтобы вершина U принадлежала циклу, необходимо, чтобы существовал путь в U , содержащий обратную дугу. Если такой путь существует, то найдется обратная дуга, ведущая в префикс U (само слово U тоже префикс U). Пусть префикс U' слова U является собственным суффиксом слова Wd для некоторой вершины W и буквы d , т. е. $Wd = W'cU'$, где c – также буква. Поскольку cU' – префикс cU и cU – вершина в \mathcal{T} , то в \mathcal{T} есть и вершина cU' . Значит, U' – обратная дуга из W с меткой d не может вести в U' по определению шага 3 алгоритма. Итак, в \mathcal{A} нет обратных дуг, ведущих в префиксы U . Лемма доказана.

Лемма 2.3. Если вершина S автомата \mathcal{A} принадлежит циклу длины p , то слово S имеет период p или $|S| < p$.

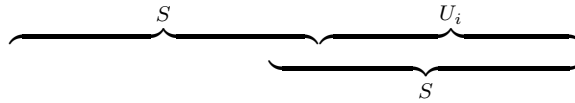
Доказательство. Пусть U – метка данного цикла, прочитанная начиная с вершины S . Тогда S является суффиксом слова SU по утверждению 5 леммы 2.1, т. е. $SU = U'S$. По известному свойству слов (см., например, [14], предложение 1.3.4) из этого равенства следует $U = PQ$, $U' = QP$ и $S = Q(PQ)^n$ ($n \geq 0$) для некоторых слов P и Q . Так как $|PQ| = |U| = p$, получаем требуемое утверждение.

Предложение 2.1. Если вершина S автомата \mathcal{A} принадлежит циклам с длинами p и q , то $|S| < p + q - \gcd(p, q)$.

Доказательство. Рассуждая от противного, предположим, что

$$|S| \geq p + q - \gcd(p, q).$$

Тогда слово S имеет периоды p and q по лемме 2.3. По классической теореме Файна–Вильфа (см., например, [14], предложение 1.3.5) S имеет также период $\gcd(p, q)$. Пусть U_1, U_2 – метки рассматриваемых циклов, прочитанные начиная с вершины S . Тогда слово S является суффиксом как SU_1 , так и SU_2 по утверждению 5 леммы 2.1. Так как $S \geq \max\{p, q\}$, и U_1 , и U_2 являются суффиксами S :



Следовательно, оба слова U_1 и U_2 имеют период $\gcd(p, q)$, а значит, являются целыми степенями одного и того же слова длины $\gcd(p, q)$. Тогда одно из них, скажем U_1 , является префиксом другого. Так как \mathcal{A} является ДКА, существует единственный путь из S с меткой U_1 , это первый из циклов. Но в данном случае либо $U_2 = U_1$, либо путь из S с меткой U_2 не является циклом. Обе альтернативы приводят к противоречию.

Следствие 2.1. *Если вершина S автомата \mathcal{A} принадлежит двум циклам длины p , то $|S| < p$.*

Для бинарного алфавита результат следствия 2.1 можно развить.

Следствие 2.2. *При $|\Sigma| = 2$ циклы длины 2 в автомате \mathcal{A} не пересекаются.*

Доказательство. Рассуждая от противного, предположим, что вершина S принадлежит двум циклам длины 2. Тогда из S исходят две дуги, и они ведут в различные вершины, не совпадающие с S . По следствию 2.1 $|S| < 2$. Если $S = \lambda$, то исходящие из S дуги ведут в вершины a и b . По лемме 2.2 $S = \lambda$ не лежит ни в каком цикле, противоречие.

Пусть теперь $|S| = 1$ (для определенности $S = a$). Рассмотрим исходящую из S дугу с меткой a . Поскольку она не ведет в S , она может вести только в вершину aa . Однако никакая дуга не может вести из вершины aa в вершину $S = a$, т. е. один из циклов не существует.

Предложение 2.2. *Метка любого цикла в автомате \mathcal{A} является примитивным (циклическим) словом. Если цикл имеет метку U , прочитанную начиная с вершины S , то слово S является суффиксом некоторой степени слова U .*

Доказательство. Как и в лемме 2.3, имеем равенство $SU = U'S$, откуда $U = PQ$, $U' = QP$ и $S = Q(PQ)^n$ ($n \geq 0$) для некоторых слов P и Q . Второе утверждение леммы немедленно следует отсюда.

Первое утверждение докажем от противного. Рассмотрим цикл в \mathcal{A} , помеченный словом X^n ($n \geq 2$), начиная с вершины S . Как мы уже доказали, S является суффиксом некоторой степени слова X^n , откуда $S = X'X^m$ для некоторого $m \geq 0$ и собственного (возможно, пустого) суффикса X' слова X . Значит, S является суффиксом слова SX . Так как путь с меткой X ведет из S в вершину, отличную от S , среди вершин автомата \mathcal{A} найдется более длинный, чем S , суффикс S' слова SX (шаг 3 алгоритма 2). Так как $S = X'X^m$,

слово S' является также суффиксом слова SX^n ; мы получаем противоречие с тем, что путь с меткой X^n ведет из S обратно в S .

Предложение 2.3. *При вершине S автомата \mathcal{A} имеется петля тогда и только тогда, когда $S = c^m$ для некоторой буквы c и $m \geq 0$, а слово c^{m+1} не является вершиной \mathcal{T} .*

Доказательство. Необходимость. По лемме 2.3 вершина S имеет период 1 или длину 0, откуда следует первое условие. Согласно шагу 3 алгоритма 2 S не имеет в \mathcal{T} исходящей дуги с меткой c , а значит, вершины c^{m+1} в \mathcal{T} нет.

Достаточность. Из условия следует, что слово $S = c^m$ является длиннейшим суффиксом слова $Sc = c^{m+1}$, являющимся вершиной. Тогда дуга S с меткой c ведет обратно в S (шаг 3 алгоритма 2).

Следствие 2.3. *Автомат \mathcal{A} имеет не более $|\Sigma|$ петель. Петля с меткой c отсутствует тогда и только тогда, когда $c^n \in M$ для некоторого n .*

Предложение 2.4. *Любое кратное ребро в \mathcal{A} имеет кратность не более $|\Sigma|-1$ и ведет в начальную вершину.*

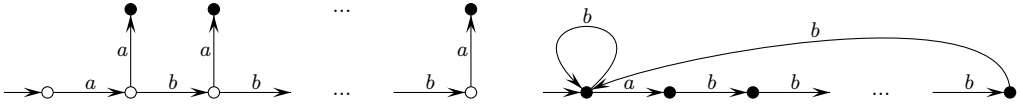
Доказательство. После шага 3 алгоритма 2 любая вершина W в \mathcal{A} имеет $|\Sigma|$ исходящих дуг, хотя бы одна из которых прямая. Значит, W имеет не более $|\Sigma|-1$ обратных дуг, все с различными метками. Дуги с метками c и d ведут в суффиксы слов Wc и Wd соответственно. Единственный общий суффикс этих слов – это λ , откуда следует требуемое утверждение.

В заключение параграфа мы приведем два примера построения КАС-автоматов и вычисления индексов роста для бесконечных серий языков. Эти примеры доказывают следующее предложение о «разнообразии» индексов роста КАС-языков в бинарном алфавите (при помощи аналогичных, но более громоздких примеров данное предложение можно обобщить на случай произвольного алфавита).

Предложение 2.5. *Пусть $|\Sigma| = 2$.*

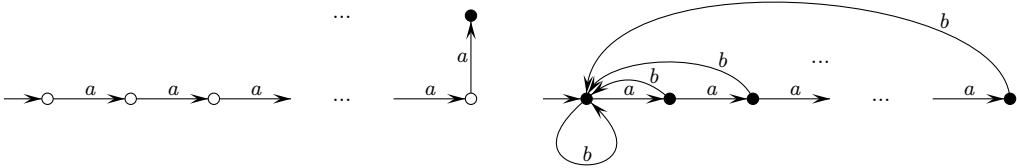
- 1) *Алгебраическое число любой степени может быть индексом роста КАС-языка над Σ .*
- 2) *Для любого $\varepsilon > 0$ существуют КАС-языки $L_1, L_2 \subseteq \Sigma^*$ такие, что $r(L_1) \in (1, 1 + \varepsilon)$, $r(L_2) \in (2 - \varepsilon, 2)$.*

Пример 2. Пусть $\tilde{M}_n = \{a^2, aba, \dots, ab^{n-1}a\}$. Соответствующие автоматы \mathcal{T} и \mathcal{A} приведены ниже:



Матрица смежности автомата \mathcal{A} имеет очень простой вид, что позволяет выписать характеристический многочлен явно. Он равен $x^n - x^{n-1} - 1$. Этот многочлен имеет единственный корень \tilde{z}_n на отрезке $[1, 2]$, и $\tilde{z}_n \rightarrow 1$ при $n \rightarrow \infty$. Так, \tilde{z}_2 равен золотому сечению, $\tilde{z}_4 \sim 1,3803$, $\tilde{z}_{1000} \sim 1,0053$.

Пример 3. Пусть $\widehat{M}_n = \{a^n\}$. Вот соответствующие автоматы \mathcal{T} и \mathcal{A} :



Как и в предыдущем примере, можно явно вычислить характеристический многочлен. Он равен $x^n - x^{n-1} - x^{n-2} - \dots - 1$, имеет единственный корень \hat{z}_n на отрезке $[1, 2]$ и $\hat{z}_n \rightarrow 2$ при $n \rightarrow \infty$. При этом \hat{z}_2 равен золотому сечению, $\hat{z}_4 \sim 1,9276$, $\hat{z}_{500} \sim 1,9998$.

3. С-графы и редукция антисловарей

При исследовании комбинаторной сложности КАС-языков при помощи автоматов разумно рассматривать не автомат целиком, а только его «существенную» часть, структура которой отвечает за поведение комбинаторной сложности, в частности за индекс роста языка. Нашей следующей целью является выделение этой существенной части, а также преобразование антисловаря, уменьшающее его размер (и соответственно размер КАС-автомата) и «незначительно» искажающее сам язык (асимптотическое поведение комбинаторной сложности, в частности индекс роста, не изменяется).

Напомним, что *компонентой сильной связности* (далее просто компонентой) орграфа G называется максимальный по включению подграф в G , от любой вершины которого существует путь до любой другой вершины. Одноэлементные компоненты и простые циклы будем называть *тривиальными* компонентами. Дадим основное определение.

Графом циклов, или *С-графом* антисловаря M , назовем подграф соответствующего автомата \mathcal{A} , полученный удалением из \mathcal{A} всех тривиальных компонент. Поскольку мы изучаем языки экспоненциальной сложности, из предложения 1.2 следует, что \mathcal{A} содержит нетривиальную компоненту, так что данное определение корректно. В примере 1 С-граф получается удалением

из \mathcal{A} вершины λ . Известное утверждение из теории графов гласит, что индекс графа равен максимуму из индексов его компонент; следовательно, для вычисления индекса роста языка можно заменить автомат \mathcal{A} на соответствующий \mathcal{C} -граф, это и есть «существенная» часть \mathcal{A} .

Теперь опишем упомянутое выше преобразование – *редукцию антисловаря*. Начнем с простого примера. Пусть для некоторого слова W антисловарь содержит все слова вида Wc , где c – буква. Тогда вершина W в \mathcal{A} не имеет исходящих дуг, т. е. образует одноэлементную компоненту. Если мы заменим в антисловаре все слова Wc на единственное слово W , то автомат \mathcal{A} претерпит единственное изменение – исчезнет вершина W . Но при переходе к \mathcal{C} -графу вершина W все равно будет удалена, т. е. данное преобразование вообще не затрагивает \mathcal{C} -граф.

В дальнейшем при перечислении неизоморфных \mathcal{C} -графов (и, следовательно, допустимых индексов роста) будет гораздо удобнее работать только с редуцированными антисловарями.

Напомним, что *реверс* слова $W = a_1 \dots a_n$ – это слово $\overline{W} = a_n \dots a_1$, *реверс* языка L определяется как $\overline{L} = \{\overline{W} \mid W \in L\}$. *Шагом редукции* антисловаря M назовем одно из следующих преобразований:

- (R1) Если все вершины КАС-автомата \mathcal{A} для антисловаря M с фиксированным префиксом W принадлежат тривиальным компонентам \mathcal{A} , то все слова из M , содержащие W , заменяются единственным словом W .
- (R2) Если все вершины КАС-автомата $\overline{\mathcal{A}}$ для антисловаря \overline{M} с фиксированным префиксом \overline{W} принадлежат тривиальным компонентам $\overline{\mathcal{A}}$, то все слова из M , содержащие W , заменяются единственным словом W .

Предложение 3.1. *Шаг редукции антисловаря сохраняет индекс роста языка с данным антисловарем.*

Доказательство. Мы докажем предложение для преобразования (R1), а затем используем симметрию. Для удобства, проведем преобразование антисловаря в два этапа. Вначале заменим все слова из M , имеющие префикс W , на единственное слово W ; затем удалим из полученного множества все слова, имеющие собственное подслово W . Язык M' , полученный на первом этапе, не обязательно антифакториальный, но мы используем замечание 2.1 и применим алгоритм 2 к M' , получая автоматы \mathcal{T}' и \mathcal{A}' . Покажем, что \mathcal{A}' совпадает с автоматом $\widehat{\mathcal{A}}$, полученным из \mathcal{A} удалением всех вершин с префиксом W .

Дерево \mathcal{T}' получено из дерева \mathcal{T} удалением поддерева с корнем W . Следовательно, множества вершин (а также множества прямых дуг) автоматов \mathcal{A}' и $\widehat{\mathcal{A}}$ совпадают. Возьмем обратную дугу (U_1, U_2) из $\widehat{\mathcal{A}}$. Согласно шагу 3

алгоритма 2, U_2 является длиннейшим (среди вершин из \mathcal{T}) суффиксом некоторого слова U_1c . Тем же свойством U_2 обладает и среди вершин из \mathcal{T}' , так как \mathcal{T}' – поддерево в \mathcal{T} . Тогда обратная дуга в \mathcal{A}' с меткой c также ведет в вершину U_2 .

Пусть теперь (U_1, U_2) – обратная дуга в \mathcal{A}' , т. е. U_2 является длиннейшим (среди вершин из \mathcal{T}') суффиксом U_1c . Если в \mathcal{T} имеется более длинный суффикс слова U_1c , то этот суффикс содержит подслово W ; значит, W является подсловом в U_1c и по утверждению 4 леммы 2.1 в \mathcal{A}' вообще нет дуги с меткой c , выходящей из U_1 . Но поскольку такая дуга (U_1, U_2) есть, U_2 – длиннейший суффикс U_1c и среди вершин из \mathcal{T} . Тогда автомат \mathcal{A} содержит дугу (U_1, U_2) , и мы доказали, что $\mathcal{A}' = \mathcal{A}$.

Все вершины, удаленные из \mathcal{A} для получения \mathcal{A}' , принадлежали к тривиальным компонентам \mathcal{A} . Значит, \mathcal{C} -графы автоматов \mathcal{A} и \mathcal{A}' совпадают.

Осталось заметить, что второй этап преобразования в точности сохраняет язык с множеством M' запрещенных слов (а значит, сохраняет сложность и индекс роста). Действительно, факториальный язык определяется множеством *минимальных* запрещенных слов, а оно не меняется на данном этапе. Для преобразования (R1) доказательство завершено.

Корректность преобразования (R2) следует из симметрии. Если язык L имеет антисловарь M , то у \overline{L} антисловарь равен \overline{M} , т. е. $C_L(n) = C_{\overline{L}}(n)$.

Замечание 3.1. По-видимому, преобразование (R1) всегда сохраняет \mathcal{C} -граф. Но вряд ли имеет смысл это доказывать, так как при переходе к реверсам \mathcal{C} -граф может изменяться, т. е. преобразование (R2) иногда уменьшает \mathcal{C} -граф (см. пример 4).

Антисловарь называется *редуцированным*, если к нему неприменим шаг редукции.

Пример 4. Пусть $\Sigma = \{a, b\}$, $M = \{a^3b, abba, babb, baabb\}$. Редукция антисловаря состоит из шага (R1), за которым следуют три шага (R2).

1) Заменим a^3b на a^3 : вершина a^3 в \mathcal{A} имеет единственную исходящую дугу, и это петля.

2) Заменим $baabb$ на $aabb$: так как слова $baabb$ и $aaabb$ запрещены, вершина $aabb$ в $\overline{\mathcal{A}}$ не имеет исходящих дуг.

3) Заменим $aabb$ и $babb$ на abb , как на шаге 2); попутно удаляется $abba$.

4) Заменим abb на bb симметрично шагу 1).

Получили редуцированный антисловарь $\{a^3, b^2\}$. Его КАС-автомат и \mathcal{C} -граф содержат четыре и три вершины соответственно, вместо одиннадцати и четырех для M .

Замечание 3.2. Для любого $n \geq 1$ слова $a^n b$, ab^n и их реверсы не могут принадлежать редуцированному антисловарю над бинарным алфавитом.

Ниже мы докажем несколько свойств S -графов редуцированных антисловарей. Эти свойства будут использованы при доказательстве классификационной теоремы в следующем разделе.

Лемма 3.1. Пусть M – редуцированный антисловарь. После шага 3 Алгоритма 2 каждый лист из \mathcal{T} достижим из некоторой вершины S -графа.

Доказательство. Предположим, что лист X из \mathcal{T} недостижим из любой вершины S -графа, и докажем, что антисловарь M нередуцированный. Имеем $X \in AD$; пусть $X = X'c$, где c – буква. Вершина X' также недостижима из любой вершины S -графа по нашему предположению. Далее, все вершины из A с суффиксом X' недостижимы из S -графа. В самом деле, в \mathcal{T} нет вершин с собственным суффиксом X из-за антифакториальности M ; значит, из любой вершины вида ZX' выходит обратная дуга в X с меткой c .

Удалим из автомата A все вершины, принадлежащие S -графу. Оставшийся автомат A' распознает те же самые слова с суффиксом X' , что и автомат A . Так как автомат A' состоит из тривиальных компонент, язык, распознаваемый A' , имеет полиномиальную сложность по предложению 1.2. Таким образом, язык L с антисловарем M содержит лишь полиномиальное количество слов с суффиксом X' . Симметрично, язык \bar{L} с антисловарем \bar{M} содержит полиномиальное количество слов с префиксом \bar{X}' .

Прочитав слово \bar{X}' из начальной вершины автомата \bar{A} , мы попадем в вершину Y , где Y – некоторый суффикс слова \bar{X}' . Но тогда в \bar{A} имеется лишь полиномиальное количество путей, начинающихся в Y . Согласно предложению 1.2 ни одна из вершин, достижимых из Y , не принадлежит к нетривиальной компоненте. Следовательно, к M применимо преобразование (R2).

Лемма 3.2. Если все слова бинарного редуцированного антисловаря M начинаются с одной и той же буквы, то вершина λ автомата A принадлежит S -графу.

Доказательство. Пусть все слова из M начинаются с буквы a . Тогда в автомате A у вершины λ есть петля с меткой b по предложению 2.3. Так как любая вершина из A достижима из вершины λ , достаточно показать, что найдется обратная дуга в λ из какой-либо другой вершины. Выберем в A вершину ab^n с максимальным n (случай $n = 0$ также возможен). Ввиду максимальной и замечания 3.2, эта вершина имеет единственную исходящую дугу в \mathcal{T} , причем с меткой a . Значит, из вершины ab^n выходит обратная дуга с меткой b . Единственный суффикс слова ab^{n+1} , принадлежащий \mathcal{T} , это λ . Значит, данная дуга ведет в вершину λ . Лемма доказана.

Вершину из \mathcal{A} назовем *крайней*, если в \mathcal{A} из нее не выходит прямых дуг. Если вершина U – крайняя, то единственная вершина в \mathcal{A} с префиксом U – это сама U . Следующее замечание прямо следует из определений.

Замечание 3.3. *Для редуцированного антисловаря M все крайние вершины автомата \mathcal{A} лежат в S -графе.*

Лемма 3.3. *Для редуцированного бинарного антисловаря M все крайние вершины автомата \mathcal{A} имеют единственную исходящую дугу.*

Доказательство. Так как из вершины U не выходит прямых дуг, ее степень исхода меньше мощности алфавита. С другой стороны, из определения S -графа и замечания 3.3 следует, что эта степень больше нуля.

Предложение 3.2. *S -граф редуцированного бинарного антисловаря содержит вершину с единственной входящей и единственной исходящей дугой.*

Доказательство. Пусть S -граф определяется редуцированным антисловарем M , $W = Ua$ – слово максимальной длины в M . Тогда вершина U – крайняя ($Ub \notin M$ по определению редуцированного антисловаря), т. е. U лежит в S -графе (замечание 3.3) и имеет единственную исходящую дугу (лемма 3.3).

Рассмотрим входящие в U дуги. Очевидно, что $|U| > 0$, а значит, в U входит прямая дуга. Пусть в \mathcal{A} найдется обратная дуга (V, U) . Согласно шагу 3 алгоритма 2, U – собственный суффикс слова Vc для некоторой буквы c . Тогда $|U| \leq |V|$, откуда $|U| = |V|$ по определению U и V также является максимальным собственным префиксом некоторого слова из M . Следовательно, V – крайняя. Заменяем U на V и повторим рассуждение. Поскольку антисловарь конечен, мы таким образом найдем крайнюю вершину, в которую либо не входит обратная дуга, либо входит обратная дуга из U . В первом случае лемма доказана, а во втором мы получаем противоречие с определением S -графа. В самом деле, в S -графе обнаружен цикл из вершин со степенью исхода 1; но такой цикл очевидно является тривиальной компонентой.

Лемма 3.4. *Если S -граф редуцированного бинарного антисловаря M сильно связан, то в нем не менее $|M|$ вершин со степенью исхода 1. Для каждой такой вершины недостающая дуга заканчивалась в листе из \mathcal{T} и была удалена на шаге 4 алгоритма 2.*

Доказательство. По лемме 3.1 в автомате, построенном на шаге 3 алгоритма 2, для любого листа X из \mathcal{T} найдется путь из некоторой вершины U S -графа в X . Для каждого листа выберем один такой путь минимальной длины. Докажем, что все выбранные пути начинаются в разных вершинах.

Рассуждаем от противного: пусть два выбранных пути начинаются в вершине U . Ввиду минимальности, первые дуги этих путей не принадлежат S -графу. Так как алфавит бинарный, из U за пределы S -графа ведет не более одной дуги. Значит, пути начинаются с одной и той же дуги, скажем (U, Y) . Тогда Y не лист (пути заканчиваются в разных листьях), т. е. Y – вершина из \mathcal{A} . Так как S -граф связан и вершина Y из него достижима, то из Y нет пути обратно в S -граф. Но тогда к M применимо преобразование (R1).

Поскольку все выбранные пути начинаются в различных вершинах, число таких вершин равно числу листьев в \mathcal{T} , т. е. $|M|$. Отсюда вытекает первое утверждение леммы.

Для доказательства второго утверждения возьмем один из выбранных путей. Пусть (U, Y) – его первая дуга. Вершина Y не лежит в S -графе ввиду минимальности; если Y лежит в \mathcal{A} , мы получаем, как и выше, что M можно редуцировать. Значит, Y – лист в \mathcal{T} , что и требовалось доказать.

Заметим, что сильно связный S -граф редуцированного бинарного антисловаря M может иметь более чем $|M|$ вершин со степенью исхода 1. В частности, такой S -граф приведен в примере 1: после шага 3 алгоритма 2 из вершин a и bba выходит дуга в лист aa .

4. Маленькие S -графы и индексы роста для бинарных антисловарей

Согласно предложению 3.1 для перечисления возможных индексов роста достаточно рассматривать языки с редуцированными антисловарями. Список «маленьких» S -графов для редуцированных бинарных антисловарей приведен в табл. 1–3. Наш основной результат состоит в доказательстве полноты этого списка.

Теорема 4.1. *Если S -граф языка с редуцированным бинарным антисловарем имеет не более четырех вершин, то он изоморфен одному из графов 1.1–1.9, 2.1–2.13 или 3.1–3.3.*

Перед доказательством теоремы заметим, что принципиально разные редуцированные антисловари могут иметь изоморфные S -графы. Например, S -граф антисловаря $\{a^3, bab\}$ изоморфен графу 2.1.

Доказательство. Суммируя доказанные ранее утверждения, получаем следующие свойства S -графа редуцированного бинарного антисловаря:

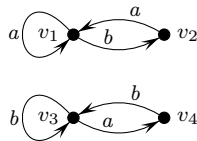
- (1) все компоненты нетривиальны (определение);
- (2) из вершины выходит одна или две дуги (определения S -графа и \mathcal{A});

- (3) есть вершина со степенью исхода и захода, равными 1 (предложение 3.2);
- (4) нет кратных дуг (предложение 2.4);
- (5) имеется не более двух петель (следствие 2.3);
- (6) циклы длины 2 не пересекаются (следствие 2.2).

Следовательно, достаточно проверить орграфы, удовлетворяющие данным свойствам. Из свойств (1)–(4) немедленно следует, что не существует S -графов с двумя вершинами, отличных от графа 1.1. Что касается S -графов с тремя вершинами, заметим вначале, что они сильно связны (одноэлементная компонента тривиальна ввиду (4)). Такой S -граф должен быть гамильтоновым, так как в соответствии с (6) в нем не более одного цикла длины 2. Согласно (1)–(3) для получения S -графа нужно добавить к гамильтонову циклу одну или две дуги. Добавлением одной дуги можно получить два неизоморфных орграфа, и это графы 2.1 и 2.2. Две добавленные дуги должны начинаться в разных вершинах ввиду (2) и не заканчиваться в третьей вершине согласно (3). Две оставшиеся возможности дают нам графы 1.2 и 1.3. Итак, мы перечислили все S -графы с двумя и тремя вершинами.

Мы видим, что выполнения свойств (1)–(6) достаточно для орграфа не более чем с тремя вершинами, чтобы являться S -графом редуцированного бинарного антисловаря. Для орграфов с четырьмя вершинами это не так, поэтому доказательство будет весьма длинным. Далее рассматриваются *только четырехвершинные орграфы*; их вершины мы обозначаем через v_1, v_2, v_3 и v_4 . Вначале заметим, что S -граф содержит от 5 до 7 дуг согласно (1)–(3), а значит, антисловарь состоит не более чем из трех слов по лемме 3.4. Докажем, что S -граф сильно связан.

Пусть орграф G обладает свойствами (1)–(6), но не является сильно связным. Тогда обе его компоненты изоморфны графу 1.1. Пусть v_1 и v_2 образуют одну компоненту, а v_3 и v_4 – другую, с петлями при вершинах v_1 и v_3 . Предположим, что G – S -граф и реконструируем антисловарь. Рассмотрим вершины G как слова. По предложению 2.1 $|v_1|, |v_3| \leq 1$; пусть a – метка петли на v_1 , а b – на v_3 (метки петель не совпадают по предложению 2.3). Тогда остальные дуги в компонентах помечены так:



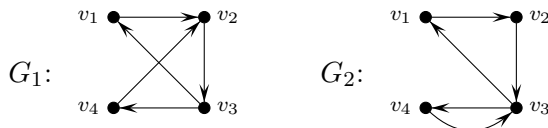
Если $v_1 = \lambda$, то остается единственный вариант: $v_3 = b$. Тогда дуга с меткой b должна вести из v_1 в v_3 , противоречие. Аналогично, $v_3 \neq \lambda$. Тогда $v_1 = a$ и $v_3 = b$. Дуга с меткой b ведет из a в суффикс слова ab ; так как эта дуга не ведет в $v_3 = b$, получаем $v_2 = ab$. Аналогично, $v_4 = ba$. Дуга

с меткой a ведет из v_2 в $v_1 = a$, в то время как $v_4 = ba$ – более длинный суффикс слова aba . Противоречие с шагом 3 алгоритма 2. Таким образом, подходящего антисловаря не существует, т. е. G не является С-графом.

Вначале перечислим С-графы, соответствующие антисловарям из одного слова. Если $M = \{W\}$, то С-граф содержит вершину λ по лемме 3.2 и длиннейший собственный префикс слова W по замечанию 3.3. Мы уже доказали, что С-граф сильно связан, значит, он совпадает с самим автоматом \mathcal{A} . В \mathcal{A} имеется $|W|$ вершин. Значит, $|W| = 4$. Пусть W начинается с буквы a . С учетом замечания 3.2 получаем шесть возможных редуцированных антисловарей. Им соответствуют С-графы 1.4–1.9.

Перейдем к антисловарям не менее чем из двух слов. Из (1), (2) и леммы 3.4 следует, что в С-графе 5 или 6 дуг. Мы перечислим все возможные орграфы и докажем, что каждый из них либо содержится в табл. 2, 3, либо не является С-графом.

Начнем с орграфов с пятью дугами. Если орграф гамильтонов, то остается только одна дуга. Все три способа добавить эту дугу, удовлетворяющих (4), представлены графами 3.1–3.3. Если же орграф не гамильтонов, то он содержит цикл длины 3 (условие (6) плюс сильная связность) и еще две дуги для присоединения последней вершины. Значит, он изоморфен графу G_1 или G_2 :



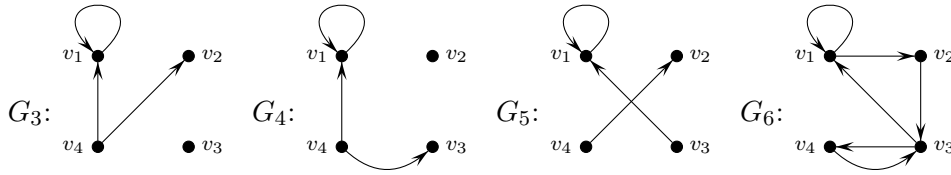
Мы докажем, что G_1 и G_2 не являются С-графами, но сначала рассмотрим орграфы с шестью дугами для составления полного списка графов «исключений».

Пусть орграф с шестью дугами гамильтонов. Если две оставшиеся дуги образуют петли, то возможны только два случая, а именно графы 2.11 и 2.12. Если петля одна, то последняя дуга выходит из одной из оставшихся трех вершин. В каждом случае есть две вершины, куда эта дуга может вести (так как она не является петлей и не повторяет дугу из гамильтонова цикла). Всего получаем шесть неизоморфных гамильтоновых орграфов с одной петлей, а именно графы 2.3, 2.5, 2.7, 2.13, и «исключения» G_3 и G_4 (см. далее).

Если С-граф с шестью дугами (неважно, гамильтонов или нет) не содержит петель, то антисловарь содержит степени букв a и b по следствию 2.3. По лемме 3.4 антисловарь состоит ровно из двух слов; пусть это a^m и b^n . В соответствующем КАС-автомате \mathcal{A} обратные дуги ведут из всех степеней a в вершину b и из всех степеней b в вершину a . Следовательно, С-граф содержит

все вершины из \mathcal{A} , кроме λ , т. е. $(m-1) + (n-1)$ вершин. Отсюда $m + n = 6$, и мы получаем два неизоморфных С-графа 2.4 и 2.10.

Остались только негамильтоновы орграфы с петлями. После удаления петли орграф должен остаться сильно связным, т. е. он будет изоморфен графу G_1 или G_2 . По условию (2) к каждому из этих графов можно добавить петлю тремя способами. Заметим, что орграфы, полученные добавлением петли к вершинам v_1 и v_4 графа G_1 , изоморфны. В итоге мы получаем пять орграфов – это графы 2.6, 2.8, 2.9 и «исключения» G_5 и G_6 .



Для доказательства теоремы осталось показать, что среди графов G_1-G_6 нет С-графов редуцированных антисловарей. Для каждого графа мы предположим, что он является С-графом редуцированного антисловаря M , рассмотрим вершины графа как слова и получим противоречие при попытке восстановить M . Сделаем предварительные замечания:

- (i) если из вершины выходят две дуги, то одна из них заведомо прямая;
- (ii) для дуги (U, V) выполняется $|V| \leq |U| + 1$; отсюда по лемме 3.4
- (iii) из вершины λ выходят две дуги.

Для графа G_1 рассуждение несложно. По следствию 2.1 $|v_2| < 3$. По предложению 2.2 v_2 является общим суффиксом меток обоих циклов, прочитанных, начиная с v_2 . Однако вторые буквы в этих метках различны, иначе мы не получим детерминированный автомат. Значит, длина общего суффикса меток циклов не превосходит единицы. Из (iii) мы получаем окончательно $|v_2| = 1$. Пусть $v_2 = a$. Отсутствующая дуга из v_2 вела в запрещенное слово (лемма 3.4); с учетом замечания 3.2 этим словом может быть только a^2 . Далее, из вершины v_3 выходит дуга с меткой a , скажем (v_3, v_4) . Согласно (iii), $v_4 \neq \lambda$, т. е. v_4 заканчивается буквой a . Тогда слово v_4a заканчивается на a^2 . Но вершина a^2 лежит в \mathcal{T} , а значит, дуга (v_4, v_2) появилась в нарушение правил шага 3 алгоритма 2. Противоречие получено.

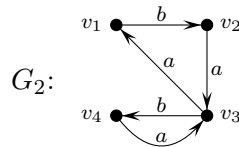
Граф G_2 оказывается самым сложным для рассмотрения. Так, он может быть С-графом нередуцированного антисловаря или компонентой С-графа редуцированного антисловаря, но не может являться С-графом редуцированного антисловаря. В последующем анализе мы постоянно используем лемму 3.4 для получения выводов двух типов:

- (*) если из вершины U не выходит дуга с меткой c , то слово Uc запрещено;
- (**) если, кроме того, выходящая из U дуга является обратной, то $Uc \in M$.

Вывод (*) часто используется в таком виде: если из U не выходит дуга с меткой c и U – суффикс V , то слово Vc запрещено.

Вначале покажем, что в G_2 нет вершины λ (это будет использовано в дальнейшем для получения противоречий). Ввиду (iii) единственным кандидатом на λ является v_3 . Пусть $v_3 = \lambda$. Тогда $v_1 = a$, $v_4 = b$ (или наоборот) по (ii). Но тогда в v_3 не ведет никакая дуга (лемма 2.2), противоречие.

Далее, есть только один (с точностью до переименования букв) способ пометить дуги в G_2 , согласующийся с предложением 2.2:



По предложению 2.1 получаем $|v_3| \leq 3$. С учетом меток дуг и предложения 2.2, $v_3 \in \{a, ba, aba\}$.

Случай 1: $v_3 = a$. Тогда v_1 является суффиксом слова a^2 . Так как $v_1 \neq \lambda$, $v_1 \neq a = v_3$, имеем $v_1 = a^2$. Далее, v_2 – длиннейший суффикс слова aab в графе, а v_4 – длиннейший суффикс ab . Так как $v_2 \neq v_4$, получаем $v_2 = aab$. Следовательно, $aabb \in M$ и одно из слов b^2 , abb (в зависимости от слова v_4) также принадлежит M . Значит, M не антифакториален, противоречие.

Случай 2: $v_3 = ba$. Тогда v_1 является суффиксом слова baa .

Случай 2.1: $v_1 = a$. Тогда слово a^2 запрещено. Следовательно, baa тоже запрещено и дуга (v_3, v_1) добавлена некорректно, противоречие.

Случай 2.2: $v_1 = a^2$. Тогда $v_4 = bab$ по (i), откуда $babb \in M$. Слово v_2 является суффиксом aab , т.е. слово $aabb$ также запрещено. Слово abb не может быть запрещено, так как $babb$ – минимальное запрещенное слово. Отсюда $aabb \in M$, и антисловарь можно редуцировать, используя слово abb .

Случай 2.3: $v_1 = baa$. Тогда слово $baaa$ запрещено. По замечанию 3.2, это слово не может лежать в редуцированном антисловаре. Это означает, с одной стороны, что $a^3 \in M$, а с другой стороны, что дуга (v_1, v_2) прямая, откуда $v_2 = baab$ и $baabb \in M$. Но в этом случае антисловарь можно редуцировать, используя слово $aabb$.

Случай 3: $v_3 = aba$. Тогда v_1 является суффиксом слова $abaa$.

Случай 3.1: $v_1 = a$ аналогичен случаю 2.1.

Случай 3.2: $v_1 = a^2$. Слово a^3 запрещено, откуда $a^3 \in M$. Далее, $v_4 = abab$ по (i), а значит, $ababb \in M$. Слово v_2 является суффиксом aab , т.е. $aabb$ запрещено. При этом abb не запрещено ввиду минимальности $ababb$. Значит, $aabb \in M$. Итак, все слова из M начинаются с буквы a , и по лемме 3.2 в S -графе есть вершина λ , противоречие.

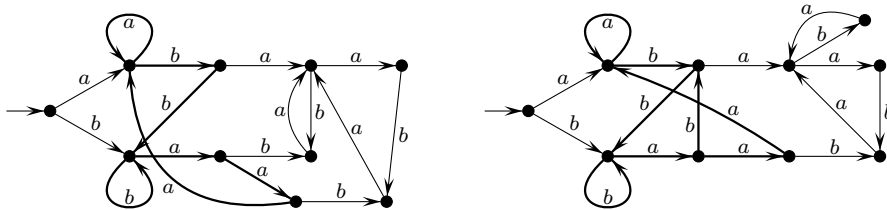
Случай 3.3: $v_1 = baa$ аналогичен случаю 2.3.

Случай 3.4: $v_1 = abaa$. Тогда слово $abaaa$ запрещено, v_2 является суффиксом $abaab$, v_4 является суффиксом $abab$, и, кроме того, $|v_2|, |v_4| \geq 2$ согласно (ii). Далее, слова v_2b и v_4b запрещены; более того, можно сделать вывод (**): если $|v_2| \geq 3$, то $v_2b \in M$ (аналогично для v_4). Таким образом, если одна из этих вершин – это ab , то мы получаем запрещенное слово abb , являющееся собственным суффиксом слова из M , что невозможно. Следовательно, $|v_2|, |v_4| \geq 3$. Если $|v_2| = |v_4| = 3$, то $aabb, babb \in M$, и антисловарь можно редуцировать, используя слово abb . Во всех оставшихся случаях мы получим противоречие, найдя в С-графе еще одну компоненту. На каждом из приведенных рисунков такая компонента выделена жирными линиями.

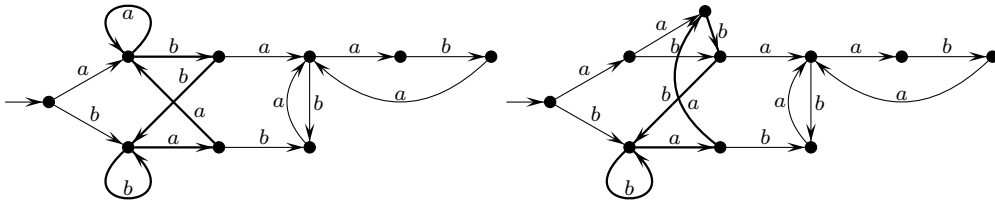
Случай 3.4.1: $v_2 = aab, v_4 = abab$. Все слова в антисловаре начинаются с a , и по лемме 3.2 в С-графе есть вершина λ .

Случай 3.4.2: $v_2 = baab, v_4 = bab$. $AD = \{abaaa, baabb, babb\}$ (рисунок слева).

Случай 3.4.3: $v_2 = baab, v_4 = abab$. $AD = \{abaaa, baabb, ababb\}$ (рисунок справа).



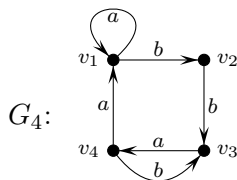
Случай 3.4.4: $v_2 = abaab, v_4 = bab$. В этом и следующем случаях запрещенное слово $abaaa$ не обязательно минимально (вывод (**)) неприменим). Если оно не минимально, то $a^3 \in M$ ($baaa$ можно редуцировать, a^2 не запрещено). Автомат, изображенный слева, получен из антисловаря $M = \{abaaa, abaabb, babb\}$, а изображенный справа – из $M = \{a^3, abaabb, babb\}$.



Случай 3.4.5: $v_2 = abaab, v_4 = abab$ аналогичен случаю 3.4.1. Доказательство для графа G_2 завершено.

Рассмотрим граф G_3 . Ввиду (i), дуга (v_1, v_2) прямая. Прямые дуги принадлежат дереву, т. е. дуга (v_4, v_2) – обратная. Тогда (v_4, v_1) – прямая; получаем $|v_2| - |v_1| = 1, |v_1| - |v_4| = 1$, откуда $|v_2| - |v_4| = 2$, противоречие с (ii).

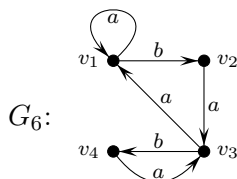
Перейдем к графу G_4 и пометим его дуги в соответствии с Предложением 2.2. Это можно сделать единственным, с точностью до переименования букв, способом:



Длина v_3 не превосходит 3 по предложению 2.1. По предложению 2.2, v_3 является суффиксом $aabb$ и одновременно суффиксом $(ab)^n$. Так как $v_3 \neq \lambda$ согласно (iii), имеем $v_3 = b$. Тогда слово b^2 запрещено, и путь из v_1 в v_3 не может быть помечен этим словом.

На граф G_5 полностью переносится рассуждение о графе G_1 .

В графе G_6 вначале пометим дуги в соответствии с предложением 2.2.



Заметим, что $v_3 \in \{a, ba, aba\}$; это доказывается так же, как для графа G_2 . По предложению 2.1, $|v_1| < 3$, откуда $v_1 \in \{\lambda, a, a^2\}$ по предложению 2.2. Дуга (v_1, v_2) является прямой согласно (i).

Случай 1: $v_1 = \lambda$. Тогда $v_2 = b$ и слово b^2 запрещено. Дуга (v_3, v_1) является обратной; согласно (i), (v_3, v_4) – прямая. Тогда $v_3bb \in M$, хотя это не минимальное запрещенное слово – противоречие.

Случай 2: $v_1 = a$. Тогда $v_3 \neq a$, $v_2 = ab$, и слово abb запрещено. Как и в случае 1, получаем $v_3bb \in M$, хотя у v_3bb есть запрещенный собственный суффикс abb , противоречие.

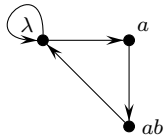
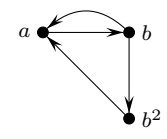
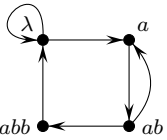
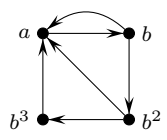
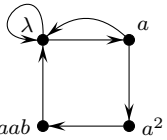
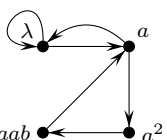
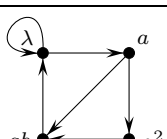
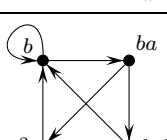
Случай 3: $v_1 = a^2$. Тогда $v_2 = aab$ и $aabb \in M$ (вывод (**)). Из (ii) и (iii) получаем $0 < |v_4| \leq 4$, откуда $v_4 \in \{b, ab, bab, abab\}$ по предложению 2.2. Если $v_4 = b$ или $v_4 = ab$, то у слова $aabb$ есть запрещенный собственный суффикс (соответственно, b^2 и abb), противоречие. В случае $v_4 = bab$ имеем $babb \in M$, а значит, M можно редуцировать при помощи слова abb . Наконец, если $v_4 = abab$, то $ababb \in M$. Тогда по лемме 3.2 в S -графе есть вершина λ ; но ни одна из вершин $v_1 - v_4$ не равна λ , противоречие.

Рассмотрение случаев и доказательство теоремы завершено.

Таблица 1. Антисловари из одного слова

№	Анти-словарь	С-граф	Характеристический многочлен	Индекс роста
1.1	a^2		$x^2 - x - 1$	1,618
1.2	a^3		$x^3 - x^2 - x - 1$	1,839
1.3	aba		$x^3 - 2x^2 + x - 1$	1,755
1.4	a^4		$x^4 - x^3 - x^2 - x - 1$	1,928
1.5	$aaba$		$x^4 - 2x^3 + x - 1$	1,867
1.6	$aabb$		$x^4 - 2x^3 + 1 = (x-1)(x^3 - x^2 - x - 1)$	1,839
1.7	$abaa$		$x^4 - 2x^3 + x - 1$	1,867
1.8	$abab$		$x^4 - 2x^3 + x^2 - 2x + 1 = (x^2 - (1+\sqrt{2})x + 1) \cdot (x^2 - (1-\sqrt{2})x + 1)$	1,883
1.9	$abba$		$x^4 - 2x^3 + x - 1$	1,867

Таблица 2. Антисловари из двух слов

№	Анти-словарь	С-граф	Характеристический многочлен	Индекс роста
2.1	a^2 aba		$x^3 - x^2 - 1$	1,466
2.2	a^2 b^3		$x^3 - x - 1$	1,325
2.3	a^2 $abba$		$x^4 - x^3 - x^2 + x - 1$	1,513
2.4	a^2 b^4		$x^4 - x^2 - x - 1 =$ $(x + 1)(x^3 - x^2 - 1)$	1,466
2.5	a^3 $aaba$		$x^4 - x^3 - x^2 - 1 =$ $(x+1)(x^3 - 2x^2 + x - 1)$	1,755
2.6	a^3 $aabb$		$x^4 - x^3 - x^2 - x + 1$	1,722
2.7	a^3 aba		$x^4 - x^3 - x - 1 =$ $(x^2 + 1)(x^2 - x - 1)$	1,618
2.8	a^3 $baba$		$x^4 - x^3 - 2x =$ $x \cdot (x^3 - x^2 - 2)$	1,696

Окончание табл. 2

№	Анти-словарь	С-граф	Характеристический многочлен	Индекс роста
2.9	a^3 bab		$x^4 - x^3 - x^2 =$ $x^2 \cdot (x^2 - x - 1)$	1,618
2.10	a^3 b^3		$x^4 - x^2 - 2x - 1 =$ $(x^2 - x - 1)(x^2 + x + 1)$	1,618
2.11	aba bab		$x^4 - 2x^3 + x^2 - 1 =$ $(x^2 - x - 1)(x^2 - x + 1)$	1,618
2.12	aba $abba$		$x^4 - 2x^3 + x^2 - 1 =$ $(x^2 - x - 1)(x^2 - x + 1)$	1,618
2.13	aba b^4		$x^4 - x^3 - x - 1 =$ $(x^2 + 1)(x^2 - x - 1)$	1,618

Таблица 3. Антисловари из трех слов

№	Анти-словарь	С-граф	Характеристический многочлен	Индекс роста
3.1	a^2 aba $abba$		$x^4 - x^3 - 1$	1,380
3.2	a^2 aba b^4		$x^4 - x - 1$	1,221
3.3	a^2 $abba$ b^4		$x^4 - x^2 - 1 =$ $\left(x^2 - \frac{1+\sqrt{5}}{2}\right)\left(x^2 - \frac{1-\sqrt{5}}{2}\right)$	1,272

Сделаем несколько заключительных замечаний. Во-первых, табл. 1–3 показывают, что количество различных индексов роста быстро растет с увеличением степени полинома. Отдельный интересный вопрос: *насколько быстро оно растет? По экспоненте? Или даже быстрее?*

Интересная проблема связана с компонентами С-графов. Как видно из доказательства теоремы, С-графы могут не быть сильно связными. На самом деле, С-граф с двумя компонентами может иметь всего пять вершин (но не четыре – см. доказательство теоремы), как, например, С-граф антисловаря $\{aabb, aabab\}$. Существуют ли С-графы с произвольным числом компонент?

Третье замечание касается реверсов антисловарей. Языки, антисловари которых являются реверсами друг друга, также являются реверсами друг друга. Значит, и комбинаторная сложность, в частности индекс роста, для них одна и та же. Однако соответствующие С-графы могут быть не только неизоморфными, но даже и не коспектральными (коспектральными называются графы с равными характеристическими многочленами). Это одно из ярких проявлений асимметричности вычислений при помощи конечного автомата. В качестве примеров укажем, что графы 1.5 и 1.7 коспектральны, но не изоморфны, а С-граф антисловаря $\{a^3, abab\}$ (реверса к 2.8) содержит пять вершин, а не четыре, т. е. характеристические многочлены имеют разную степень.

Предыдущее замечание показывает, что индекс роста часто определяется собственным делителем характеристического многочлена С-графа, а не самим этим многочленом. Характерный пример дают графы 2.7, 2.9–2.13: их индекс роста определяется делителем $x^2 - x - 1$ их характеристических многочленов, который сам является характеристическим многочленом (графа 1.1). При этом сам граф 1.1 является делителем (в смысле теории графов, см., например, [11]) некоторых из этих графов и не является делителем других. Так что вопрос сведения С-графов к «неделимым» графам представляется интересным, но весьма сложным.

1. MORSE M., HEDLUND G. A. Symbolic dynamics // Amer. J. Math. 1983. Vol. 60. P. 815–866.
2. EHRENFEUCHT A., ROZENBERG G. A limit theorem for sets of subwords in deterministic TOL languages // Inform. Process. Lett. 1973. Vol. 2. P. 70–73.
3. BRANDENBURG F.-J. Uniformly growing k -th power free homomorphisms // Theoret. Comput. Sci. 1983. Vol. 23. P. 69–82.
4. ГОВОРОВ В. Е. Градуированные алгебры // Матем. заметки. 1972. Т. 12. С. 197–204.

5. SHUR A. M. Combinatorial complexity of regular languages // Computer Science – Theory and Applications. [Lecture Notes in Computer Science; Vol. 5010]. B.; Heidelberg: Springer, 2008. P. 289–301.
6. SHUR A. M. Polynomial languages with finite antidictionaries // TUCS Tech. Report. 2005. No 721. November.
7. EHRENFEUCHT A., ROZENBERG G. On subword complexities of homomorphic images of languages // RAIRO Inform. Theor. 1982. Vol. 16. P. 303–316.
8. ШУР А. М. Комбинаторная сложность рациональных языков // Дискрет. анализ и исследование операций. Сер. 1. 2005. Т. 12, № 2. С. 78–99.
9. KOBAYASHI Y. Repetition-free words // Theoret. Comput. Sci. 1986. Vol. 44. P. 175–197.
10. DEWDNEY A. K. DeBruijn graphs and sequences // Handbook of graph theory / eds. J. L. Gross, J. Yellen. Boca Raton, Fla.: CRC Press, 2004. P. 253–260.
11. CVETKOVIĆ D. M., DOOB M., SACHS H. Spectra of graphs. Theory and applications. 3rd edition. Heidelberg: Johann Ambrosius Barth, 1995.
12. SHUR A. M. Comparing complexity functions of a language and its extendable part // Theor. Inform. Appl. 2008. Vol. 42, No. 3. P. 647–655.
13. CROCHEMORE M., MIGNOSI F., RESTIVO A. Automata and forbidden words // Inform. Process. Lett. 1998. Vol. 67, № 3. P. 111–117.
14. LOTHAIRE M. Combinatorics on Words. Reading, Mass.: Addison-Wesley, 1983.

Статья поступила 17.04.2008, окончательный вариант 04.04.2008