ФГБОУ ВО "Московский государственный университет им. М.В.Ломоносова"
Механико-математический факультет
Кафедра математической теории интеллектуальных систем

На правах рукописи

Дергач Пётр Сергеевич

Алфавитное кодирование регулярных языков с полиномиальной функцией роста.

Специальность 01.01.09 — дискретная математика и
математическая кибернетика

Диссертация

на соискание ученой степени
кандидата физико-математических наук

Научный руководитель:
академик, доктор физико-математических наук,
профессор В.Б.Кудрявцев

Москва 2016

Содержание

Введение

Теория кодирования по праву занимает важное место среди разделов дискретной математики. Можно считать, что современная теория кодирования началась с работы Клода Э. Шеннона - американского математика, первым решившего проблему избыточности передачи информации по каналу связи(см. [1-2]). В ней Шеннон, ссылаясь на своего коллегу, Белла Ричарда Хэмминга, предложил линейный код, исправляющий одну ошибку. Справедливости ради следует упомянуть, что такой же код был независимо получен другим математиком из Швейцарии - Марселем Дж. Голеем(см. [3]). Но исторически так сложилось, что эти коды в дальнейшем стали называть *кодами Хемминга.* После этого в 1949 году Леоном Г. Крафтом для класса префиксных кодов было получено знаменитое *неравенство Крафта-Макмиллана*(см. [4]). Своим именем оно также обязано американскому ученому и политику Броквею МакМиллану, который в 1956 году доказал выполнение этого неравенства для разделимых кодов(см. [5]). Затем американский ученый Дэвид Хаффман, будучи аспирантом создает первый код оптимального сжатия, широко известный теперь как *код Хаффмана*(см. [6]). В 1954 году американцы Ирвинг С. Рид и Дэвид Е. Мюллер придумывают *блочные коды Рида-Маллера*(см. [7]). Далее французский математик Алексис Хоквингемм(см. [8]) и независимо

от него чуть позже американцы Радж Чандра Боуз и Двайджендра Камар Рей-Чоудхури(см. [9]) изобретают *коды Боуза-Чоудхури-Хоквингема* (сокращенно *БЧХ-коды).* В том же 1960 году уже упоминавшийся Ирвинг С. Рид и еще одного американец Густав Соломон представляют общественности *коды Рида-Соломона*(см. [10]). Затем результаты в области теории кодирования идут по нарастающей. К наиболее важным из них, пожалуй, стоит отнести

- создание *LDPC-кодов*(см. [11]) (Роберт Галлагер, 1962);
- возникновение *алгоритма Витерби* по декодированию сверточных кодов(см. [12]) (Эндрю Витерби, 1967);
- создание *арифметического кодирования*(см. [13]) (Йорма Риссанен, 1976);
- появление *алгоритмов сжатия* LZ-77(см. [14]) и LZ-78(см. [15]) (Якоб Зив, Авраам Лемпел, 1977-1978);
- появление *турбо-кодов*(см. [16]) (Клод Берроу , Алэйн Главиукс и П.Ситимашимой, 1993);
- открытие *преобразования Барроуза-Уилера*(см. [17]) (Майкл Барроуз, Дэвид Уилер, 1994);
- изобретение *полярных кодов*(см. [18]) (Эрдал Арикан, 2009).

Параллельно с теорией кодирования развивалась и *теория формальных грамматик.* Самым важным результатом этой теории, без

сомнения, является создание в 1956 году *иерархии Холмского*(см. [19]), названной так в честь американского философа и лингвиста - Аврама Ноама Хомского. Согласно этой иерархии все формальные грамматики можно разделить на 4 типа:

- *рекурсивно-перечислимые грамматики;*
- *контекстно-зависимые грамматики;*
- *контекстно-свободные грамматики;*
- *регулярные грамматики.*

Для каждой из этих грамматик известно, какими моделями можно их распознавать. Так, рекурсивно-перечислимые грамматики распознаются *машинами Тьюринга,* контекстно-зависимые грамматики - *линейными ограниченными автоматами,* контекстно-свободные грамматики - *автоматами с магазинной памятью,* и, наконец, регулярные грамматики - *абстрактными конечными автоматами*(см.,например, [20]).

Диссертация посвящена одному из основных типов кодирования - *алфавитному кодированию.* Основоположником этого направления в России можно считать русского математика из Нижнего Новгорода Александра Александровича Маркова. Обязательно следует упомянуть здесь его работу "Основания общей теории кодов", которая, несомненно, внесла большой вклад в развитие теории ко-

дирования в России(см. [21]). В 1984 году Ал. А. Марков успешно защищает в Московском государственном университете докторскую диссертацию по теме "Вопросы взаимной однозначности и сложности в алфавитном кодировании". Там ставится и успешно решается вопрос об алгоритмической разрешимости проблемы однозначности алфавитного декодирования для класса регулярных языков из классификации Хомского(см. [22]). В дальнейшем, для краткости, мы будем обозначать эту проблему как *проблему ОАД.* Для случая, когда регулярный язык совпадает с множеством всех слов входного алфавита, этот результат широко известен и называется Теоремой Маркова(см. [23]). Также известно, что в классе контекстно-свободных грамматик проблема ОАД при ряде дополнительных незначительных ограничений алгоритмически неразрешима(см. [24]). Тем не менее, оставался открытым вопрос о практической реализации решения проблемы ОАД в классе регулярных языков. В связи с этим автором диссертации была разработана специальная автоматно-алгебраическая техника, дающая альтернативное доказательство алгоритмической разрешимости проблемы ОАД и обеспечивающая необходимые оценки сложности алгоритма в терминах абстрактных конечных автоматов. Для этого сначала был введен и исследован подкласс регулярных языков - *класс тонких языков* $\mathfrak{T}(A)$. Этот класс можно рассматривать, как

класс языков в алфавите A, которые

- регулярны;
- имеют не более чем линейную функцию роста.

Было дано описание структуры таких языков, предложен способ их канонического представления и приведены оценки на его сложность. Все эти результаты позволили получить хорошие оценки на сложность алгоритма, доставляющего решение проблемы ОАД в классе $\mathfrak{T}(A)$. После этого исследование было перенесено на гораздо более широкий класс регулярных языков - *класс* $RP(A)$. Под этим классом мы здесь понимаем класс языков в алфавите A, которые

- регулярны;
- имеют полиномиальную функцию роста.

Также автором был предложен алгоритм решения проблемы ОАД для этого класса, работающий достаточно быстро для того, чтобы его можно было реализовать на практике. Прежде чем переходить к описанию такой реализации в диссертации нужно было еще решить пару проблем - проблему сравнения полезности произвольной пары функций алфавитного кодирования (или, сокращенно, *проблему* $ВКД_1$*)* и проблему сравнения полезности произвольной пары регулярных языков (или, сокращенно, *проблему* $ВКД_2$*)*. Удалось доказать, что первая проблема всегда алгоритмически разреши-

ма. Для второй проблемы алгоритмическая разрешимость показана для случая, когда мощность входного алфавита равна двум. Чтобы полученные результаты не выглядели оторванными от реальности, их практическое применение демонстрируется на примере разработанной автором модели передачи информации на большие расстояния по оптическому волокну, в которой применяется алфавитное кодирование в классах $\mathfrak{T}(A)$ и $RP(A)$. Исследован вопрос о надежности такой системы. Автор работы особо подчеркивает, что эта модель не претендует на внедрение в стандарты шифрования и ее основной задачей является иллюстрация одного из способов применения полученных теоретических результатов в реальной жизни. В рамках этого вопроса в конце каждой главы дается описание применения полученных теоретических результатов на практике.

Цели и задачи работы. Основной целью работы является разработка нового *автоматно-алгебраического подхода* к решению проблемы ОАД, доставляющего полиномиальные верхние оценки на сложность решения проблемы в классах $\mathfrak{T}(A)$ и $RP(A)$. Так же необходимо разработать модель, демонстрирующую практическую ценность применения этого подхода. Поэтапное достижение основной цели ставит перед автором следующие задачи.

- Разработать общий автоматно-алгебраический подход к решению проблемы ОАД.

- Привести для языков из класса $\mathfrak{T}(A)$ полное неизбыточное описание в терминах конечных объединений *прогрессивных множеств.*

- Привести для языков из класса $RP(A)$ полное неизбыточное описание в терминах конечных объединений *множеств правильного линейного вида.*

- Применить общий подход решения проблемы ОАД к классам $\mathfrak{T}(A), RP(A)$ и, используя информацию об их структуре, улучшить его, получив соответствующие оценки на сложность решающего алгоритма.

- Найти алгоритмическое решение проблем ВКД_1 и ВКД_2.

- Показать, что процедура алфавитного декодирования в классах $\mathfrak{T}(A), RP(A)$ имеет простую сложность.

Научная новизна. Результаты являются новыми, получены автором самостоятельно. Основные результаты:

- Предложен новый автоматно-алгебраический подход к решению проблемы ОАД.

- Приведена классификация внутренней структуры $\mathfrak{T}(A)$ в терминах конечных объединений прогрессивных множеств.

- Приведена классификация внутренней структуры $RP(A)$ в

терминах конечных объединений множеств правильного линейного вида.

- Показана полиномиальность верхних оценок на сложность решения проблемы ОАД новым автоматно-алгебраическим подходом.
- Показано, что проблема ВКД_1 алгоритмически разрешима.
- Показано, что проблема ВКД_2 алгоритмически разрешима в случае, когда мощность входного алфавита равна двум.
- Предложена модель, демонстрирующая практическую значимость полученных результатов.
- Исследован вопрос о сложности процедуры алфавитного декодирования в классах $\mathfrak{T}(A)$ и $RP(A)$.

Теоретическая и практическая значимость. Работа имеет теоретический характер и может быть интерпретирована как продолжение исследований Ал. А. Маркова в области алфавитного кодирования. С помощью результатов работы можно эффективно проверять свойство однозначности алфавитного кодирования в языках из классов $\mathfrak{T}(A)$ и $RP(A)$. Также эти результаты могут быть применены в области исследований по сжатию информации. Разработана модель, демонстрирующая практическую ценность полученных теорем и при передаче информации по оптическому ка-

налу связи с целью сокрытия внутренней логики соответствующего кодирования.

Методология и методы исследования. В работе применены методы дискретной математики, теории чисел и теории автоматов.

Положения, выносимые на защиту:

1. Получение верхних полиномиальных оценок на сложность решения проблемы однозначности алфавитного кодирования в классах $\mathfrak{T}(A)$ и $RP(A)$.

2. Построение полной неизбыточной классификации внутренней структуры элементов из класса $\mathfrak{T}(A)$ в терминах конечных объединений прогрессивных множеств.

3. Построение полной неизбыточной классификации внутренней структуры элементов из класса $RP(A)$ в терминах конечных объединений множеств правильного линейного вида.

Структура и объем диссертации. Диссертация состоит из введения, раздела благодарностей, 6 глав, заключения, краткого списка обозначений и библиографии. Общий объем диссертации - 212 страниц, из них 199 страниц текста. Библиография включает в себя 40 наименований на 5 страницах.

Краткое содержание диссертации.

Во введении приведен исторический обзор по теме диссертации, поставлены ее основные цели и задачи, обоснованы научная

новизна и практическая значимость работы, сформулированы методология и выносимые на защиту положения, описана структура и краткое содержание диссертации.

В первой главе изложено новое автоматно-алгебраическое доказательство алгоритмической разрешимости проблемы ОАД в случае, когда кодируемое множество слов является произвольным регулярным множеством. Приводятся необходимые определения. Также описаны следствия из этого доказательства для случая, когда множества имеют полиномиальную функцию роста. В заключение, приводятся некоторые примеры, показывающие существенную неулучшаемость предложенного алгоритма.

Во второй главе описывается класс языков $\mathfrak{T}(A)$, приводится критериальное описание его элементов в терминах прогрессивных множеств. Особо выделяется случай, когда ограничивающая константа равна 1. Такие языки называются 1-тонкими и соответствующий им класс обозначается через $\mathfrak{T}_1(A)$. Для класса $\mathfrak{T}_1(A)$, так же как и для класса $\mathfrak{T}(A)$, приводится критериальное описание, использующее такие понятия, как спектральная независимость и общепрогрессивное множество.

В третьей главе описывается класс языков $RP(A)$, приводится критериальное описание его элементов в терминах множеств правильного линейного вида. Кроме того, выявляется связь введенного

ранее в главе 2 класса $\mathfrak{T}(A)$ с классом регулярных языков с не более чем линейной функцией роста.

В четвертой главе приводится решение проблемы ОАД для класса $\mathfrak{T}(A)$ тонких языков в некотором произвольном алфавите A.

В пятой главе приводится решение проблемы ОАД для класса $RP(A)$ регулярных языков с полиномиальной функцией роста в некотором произвольном алфавите A.

В шестой главе изучаются две проблемы вложения: проблема вложения классов допустимых регулярных языков, задаваемых функциями алфавитного кодирования (сокращенно - проблема ВКД_1) и проблема вложения классов допустимых функций алфавитного кодирования, задаваемых регулярными языками (сокращенно - проблема ВКД_2). Исследуется алгоритмическая разрешимость этих проблем.

Благодарности

Автор выражает глубокую благодарность своему научному руководителю академику профессору *Валерию Борисовичу Кудрявцеву* за постановку задачи, постоянную поддержку и за доброжелательное внимание к работе. Автор выражает благодарность сотрудникам кафедры Математической теории интеллектуальных систем докторам физико-математических наук, профессорам в лице *Алешина Станислава Владимировича*, *Буевича Вячеслава Александровича*, *Гасанова Эльяра Эльдаровича*, *Подколзина Александра Сергеевича* и кандидатов физико-математических наук в лице *Часовских Анатолия Александровича*, *Галатенко Алексея Владимировича*, *Пантелеева Павла Анатольевича* , *Жука Дмитрия Николаевича* и *Бокова Григория Владимировича* за полезные обсуждения по развитию научной работы и конструктивную критику. Автор выражает глубокую благодарность своим родителям *Дергачу Сергею Петровичу* и *Дергач Валентине Юрьевне* за поддержку и терпение.

Глава 1

Аннотация

В этой главе изложено новое автоматно-алгебраическое доказательство алгоритмической разрешимости проблемы ОАД в случае, когда кодируемое множество слов является произвольным регулярным множеством. Также приведены следствия из этого доказательства для случая, когда множества имеют полиномиальную функцию роста. В заключение, приводятся некоторые примеры, показывающие существенную неулучшаемость предложенного алгоритма.

1. Основные понятия и результаты

Абстрактным конечным автоматом называется набор

$$V = (A, Q, B, \varphi, \psi),$$

где A, Q, B - конечные множества, φ - функция, определенная на множестве $Q \times A$ и принимающая значения из Q, ψ - функция, определенная на множестве $Q \times A$ и принимающая значения из B. Множества A, Q, B называются соответственно *входным алфавитом*, *алфавитом состояний* и *выходным алфавитом* автомата V. Функция φ называется *функцией переходов*, а функция ψ - *функцией выходов* автомата V. *Входными словами* автомата V, $V = (A, Q, B, \varphi, \psi)$ называем произвольные конечные последовательности символов алфавита A. Для удобства рассматриваем при

этом также "пустое"слово, не имеющее ни одного символа и обозначаемое Λ. *Выходными словами* алфавита V называем конечные последовательности символов алфавита B, *словами состояний* - конечные последовательности символов алфавита Q (в обоих случаях допускается и пустое слово Λ). Для каждого состояния автомата V можно рассмотреть набор $(A, Q, B, \varphi, \psi, q)$, определяющий автомат V с выделенным начальным состоянием q. Такие наборы $(A, Q, B, \varphi, \psi, q)$ называются *инициальными абстрактными конечными автоматами.* Для краткости обозначаем их через V_q. Для произвольного $n \in \mathbb{N}$ через $K(A, B, n)$ обозначаем множество всех инициальных абстрактных конечных автоматов с входным алфавитом A, выходным алфавитом B и алфавитом состояний мощности n. Через $K_{\leq}(A, B, n)$ обозначаем множество всех инициальных абстрактных конечных автоматов с входным алфавитом A, выходным алфавитом B и алфавитом состояний мощности не выше n.

Для произвольного $V_q = (A, Q, B, \varphi, \psi, q)$ через $[V_q]$ обозначаем множество

$$\{V_q{}' = (A, Q, B, \varphi, \psi, q') \mid q' \in Q\}$$

инициальных абстрактных конечных автоматов, полученных из V_q изменением начального состояния. Сам автомат V_q тоже входит в $[V_q]$.

Введем ряд понятий, связанных со словами. Пусть C - некоторое

конечное множество. Если $\gamma = c(1)\dots c(n)$ - конечная последовательность символов $c(1),\dots,c(n)$ алфавита C, то говорим, что γ есть *слово в алфавите* C. Число n называем *длиной* слова γ и обозначаем через $|\gamma|$. Длина пустого слова, по определению, равна 0. Пусть $\alpha = c(1)\dots c(n)$, $\beta = c'(1)\dots c'(m)$ - два слова в алфавите C. Говорим, что *слово* $c''(1)\dots c''(n+m)$ *получено приписыванием слова* β *к слову* α, если

$$c''(i) = \begin{cases} c(i), & \text{если } 1 \le i \le n; \\ c'(i-n), & \text{если } n+1 \le i \le n+m. \end{cases}$$

Для краткости обозначаем слово $c''(1)\dots c''(n+m)$ через $\alpha\beta$. Если γ и δ - слова, причем $\gamma = \delta\delta'$ для некоторого слова δ', то говорим, что δ - *префикс* слова γ, δ' - *постфикс* слова γ. Множество всех слов в алфавите A обозначаем через A^*. Префикс слова γ, имеющий длину l, обозначаем $[_l(\gamma)$. Постфикс слова γ, имеющий длину l, обозначаем через $]_l(\gamma)$. Через $\gamma_{l,m}$ обозначаем слово $]_{m-l}([_m(\gamma))$, где $|\gamma| \ge m > l \ge 1$. Для произвольных $n \in \mathbb{N}$, $P \subseteq A^*$ через $P_{\le}(n)$ обозначаем множество слов из P, длина которых не превосходит n.

Функции переходов и выходов алфавита $V = (A, Q, B, \varphi, \psi)$ доопределим на множестве $Q \times A^*$ (сохраним за ними те же обозначения). Именно, полагаем по определению

$$\varphi(q, \Lambda) = q, \ \ \varphi(q, \alpha a) = \varphi(\varphi(q, \alpha), a),$$

$$\psi(q,\Lambda)=\Lambda,\ \psi(q,\alpha a)=\psi(\varphi(q,\alpha)a),$$

где $q\in Q,\ \alpha\in A^*,\ a\in A$.

Пусть $V_q=(A,Q,B,\varphi,\psi,q)$ - инициальный абстрактный конечный автомат, $B'\subseteq B$. Множество

$$\{\alpha\mid\alpha\in A^*,\ \psi(q,\alpha)\in B'\}$$

называем *распознаваемым в конечном автомате* V_q *с помощью подмножества* B' выходных символов и обозначаем его через $B'(V_q)$. Говорим также, что автомат V_q *распознает* $B'(V_q)$ *посредством* B'. Пусть $P\subseteq A^*\setminus\{\Lambda\}$. Если существует конечный автомат V_q, распознающий множество P с помощью некоторого подмножества $B'\subseteq B$, то множество P будем называть *распознаваемым*. Заметим, что если ограничиться случаем $B=\{0,1\}$ и $B'=\{1\}$, то класс распознаваемых множеств от этого не изменится. Подробнее об этом можно прочитать в [26]. Множество $\{0,1\}$ обозначаем для краткости через E_2.

Недетерминированным конечным автоматом называется набор

$$V=(A,Q,B,\gamma),$$

где A,Q,B - конечные множества, γ - функция, определенная на множестве $Q\times A$ и принимающая в качестве своих значений подмножества множества $Q\times B$. Если для каждой пары (q,a), где $q\in Q$ и $a\in A$ значение $\gamma(q,a)$ есть одноэлементное множество (q',b), то можно определить функции φ,ψ:

$$\varphi(q,a)=q',\ \psi(q,a)=b.$$

В указанном смысле абстрактный конечный автомат можно рассматривать как частный случай недетерминированного конечного автомата.

Понятие инициального недетерминированного конечного автомата, аналогичное понятию инициального абстрактного конечного автомата, возникает, если в набор (A,Q,B,γ) добавляется некоторое выделенное подмножество $Q'\subseteq Q$. Полученный автомат обозначаем для краткости через $V_{Q'}$.

Пусть $\alpha=a(1)\dots a(s)$ - слово в алфавите A. Определим класс $\tilde{\gamma}(Q',\alpha)$ последовательностей

$$(q(1),b(1)),\dots,(q(s),b(s)).$$

Каждая последовательность этого класса удовлетворяет следующим условиям:

(1) $q(1)\in Q'$;

(2) $(q(i+1),b(i))\in\gamma(q(i),a(i)), i=1,\dots,s$ (в случае $i=s$ рассматривается значение $q(s+1)$, не включаемое в последовательность $\tilde{\gamma}(Q',\alpha)$).

Пусть $V_{Q'}=(A,Q,B,\gamma,Q')$ - инициальный недетерминированный конечный автомат. Его можно рассматривать как модель устройства, применяемого для распознавания входных последовательностей. При этом выделяется подмножество B' множества B

выходных символов. Множество непустых входных слов, *распознаваемых автоматом*, состоит из тех и только тех слов $\alpha \in A^*$, для которых в $\tilde{\gamma}(Q', \alpha))$ существует последовательность

$$(q(1), b(1)), \ldots, (q(s), b(s)),$$

удовлетворяющая условию $b(s) \in B'$. Обозначаем это множество через $B'(V_{Q'})$. Говорим также, что *множество* $B'(V_{Q'})$ *распознаваемо в недетерминированном конечном автомате* $V_{Q'}$ *посредством* подмножества B' выходных символов.

Пусть A, B - конечные непустые алфавиты и $n \in \mathbb{N}$. Класс инициальных недетерминированных конечных автоматов

$$(A, Q, B, \gamma, Q'),$$

в которых $|Q| = n$, обозначаем для краткости через $\tilde{K}(A, B, n)$.

Пусть $A = \{a_1, \ldots, a_r\}$ - произвольный конечный непустой алфавит. Пусть P_1, P_2 - непустые множества слов в алфавите A. Определим следующие операции над P_1 и P_2:

1. *Объединение* множеств P_1 и P_2 (обозначаем $P_1 \bigcup P_2$) есть множество всех слов вида α, где $\alpha \in P_1$ или $\alpha \in P_2$.

2. *Конкатенация* множеств P_1 и P_2 (обозначаем $P_1 \cdot P_2$) есть множество всех слов вида $\alpha_1\alpha_2$, где $\alpha_1 \in P_1,\ \alpha_2 \in P_2$.

3. *Итерация* множества P_1 (обозначаем $(P_1)^*$) есть множество всех слов вида $\alpha_1 \ldots \alpha_k$, где $\alpha_1 \in P_1,\ \ldots, \alpha_k \in P_1,\ k \geq 1$. Иногда

будем также считать, что $k \geq 0$, где для $k = 0$ имеем $\alpha_1 \dots \alpha_k = \Lambda$. Ясно, что этот подход отличается от стандартного только ответом на вопрос - добавлять ли пустое пустое слово в итерацию.

Введем понятие регулярного множества в алфавите A. Называем множество $P,\ P \subseteq A^*$ *регулярным в алфавите* A, если его можно получить из пустого множества и одноэлементных однобуквенных множеств $\{a\},\ a \in A$, применением конечного числа конкатанаций, объединений и итераций. Более подробно, определение регулярных множеств таково:

(1) $\emptyset$ - регулярное множество в алфавите A;

(2) $\{a\}$, где a - произвольная буква алфавита A, - регулярное множество в алфавите A;

(3) Если P_1, P_2 - регулярные множества в алфавите A, то и множества $P_1 \cup P_2,\ P_1 \cdot P_2,\ (P_1)^*$ - регулярные множества в алфавите A;

(4) Регулярность произвольного множества в алфавите A устанавливается в соответствиями с пунктами (1)-(3) за конечное число шагов.

Множество регулярных множеств в алфавите A обозначаем через $R(A)$.

Введем понятие регулярного выражения в алфавите A. Регулярное выражение в алфавите A представляет собой слово в алфавите

$A \cup \{\vee, \cdot, (,), \lambda, ^*\}$, определяемое следующим образом:

(1) λ - регулярное выражение в алфавите A;

(2) Буквы алфавита A - регулярные выражения в алфавите A;

(3) Если $\mathfrak{P}_1,\ \mathfrak{P}_2$ - регулярные выражения в алфавите A, то и $(\mathfrak{P}_1 \vee \mathfrak{P}_2)$, $(\mathfrak{P}_1 \cdot \mathfrak{P}_2)$, $(\mathfrak{P}_1)^*$ - регулярные выражения в алфавите A;

(4) Регулярность произвольного выражения в алфавите A устанавливается в соответствиями с пунктами (1)-(3) за конечное число шагов.

Сопоставим индуктивно каждому регулярному выражению $\mathfrak{P}$ в алфавите A множество $|\mathfrak{P}|$ в алфавите A:

(1) Множество $\emptyset$ - в случае $\mathfrak{P} = \lambda$;

(2) Множество $\{a\}$ - в случае $\mathfrak{P} = a,\ a \in A$;

(3) Множество $|\mathfrak{P}_1| \cup |\mathfrak{P}_2|$ - в случае $\mathfrak{P} = (\mathfrak{P}_1 \vee \mathfrak{P}_2)$;

(4) Множество $|\mathfrak{P}_1| \cdot |\mathfrak{P}_2|$ - в случае $\mathfrak{P} = (\mathfrak{P}_1 \cdot \mathfrak{P}_2)$;

(5) Множество $(|\mathfrak{P}_1|)^*$ - в случае $\mathfrak{P} = (\mathfrak{P}_1)^*$.

Говорим, что множество $|\mathfrak{P}|$ *представимо регулярным выражением* $\mathfrak{P}$. Очевидно, что множества, представимые регулярными выражениями, регулярны. Верно и обратное. Любое регулярное множество, в силу определения, представимо хотя бы одним регулярным выражением. В некотором смысле можно считать, что регулярное выражение - способ построения регулярных множеств из пустого множества и букв алфавита.

Зафиксируем два конечных непустых алфавита A и B. Пусть есть какое-то отображение $f : A \to B^* \setminus \{\Lambda\}$:

$$f(a_1) = \beta_1$$
$$f(a_2) = \beta_2$$
$$\ldots$$
$$f(a_r) = \beta_r.$$

Это отображение называем *схемой кодирования из алфавита A в алфавит B*. Множество всех схем кодирования из алфавита A в алфавит B обозначаем через $F(A, B)$. *Длиной схемы кодирования* называем число $|\beta_1| + \ldots + |\beta_r|$ и обозначаем его через L_f. *Сложностью схемы кодирования* называем число $\max\limits_{1 \le i \le r} |\beta_i|$ и обозначаем его через l_f.

Доопределим отображение f до отображения $\tilde{f} : A^* \to B^*$ следующим образом:

$$\tilde{f}(\Lambda) = \Lambda,$$
$$\tilde{f}(a_{i_1} a_{i_2} \ldots a_{i_n}) = \beta_{i_1} \beta_{i_2} \ldots \beta_{i_n}.$$

Отображение $\tilde{f}$ называем *алфавитным кодированием из алфавита A в алфавите B*.

Для произвольного $P \subseteq A^*$ через $\tilde{f}(P)$ обозначаем множество

$$\tilde{f}(P) := \{\tilde{f}(\alpha) \mid \alpha \in P\}.$$

Для произвольного $P \subseteq A^*$ обозначаем через $(\tilde{f})_P$ функцию

$$(\tilde{f})_P : P \to B^*,$$

полученную из $\tilde{f}$ сужением на P. Пусть $f \in F(A, B)$ - схема кодирования. Обозначаем через $I(f)$ множество

$$I(f) := \{P \subseteq A^* \setminus \{\lambda\} \mid (\tilde{f})_P - \text{инъекция}\},$$

называемое *классом допустимых языков для схемы* f.

Проблемой проверки однозначности алфавитного декодирования в классе регулярных языков (или сокращенно - проблемой $ОАД_1$*)* называем проблему распознавания свойства

$$P \in I(f)$$

для произвольных $f \in F(A, B)$ и $P \in R(A)$.

Пусть есть некоторое регулярное множество P в алфавите A и некоторое алфавитное кодирование $\tilde{f}$. Пусть $\beta \in \tilde{f}(P)$. Тогда $\alpha \in P$ называется *расшифровкой* β *при алфавитном кодировании* $\tilde{f}$ *на регулярном множестве* P или просто *расшифровкой* β, если $f(\alpha) = \beta$. Таких расшифровок может быть несколько. Если для любых различных слов $\alpha_1, \alpha_2 \in P$ выполняется $\tilde{f}(\alpha_1) \neq \tilde{f}(\alpha_2)$, то *декодирование однозначно на* P *по* $\tilde{f}$. Также говорим, что $\tilde{f}$ *биективно на* P.

Через $\mathbb{N}_0$ обозначаем множество $\mathbb{N} \cup \{0\}$.

Теорема 1.1 *Пусть A, B - некоторые конечные непустые алфавиты, $f \in F(A, B)$, $P \subseteq A^*$. Пусть, кроме того, для некоторых фиксированных $m, k \in \mathbb{N}$ имеем:*

1) существует $V \in K(A, E_2, k)$, для которого $1(V) = P$;

2) для каждого $V_1 \in [V]$ существует $W_1 \in K_{\leq}(B, E_2, m)$, для которого $1(W_1) = \tilde{f}(1(V_1))$.

Тогда $P \in I(f)$ если и только если $P_{\leq}(k + m^2 + l_f) \in I(f)$.

Теорема 1.2 *Пусть A, B - некоторые конечные непустые алфавиты, $f \in F(A, B)$, $P \subseteq A^*$. Пусть, кроме того, P представимо в виде конечного объединения множеств P_i таких, что для некоторых фиксированных $m, k \in \mathbb{N}$ при всех i имеем:*

1) существует $V_i \in K_{\leq}(A, E_2, k)$, для которого $1(V_i) = P_i$;

2) для каждого $V \in [V_i]$ существует $W \in K_{\leq}(B, E_2, m)$, для которого $1(W) = \tilde{f}(1(V))$.

Тогда $P \in I(f)$ если и только если $P_{\leq}(k^2 + m^2 + l_f) \in I(f)$.

Теорема 1.3 *Пусть A, B - некоторые конечные непустые алфавиты, $f \in F(A, B)$, $P \subseteq A^*$. Пусть, кроме того, P представимо в виде конечного объединения множеств P_i таких, что для некоторых фиксированных $m, k \in \mathbb{N}$ при всех i имеем:*

1) существует $V_i \in K_{\leq}(A, E_2, k)$, для которого $1(V_i) = P_i$;

2) для каждого $V \in [V_i]$ существует конечное множество авто-

матов

$$W_i \in K_{\leq}(B, E_2, m),\ 1 \leq i \leq s,\ s \in \mathbb{N}$$

таких, что $\bigcup\limits_{i=1}^{s} 1(W_i) = \tilde{f}(1(V))$.
Тогда $P \in I(f)$ *если и только если* $P_{\leq}(k^2 + m^2 + l_f) \in I(f)$.

2. Доказательство вспомогательных утверждений

Лемма 1. *Пусть* A *- конечный непустой алфавит,* $n \in \mathbb{N}$ *и* $V_{Q'} \in \tilde{K}(A, E_2, n)$. *Тогда для некоторого* $m \in \mathbb{N}$, $m \leq 2^n$ *существует автомат* $V_q \in K(A, E_2, m)$ *такой, что*

$$1(V_{Q'}) = 1(V_q).$$

Доказательство этого факта приведено в [26].

Лемма 2(О склейке). *Пусть* $m, n \in \mathbb{N}$, A *- конечный алфавит и* δ_1, δ_2, ξ_1, ξ_2, β *- слова (возможно, пустые) в алфавите* A. *И пусть*

$$V_1 \in K(A, E_2, m), V_2 \in K(A, E_2, n);$$

$$1(V_1) = P_1, 1(V_2) = P_2;$$

$$\delta_1\beta\delta_2 \in P_1,\ \xi_1\beta\xi_2 \in P_2.$$

Тогда существует слово β' *в алфавите* A *такое, что*

$$\delta_1\beta'\delta_2 \in P_1,\ \xi_1\beta'\xi_2 \in P_2,\ |\beta'| \leq mn.$$

Доказательство. Пусть

$$V_1 = (A, Q_1, B, \varphi_1, \psi_1, q_1), V_2 = (A, Q_2, B, \varphi_2, \psi_2, q_2).$$

Здесь $|Q_1| = m$, $|Q_2| = n$. Допустим, что $|\beta| \geq |Q_1||Q_2| = mn$. Рассмотрим множество T пар состояний автоматов V_1, V_2 такое, что

$$T = \{(\varphi_1(q_1, \delta_1[_l(\beta)), \varphi_2(q_2, \xi_1[_l(\beta))) \mid 0 \leq l < |\beta|\}.$$

Так как $T \subseteq Q_1 \times Q_2$, то

$$|T| \leq |Q_1||Q_2|.$$

Из принципа Дирихле заключаем, что существуют

$$0 \leq l_1 < l_2 \leq |Q_1||Q_2|,\ l_1, l_2 \in \mathbb{N}_0,$$

для которых верно

$$\varphi_1(q_1, \delta_1[_{l_1}(\beta)) = \varphi_1(q_1, \delta_1[_{l_2}(\beta)),$$

$$\varphi_2(q_2, \xi_1[_{l_1}(\beta)) = \varphi_2(q_2, \xi_1[_{l_2}(\beta)).$$

Пусть β_1 и β_2 удовлетворяют соотношениям

$$\beta_1 = [_{l_1}(\beta),\ \beta = [_{l_2}(\beta)\beta_2.$$

Здесь $|\beta_2| > 0$, так как

$$l_2 \leq |Q_1||Q_2|,\ |\beta| > |Q_1||Q_2|.$$

Тогда

$$\psi_1(q_1, \delta_1\beta_1\beta_2\delta_2) = \psi_1(\varphi_1(q_1, \delta_1\beta_1), \beta_2\delta_2) =$$

$$= \psi_1(\varphi_1(q_1, \delta_1[_{l_1}(\beta)), \beta_2\delta_2) = \psi_1(\varphi_1(q_1, \delta_1[_{l_2}(\beta)), \beta_2\delta_2) =$$

$$= \psi_1(\varphi_1(q_1, \delta_1), [_{l_2}(\beta)\beta_2\delta_2) = \psi_1(\varphi_1(q_1, \delta_1), \beta\delta_2) = \psi_1(q_1, \delta_1\beta\delta_2) = 1.$$

Аналогично выводим, что

$$\psi_2(q_2, \xi_1\beta_1\beta_2\xi_2) = \psi_2(\varphi_2(q_2, \xi_1\beta_1), \beta_2\xi_2) =$$

$$= \psi_2(\varphi_2(q_2, \xi_1[_{l_1}(\beta)), \beta_2\xi_2) = \psi_2(\varphi_2(q_2, \xi_1[_{l_2}(\beta)), \beta_2\xi_2) =$$

$$= \psi_2(\varphi_2(q_2, \xi_1), [_{l_2}(\beta)\beta_2\xi_2) = \psi_2(\varphi_2(q_2, \xi_1), \beta\xi_2) = \psi_2(q_2, \xi_1\beta\xi_2) = 1.$$

Отсюда, по опредению, немедленно следует, что

$$\delta_1\beta_1\beta_2\delta_2 \in B_1,\ \xi_1\beta_1\beta_2\xi_2 \in B_2.$$

При этом, $|\beta_1\beta_2| < |\beta|$. Проводя необходимое количество раз изложенную выше процедуру сокращения $\beta \to \beta_1\beta_2$, окончательно получаем слово β' длины не большей $|Q_1||Q_2| = mn$, для которого

$$\delta_1\beta'\delta_2 \in B_1,\ \xi_1\beta'\xi_2 \in B_2.$$

Лемма 2 доказана.

Лемма 3(О разрезе). *Пусть $n \in \mathbb{N}$, A - конечный непустой алфавит, $V \in K(A, E_2, n)$, $1(V) = P$, $\alpha \in A^*$ и $s \in \mathbb{N}$, причем $s < |\alpha|$. Тогда существуют $V_1, V_2 \in K(A, E_2, n)$ такие, что*

$$1(V_1) \cdot 1(V_2) \subseteq P,\ [_s(\alpha) \in 1(V_1),\ ]_{|\alpha|-s}(\alpha) \in 1(V_2), V_2 \in [V].$$

Доказательство. Пусть

$$V = (A, Q, E_2, \varphi, \psi, q_0).$$

Определяем искомую пару инициальных абстрактных конечных автоматов

$$V_1 = (A, Q, E_2, \varphi_1, \psi_1, q_0), V_2 = (A, Q, E_2, \varphi_2, \psi_2, q_1)$$

следующим образом:

$$q_1 = \varphi(q_0, [_s(\alpha));$$

$$\psi_1(q, a) = \begin{cases} 1, & \text{при } \varphi(q, a) = q_1, \\ 0, & \text{иначе;} \end{cases}$$

$$\psi_2(q, a) = \psi(q, a);$$

$$\varphi_1(q, a) = \varphi_2(q, a) = \varphi(q, a).$$

Покажем, что все условия леммы выполнены. Прежде всего, замечаем, что $V_1, V_2 \in K(A, E_2, n)$ и $V_2 \in [V]$. Обозначаем через P_1 и P_2 множества, распознаваемые посредством $\{1\}$ автоматами V_1 и V_2 соответственно. Замечаем, что $\psi_1(q_0, [_s(\alpha)) = 1$, так как $\varphi(q_0, [_s(\alpha)) = q_1$. Значит $[_s(\alpha) \in P_1$. Далее, так как $\alpha \in P$, то

$$1 = \psi(q_0, \alpha) = \psi\big(\varphi(q_0, [_s(\alpha)),]_{|\alpha|-s}(\alpha)\big) = \psi(q_1,]_{|\alpha|-s}(\alpha)).$$

Поэтому $]_{|\alpha|-s}(\alpha) \in P_2$. Значит $\alpha = [_s(\alpha) \cdot]_{|\alpha|-s}(\alpha) \in P_1 \cdot P_2$.

Докажем теперь, что $P_1 \cdot P_2 \subseteq P$. Пусть $\alpha_1 \in P_1, \alpha_2 \in P_2$. Тогда

$$\varphi(q_0, \alpha_1) = q_1, \quad \psi(q_1, \alpha_2) = 1.$$

Значит

$$\psi(q_0, \alpha_1\alpha_2) = \psi(\varphi(q_0, \alpha_1), \alpha_2) = \psi(q_1, \alpha_2) = 1,$$

то есть $\alpha_1\alpha_2 \in P$. Поэтому $P_1 \cdot P_2 \subseteq P$. Все условия леммы выполнены. Лемма 3 доказана.

Лемма 4. (Об образе) *Пусть A и B - некоторые непустые конечные алфавиты и $f \in F(A, B)$, $V_{q_0} \in K(A, E_2, n)$, $1(V_{q_0}) = P$. Тогда существуют $m \in \mathbb{N}$, $V_{q_0'} \in K(B, E_2, m)$ такие, что*

$$m \leq 2^{|Q|(L_f - |A|+1)},\ 1(V_{q_0'}) = \tilde{f}(P).$$

Доказательство. Пусть $A = \{a_1, \ldots, a_r\}$. Заменим в автомате V_{q_0} входной алфавит A на алфавит $\tilde{f}(A)$ и поправим функции перехода и выхода, заменив в них все a_i на $\tilde{f}(a_i)$:

$$\varphi'(q, \tilde{f}(a_i)) = \varphi(q, a_i),\ \psi'(q, \tilde{f}(a_i)) = \psi(q, a_i).$$

Получаем некоторый инициальный абстрактный конечный автомат

$$W_{q_0} \in K(\tilde{f}(A), E_2, m).$$

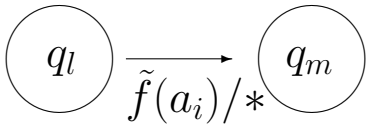
Далее, рассматриваем все наборы

$$(q_l, q_m, *, f(a_i)),$$

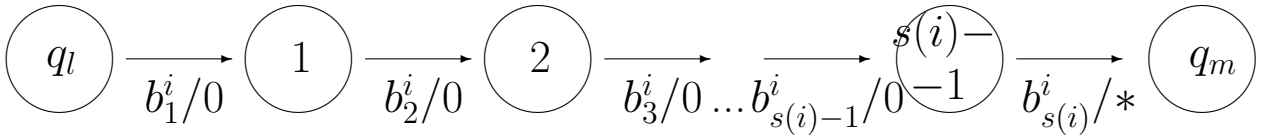
где q_l, q_m - состояния автомата W_{q_0}, $* \in E_2$ и $f(a_i) = b_1^i \ldots b_{s(i)}^i$, для которых выполнено

$$\varphi'(q_l, f(a_i)) = q_m,\ \psi'(q_l, f(a_i)) = *.$$

Для каждого такого набора поменяем в функциях φ' и ψ' соответствующий переход на последовательность состояний с побуквенными переходами $b^i_1, \ldots, b^i_{s(i)}$ следующим образом:



переходит в:



Получаем некоторый инициальный недетерминированный конечный автомат $\tilde{V}_{q_0} \in K(B, E_2, m)$, представляющий по $\{1\}$ множество $\tilde{f}(A)$. Найдем мощность m множества состояний автомата $\tilde{V}_{q_0}$. От автомата W_{q_0} остается $|Q|$ состояний. Для каждого перехода

$$\varphi'(q_l, \tilde{f}(a_i)) = q_m, \ \psi'(q_l, \tilde{f}(a_i)) = *$$

к ним добавляется $s(i) - 1$ состояний. Если зафиксировать q_l, то всего таких переходов будет $r = |A|$ штук - по одному для каждого $\tilde{f}(a_i)$. В итоге, для каждого q_l получаем дополнительные

$$(s(1) - 1) + (s(2) - 1) + \ldots + (s(r) - 1) = L_f - r$$

состояний. Поэтому

$$m = |Q| + |Q|(L_f - r) = |Q|(L_f - |A| + 1).$$

Доказательство леммы завершает применение леммы 1. Лемма 4 доказана.

Лемма 5. (О минимизации) *Пусть выполнены следующие условия:*

1) A и B - непустые конечные алфавиты;

2) m, n - фиксированные натуральные числа;

3) $f \in F(A, B)$, $V_q \in K(A, E_2, n)$, $1(V_q) = P$;

4) для каждого $V_{q'} \in [V_q]$ существует $W_{q'} \in K_{\leq}(B, E_2, m)$, для которого $1(W_{q'}) = \tilde{f}(1(V_{q'}))$;

5) $\alpha_1, \alpha_2 \in P$, $\alpha_1 \neq \alpha_2$, $\tilde{f}(\alpha_1) = \tilde{f}(\alpha_2)$.

Тогда существуют $\alpha'_1, \alpha'_2 \in P_{\leq}(n + m^2 + l_f)$ такие, что

$$\alpha'_1 \neq \alpha'_2,\ \tilde{f}(\alpha'_1) = \tilde{f}(\alpha'_2).$$

Доказательство. Пусть γ - наибольший по длине общий префикс слов α_1 и α_2. Слово γ, вообще говоря, может быть пустым. Так как $\alpha_1 \neq \alpha_2$, то эти слова можно представить в виде

$$\alpha_1 = \gamma a_1 \beta_1,\ \alpha_2 = \gamma a_2 \beta_2,$$

где a_1, a_2 - буквы алфавита A и $a_1 \neq a_2$. Здесь мы воспользовались еще и тем фактом, что $\tilde{f}(\alpha_1) = \tilde{f}(\alpha_2)$ и значит α_1 не может быть префиксом α_2, а α_2 - префиксом α_1. Если $\gamma \neq \Lambda$, то по лемме 3 о разрезе слова α_1, сделанного по автомату V_q и числу $s = |\gamma|$ получаем, что найдутся $V_1, \tilde{V}_1 \in K(A, E_2, n)$, для которых

$$1(V_1) \cdot 1(\tilde{V}_1) \subseteq P,\ \gamma \in 1(V_1),\ a_1\beta_1 \in 1(\tilde{V}_1), \tilde{V}_1 \in [V_q].$$

Случай, когда $\gamma = \Lambda$, разберем позже. Теперь применяем лемму

3 к слову $a_1\beta_1$ по автомату $\tilde{V}_1$ и числу $s = 1$. По ней найдутся $V_2, V_3 \in K(A, E_2, n)$ такие, что

$$1(V_2) \cdot 1(V_3) \subseteq 1(\tilde{V}_1),\ a_1 \in 1(V_2),\ \beta_1 \in 1(V_3), V_3 \in [\tilde{V}_1].$$

Окончательно получаем, что найдутся $V_1, V_2, V_3 \in K(A, E_2, n)$, для которых

$$1(V_1) \cdot 1(V_2) \cdot 1(V_3) \subseteq P,\ \gamma \in 1(V_1),\ a_1 \in 1(V_2),\ \beta_1 \in 1(V_3), V_3 \in [V_q].$$

Обозначаем

$$\tilde{m} := 2^{n(L_f - |A| + 1)}.$$

Из леммы 4 следует, что найдутся автоматы

$$V_1', V_2' \in K_{\leq}(B, E_2, \tilde{m})$$

такие, что

$$1(V_1') = \tilde{f}(1(V_1)),\ 1(V_2') = \tilde{f}(1(V_2)).$$

Кроме того, $V_3 \in [V_q]$, а это значит, что существует автомат

$$V_3' \in K_{\leq}(B, E_2, m),$$

для которого

$$1(V_3') = \tilde{f}(1(V_3)).$$

Применяя аналогичное рассуждение для слова α_2, строим автоматы $W_1, W_2, W_3 \in K(A, E_2, n)$ такие, что

$$1(W_1) \cdot 1(W_2) \cdot 1(W_3) \subseteq P,\ \gamma \in 1(W_1),$$

$$a_2 \in 1(W_2),\ \beta_2 \in 1(W_3), W_3 \in [V_q].$$

Также возникают автоматы

$$W_1', W_2' \in K(B, E_2, \tilde{m}),\ W_3' \in K_{\leq}(B, E_2, m),$$

для которых

$$1(W_1') = \tilde{f}(1(W_1)),\ 1(W_2') = \tilde{f}(1(W_2)),\ 1(W_3') = \tilde{f}(1(W_3)).$$

Отметим, что автоматы V_1 и W_1 совпадают, так как получены из автомата V_q применением леммы 3 о разрезе по одинаковому слову γ. Кроме того, q - начальное состояние у V_1.

Пусть $|\gamma| > n$ и φ_1, ψ_1 - функция переходов и функция выходов автомата V_1 соответственно. Рассмотрим множество

$$\{\varphi_1(q, [_l(\gamma)) \mid 0 \leq l \leq n\}.$$

Его мощность не превосходит мощности множества всех состояний автомата V_1, которая равна n. Поэтому из принципа Дирихле заключаем, что найдутся

$$0 \leq l_1 < l_2 \leq n,\ l_1, l_2 \in \mathbb{N}_0$$

такие, что

$$\varphi(q, [_{l_1}(\gamma)) = \varphi(q, [_{l_2}(\gamma)).$$

Рассмотрим слово $\gamma' = [_{l_1}(\gamma)\cdot]_{|\gamma|-l_2}(\gamma)$. Тогда

$$\psi_1(q, \gamma') = \psi_1(\varphi_1(q, [_{l_1}(\gamma)),]_{|\gamma|-l_2}(\gamma)) =$$

$$= \psi_1(\varphi_1(q, [l_2(\gamma)),]_{|\gamma|-l_2(\gamma)}) = \psi_1(q, [l_2(\gamma)\cdot]_{|\gamma|-l_2(\gamma)}) = \psi_1(q, \gamma) = 1.$$

Последнее равенство верно, так как $\gamma \in 1(V_1)$. Поэтому $\gamma' \in 1(V_1)$ и

$$\gamma' a_1 \beta_1 \in 1(V_1) \cdot 1(V_2) \cdot 1(V_3) \subseteq P,$$

$$\gamma' a_2 \beta_2 \in 1(W_1) \cdot 1(W_2) \cdot 1(W_3) \subseteq P.$$

При этом

$$\gamma' a_1 \beta_1 \neq \gamma' a_2 \beta_2,$$

так как $a_1 \neq a_2$. Кроме того,

$$\tilde{f}(\gamma' a_1 \beta_1) = \tilde{f}(\gamma' a_2 \beta_2),$$

ведь $\tilde{f}(\gamma a_1 \beta_1) = \tilde{f}(\gamma a_2 \beta_2)$. Наконец,

$$|\gamma'| = l_1 + |\gamma| - l_2 < |\gamma|.$$

Исходя из всего вышеизложенного изначально можем теперь считать, что $|\gamma| \leq n$. Пусть

$$\tilde{f}(\beta_1) = \nu_1, \ \tilde{f}(\beta_2) = \nu_2 \nu_1.$$

Так как

$$\tilde{f}(\gamma)\tilde{f}(a_1)\nu_1 = \tilde{f}(\gamma)\tilde{f}(a_1)\tilde{f}(\beta_1) = \tilde{f}(\gamma a_1 \beta_1) =$$

$$= \tilde{f}(\gamma a_2 \beta_2) = \tilde{f}(\gamma)\tilde{f}(a_2)\tilde{f}(\beta_2) = \tilde{f}(\gamma)\tilde{f}(a_2)\nu_2 \nu_1,$$

то

$$\tilde{f}(a_1) = \tilde{f}(a_2)\nu_2.$$

Применяя лемму 2 о склейке для автоматов

$$V_3', W_3' \in K_{\leq}(B, E_2, m)$$

и слов $\tilde{f}(\beta_1), \tilde{f}(\beta_2)$ по их общей части ν_1, получаем существование такого слова ν_1' в алфавите B, для которого выполнено

$$\nu_1' \in 1(V_3'),\ \nu_2\nu_1' \in 1(W_3'),\ |\nu_1'| \leq m^2.$$

Так как

$$1(V_3') = \tilde{f}(1(V_3)),\ 1(W_3') = \tilde{f}(1(W_3)),$$

то существуют

$$\beta_1' \in 1(V_3),\ \beta_2' \in 1(W_3)$$

такие, что

$$\tilde{f}(\beta_1') = \nu_1',\ \tilde{f}(\beta_2') = \nu_2\nu_1'.$$

При этом

$$\gamma a_1 \beta_1' \in 1(V_1) \cdot 1(V_2) \cdot 1(V_3) \subseteq P,$$

$$\gamma a_2 \beta_2' \in 1(W_1) \cdot 1(W_2) \cdot 1(W_3) \subseteq P.$$

Кроме того,

$$\tilde{f}(\gamma a_1 \beta_1') = \tilde{f}(\gamma)\tilde{f}(a_1)\nu_1' = \tilde{f}(\gamma)\tilde{f}(a_2)\nu_2\nu_1' =$$

$$= \tilde{f}(\gamma)\tilde{f}(a_2)\tilde{f}(\beta_2') = \tilde{f}(\gamma a_2 \beta_2').$$

Наконец,

$$\gamma a_1 \beta_1' \neq \gamma a_2 \beta_2',$$

так как $a_1 \neq a_2$. Исходя из всего вышеизложенного изначально можем теперь считать, что $|\nu_1| \leq m^2$.

Далее, так как $\tilde{f}(a_1) = \tilde{f}(a_2)\nu_2$, то

$$|\nu_2| = |\tilde{f}(a_1)| - |\tilde{f}(a_2)| \leq l_f - 1.$$

Значит

$$|\gamma a_1 \beta_1| = |\gamma| + |a_1| + |\beta_1| \leq n+1+|\tilde{f}(\beta_1)| = n+1+|\nu_1| \leq n+1+m^2,$$

$$|\gamma a_2 \beta_2| = |\gamma| + |a_2| + |\beta_2| \leq n+1+|\tilde{f}(\beta_2)| = n+1+|\nu_2\nu_1| \leq$$

$$\leq n+1+m^2+l_f-1 = n+m^2+l_f.$$

Получили требуемую оценку. Для случая, когда $\gamma = \Lambda$ рассуждения проходят аналогично с той лишь разницей, что в них нет автоматов V_1, W_1, V_1', W_1'. И итоговая оценка станет равна $m^2 + l_f$. Утверждение леммы доказано.

3. Доказательство основных утверждений

Теорема 1.1 *Пусть A, B - некоторые конечные непустые алфавиты, $f \in F(A,B)$, $P \subseteq A^*$. Пусть, кроме того, для некоторых фиксированных $m, k \in \mathbb{N}$ имеем:*

1) существует $V \in K(A, E_2, k)$, для которого $1(V) = P$;

2) для каждого $V_1 \in [V]$ существует $W_1 \in K_{\leq}(B, E_2, m)$, для которого $1(W_1) = \tilde{f}(1(V_1))$.

Тогда $P \in I(f)$ если и только если $P_{\leq}(k+m^2+l_f) \in I(f)$.

Доказательство. Нетрудно видеть, что утверждение теоремы 1.1 непосредственно вытекает из леммы 5. Теорема доказана.

Замечание. Из леммы 4 следует, что в качестве числа m из формулировки теоремы 1 всегда можно взять число $2^{k(L_f-|A|+1)}$. Ясно, что это очень грубая оценка. Для класса тонких языков и для класса регулярных языков с полиномиальной функцией роста ее можно значительно понизить. Об этом можно прочитать в главах 4 и 5.

Теорема 1.2 *Пусть A, B - некоторые конечные непустые алфавиты, $f \in F(A,B)$, $P \subseteq A^*$. Пусть, кроме того, P представимо в виде конечного объединения множеств P_i таких, что для некоторых фиксированных $m, k \in \mathbb{N}$ при всех i имеем:*
1) существует $V_i \in K_{\leq}(A, E_2, k)$, для которого $1(V_i) = P_i$;
2) для каждого $V \in [V_i]$ существует $W \in K_{\leq}(B, E_2, m)$, для которого $1(W) = \tilde{f}(1(V))$.
Тогда $P \in I(f)$ если и только если $P_{\leq}(k^2 + m^2 + l_f) \in I(f)$.

Доказательство. Ясно, что из условия $P \in I(f)$, очевидно следует $P_{\leq}(k^2+m^2+l_f) \in I(f)$. Допустим теперь, что $P \notin I(f)$. Тогда существуют $\alpha_1, \alpha_2 \in P$ такие, что $\alpha_1 \neq \alpha_2$ и $\tilde{f}(\alpha_1) = \tilde{f}(\alpha_2)$. Так как P распадается в конечное объединение P_i, то для некоторых $r, s \in \mathbb{N}$ имеем $\alpha_1 \in P_r$ и $\alpha_2 \in P_s$. При этом r и s могут совпадать. Далее рассуждение почти полностью повторяет идею доказатель-

ства теоремы 1 с использованием леммы 5 о минимизации. Единственным отличием является то, что оценка на длину общего для α_1 и α_2 префикса γ теперь равна не k, а k^2. Это связано с тем, что верхний (по α_1) и нижний (по α_2) автоматы для этого префикса теперь могут быть разными. Теорема 1.2 доказана.

Замечание. Эта теорема будет использована в главе 4 для решения проблемы ОАД_2 - проблемы ОАД для класса тонких языков.

Теорема 1.3 *Пусть A, B - некоторые конечные непустые алфавиты, $f \in F(A,B)$, $P \subseteq A^*$. Пусть, кроме того, P представимо в виде конечного объединения множеств P_i таких, что для некоторых фиксированных $m, k \in \mathbb{N}$ при всех i имеем:*

1) существует $V_i \in K_{\leq}(A, E_2, k)$, для которого $1(V_i) = P_i$;

2) для каждого $V \in [V_i]$ существует конечное множество автоматов

$$U_i \in K_{\leq}(B, E_2, m),\ 1 \leq i \leq s,\ s \in \mathbb{N}$$

таких, что $\bigcup\limits_{i=1}^{s} 1(U_i) = \tilde{f}(1(V))$.

Тогда $P \in I(f)$ если и только если $P_{\leq}(k^2 + m^2 + l_f) \in I(f)$.

Доказательство. Так же, как теорема 1.2 является обобщением теоремы 1.1, так и теорема 1.3 является обобщением теоремы 1.2. Поэтому доказательства теорем 1.2 и 1.3 очень похожи.

Если $P \in I(f)$, то очевидно, что $P_{\leq}(k^2 + m^2 + l_f) \in I(f)$. Допустим теперь, что $P \notin I(f)$. Тогда существует пара несовпадающих

слов $\alpha_1, \alpha_2 \in P$, для которых $\tilde{f}(\alpha_1) = \tilde{f}(\alpha_2)$. Так как P распадается в конечное объединение P_i, то для некоторых $r, s \in \mathbb{N}$ имеем $\alpha_1 \in P_r$ и $\alpha_2 \in P_s$. Последующее доказательство отличается от доказательства теоремы 1.2 только тем, как получается оценка на длину слова ν_1 в обозначениях из леммы 5. Мы знаем, что

$$\beta_1 \in 1(V_3), V_3 \in [V_r],$$

$$\beta_2 \in 1(W_3), W_3 \in [V_s].$$

Из условия 2) формулировки теоремы 3 известно, что

$$\tilde{f}(\beta_1) \in 1(U_i),$$

$$\tilde{f}(\beta_2) \in 1(U_j)$$

для некоторых $1 \leq i \leq s$, $1 \leq j \leq s$. Применяя лемму 2 о склейке для автоматов $U_i, U_j \in K_{\leq}(B, E_2, m)$ и слов $\tilde{f}(\beta_1), \tilde{f}(\beta_2)$ по их общей части ν_1, получаем, что найдется слово ν_1' в алфавите B, для которого

$$\nu_1' \in 1(U_i),\ \nu_2\nu_1' \in 1(U_j),\ |\nu_1'| \leq m^2.$$

Так как

$$1(U_i) \subseteq \tilde{f}(1(V_3)),\ 1(U_j) \subseteq \tilde{f}(1(W_3)),$$

то существуют $\beta_1' \in 1(V_3)$, $\beta_2' \in 1(W_3)$, для которых

$$\tilde{f}(\beta_1') = \nu_1',\ \tilde{f}(\beta_2') = \nu_2\nu_1'.$$

При этом

$$\gamma a_1 \beta_1' \in 1(V_1) \cdot 1(V_2) \cdot 1(V_3) \subseteq P,$$

$$\gamma a_2 \beta_2' \in 1(W_1) \cdot 1(W_2) \cdot 1(W_3) \subseteq P.$$

Кроме того,

$$\tilde{f}(\gamma a_1 \beta_1') = \tilde{f}(\gamma)\tilde{f}(a_1)\nu_1' = \tilde{f}(\gamma)\tilde{f}(a_2)\nu_2\nu_1' =$$

$$= \tilde{f}(\gamma)\tilde{f}(a_2)\tilde{f}(\beta_2') = \tilde{f}(\gamma a_2 \beta_2').$$

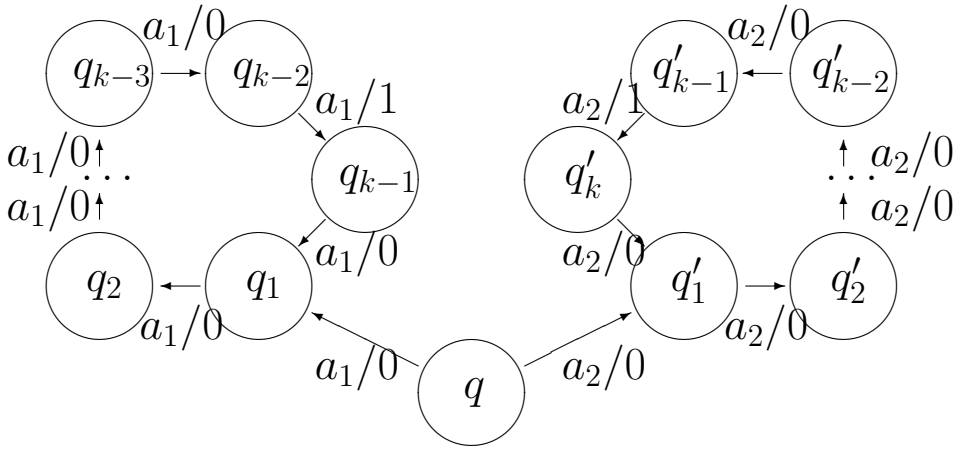
Наконец, $\gamma a_1 \beta_1' \neq \gamma a_2 \beta_2'$, так как $a_1 \neq a_2$. Исходя из всего вышеизложенного изначально можем теперь считать, что $|\nu_1| \leq m^2$. Далее рассуждения полностью повторяют доказательство теоремы 1.2. Теорема 1.3 доказана.

Замечание. Эта теорема будет использована в главе 5 для решения проблемы ОАД_3 - проблемы ОАД для класса регулярных языков с полиномиальной функцией роста.

4. Заключение главы 1

Необходимо отметить, что изложенный метод построения оценки сверху на длину минимально возможной склеивающей пары в некоторых случаях можно оптимизировать. Это происходит из-за того, что при применении леммы 3 о разрезе некоторые состояния в полученных автоматах могут стать недостижимыми из начального состояния и их можно выкинуть. Проиллюстрируем это на примере.

Пусть $A = \{a_1, b_1\}$, $B = \{b\}$, $k \in \mathbb{N}$, $P = (a_1^{k-1})^* \cup (a_2^k)^*$, $f(a_1) = f(a_2) = b$. Тогда минимальной склеиваемой парой будет $\alpha_1 = (a_1^{k-1})^k$, $\alpha_2 = (a_2^k)^{k-1}$. То есть реальная оценка равна $k(k-1)$. Автомат $V_q \in K(A, E_2, n)$, распознающий P посредством $\{1\}$, изображен на рисунке:



Кроме того, есть еще одно особое состояние q_λ, в которое ведут все стрелки, не отмеченные на рисунке. Выходное значение на этих стрелках равно 0. Попав в это состояние, входное слово из него уже не выходит. Видно, что количество состояний в приведенном автомате равно $1+(k-1)+k+1$, то есть $2k+2$. И меньшим количеством состояний в автомате множество P представить нельзя. Но если применить к автомату V_q лемму о разрезе для слова $(a_1^{k-1})^k$ по его первой букве, то в автомате V_3 из леммы 5 будет k состояний. Аналогично, если применить к автомату V_q лемму о разрезе для слова $(a_2^k)^{k-1}$ по его первой букве, то в автомате W_3 из леммы 5 бу-

дет $k+1$ состояний. Поэтому общая оценка метода, приведенного в лемме 5, превратится в $l_f + k(k+1)$, то есть $k(k+1)+1$. Это лишь ненамного хуже реальной оценки. Чтобы еще ближе к ней подойти, надо дополнительно учесть тот факт, что при применении леммы 2 о склейке мы никогда не можем встретить особого состояния q_λ. Приведенный пример показывает, что, вообще говоря, приведенный метод оценки в общем случае существенно улучшить нельзя.

Теперь сформулируем модель передачи информации по каналу связи, на примере которой будем иллюстрировать практическую ценность полученных результатов. Допустим, что нам нужно быстро, недорого и надежно передавать информацию на большие расстояния. Для этого традиционно используются оптоволоконные системы связи. Если не вдаваться в технические детали, то идея таких систем заключается в том, что информация по оптическому кабелю передается по цепочке между передаточными станциями, способными считывать, посылать и восстанавливать искаженный сигнал. Использовать эти технологии может каждый, но предварительно их нужно купить у разработчика. А интерес разработчика состоит в том, чтобы используемая им система кодирования сигнала оставалась в тайне от людей, пытающихся изготовить аналогичный продукт самостоятельно. Ведь иначе их продукция будет не востребована. Традиционно при передаче по оптоволокну

используются БЧХ коды, коды Рида-Соломона, турбо-коды, сверточные коды и так далее. Прочитать об этих технологиях можно, например, в [27-34]. Мы же предлагаем добавить в этот ряд еще и алфавитное кодирование.

Переходим теперь к конкретному описанию модели. Разработчик решает, каким образом он будет кодировать свои сообщения. Здесь могут быть применены любые из упомянутых ранее технологий. При этом нужно еще и стремиться к тому, чтобы закодированное сообщение (в алфавите X) лежало в некотором языке из алфавита X. Вообще говоря, за этот язык всегда можно взять весь X^*. Однако из соображений быстроты и безопасности реализации рекомендуется использовать в качестве такого языка какой-то более узкий и хранящийся в тайне язык $P \subsetneq X^*$. Ниже будет объяснено, почему P следует брать из класса регулярных языков $RP(X)$ с полиномиальной функцией роста. После этого разработчик выбирает некоторый алфавит Y и функцию $f \in F(X, Y)$ алфавитного кодирования, для которой выполнено свойство

$$P \in I(f). \tag{1}$$

И для передачи сообщения $\alpha \in P$ по каналу связи посылается сообщение $\tilde{f}(\alpha)$. После этого на приемном пункте, исходя из известных системе P и f, сообщение можно декодировать и найти α. Дальше это сообщение еще раз декодируется уже в изначальный

текст передаваемого сообщения. Здесь нужно сделать оговорку о том, что для борьбы с искажениями в канале связи может потребоваться еще и закодировать каким-то из традиционных способов сообщение $\tilde{f}(\alpha)$. И уже его посылать по каналу. Этот аспект давно известен и подробно освещается в стандартах $G.975$ и $G.975.1$ от международного консультационного комитета по телефонии и телеграфии $ITU\text{-}T$. Мы же об этом далее говорить не будем и уделим наше основное внимание вопросу выбора P и f.

Прежде всего объясним, зачем вообще нужно выбирать какой-то язык $P \subsetneq X^*$. Если мы ограничимся только случаем $P = X^*$, то, во-первых, сильно сузим класс функций f, удовлетворяющих ограничению (1). Во-вторых, как показано в этой главе, процедура декодирования для языка X^* хоть и, в теории, возможна но может быть крайне затруднена экспоненциальным перебором возможных вариантов. Для нас такая ситуация является неприемлимой. В третьих, если сторонний конкурент(или потребитель нашего продукта) каким-либо образом выяснит, чему равно f, то он тут же сможет на равных правах конкурировать с нами. Если же он знает f, но не знает P, то система все еще находится в относительной безопасности. Далее, на практике по каналу связи сообщения традиционно передаются в алфавите $E_2 = \{0, 1\}$. Поэтому и нам, из соображений совместимости используемой аппаратуры, необходи-

мо взять $X = Y = E_2$. Кроме того, для пары P, f должно быть выполнено свойство (1). Это необходимо хотя бы уже потому, что мы должны уметь правильно декодировать сообщение $\tilde{f}(\alpha)$. При этом из соображений безопасности пару P, f нужно время от времени менять. Значит должен существовать алгоритм проверки ограничения (1). Результаты главы 1 наводят нас на мысль о том, что для этого надо брать $P \in R(X)$. Попытки рассмотреть более широкие классы языков сразу наталкиваются на непреодолимые трудности. Так, для класса контекстно-свободных языков в работе [25] показано, что проблема распознавания проверки взаимной однозначности алфавитного кодирования алгоритмически неразрешима. Тем самым, идея использования класса $R(X)$ выглядит вполне разумно. Из теоремы 1.1 следует, что проблема распознавания проверки однозначности алфавитного декодирования по произвольной паре

$$(\tilde{f}, P) \in F(A, B) \times R(A)$$

алгоритмически разрешима. Обсудим теперь, насколько изложенный метод решения применим на практике. Хотелось бы, чтобы его сложность была ограничена сверху полиномом от входных данных, задающих P и f. Для этого было бы достаточно выполнения двух условий:

1) оценки $n + m^2 + l_f$, $n^2 + m^2 + l_f$ должны быть ограничены сверху полиномом от входных данных, задающих P и f;

2) количество слов в множестве $P_{\leq}(k)$ должно быть ограничено сверху полиномом от k.

Второе условие также можно переформулировать следующим образом:

2') язык P должен иметь полиномиальную функцию роста (определение функции роста будет приведено в главе 2).

Итак, мы показали что для применения изложенного метода кодирования на практике необходимо, чтобы язык P был регулярен и имел полиномиальную функцию роста. Будем теперь двигаться в этом направлении и постепенно доберемся до описания свойств соответствующего класса $RP(X)$. Но сначала в главе 2 будет рассмотрен некоторый простой подкласс таких языков.

Глава 2

Аннотация

В этой главе рассматриваются регулярные языки в произвольном конечном алфавите A, в которых количество различных слов одинаковой длины ограничено сверху независящей от этой длины константой. Такие языки в дальнейшем называются тонкими и и соответствующий им класс обозначается через $\mathfrak{T}(A)$. Это промежуточный шаг на пути к исследованию класса $RP(A)$ регулярных языков в алфавите A с полиномиальной функцией роста. Для языков из класса $\mathfrak{T}(A)$ приводится критериальное описание в терминах прогрессивных множеств. Особо выделяется случай, когда ограничивающая константа равна 1. Такие языки называются 1-тонкими и соответствующий им класс обозначается через $\mathfrak{T}_1(A)$. Для класса $\mathfrak{T}_1(A)$, так же как и для класса $\mathfrak{T}(A)$, приводится критериальное описание, использующее такие понятия, как спектральная независимость и общепрогрессивное множество. Полученные результаты используются в главе 4 для решения проблемы ОАД_2 - проблемы ОАД для класса $\mathfrak{T}(A)$.

1. Основные понятия и результаты

Здесь приведены только те определения, которых еще не было до этого. Если при чтении главы какие-то термины не ясны и их нет в этом разделе, то их определения можно найти в аналогичных разделах предыдущих глав.

Пусть β - непустое слово в алфавите A. Если существует слово α в алфавите A, для которого при некотором $k \in \mathbb{N}$ имеем $\beta = \alpha^k$, то называем слово α *измельчением* слова β. Здесь через α^k обозначена конкатенация k слов α. Если $k > 1$, то такое измельчение называется *собственным.* Если у β есть собственное измельчение,

то говорим, что β *измельчимо*. Иначе говорим, что β *неизмельчимо*. Называем слово α *минимальным измельчением* слова β, если его длина минимальна среди всех длин его измельчений. Ясно, что у любого непустого слова $\beta \in A^*$ существует хотя бы одно измельчение, так как при $k = 1$ таким измельчением будет само слово β. Значит у него существует и минимальное измельчение. Называем два произвольных непустых слова $\beta_1, \beta_2 \in A^*$ *соизмеримыми,* если у них одинаковое минимальное измельчение. В противном случае называем эти слова *несоизмеримыми.* Для произвольного $P \subseteq A^*$ говорим, что P *измеримо,* если любая пара слов из P соизмерима.

Введем понятие спектра множеств. Пусть $P \subseteq A^*$ - произвольное множество слов. *Спектром* этого множества называем множество

$$Sp(P) := \big\{|\alpha|\big|\alpha \in P\big\}.$$

Пусть $P_1, P_2 \subseteq A^*$. Будем говорить, что эти множества *спектрально независимы*, если их спектры не пересекаются. Пусть $P_1, \ldots, P_r \subseteq A^*$, $r \geq 1$. Будем говорить, что эти множества *спектрально независимы в совокупности*, если любые два из них спектрально независимы. В противном случае говорим, что эти множества *спектрально зависимы в совокупности.*

Для произвольного множества $P \subseteq A^*$ через $|P|$ обозначаем его мощность. Если множество P конечно, то пишем $|P| < \infty$.

Введем понятие 1-тонкого множества в алфавите A. Регулярное

множество $P \subseteq A^*$ называем *1-тонким в алфавите* A, если для произвольной пары слов $\alpha, \beta \in P$ из условия $|\alpha| = |\beta|$ следует условие $\alpha = \beta$. Другими словами, в P не должно быть двух несовпадающих слов одинаковой длины.

Пусть $(\alpha, \beta, \gamma, k, m) \in (A^*)^3 \times \mathbb{N} \times (\mathbb{N}_0)$. Говорим, что $(\alpha, \beta, \gamma, m, n)$ - *порождающий след*, если выполнено одно из двух условий:

1. $\beta = \gamma = \lambda$, $k = 1, m = 0$;

2. $\beta \neq \lambda$, у α и β нет одинаковых непустых постфиксов, β неизмельчимо и не является префиксом γ.

Говорим, что множество $P \subseteq A^*$ является *прогрессивным*, если оно представимо с помощью регулярного выражения

$$\alpha \cdot (\beta^k)^* \cdot \beta^m \cdot \gamma$$

для некоторого порождающего следа $(\alpha, \beta, \gamma, k, m)$. В этом случае говорим также, что множество P *имеет порождающий след* $(\alpha, \beta, \gamma, k, m)$. Упорядоченную тройку (α, β, γ) называем *основанием* множества P. Позже будет доказано, что у любого прогрессивного множества может быть только один порождающий след, а значит и только одно основание. Называем множество $P \subseteq A^*$ *общепрогрессивным*, если оно является конечным объединением прогрессивных множеств с одинаковым основанием.

Пусть A - непустой конечный алфавит и $s \in \mathbb{N}$. По аналогии с

понятием 1-тонкого множества вводим понятие s-тонкого множества в алфавите A. Регулярное множество P, $P \subseteq A^*$, называем *s-тонким в алфавите* A, если

1) существует s попарно различных слов $\beta_1, \ldots, \beta_s \in P$ таких, что

$$|\beta_1| = |\beta_2| = \ldots = |\beta_s|;$$

2) не существует $s+1$ попарно различных слов $\alpha_1, \ldots, \alpha_{s+1} \in P$ таких, что

$$|\alpha_1| = |\alpha_2| = \ldots = |\alpha_{s+1}|.$$

Другими словами, в P должно быть s несовпадающих слов одинаковой длины, но не должно быть $s+1$ несовпадающих слов одинаковой длины. При $s = 1$ мы получаем определение класса 1-тонких множеств, приведенное выше.

Для каждого $s \in \mathbb{N}$ обозначаем через $\mathfrak{T}_s(A)$ множество всех s-тонких множеств в алфавите A. Через $\mathfrak{T}(A)$ обозначаем множество

$$\mathfrak{T}(A) := \bigcup_{i=1}^{\infty} \mathfrak{T}_s(A).$$

Называем это множество *классом тонких множеств*, а его элементы - *тонкими множествами.*

Пусть $P \subseteq A^*$. Через $T_n(P)$ обозначаем мощность множества $P_{\leq}(n)$:

$$T_n(P) := |P_{\leq}(n)|.$$

Через T_P обозначаем функцию $T_P : \mathbb{N} \to \mathbb{N}_0$, где

$$T_P(n) := T_n(P)$$

для всех $n \in \mathbb{N}$. Называем T_P *функцией роста* для P. Говорим, что P *имеет константную функцию роста* и пишем $T_P \in Const$, если функция T_P ограничена сверху каким-нибудь полиномом нулевой степени (то есть, константой). Говорим, что P *имеет линейную функцию роста* и пишем $T_P \in Lin$, если функция T_P ограничена сверху каким-нибудь полиномом первой степени и при этом не ограничена сверху никаким полиномом нулевой степени (то есть, константой).

Пусть $a, b \in \mathbb{N}$. Если a делится нацело на b, то пишем $b|a$. Через $Z_{a,b}$ обозначаем множество:

$$Z_{a,b} := \{n \in \mathbb{N} \mid n \geq a,\ b|n\}.$$

Проблемой проверки однозначности алфавитного декодирования в классе тонких языков в алфавите A (или сокращенно - проблемой $ОАД_2$) называем проверку свойства

$$P \in I(f)$$

для произвольных $f \in F(A, B)$ и $P \in \mathfrak{T}(A)$.

Теорема 2.1 *Имеют место следующие утверждения:*

а) любое конечное объединение спектрально независимых в сово-

купности общепрогрессивных множеств является 1-тонким множеством;

б) любое 1-тонкое множество представимо в виде конечного объединения спектрально независимых в совокупности общепрогрессивных множеств.

Теорема 2.2 *Имеют место следующие утверждения:*

а) любое конечное объединение попарно непересекающихся прогрессивных множеств является тонким множеством;

б) любое тонкое множество представимо в виде конечного объединения попарно непересекающихся прогрессивных множеств.

2. Доказательство вспомогательных утверждений

Лемма 1. *Пусть A - некоторый конечный алфавит и P - регулярное множество в алфавите A. Тогда оно представимо регулярным выражением вида*

$$\bigvee_{i=1}^{k} \alpha_{i,1} \cdot (\mathfrak{P}_{i,1})^* \cdot \alpha_{i,2} \cdot \ldots \cdot \alpha_{i,s(i)-1} \cdot (\mathfrak{P}_{i,s(i)-1})^* \cdot \alpha_{i,s(i)},$$

где $k, s(1), \ldots, s(k)$ - некоторые натуральные числа, $\alpha_{1,1}, \ldots, \alpha_{k,s(k)}$ - некоторые слова в алфавите A, $\mathfrak{P}_{1,1}, \ldots, \mathfrak{P}_{k,s(k)-1}$ - некоторые регулярные выражения в алфавите A.

Доказательство. Будем доказывать утверждение индукцией по минимальной длине l вывода P из $\emptyset$ и букв алфавита A с помощью операций $\cup, \cdot, ^*$.

База индукции ($l = 0$).

Если $P = \emptyset$, то P представимо регулярным выражением λ. Здесь $k = 1$, $s(1) = 1$, $\alpha_{1,1} = \lambda$. Если же $P = \{a\}$, где $a \in A$, то P представимо регулярным выражением a. Здесь $k = 1$, $s(1) = 1$, $\alpha_{1,1} = a$.

Переход индукции ($1, \ldots, l \Rightarrow l + 1$).

Разбираем случаи в зависимости от того, какая из операций $\cup, \cdot, ^*$ применена последней.

Случай 1. Минимальная длина вывода P равна l и $P = P_1^*$. Тогда минимальная длина вывода P_1 меньше l и по предположению индукции P_1 представимо регулярным выражением

$$\mathfrak{P}_1 = \bigvee_{i=1}^{k} \alpha_{i,1} \cdot (\mathfrak{P}_{i,1})^* \cdot \alpha_{i,2} \cdot \ldots \cdot \alpha_{i,s(i)-1} \cdot (\mathfrak{P}_{i,s(i)-1})^* \cdot \alpha_{i,s(i)}.$$

Значит P представимо регулярным выражением $(\mathfrak{P}_1)^*$. Осталось взять $k = 1$, $s(1) = 2$, $\alpha_{1,1} = \alpha_{1,2} = \lambda, \mathfrak{P}_{1,1} = \mathfrak{P}_1$.

Случай 2. Минимальная длина вывода P равна l и $P = P_1 \cup P_2$. Тогда минимальная длина вывода P_1 и P_2 меньше l и по предположению индукции P_1 и P_2 представимы регулярными выражениями

$$\mathfrak{P}_1 = \bigvee_{i=1}^{k} \alpha_{i,1} \cdot (\mathfrak{P}_{i,1})^* \cdot \alpha_{i,2} \cdot \ldots \cdot \alpha_{i,s(i)-1} \cdot (\mathfrak{P}_{i,s(i)-1})^* \cdot \alpha_{i,s(i)},$$

$$\mathfrak{P}_2 = \bigvee_{i=1}^{m} \beta_{i,1} \cdot (\mathfrak{C}_{i,1})^* \cdot \beta_{i,2} \cdot \ldots \cdot \beta_{i,t(i)-1} \cdot (\mathfrak{C}_{i,t(i)-1})^* \cdot \beta_{i,t(i)}$$

соответственно. Значит, P представимо регулярным выражением

$$(\mathfrak{P}_1 \vee \mathfrak{P}_2) = \bigvee_{i=1}^{k+m} \gamma_{i,1} \cdot (\mathfrak{I}_{i,1})^* \cdot \gamma_{i,2} \cdot \ldots \cdot \gamma_{i,u(i)-1} \cdot (\mathfrak{I}_{i,u(i)-1})^* \cdot \gamma_{i,u(i)},$$

где

$$u(i) = \begin{cases} s(i), & \text{если } k \geq i \geq 1; \\ t(i-k), & \text{если } k+m \geq i \geq k+1, \end{cases}$$

$$\gamma_{i,j} = \begin{cases} \alpha_{i,j}, & \text{если } k \geq i \geq 1, u(i) \geq j \geq 1; \\ \beta_{i-k,j}, & \text{если } k+m \geq i \geq k+1, u(i) \geq j \geq 1, \end{cases}$$

$$\mathfrak{I}_{i,j} = \begin{cases} \mathfrak{P}_{i,j}, & \text{если } k \geq i \geq 1, u(i)-1 \geq j \geq 1; \\ \mathfrak{C}_{i-k,j}, & \text{если } k+m \geq i \geq k+1, u(i)-1 \geq j \geq 1. \end{cases}$$

Случай 3. Минимальная длина вывода P равна l и $P = P_1 \cdot P_2$. Тогда минимальная длина вывода P_1 и P_2 меньше l и по предположению индукции P_1 и P_2 представимы регулярными выражениями

$$\mathfrak{P}_1 = \bigvee_{i=1}^{k} \alpha_{i,1} \cdot (\mathfrak{P}_{i,1})^* \cdot \alpha_{i,2} \cdot \ldots \cdot \alpha_{i,s(i)-1} \cdot (\mathfrak{P}_{i,s(i)-1})^* \cdot \alpha_{i,s(i)},$$

$$\mathfrak{P}_2 = \bigvee_{i=1}^{m} \beta_{i,1} \cdot (\mathfrak{C}_{i,1})^* \cdot \beta_{i,2} \cdot \ldots \cdot \beta_{i,t(i)-1} \cdot (\mathfrak{C}_{i,t(i)-1})^* \cdot \beta_{i,t(i)}$$

соответственно. Значит P представимо регулярным выражением

$$(\mathfrak{P}_1 \cdot \mathfrak{P}_2) = \bigvee_{i=1}^{km} \gamma_{i,1} \cdot (\mathfrak{I}_{i,1})^* \cdot \gamma_{i,2} \cdot \ldots \cdot \gamma_{i,u(i)-1} \cdot (\mathfrak{I}_{i,u(i)-1})^* \cdot \gamma_{i,u(i)},$$

где

$$u(i) = s\left(\left[\frac{i-1}{m}\right]+1\right) + t\left(i - \left[\frac{i-1}{m}\right]m\right) - 1 \quad \text{при } km \geq i \geq 1,$$

$$\gamma_{i,j} = \begin{cases} \alpha_{\left[\frac{i-1}{m}\right]+1,j} & \text{при } km \geq i \geq 1, s\left(\left[\frac{i-1}{m}\right]+1\right) - 1 \geq j \geq 1; \\ \alpha_{\left[\frac{i-1}{m}\right]+1,j} \cdot \beta_{i-\left[\frac{i-1}{m}\right]m,1} & \text{при } km \geq i \geq 1, j = s\left(\left[\frac{i-1}{m}\right]+1\right); \\ \beta_{i-\left[\frac{i-1}{m}\right]m,j-s\left(\left[\frac{i-1}{m}\right]+1\right)+1} & \text{иначе,} \end{cases}$$

$$\mathfrak{I}_{i,j} = \begin{cases} \mathfrak{P}_{\left[\frac{i-1}{m}\right]+1,j} & \text{при } km \geq i \geq 1, s\left(\left[\frac{i-1}{m}\right]+1\right) - 1 \geq j \geq 1; \\ \mathfrak{C}_{i-\left[\frac{i-1}{m}\right]m,j-s\left(\left[\frac{i-1}{m}\right]+1\right)+1} & \text{иначе.} \end{cases}$$

Таким образом, утверждение индукции, а с ним и лемма 1 доказана.

Лемма 2. *Пусть A - конечный алфавит, $\alpha, \beta \in A^* \setminus \{\Lambda\}$ и $\alpha^k = \beta^m$ для некоторых $k, m \in \mathbb{N}$. Тогда существует $\nu \in A^* \setminus \{\Lambda\}$ такое, что*

$$|\nu| = НОД(|\alpha|, |\beta|),\ \alpha = \nu^{\frac{|\alpha|}{|\nu|}},\ \beta = \nu^{\frac{|\beta|}{|\nu|}}.$$

Доказательство. Пусть $|\alpha| = a$, $|\beta| = b$, НОД$(|\alpha|, |\beta|) = c$. Через S обозначим множество

$$S := \left\{ a \cdot i \,\middle|\, 1 \leq i \leq \frac{b}{c}, i \in \mathbb{N} \right\}.$$

Так как для любого $x \in S$ имеем $x \equiv 0 \pmod{c}$ и $b \equiv 0 \pmod{c}$, то числа из S могут принимать по модулю b только остатки, кратные c. Всего таких остатков $\frac{b}{c}$, как и чисел в S. Заметим, что никакие два разных числа $a \cdot i_1, a \cdot i_2$ из S не могут давать одинаковые остатки по модулю b, так как иначе их разность $a \cdot (i_1 - i_2)$ делилась бы на b, а это, очевидно, невозможно в силу того факта, что НОД$(a, \frac{b}{c}) = 1$ и $\frac{b}{c} > i_1 - i_2$. Значит существует такое $x \in S$, что $x \equiv c \pmod{b}$. Другими словами, существуют $s, t \in \mathbb{N}_0$, для которых

$$1 \leq s \leq \frac{b}{c},\ 0 \leq t < \frac{a}{c},\ as = c + bt.$$

Обозначим через δ слово

$$\delta := \alpha^{2ks} = \beta^{2ms}.$$

При всех $1 \leq i \leq |\delta| - a$ верно, что $\delta_{i-1,i} = \delta_{i+a-1,i+a}$. Аналогично, при всех $b + 1 \leq i \leq |\delta|$ верно, что $\delta_{i-1,i} = \delta_{i-b-1,i-b}$. Поэтому для любого $1 \leq i \leq c$ получаем:

$$\delta_{i-1,i} = \delta_{i+as-1,i+as} = \delta_{i+as-bt-1,i+as-bt} = \delta_{i+c-1,i+c}.$$

Здесь мы неявно воспользовались тем, что

$$i + as \leq c + as \leq a + as \leq 2aks = |\delta|.$$

Значит для всех $|\delta| - 2c \geq i \geq 0$ получаем

$$\delta_{i,i+c} = \delta_{i+c,i+2c}.$$

Так как a делится на c, то

$$\alpha = \delta_{0,a} = \delta_{0,c}^{\frac{a}{c}}.$$

Аналогично

$$\beta = \delta_{0,b} = \delta_{0,c}^{\frac{b}{c}}.$$

Осталось положить

$$\nu := \delta_{0,c}.$$

Утверждение леммы 2 доказано.

Лемма 3. *Пусть* $x_1, \ldots, x_k$ *- натуральные числа и*

$$r := \textit{НОД}(x_1, \ldots, x_k),\ H := \{a_1 \cdot x_1 + \ldots + a_k \cdot x_k \,|a_1, \ldots, a_k \in \mathbb{N}_0\}.$$

Тогда существует $n_0 \in \mathbb{N}$ *такое, что* $Z_{n_0,r} \subseteq H$.

Доказательство. Будем доказывать утверждение индукцией по k.

База индукции ($k = 1, 2$).

При $k = 1$ имеем

$$r = x_1,\ H = \{a \cdot x_1 | a \in \mathbb{N}_0\}.$$

Поэтому для любого $n \in \mathbb{N}$ из условия $r|n$ следует, что $n \in H$. Значит за n_0 можно взять, например, 1.

При $k = 2$ рассматриваем множество

$$S := \left\{ a \cdot x_1 \,\middle|\, 1 \leq a \leq \frac{x_2}{r} \right\}.$$

Так как $r|x_2$ и для любого $x \in S$ верно, что $r|x$, то числа из S могут принимать по модулю x_2 только остатки, кратные r. Всего таких остатков $\frac{x_2}{r}$, как и чисел в S. Заметим, что никакие два разных числа $i_1 \cdot x_1, i_2 \cdot x_1$ из S не могут давать одинаковые остатки по модулю x_2, так как иначе их разность $(i_1 - i_2) \cdot x_1$ делилась бы на x_2, что, очевидно, невозможно в силу $\frac{x_2}{r} > i_1 - i_2$. Значит в S встречаются все остатки по модулю x_2, кратные r. Положим

$$n_0 := \frac{x_2}{r} \cdot x_1.$$

Тогда для всех $n \in Z_{n_0,r}$ существует $x \in S$ такое, что

$$x \equiv n \pmod{x_2}.$$

Отсюда тривиально получаем $Z_{n_0,r} \subseteq H$.

Переход индукции $(1, \ldots, k \Rightarrow k+1)$.

Пусть $x_1, \ldots, x_{k+1}$ - натуральные числа и

$$H := \{a_1 \cdot x_1 + \ldots + a_{k+1} \cdot x_{k+1} | a_1, \ldots, a_{k+1} \in \mathbb{N}_0\},$$

$$H_1 := \{a_1 \cdot x_1 + \ldots + a_k \cdot x_k | a_1, \ldots, a_k \in \mathbb{N}_0\},$$

$$r := \text{НОД}(x_1, \ldots, x_{k+1}), r_1 := \text{НОД}(x_1, \ldots, x_k).$$

По предположению индукции существует $n_0 \in \mathbb{N}$, для которого

$$Z_{n_0,r_1} \subseteq H_1.$$

Пусть p - произвольное натуральное число такое, что $p \geq n_0$ и НОД$(p, x_{k+1}) = 1$. Например, в качестве такого p можно взять произвольное простое число, большее n_0 и x_{k+1}. Обозначаем через H_2 множество

$$H_2 := \{a_1 \cdot r_1 \cdot p + a_2 \cdot x_{k+1} | a_1, a_2 \in \mathbb{N}_0\}.$$

Заметим, что

$$\text{НОД}(r_1 \cdot p, x_{k+1}) = \text{НОД}(r_1, x_{k+1}) =$$

$$= \text{НОД}(\text{НОД}(x_1, \ldots, x_k), x_{k+1}) = r.$$

Отсюда и из предположения индукции, примененного к числам $r_1 \cdot p$ и x_{k+1} следует существование $n_1 \in \mathbb{N}$, для которого выполнено

$$Z_{n_1,r} \subseteq H_2.$$

Так как $r_1 \cdot p \geq n_0$, то $r_1 \cdot p \in Z_{n_0,r_1} \subseteq H_1$. Значит $r_1 \cdot p \in H$ и $x_{k+1} \in H$. Поэтому $H_2 \subseteq H$ и $Z_{n_1,r} \subseteq H$. Утверждение индукции, а вместе с ним и лемма 3, доказаны.

Лемма 4. *Пусть A - некоторый конечный алфавит, $R \subseteq A^*$ и R^* - 1-тонкое множество. Тогда существует конечное множество $R_1 \subseteq A^*$ и существуют $\delta, \gamma \in A^*$ такие, что*

$$R^* = R_1 \cup (\{\delta\} \cdot \{\gamma\}^*).$$

Доказательство. Если $|R| = 1$, то $R = \{\alpha\}$ для некоторого слова $\alpha \in A^*$. Тогда полагаем $R_1 = \emptyset$, $\delta = \lambda$, $\gamma = \alpha$.

Пусть $|R| > 1$. Так как $\{R \setminus \{\Lambda\}\}^* = R^*$, то можно считать, что $\Lambda \notin R$. Будем строить последовательность множеств

$$M_1 \subset M_2 \subset \ldots \subset M_l \subset R$$

такую, что:

1. $M_i = \{\alpha_1, \alpha_2, \ldots, \alpha_{i+1}\}$ для всех $1 \leq i \leq l$;

2. $\alpha_i = \nu_j^{a_i(j)}$ для всех $1 \leq j \leq l$ и $1 \leq i \leq j+1$, где $\nu_j \in A^* \setminus \{\Lambda\}$ и $a_i(j) \in \mathbb{N}$;

3. НОД$(a_1(j), \ldots, a_{j+1}(j)) = 1$ для всех $1 \leq j \leq l$.

Построим M_1. Так как $|R| > 1$, то существуют

$$\alpha, \beta \in R,\ \alpha \neq \beta.$$

Обозначаем

$$a := |\alpha|,\ b := |\beta|.$$

Так как $\alpha^b, \beta^a \in R^*$ и $|\alpha^b| = |\beta^a| = ab$, то $\alpha^b = \beta^a$. По лемме 2 существует $\nu \in A^* \setminus \{\Lambda\}$ такое, что

$$|\nu| = \text{НОД}(a, b),\ \alpha = \nu^{\frac{a}{|\nu|}},\ \beta = \nu^{\frac{b}{|\nu|}}.$$

Обозначаем

$$m := \frac{a}{|\nu|},\ n := \frac{b}{|\nu|}.$$

Заметим, что НОД$(m, n) = 1$. Осталось положить

$$\alpha_1 := \alpha,\ \alpha_2 := \beta,\ \nu_1 := \nu,\ a_1(1) := m,\ a_2(1) := n.$$

Пусть мы уже построили $M_1, \ldots, M_k$. Если для любого $\xi \in R$ число $|\xi|$ делится на число $|\nu_k|$, то $l = k$ и построение закончено. Пусть существует $\xi \in R$ такое, что

$$\text{НОД}(|\xi|, |\nu_k|) < |\nu_k|.$$

Обозначаем

$$c := |\xi|.$$

Так как $\alpha^c, \xi^a \in R^* \setminus \{\Lambda\}$ и $|\alpha^c| = |\xi^a| = ac$, то $\alpha^c = \xi^a$. Кроме того, $\alpha = \alpha_1 = \nu_k^{a_1(k)}$, поэтому $\xi^a = \nu_k^{c \cdot a_1(k)}$. Из леммы 2 следует существование $\sigma \in A^* \setminus \{\Lambda\}$, для которого

$$|\sigma| = \text{НОД}(c, |\nu_k|),\ \xi = \sigma^{\frac{c}{|\sigma|}},\ \nu_k = \sigma^{\frac{|\nu_k|}{|\sigma|}}.$$

Обозначаем

$$u := \frac{c}{|\sigma|},\ v := \frac{|\nu_k|}{|\sigma|}.$$

Получаем, что

$$\alpha_1 = \nu_k^{a_1(k)} = \sigma^{v \cdot a_1(k)},$$

$$\alpha_2 = \nu_k^{a_2(k)} = \sigma^{v \cdot a_2(k)},$$

$$\ldots$$

$$\alpha_{k+1} = \nu_k^{a_{k+1}(k)} = \sigma^{v \cdot a_{k+1}(k)},$$

$$\xi = \sigma^u.$$

Кроме того,

$$\text{НОД}(v \cdot a_1(k), v \cdot a_2(k), \ldots, v \cdot a_{k+1}(k), u) =$$

$= \text{НОД}(\text{НОД}(v \cdot a_1(k), v \cdot a_2(k), \ldots, v \cdot a_{k+1}(k)), u) = \text{НОД}(v, u) = 1.$

Осталось положить

$$\alpha_{k+2} = \xi,$$

$$\nu_{k+1} = \sigma,$$

$$a_1(k+1) = v \cdot a_1(k),$$

$$\ldots$$

$$a_{k+1}(k+1) = v \cdot a_{k+1}(k),$$

$$a_{k+2}(k+1) = u.$$

При этом

$$|\nu_{k+1}| = |\sigma| = \text{НОД}(c, |\nu_k|) = \text{НОД}(|\xi|, |\nu_k|) < |\nu_k|.$$

Значит процесс построения множеств M_i закончится и для некоторого l все длины слов из R будут делиться нацело на $|\nu_l|$.

Возьмем произвольное $\rho \in R$. Обозначаем

$$d := |\rho|.$$

Так как $\rho^a, \alpha^d \in R^* \setminus \{\Lambda\}$ и $|\rho^a| = |\alpha^d| = ad$, то $\rho^a = \alpha^d = \nu_l^{d \cdot a_1(l)}$. Так как

$$a = |\alpha| = |\nu_l^{a_1(l)}| = |\nu_l| \cdot a_1(l),$$

то $\rho^{|\nu_l| \cdot a_1(l)} = \nu_l^{d \cdot a_1(l)}$ и значит $\rho^{|\nu_l|} = \nu_l^d$. Осталось заметить, что d делится нацело на $|\nu_l|$. Отсюда окончательно получаем

$$\rho = \nu_l^{\frac{d}{|\nu_l|}}.$$

Значит $R \subseteq \{\nu_l\}^*$. Тогда и $R^* \subseteq \{\nu_l\}^*$. Но

$$\nu_l^{a_1(l)}, \ldots, \nu_l^{a_{l+1}(l)} \in R \ \text{ и } \ \text{НОД}(a_1(l), \ldots, a_{l+1}(l)) = 1.$$

По лемме 3, примененной к числам $a_1(l), \ldots, a_{l+1}(l)$ получаем существование $n_0 \in \mathbb{N}$, для которого

$$Z_{n_0,1} \subseteq \{a_1 \cdot a_1(l) + \ldots + a_{l+1} \cdot a_{l+1}(l) \mid a_1, \ldots, a_{l+1} \in \mathbb{N}_0\}.$$

Поэтому

$$\{\nu_l^{n_0}\} \cdot \{\nu_l\}^* \subseteq \left\{\nu_l^{a_1(l)}, \ldots, \nu_l^{a_{l+1}(l)}\right\}^*.$$

С другой стороны,

$$\{\nu_l^{a_1(l)}, \ldots, \nu_l^{a_{l+1}(l)}\}^* \subseteq R^*.$$

Получаем, что

$$\{\nu_l^{n_0}\} \cdot \{\nu_l\}^* \subseteq R^* \subseteq \{\nu_l\}^*.$$

Значит

$$R^* = (R^* \setminus (\{\nu_l^{n_0}\} \cdot \{\nu_l\}^*)) \cup (\{\nu_l^{n_0}\} \cdot \{\nu_l\}^*).$$

При этом

$$|R^* \setminus (\{\nu_l^{n_0}\} \cdot \{\nu_l\}^*)| \leq |\{\nu_l\}^* \setminus (\{\nu_l^{n_0}\} \cdot \{\nu_l\}^*)| < \infty.$$

Осталось взять

$$R_1 := R^* \setminus (\{\nu_l^{n_0}\} \cdot \{\nu_l\}^*), \ \delta := \nu_l^{n_0}, \ \gamma := \nu_l.$$

Утверждение леммы 4 доказано.

Лемма 5. *Пусть A - конечный алфавит, $\alpha, \beta, \gamma \in A^* \setminus \{\Lambda\}$ и $R := \{\alpha\}^* \cdot \{\beta\} \cdot \{\gamma\}^*$ - 1-тонкое множество. Тогда существуют $\delta, \nu, \mu \in A^* \setminus \{\Lambda\}$ такие, что*

$$R = \{\delta\}^* \cdot \{\nu\}^* \cdot \{\mu\}.$$

Доказательство. Будем доказывать утверждение индукцией по

$$k := |\beta|.$$

База индукции ($k = 1$).

Здесь $\beta = a$ для некоторого $a \in A$. Тогда

$$R = \{\alpha\}^* \cdot \{a\} \cdot \{\gamma\}^* \text{ и } \alpha^{|\gamma|} \cdot a, a \cdot \gamma^{|\alpha|} \in R.$$

Но $|\alpha^{|\gamma|} \cdot a| = |a \cdot \gamma^{|\alpha|}| = |\alpha| \cdot |\gamma| + 1$. Значит

$$\alpha^{|\gamma|} \cdot a = a \cdot \gamma^{|\alpha|}.$$

Поэтому в алфавите A существует слово γ_1 такое, что $\gamma = \gamma_1 \cdot a$. Отсюда

$$R = \{\alpha\}^* \cdot \{a\} \cdot \{\gamma_1 \cdot a\}^* = \{\alpha\}^* \cdot \{a \cdot \gamma_1\}^* \cdot \{a\}.$$

Осталось положить

$$\delta := \alpha, \ \nu := a \cdot \gamma_1, \ \mu := a.$$

Очевидно, что $\delta, \nu, \mu \in A^* \setminus \{\Lambda\}$.

Переход индукции ($k \Rightarrow k + 1$).

Пусть $|\beta| = k+1$. Тогда для некоторых $a \in A$ и $\beta_1 \in A^* \setminus \{\Lambda\}$ имеем

$$\beta = \beta_1 \cdot a,\ |\beta_1| = k,$$

$$R = \{\alpha\}^* \cdot \{\beta_1 \cdot a\} \cdot \{\gamma\}^*,$$

$$\alpha^{|\gamma|} \cdot \beta_1 \cdot a, \beta_1 \cdot a \cdot \gamma^{|\alpha|} \in R.$$

Так как

$$|\alpha^{|\gamma|} \cdot \beta_1 \cdot a| = |\beta_1 \cdot a \cdot \gamma^{|\alpha|}| = |\alpha| \cdot |\gamma| + k + 1,$$

то

$$\alpha^{|\gamma|} \cdot \beta_1 \cdot a = \beta_1 \cdot a \cdot \gamma^{|\alpha|}.$$

Поэтому существует $\gamma_1 \in A^*$, для которого $\gamma = \gamma_1 \cdot a$. Отсюда

$$R = \{\alpha\}^* \cdot \{\beta_1\} \cdot \{a\} \cdot \{\gamma_1 \cdot a\}^* = \{\alpha\}^* \cdot \{\beta_1\} \cdot \{a \cdot \gamma_1\}^* \cdot \{a\}.$$

Пусть

$$S := \{\alpha\}^* \cdot \{\beta_1\} \cdot \{a \cdot \gamma_1\}^*,$$

$$\alpha_1, \alpha_2 \in S,\ |\alpha_1| = |\alpha_2|.$$

Тогда

$$\alpha_1 \cdot a, \alpha_2 \cdot a \in R,\ |\alpha_1 \cdot a| = |\alpha_2 \cdot a|.$$

Так как R - 1-тонкое множество, то $\alpha_1 \cdot a = \alpha_2 \cdot a$ и $\alpha_1 = \alpha_2$. Поэтому S - 1-тонкое множество. По предположению индукции существуют

$$\delta_1, \nu_1, \mu_1 \in A^* \setminus \{\Lambda\}$$

такие, что

$$S = \{\delta_1\}^* \cdot \{\nu_1\}^* \cdot \{\mu_1\}.$$

Поэтому

$$R = \{\delta_1\}^* \cdot \{\nu_1\}^* \cdot \{\mu_1\} \cdot \{a\}.$$

Осталось положить

$$\delta := \delta_1,\ \nu := \nu_1,\ \mu := \mu_1 \cdot a.$$

Очевидно, что $\delta, \nu, \mu \in A^* \setminus \{\Lambda\}$. Утверждение индукции и лемма 5 доказаны.

Лемма 6. *Пусть A - конечный алфавит, $\alpha, \beta \in A^* \setminus \{\Lambda\}$ и множество $R = \{\alpha\}^* \cdot \{\beta\}^*$ - 1-тонкое множество. Тогда существует конечное множество $R_1 \subseteq A^*$ и существуют $\delta, \gamma \in A^* \setminus \{\Lambda\}$ такие, что*

$$R = R_1 \cup (\{\delta\} \cdot \{\gamma\}^*).$$

Доказательство. Обозначаем

$$a := |\alpha|,\ b := |\beta|.$$

Замечаем, что $\alpha^b, \beta^a \in R$, $|\alpha^b| = |\beta^a| = ab$. Поэтому $\alpha^b = \beta^a$ и по лемме 2 существует $\nu \in A^* \setminus \{\Lambda\}$, для которого

$$|\nu| = \text{НОД}(a, b),\ \alpha = \nu^{\frac{a}{|\nu|}},\ \beta = \nu^{\frac{b}{|\nu|}}.$$

Обозначаем

$$m := \frac{a}{|\nu|},\ n := \frac{b}{|\nu|}.$$

Тогда НОД$(m, n) = 1$ и $R = \{\nu^m\}^* \cdot \{\nu^n\}^*$. Пусть

$$L := \{a_1 \cdot m + a_2 \cdot n | a_1, a_2 \in \mathbb{N}_0\}.$$

По лемме 3 существует $n_0 \in \mathbb{N}$ такое, что $Z_{n_0,1} \subseteq L$. Значит

$$\{\nu^{n_0}\} \cdot \{\nu\}^* \subseteq \{\nu^m\}^* \cdot \{\nu^n\}^*.$$

С другой стороны,

$$\{\nu^m\}^* \cdot \{\nu^n\}^* \subseteq \{\nu\}^*.$$

Тогда

$$\{\nu^{n_0}\} \cdot \{\nu\}^* \subseteq R \subseteq \{\nu\}^*.$$

Поэтому

$$R = (R \setminus (\{\nu^{n_0}\} \cdot \{\nu\}^*)) \cup (\{\nu^{n_0}\} \cdot \{\nu\}^*).$$

При этом

$$|R \setminus (\{\nu^{n_0}\} \cdot \{\nu\}^*)| \leq |\{\nu\}^* \setminus (\{\nu^{n_0}\} \cdot \{\nu\}^*)| < \infty.$$

Осталось положить

$$R_1 := R \setminus (\{\nu^{n_0}\} \cdot \{\nu\}^*),\ \delta := \nu^{n_0},\ \gamma := \nu.$$

Очевидно, что $\delta, \gamma \in A^* \setminus \{\Lambda\}$. Утверждение леммы 6 доказано.

Лемма 7. *У любого прогрессивного множества есть только один порождающий след.*

Доказательство. Будем доказывать утверждение леммы от противного. Пусть у некоторого прогрессивного множества P, $P \subseteq A^*$ есть два разных порождающих следа $(\alpha_1, \beta_1, \gamma_1, k_1, m_1)$, $(\alpha_2, \beta_2, \gamma_2, k_2, m_2)$, то есть

$$P = |\alpha_1 \cdot (\beta_1^{k_1})^* \cdot \beta_1^{m_1} \cdot \gamma_1| = |\alpha_2 \cdot (\beta_2^{k_2})^* \cdot \beta_2^{m_2} \cdot \gamma_2|.$$

Будем по множеству P восстанавливать значения элементов порождающих следов и покажем, что эти следы совпадают.

Если P конечно, то $\beta_1 = \beta_2 = \Lambda$ и $\gamma_1 = \gamma_2 = \Lambda$, $k_1 = k_2 = 1$, $m_1 = m_2 = 0$. Значит

$$P = |\alpha_1 \cdot (\beta_1^{k_1})^* \cdot \beta_1^{m_1} \cdot \gamma_1| = |\alpha_1| = \{\alpha_1\}.$$

Аналогично

$$P = |\alpha_2 \cdot (\beta_2^{k_2})^* \cdot \beta_2^{m_2} \cdot \gamma_2| = |\alpha_2| = \{\alpha_2\}.$$

Значит $\alpha_1 = \alpha_2$, то есть в этом случае порождающие следы совпадают.

Пусть теперь P бесконечно. Тогда $\beta_1 \neq \Lambda$ и $\beta_2 \neq \Lambda$. Обозначаем через α самое маленькое по длине слово из P и через β - самое маленькое по длине слово из $P \setminus \{\alpha\}$. Тогда

$$\alpha = \alpha_1 \cdot \beta_1^{m_1} \cdot \gamma_1, \ \beta = \alpha_1 \cdot \beta_1^{k_1} \cdot \beta_1^{m_1} \cdot \gamma_1.$$

Заметим, что оба слова имеют окончания $\beta_1^{m_1} \cdot \gamma_1$. Если $\alpha_1 = \Lambda$, то слово α является окончанием слова β. Пусть $\alpha_1 \neq \Lambda$. Тогда перед

этим окончанием в α расположена последняя буква слова α_1, а в слове β - последняя буква слова β_1. По определению порождающего следа эти буквы различны. Таким образом, если слово α является окончанием слова β, то $\alpha_1 = \Lambda$. Если же это не так, то в слове α можно отбросить максимальное по длине общее окончание слов α, β и получить слово α_1. В любом случае, по словам α и β однозначно восстанавливается слово α_1. Применяя то же рассуждение и для второго порождающего следа, заключаем

$$\alpha_1 = \alpha_2.$$

Так как мы знаем слово α_1, то мы можем откинуть его от начала слова α и получить слово $\beta_1^{m_1} \cdot \gamma_1$. Если его откинуть теперь от окончания слова β, то останется слово $\alpha_1 \cdot \beta_1^{k_1}$. Осталось убрать из его начала слово α_1 и получится слово $\beta_1^{k_1}$. Проводя аналогичное рассуждение для другого порождающего следа, выводим

$$\beta_1^{k_1} = \beta_2^{k_2}.$$

По лемме 2 существует $\nu \in A^* \setminus \{\Lambda\}$, для которого

$$|\nu| = \text{НОД}(|\beta_1|, |\beta_2|),\ \beta_1 = \nu^{\frac{|\beta_1|}{|\nu|}},\ \beta_2 = \nu^{\frac{|\beta_2|}{|\nu|}}.$$

Но слова β_1, β_2 неизмельчимы. Значит $\beta_1 = \nu = \beta_2$. Но $\beta_1^{k_1} = \beta_2^{k_2}$, поэтому $k_1 = k_2$. Далее, нам известно слово $\beta_1^{m_1} \cdot \gamma_1$. По определению порождающего следа слово β_1 не является началом слова γ_1. Поэтому слово γ_1 получается из известного нам слова $\beta_1^{m_1} \cdot \gamma_1$ от-

брасыванием из его начала слова β_1 до тех пор, пока это возможно. Количество таких отбрасываний однозначно задает нам значение m_1. Аналогично, для второго порождающего следа верно, что

$$\gamma_1 = \gamma_2,\ m_1 = m_2.$$

То есть и в этом случае порождающие следы совпадают. Полученное противоречие завершает доказательство леммы 7.

Лемма 8. *Пусть A - конечный алфавит, $\alpha_1, \alpha_2, \alpha_3 \in A^*$ и множество P равно $\{\alpha_1\} \cdot \{\alpha_2\}^* \cdot \{\alpha_3\}$. Тогда P - прогрессивное множество.*

Доказательство. Будем искать представление P в виде

$$\{\alpha\} \cdot \{\beta^k\}^* \cdot \{\beta^m\} \cdot \{\gamma\},$$

где $(\alpha, \beta, \gamma, k, m)$ - порождающий след.

Если P конечно, то $\alpha_2 = \Lambda$. Значит, $P = \{\alpha_1\gamma_1\}$. Осталось положить

$$\alpha := \alpha_1\gamma_1,\ \beta := \lambda,\ \gamma := \lambda,\ k := 1,\ m := 0.$$

Пусть P бесконечно. Тогда $\alpha_2 \neq \Lambda$.

Если $\alpha_1 = \Lambda$, то $P = \{\alpha_2\}^* \cdot \{\alpha_3\}$. Рассмотрим множество

$$L := \left\{\delta \in A^* \setminus \{\Lambda\} \,\middle|\, \delta^k = \alpha_2, k \geq 2\right\}.$$

Пусть ν - минимальный по длине элемент множества L. Очевидно, что ν неизмельчимо. Заметим далее, что $\nu \neq \Lambda$ и для некоторого

$r \in \mathbb{N}$, $r \geq 2$ верно $\alpha_2 = \nu^r$. Поэтому

$$P = \{\nu^r\}^* \cdot \{\alpha_3\}.$$

Выделяя из начала слова α_3 слова ν до тех пор, пока это возможно, получаем равенство $\alpha_3 = \nu^s \alpha_4$. Здесь $s \in \mathbb{N}_0$ и слово ν не является началом слова α_4. Итак,

$$P = \{\nu^r\}^* \cdot \{\nu^s\} \cdot \{\alpha_4\}.$$

Осталось положить

$$\alpha := \Lambda,\ \beta := \nu,\ \gamma := \alpha_4,\ k := r,\ m := s.$$

Все свойства из определения порождающего следа выполнены.

Пусть теперь $\alpha_1 \neq \Lambda$. Обозначим через δ наибольшее общее окончание слов α_1 и α_2. Пусть $\alpha_1 = \mu\delta$, $\alpha_2 = \rho\delta$ и у слов μ, ρ нет одинаковых непустых окончаний. Тогда

$$P = \{\mu\delta\} \cdot \{\rho\delta\}^* \cdot \{\alpha_3\} = \{\mu\} \cdot \{\delta\rho\}^* \cdot \{\delta\alpha_3\}.$$

С оставшейся частью $\{\delta\rho\}^* \cdot \{\delta\alpha_3\}$ поступаем так же, как и в предыдущем случае с $\{\alpha_2\}^* \cdot \{\alpha_3\}$. А именно, находим минимальное неизмельчимое слово $\nu \neq \Lambda$, для которого $\delta\rho = \nu^t$, $t \geq 2$. Так как слово ν - окончание слова $\delta\rho$, то и у пары μ, ν нет одинаковых окончаний. Теперь выделим из начала слова $\delta\alpha_3$ все слова ν. Имеем $\delta\alpha_3 = \nu^u \alpha_4$, где $u \in \mathbb{N}_0$ и слово ν не является началом слова α_4. Итак,

$$P = \{\mu\} \cdot \{\nu^t\}^* \cdot \{\nu^u\} \cdot \{\alpha_4\}.$$

Осталось положить

$$\alpha := \mu,\ \beta := \nu,\ \gamma := \alpha_4,\ k := t,\ m := u.$$

Все свойства из определения порождающего следа выполнены. Утверждение леммы 8 доказано.

Лемма 9. *Пусть P_1, P_2 -бесконечные спектрально зависимые прогрессивные множества и $P = P_1 \cup P_2$ - 1-тонкое множество. Тогда основания множеств P_1 и P_2 совпадают.*

Доказательство. Для начала заметим, что по лемме 7 у любого прогрессивного множества есть только один порождающий след. Поэтому и основания таких множеств определены однозначно.

Пусть P_1 имеет порождающий след $(\alpha_1, \beta_1, \gamma_1, k_1, m_1)$ и P_2 имеет порождающий след $(\alpha_2, \beta_2, \gamma_2, k_2, m_2)$. Это значит, что

$$P_1 = |\alpha_1 \cdot (\beta_1^{k_1})^* \cdot \beta_1^{m_1} \cdot \gamma_1|,\ P_2 = |\alpha_2 \cdot (\beta_2^{k_2})^* \cdot \beta_2^{m_2} \cdot \gamma_2|.$$

Из бесконечности P_1 и P_2 получаем $\beta_1 \neq \Lambda$, $\beta_2 \neq \Lambda$. Так как P_1 и P_2 спектрально зависимы, то

$$Sp(P_1) \cap Sp(P_2) \neq \emptyset.$$

Поэтому существуют $\nu_1 \in P_1$, $\nu_2 \in P_2$, для которых $|\nu_1| = |\nu_2|$. Так как

$$\nu_1, \nu_2 \in P_1 \cup P_2 = P$$

и P - 1-тонкое множество, то $\nu_1 = \nu_2$. Тогда для некоторых чисел

$a_1, a_2 \in \mathbb{N}_0$ выполнено

$$\alpha_1 \cdot \beta_1^{a_1 \cdot k_1 + m_1} \cdot \gamma_1 = \nu_1 = \nu_2 = \alpha_2 \cdot \beta_2^{a_2 \cdot k_2 + m_2} \cdot \gamma_2.$$

Докажем от противного, что $\alpha_1 = \alpha_2$. Пусть это не так и

$$l_1 := |\alpha_1|,\ l_2 := |\alpha_2|.$$

Тогда $l_1 \neq l_2$. Без ограничения общности можем считать, что $l_1 < l_2$.

Обозначаем через μ, δ слова

$$\mu := \alpha_1 \cdot \beta_1^{(a_1 + 2 \cdot k_2 \cdot |\beta_2| \cdot l_2) \cdot k_1 + m_1} \cdot \gamma_1,\ \delta := \alpha_2 \cdot \beta_2^{(a_2 + 2 \cdot k_1 \cdot |\beta_1| \cdot l_2) \cdot k_2 + m_2} \cdot \gamma_2.$$

Ясно, что $\mu \in P_1$, $\delta \in P_2$ и

$$|\mu| = |\nu_1| + 2 \cdot l_2 \cdot k_1 \cdot k_2 \cdot |\beta_1| \cdot |\beta_2| = |\nu_2| + 2 \cdot l_2 \cdot k_1 \cdot k_2 \cdot |\beta_1| \cdot |\beta_2| = |\delta|.$$

Так как P - 1-тонкое множество, то $\mu = \delta$. Пусть

$$a := \delta_{l_2 - 1, l_2}.$$

Это последняя буква слова α_2. С другой стороны,

$$a = \mu_{l_2 - 1, l_2}.$$

При этом

$$\left|\beta_1^{(a_1 + 2 \cdot k_2 \cdot |\beta_2| \cdot l_2) \cdot k_1 + m_1}\right| = l_1 + |\beta_1| \cdot ((a_1 + 2 \cdot k_2 \cdot |\beta_2| \cdot l_2) \cdot k_1 + m_1) =$$

$$= l_1 + |\beta_1| \cdot m_1 + |\beta_1| \cdot k_1 \cdot a_1 + 2 \cdot |\beta_1| \cdot k_1 \cdot k_2 \cdot |\beta_2| \cdot l_2 \geq l_2 + |\beta_1| \cdot |\beta_2|.$$

Поэтому

$$l_2 + |\beta_1| \cdot |\beta_2| \leq l_1 + \left|\beta_1^{(a_1 + 2 \cdot k_2 \cdot |\beta_2| \cdot l_2) \cdot k_1 + m_1}\right| = \left|\alpha_1 \cdot \beta_1^{(a_1 + 2 \cdot k_2 \cdot |\beta_2| \cdot l_2) \cdot k_1 + m_1}\right|.$$

Но $|\alpha_1| = l_1 < l_2$. Значит слово $\mu_{l_2-1,l_2+|\beta_1|\cdot|\beta_2|}$ является частью слова $\beta_1^{(a_1+2\cdot k_2\cdot|\beta_2|\cdot l_2)\cdot k_1+m_1}$. Тогда

$$\mu_{l_2+|\beta_1|\cdot|\beta_2|-1,l_2+|\beta_1|\cdot|\beta_2|} =$$

$$= \mu_{l_2+|\beta_1|\cdot(|\beta_2|-1)-1,l_2+|\beta_1|\cdot(|\beta_2|-1)} = \ldots = \mu_{l_2-1,l_2} = a.$$

Кроме того,

$$\mu_{l_2+|\beta_1|\cdot|\beta_2|-1,l_2+|\beta_1|\cdot|\beta_2|} = \delta_{l_2+|\beta_1|\cdot|\beta_2|-1,l_2+|\beta_1|\cdot|\beta_2|} =$$

$$= \delta_{l_2+(|\beta_1|-1)\cdot|\beta_2|-1,l_2+(|\beta_1|-1)\cdot|\beta_2|} = \ldots =$$

$$= \delta_{l_2+|\beta_2|-1,l_2+|\beta_2|} = (\alpha_2\beta_2)_{l_2+|\beta_2|-1,l_2+|\beta_2|} = (\beta_2)_{|\beta_2|-1,|\beta_2|}.$$

Значит последние буквы слов α_2 и β_2 совпадают. Это противоречит определению порождающего следа для $(\alpha_2, \beta_2, \gamma_2, k_2, m_2)$. Итак, мы показали, что

$$\alpha_1 = \alpha_2.$$

Так как $\alpha_1 = \alpha_2$ и $\mu = \delta$, то

$$\beta_1^{(a_1+2\cdot k_2\cdot|\beta_2|\cdot l_2)\cdot k_1+m_1}\cdot\gamma_1 = \beta_2^{(a_2+2\cdot k_1\cdot|\beta_1|\cdot l_2)\cdot k_2+m_2}\cdot\gamma_2.$$

Обозначим это слово через ρ. Так как

$$|\rho| \geq \left|\beta_1^{(a_1+2\cdot k_2\cdot|\beta_2|\cdot l_2)\cdot k_1+m_1}\right| \geq |\beta_1|\cdot|\beta_2|,$$

то у ρ есть префикс длины $|\beta_1|\cdot|\beta_2|$. Он равно одновременно $\beta_1^{|\beta_2|}$ и $\beta_2^{|\beta_1|}$. По лемме 2 существует $\delta \in A^* \setminus \{\Lambda\}$, для которого

$$|\delta| = \text{НОД}(|\beta_1|, |\beta_2|),\ \beta_1 = \delta^{\frac{|\beta_1|}{|\delta|}},\ \beta_2 = \delta^{\frac{|\beta_2|}{|\delta|}}.$$

Но по определению порождающего следа слова β_1, β_2 неизмельчимы. Значит

$$\beta_1 = \nu = \beta_2.$$

Теперь пользуемся определением порождающего следа и замечаем, что слово γ_1 получается из слова ρ отбрасыванием из его начала слова β_1 до тех пор, пока это возможно. Аналогично, слово γ_2 получается из слова ρ отбрасыванием из его начала слова β_2 до тех пор, пока это возможно. Так как $\beta_1 = \beta_2$, то

$$\gamma_1 = \gamma_2.$$

Мы доказали, что $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$ и $\gamma_1 = \gamma_2$. Значит основания множеств P_1 и P_2 совпадают. Утверждение леммы 9 доказано.

Лемма 10. *Пусть P_1, P_2 - спектрально независимые 1-тонкие множества. Тогда $P = P_1 \cup P_2$ - тоже 1-тонкое множество.*

Доказательство. Пусть $\alpha, \beta \in P$ и $|\alpha| = |\beta|$. Из спектральной независимости P_1 и P_2 следует, что или оба эти слова лежат в P_1, или же оба они лежат в P_2. Так как P_1 и P_2 - 1-тонкие множества, то получаем отсюда, что $\alpha = \beta$. Поэтому P - 1-тонкое множество. Утверждение леммы 10 доказано.

Лемма 11. *Пусть A - конечный алфавит, $R_1 \subseteq A^*$, $R_2 \subseteq A^*$, $R = R_1 \cup R_2$ и $R \in \mathfrak{T}(A)$. Тогда $R_1 \in \mathfrak{T}(A)$ и $R_2 \in \mathfrak{T}(A)$.*

Доказательство. Так как R - тонкое множество, то для неко-

торого $s \in \mathbb{N}$ верно, что $R \in \mathfrak{T}_s(A)$. Напоминаем, что $\mathfrak{T}_s(A)$ - это обозначение для множества всех s-тонких множеств в алфавите A.

Пусть $\alpha_1, \alpha_2, \ldots, \alpha_{s+1} \in R_1$ и $|\alpha_1| = |\alpha_2| = \ldots = |\alpha_{s+1}|$. Тогда

$$\alpha_1, \alpha_2, \ldots, \alpha_{s+1} \in R.$$

Так как R - s-тонкое множество, то существуют $i, j \in \mathbb{N}$ такие, что

$$\alpha_i = \alpha_j,$$

$$1 \leq i < j \leq s+1.$$

Таким образом, в R_1 не существует более чем s несовпадающих слов одинаковой длины. Обозначим через s_1 максимальное количество попарно различных слов из множества R_1, имеющих одинаковую длину. Мы показали, что такое s_1 существует и не превосходит s. Осталось заметить, что, в силу максимальности выбора числа s_1, в R_1 нет $s_1 + 1$ несовпадающих слов одинаковой длины. Поэтому $R_1 \in \mathfrak{T}_{s_1}(A)$. Аналогично показывается, что существует $s_2 \in \mathbb{N}$, $s_2 \leq s$, для которого $R_2 \in \mathfrak{T}_{s_2}(A)$. Утверждение леммы 11 доказано.

Лемма 12. *Пусть A - конечный алфавит, есть множества $R, R_1, R_2 \subseteq A^*$, $R = R_1 \cdot R_2$ и $R \in \mathfrak{T}_s(A)$. Тогда для некоторых $s_1, s_2 \leq s$ верно, что $R_1 \in \mathfrak{T}_{s_1}(A)$ и $R_2 \in \mathfrak{T}_{s_2}(A)$.*

Доказательство. Пусть

$$\alpha_1, \alpha_2, \ldots, \alpha_{s+1} \in R_1, \ \beta \in R_2 \ \text{ и } \ |\alpha_1| = |\alpha_2| = \ldots = |\alpha_{s+1}|.$$

Тогда

$$\alpha_1 \cdot \beta, \alpha_2 \cdot \beta, \ldots, \alpha_{s+1} \cdot \beta \in R.$$

Так как R - s-тонкое множество, то существуют $i, j \in \mathbb{N}$ такие, что

$$\alpha_i \cdot \beta = \alpha_j \cdot \beta,$$

$$1 \leq i < j \leq s + 1.$$

То есть, $\alpha_i = \alpha_j$. Таким образом, в R_1 не существует более s несовпадающих слов одинаковой длины. Обозначим через s_1 максимальное количество несовпадающих слов из множества R_1, имеющих одинаковую длину. Мы показали, что такое s_1 существует и не превосходит s. Осталось заметить, что в силу максимальности выбора числа s_1 в R_1 нет s_1+1 попарно различных слов одинаковой длины. Поэтому $R_1 \in \mathfrak{T}_{s_1}(A)$. Аналогично показывается, что существует $s_2 \in \mathbb{N}$, $s_2 \leq s$ такое, что $R_2 \in \mathfrak{T}_{s_2}(A)$. Утверждение леммы 12 доказано.

Лемма 13. *Пусть A - конечный алфавит, $R \subseteq A^*$, $s \in \mathbb{N}$ и $R^* \in \mathfrak{T}_s(A)$. Тогда существует конечное множество $R_1 \subseteq A^*$ и существуют слова $\delta, \gamma \in A^*$ такие, что*

$$R^* = R_1 \vee (\{\delta\} \cdot \{\gamma\}^*).$$

Доказательство. Доказательство леммы 13 будет похоже на доказательство леммы 4. Пусть $|R| = 1$, то есть $R = \{\mu\}$ для некото-

рого $\mu \in A^*$. Тогда полагаем $R_1 = \emptyset$, $\delta = \lambda$, $\gamma = \mu$.

Пусть $|R| > 1$. Так как $\{R \setminus \{\Lambda\}\}^* = R^*$, то можно считать, что $\Lambda \notin R$. Будем строить последовательность множеств

$$M_1 \subset M_2 \subset \ldots \subset M_l \subset R$$

такую, что:

1. $M_i = \{\alpha_1, \alpha_2, \ldots, \alpha_{i+1}\}$ для всех $1 \leq i \leq l$;

2. $\alpha_i = \nu_j^{a_i(j)}$ для всех $1 \leq j \leq l$ и $1 \leq i \leq j+1$, где $\nu_j \in A^* \setminus \{\Lambda\}$ и $a_i(j) \in \mathbb{N}$;

3. НОД$(a_1(j), \ldots, a_{j+1}(j)) = 1$ для всех $1 \leq j \leq l$.

Построим M_1. Так как $|R| > 1$, то существуют

$$\alpha, \beta \in R,\ \alpha \neq \beta.$$

Обозначаем

$$a := |\alpha|,\ b := |\beta|.$$

Тогда для всех $0 \leq i \leq s$ верно

$$\alpha^{bi} \cdot \beta^{a(s-i)} \in R^*,$$

$$|\alpha^{bi} \cdot \beta^{a(s-i)}| = abs.$$

Значит существуют $t, u \in \mathbb{N}_0$, $0 \leq t < u \leq s$, для которых

$$\alpha^{bt} \cdot \beta^{a(s-t)} = \alpha^{bu} \cdot \beta^{a(s-u)}.$$

Отсюда получаем $\beta^{a(u-t)} = \alpha^{b(u-t)}$, то есть $\alpha^b = \beta^a$. По лемме 2 существует $\nu \in A^* \setminus \{\Lambda\}$ такое, что

$$|\nu| = \text{НОД}(a, b),\quad \alpha = \nu^{\frac{a}{|\nu|}},\quad \beta = \nu^{\frac{b}{|\nu|}}.$$

Обозначаем

$$m := \frac{a}{|\nu|},\ n := \frac{b}{|\nu|}.$$

При этом НОД$(m, n) = 1$. Осталось положить

$$\alpha_1 := \alpha,\ \alpha_2 := \beta,\ \nu_1 := \nu,\ a_1(1) := m,\ a_2(1) := n.$$

Пусть мы уже построили $M_1, \ldots, M_k$. Если для любого $\xi \in R$ число $|\xi|$ делится на $|\nu_k|$, то $l = k$ и построение закончено. Пусть существует $\xi \in R$ такое, что

$$\text{НОД}(|\xi|, |\nu_k|) < |\nu_k|.$$

Обозначаем

$$c := |\xi|.$$

Тогда для всех $0 \leq i \leq s$ верно

$$\alpha^{ci} \cdot \xi^{a(s-i)} \in R^* \setminus \{\Lambda\},$$

$$|\alpha^{ci} \cdot \xi^{a(s-i)}| = acs.$$

Значит существуют $t, u \in \mathbb{N}_0$, $0 \leq t < u \leq s$, для которых

$$\alpha^{ct} \cdot \xi^{a(s-t)} = \alpha^{cu} \cdot \xi^{a(s-u)}.$$

Отсюда получаем $\xi^{a(u-t)} = \alpha^{c(u-t)}$, то есть $\alpha^c = \xi^a$. Но

$$\alpha = \alpha_1 = \nu_k^{a_1(k)},$$

поэтому $\xi^a = \nu_k^{c \cdot a_1(k)}$. По лемме 2 существует $\sigma \in A^* \setminus \{\Lambda\}$ такое, что

$$|\sigma| = \text{НОД}(c, |\nu_k|),\quad \xi = \sigma^{\frac{c}{|\sigma|}},\quad \nu_k = \sigma^{\frac{|\nu_k|}{|\sigma|}}.$$

Обозначаем

$$u := \frac{c}{|\sigma|},\ v := \frac{|\nu_k|}{|\sigma|}.$$

При этом выполнено

$$\alpha_1 = \nu_k^{a_1(k)} = \sigma^{v \cdot a_1(k)},$$

$$\alpha_2 = \nu_k^{a_2(k)} = \sigma^{v \cdot a_2(k)},$$

$$\dots,$$

$$\alpha_{k+1} = \nu_k^{a_{k+1}(k)} = \sigma^{v \cdot a_{k+1}(k)},$$

$$\xi = \sigma^u.$$

Кроме того,

$$\text{НОД}(v \cdot a_1(k), v \cdot a_2(k), \dots, v \cdot a_{k+1}(k), u) =$$

$$= (\text{НОД}(v \cdot a_1(k), v \cdot a_2(k), \dots, v \cdot a_{k+1}(k)), u) = \text{НОД}(v, u) = 1.$$

Осталось положить

$$a_1(k+1) := v \cdot a_1(k),$$

$$\dots,$$

$$a_{k+1}(k+1) := v \cdot a_{k+1}(k),$$

$$a_{k+2}(k+1) := u,$$

$$\alpha_{k+2} := \xi,$$

$$\nu_{k+1} := \sigma.$$

Замечаем, что

$$|\nu_{k+1}| = |\sigma| = \text{НОД}(c, |\nu_k|) = \text{НОД}(|\xi|, |\nu_k|) < |\nu_k|.$$

Значит для некоторого $l \in \mathbb{N}$ все длины слов из R будут делиться нацело на $|\nu_l|$ и процесс построения множеств M_i закончится. Возьмем произвольное $\rho \in R$. Обозначаем

$$d := |\rho|.$$

Тогда для всех $0 \leq i \leq s$ верно

$$\rho^{ai} \cdot \alpha^{d(s-i)} \in R^* \setminus \{\Lambda\},$$

$$|\rho^{ai} \cdot \alpha^{d(s-i)}| = ads.$$

Значит существуют $t, u \in \mathbb{N}_0$, $0 \leq t < u \leq s$, для которых

$$\rho^{at} \cdot \alpha^{d(s-t)} = \rho^{au} \cdot \alpha^{d(s-u)}.$$

Отсюда получаем, что $\alpha^{d(u-t)} = \rho^{a(u-t)}$, то есть $\rho^a = \alpha^d = \nu_l^{d \cdot a_1(l)}$. Но

$$a = |\alpha| = \left|\nu_l^{a_1(l)}\right| = |\nu_l| \cdot a_1(l).$$

Поэтому $\rho^{|\nu_l| \cdot a_1(l)} = \nu_l^{d \cdot a_1(l)}$, откуда $\rho^{|\nu_l|} = \nu_l^d$. Осталось вспомнить, что d делится нацело на $|\nu_l|$. Отсюда окончательно получаем

$$\rho = \nu_l^{\frac{d}{|\nu_l|}}.$$

Значит $R \subseteq \{\nu_l\}^*$. Отсюда следует, что и $R^* \subseteq \{\nu_l\}^*$. Но

$$\nu_l^{a_1(l)}, \ldots, \nu_l^{a_{l+1}(l)} \in R \ \text{ и } \ \text{НОД}(a_1(l), \ldots, a_{l+1}(l)) = 1.$$

По лемме 3, примененной к числам $a_1(l), \ldots, a_{l+1}(l)$, получаем существование $n_0 \in \mathbb{N}$, для которого

$$Z_{n_0,1} \subseteq \{a_1 \cdot a_1(l) + \ldots + a_{l+1} \cdot a_{l+1}(l) | a_1, \ldots, a_{l+1} \in \mathbb{N}_0\}.$$

Значит

$$\{\nu_l^{n_0}\} \cdot \{\nu_l\}^* \subseteq \{\nu_l^{a_1(l)}, \ldots, \nu_l^{a_{l+1}(l)}\}^*.$$

С другой стороны,

$$\{\nu_l^{a_1(l)}, \ldots, \nu_l^{a_{l+1}(l)}\}^* \subseteq R^*.$$

Получаем, что

$$\{\nu_l^{n_0}\} \cdot \{\nu_l\}^* \subseteq R^* \subseteq \{\nu_l\}^*.$$

То есть

$$R^* = (R^* \setminus (\{\nu_l^{n_0}\} \cdot \{\nu_l\}^*)) \vee (\{\nu_l^{n_0}\} \cdot \{\nu_l\}^*).$$

При этом

$$|R^* \setminus (\{\nu_l^{n_0}\} \cdot \{\nu_l\}^*)| \leq |\{\nu_l\}^* \setminus (\{\nu_l^{n_0}\} \cdot \{\nu_l\}^*)| < \infty.$$

Осталось положить

$$R_1 := R^* \setminus (\{\nu_l^{n_0}\} \cdot \{\nu_l\}^*),\ \delta := \nu_l^{n_0},\ \gamma := \nu_l.$$

Утверждение леммы 13 доказано.

Лемма 14. *Пусть A - конечный алфавит, $\alpha, \beta, \gamma \in A^* \setminus \{\Lambda\}$, $R = \{\alpha\}^* \cdot \{\beta\} \cdot \{\gamma\}^*$ и $R \in \mathfrak{T}_s(A)$ для некоторого $s \in \mathbb{N}$. Тогда существуют $\delta, \nu, \mu \in A^* \setminus \{\Lambda\}$ такие, что*

$$R = \{\delta\}^* \cdot \{\nu\}^* \cdot \{\mu\}.$$

Доказательство. Доказательство леммы 14 будет похоже на доказательство леммы 5.

Будем доказывать утверждение индукцией по

$$k = |\beta|.$$

База индукции ($k = 1$).

Пусть $|\beta| = 1$, то есть $\beta = a$ для некоторого $a \in A$. Тогда

$$R = \{\alpha\}^* \cdot \{a\} \cdot \{\gamma\}^*.$$

Для всех $0 \leq i \leq s$ верно

$$\alpha^{|\gamma|i} \cdot a \cdot \gamma^{|\alpha|(s-i)} \in R,$$

$$\left|\alpha^{|\gamma|i} \cdot a \cdot \gamma^{|\alpha|(s-i)}\right| = |\alpha| \cdot |\gamma| \cdot s + 1.$$

Значит существуют $t, u \in \mathbb{N}_0$ такие, что

$$0 \leq t < u \leq s,$$

$$\alpha^{|\gamma|t} \cdot a \cdot \gamma^{|\alpha|(s-t)} = \alpha^{|\gamma|u} \cdot a \cdot \gamma^{|\alpha|(s-u)}.$$

Отсюда заключаем

$$a \cdot \gamma^{|\alpha|(u-t)} = \alpha^{|\gamma|(u-t)} \cdot a.$$

Поэтому существует $\gamma_1 \in A^*$, для которого $\gamma = \gamma_1 \cdot a$. Значит

$$R = \{\alpha\}^* \cdot \{a\} \cdot \{\gamma_1 \cdot a\}^* = \{\alpha\}^* \cdot \{a \cdot \gamma_1\}^* \cdot \{a\}.$$

Осталось положить

$$\delta := \alpha,\ \nu := a \cdot \gamma_1,\ \mu := a.$$

Очевидно, что $\delta, \nu, \mu \in A^* \setminus \{\Lambda\}$.

Переход индукции ($k \Rightarrow k+1$).

Пусть $|\beta| = k+1$, то есть $\beta = \beta_1 \cdot a$ для некоторых $a \in A$ и $\beta_1 \in A^* \setminus \{\Lambda\}$. Тогда

$$R = \{\alpha\}^* \cdot \{\beta_1 \cdot a\} \cdot \{\gamma\}^*$$

и для всех $0 \le i \le s$ верно

$$\alpha^{|\gamma|i} \cdot \beta_1 \cdot a \cdot \gamma^{|\alpha|(s-i)} \in R,$$

$$|\alpha^{|\gamma|i} \cdot \beta_1 \cdot a \cdot \gamma^{|\alpha|(s-i)}| = |\alpha| \cdot |\gamma| \cdot s + k + 1.$$

Значит существуют $t, u \in \mathbb{N}_0$ такие, что

$$\alpha^{|\gamma|t} \cdot \beta_1 \cdot a \cdot \gamma^{|\alpha|(s-t)} = \alpha^{|\gamma|u} \cdot \beta_1 \cdot a \cdot \gamma^{|\alpha|(s-u)},$$

$$0 \le t < u \le s.$$

Поэтому

$$\beta_1 \cdot a \cdot \gamma^{|\alpha|(u-t)} = \alpha^{|\gamma|(u-t)} \cdot \beta_1 \cdot a.$$

Отсюда следует существование $\gamma_1 \in A^*$, для которого выполнено $\gamma = \gamma_1 \cdot a$. Тогда

$$R = \{\alpha\}^* \cdot \{\beta_1\} \cdot \{a\} \cdot \{\gamma_1 \cdot a\}^* = \{\alpha\}^* \cdot \{\beta_1\} \cdot \{a \cdot \gamma_1\}^* \cdot \{a\}.$$

Обозначаем через S множество

$$S := \{\alpha\}^* \cdot \{\beta_1\} \cdot \{a \cdot \gamma_1\}^*.$$

Тогда $R = S \cdot \{a\}$. Пусть

$$\alpha_1, \alpha_2, \ldots, \alpha_{s+1} \in S, \quad |\alpha_1| = |\alpha_2| = \ldots |\alpha_{s+1}|.$$

Верно, что

$$\alpha_1 \cdot a, \alpha_2 \cdot a, \ldots, \alpha_{s+1} \cdot a \in R,$$

$$|\alpha_1 \cdot a| = |\alpha_2 \cdot a| = \ldots |\alpha_{s+1} \cdot a|.$$

Так как $R \in \mathfrak{T}_s(A)$, то существуют $t, u \in \mathbb{N}$, для которых

$$1 \leq t < u \leq s + 1,$$

$$\alpha_t \cdot a = \alpha_u \cdot a.$$

Поэтому $\alpha_t = \alpha_u$. С другой стороны, так как $R \in \mathfrak{T}_s(A)$, то существует s попарно различных слов $\lambda_1, \lambda_2, \ldots, \lambda_s \in R$ таких, что

$$|\lambda_1| = |\lambda_2| = \ldots = |\lambda_s| = p$$

для некоторого $p \in \mathbb{N}$. Но $R = S \cdot \{a\}$. Значит

$$]_1(\lambda_1) =]_1(\lambda_2) = \ldots =]_1(\lambda_s) = a \ \text{ и } \ p > 1.$$

Обозначаем

$$\delta_1 := [_{p-1}(\lambda_1) = \delta_1, \ \delta_2 := [_{p-1}(\lambda_2), \ \ldots \ , \delta_s := [_{p-1}(\lambda_s).$$

Тогда

$$\lambda_1 = \delta_1 \cdot a, \ \lambda_2 = \delta_2 \cdot a, \ \ldots, \lambda_s = \delta_s \cdot a,$$

$$\delta_1, \delta_2, \ldots, \delta_s \in S.$$

Из попарного различия слов $\lambda_1, \lambda_2, \ldots, \lambda_s$ заключаем, что и слова $\delta_1, \delta_2, \ldots, \delta_s$ попарно различны. Но $|\delta_1| = |\delta_2| = \ldots = |\delta_s| = p - 1$. Получили s попарно различных слов из множества S одинаковой длины.

Объединяя полученные результаты, получаем $S \in \mathfrak{T}_s(A)$. По предположению индукции существуют $\delta_1, \nu_1, \mu_1 \in A^* \setminus \{\Lambda\}$, для которых

$$S = \{\delta_1\}^* \cdot \{\nu_1\}^* \cdot \{\mu_1\}.$$

Поэтому

$$R = \{\delta_1\}^* \cdot \{\nu_1\}^* \cdot \{\mu_1\} \cdot \{a\}.$$

Осталось положить

$$\delta := \delta_1,\ \nu := \nu_1,\ \mu := \mu_1 \cdot a.$$

Очевидно, что $\delta, \nu, \mu \in A^* \setminus \{\Lambda\}$. Утверждение индукции и лемма 14 доказаны.

Лемма 15. *Пусть A - конечный алфавит, есть непустые слова $\alpha, \beta \in A^* \setminus \{\Lambda\}$ и $R = \{\alpha\}^* \cdot \{\beta\}^*$, $R \in \mathfrak{T}_s(A)$. Тогда существует конечное множество $R_1 \subseteq A^*$ и существуют слова $\delta, \gamma \in A^* \setminus \{\Lambda\}$ такие, что*

$$R = R_1 \cup (\{\delta\} \cdot \{\gamma\}^*).$$

Доказательство. Пусть

$$a := |\alpha|,\ b := |\beta|.$$

Тогда для всех $0 \leq i \leq s$ верно

$$\alpha^{bi} \cdot \beta^{a(s-i)} \in R^*,$$

$$\left|\alpha^{bi} \cdot \beta^{a(s-i)}\right| = abs.$$

Значит существуют t, u, $0 \leq t < u \leq s$ такие, что

$$\alpha^{bt} \cdot \beta^{a(s-t)} = \alpha^{bu} \cdot \beta^{a(s-u)}.$$

Отсюда получаем $\beta^{a(u-t)} = \alpha^{b(u-t)}$, то есть $\alpha^b = \beta^a$. По лемме 2 существует $\nu \in A^* \setminus \{\Lambda\}$, для которого

$$|\nu| = \text{НОД}(a, b),\quad \alpha = \nu^{\frac{a}{|\nu|}},\quad \beta = \nu^{\frac{b}{|\nu|}}.$$

Обозначаем

$$m := \frac{a}{|\nu|},\ n := \frac{b}{|\nu|}.$$

Тогда $\text{НОД}(m, n) = 1$ и

$$R = \{\nu^m\}^* \cdot \{\nu^n\}^*.$$

Пусть

$$H := \{a_1 \cdot m + a_2 \cdot n | a_1, a_2 \in \mathbb{N}_0\}.$$

По лемме 3 существует $n_0 \in \mathbb{N}$ такое, что

$$Z_{n_0,1} \subseteq H.$$

Значит

$$\{\nu^{n_0}\} \cdot \{\nu\}^* \subseteq \{\nu^m\}^* \cdot \{\nu^n\}^*.$$

С другой стороны,

$$\{\nu^m\}^* \cdot \{\nu^n\}^* \subseteq \{\nu\}^*.$$

Получаем

$$\{\nu^{n_0}\} \cdot \{\nu\}^* \subseteq R \subseteq \{\nu\}^*.$$

То есть

$$R = (R \setminus (\{\nu^{n_0}\} \cdot \{\nu\}^*)) \vee (\{\nu^{n_0}\} \cdot \{\nu\}^*).$$

При этом

$$|R \setminus (\{\nu^{n_0}\} \cdot \{\nu\}^*)| \leq |\{\nu\}^* \setminus (\{\nu^{n_0}\} \cdot \{\nu\}^*)| < \infty.$$

Осталось положить

$$R_1 := R \setminus (\{\nu^{n_0}\} \cdot \{\nu\}^*),\ \delta := \nu^{n_0},\ \gamma := \nu.$$

Очевидно, что $\delta, \gamma \in A^* \setminus \{\Lambda\}$. Утверждение леммы 15 доказано.

Лемма 16. *Пусть A - конечный алфавит. Любое конечное объединение прогрессивных множеств в алфавите A представимо в виде конечного объединения непересекающихся прогрессивных множеств в алфавите A.*

Доказательство. Пусть для некоторого $k \in \mathbb{N}$ верно, что

$$R = \bigcup_{i=1}^{k} R_i$$

и все R_i - прогрессивные множества.Тогда

$$R = R_1 \bigsqcup (R_2 \setminus R_1) \bigsqcup ((R_3 \setminus R_2) \setminus R_1) \bigsqcup$$

$$\bigsqcup \ldots \bigsqcup ((\ldots (R_k \setminus R_{k-1}) \setminus \ldots \setminus R_1).$$

Осталось показать, что разность двух прогрессивных множеств тоже будет прогрессивным множеством. Пусть

$$R_1 = \{\alpha_1\} \cdot \{\beta_1\}^* \cdot \{\gamma_1\},$$

$$R_2 = \{\alpha_2\} \cdot \{\beta_2\}^* \cdot \{\gamma_2\}$$

для некоторых $\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2 \in A^*$. Либо $R_1 \bigcap R_2 = \emptyset$, либо для некоторого $s \in \mathbb{N}$ имеем $R_1 \bigcap R_2 = \{\alpha_1\} \cdot \{\beta_1^s\}^* \cdot \{\gamma_1\}$. В этом случае

$$R_1 \setminus R_2 = \bigsqcup_{i=1}^{s-1} \{\alpha_1 \cdot \beta_1^i\} \cdot \{\beta_1^s\}^* \cdot \{\gamma_1\}.$$

Осталось заметить, что при всех $1 \leq i \leq s-1$ множество

$$\{\alpha_1 \cdot \beta_1^i\} \cdot \{\beta_1^s\}^* \cdot \{\gamma_1\}$$

является прогрессивным. Утверждение леммы 16 доказано.

3. Доказательство основных утверждений

Теорема 2.1 *Имеют место следующие утверждения:*

а) любое конечное объединение спектрально независимых в совокупности общепрогрессивных множеств является 1-тонким множеством;

б) любое 1-тонкое множество представимо в виде конечного объединения спектрально независимых в совокупности общепрогрессивных множеств.

Доказательство. Докажем сначала первую часть утверждения теоремы. Из леммы 10 следует, что любое конечное объединение спектрально независимых в совокупности 1-тонких множеств является 1-тонким множеством. Осталось доказать, что любое общепрогрессивное множество будет 1-тонким. Пусть P - общепрогрессивное множество. Оно является конечным объединением прогрессивных множеств с одинаковым основанием, то есть

$$P = \bigcup_{i=1}^{k} \left|\alpha \cdot (\beta^{k_i})^* \cdot \beta^{m_i} \cdot \gamma\right|$$

для некоторых $\alpha, \beta, \gamma \in A^*$, $k_i \in \mathbb{N}$, $m_i \in \mathbb{N}_0$. Поэтому

$$P \subseteq |\alpha \cdot (\beta)^* \cdot \gamma|.$$

Значит P - 1-тонкое множество. Первая часть утверждения теоремы 1 доказана.

Докажем вторую часть теоремы. Пусть P - 1-тонкое множество. Оно регулярно. Из леммы 1 получаем, что для некоторых чисел $k, s(1), \ldots, s(k) \in \mathbb{N}$, слов $\alpha_{1,1}, \ldots, \alpha_{k,s(k)} \in A^*$ и регулярных выражений $\mathfrak{P}_{1,1}, \ldots, \mathfrak{P}_{k,s(k)-1}$ верно

$$P = \bigcup_{i=1}^{k} |\alpha_{i,1} \cdot (\mathfrak{P}_{i,1})^* \cdot \alpha_{i,2} \cdot \ldots \cdot \alpha_{i,s(i)-1} \cdot (\mathfrak{P}_{i,s(i)-1})^* \cdot \alpha_{i,s(i)}|.$$

Поэтому

$$P = \bigcup_{i=1}^{k} \{\alpha_{i,1}\} \cdot |\mathfrak{P}_{i,1}|^* \cdot \{\alpha_{i,2}\} \cdot \ldots \cdot \{\alpha_{i,s(i)-1}\} \cdot |\mathfrak{P}_{i,s(i)-1}|^* \cdot \{\alpha_{i,s(i)}\}. \quad (1)$$

Так как P - 1-тонкое множество, то все $|\mathfrak{P}_{i,j}|^*$ - 1-тонкие множества. Из леммы 4 следует, что

$$|\mathfrak{P}_{i,j}|^* = R_{i,j} \cup (\{\delta_{i,j}\} \cdot \{\gamma_{i,j}\}^*) \quad (2)$$

для некоторых конечных множеств $R_{i,j} \subseteq A^*$ и некоторых слов $\delta_{i,j}, \gamma_{i,j} \in A^*$. Но для любых $\alpha, \beta, \gamma \in A^*$ имеем

$$(\{\alpha\} \cup \{\beta\}) \cdot \{\gamma\} = \{\alpha\gamma \cup \alpha\gamma\}, \quad \{\gamma\} \cdot (\{\alpha\} \cup \{\beta\}) = \{\gamma\alpha \cup \gamma\alpha\}. \quad (3)$$

Подставляя (2) в (1) и применяя (3), получаем:

$$P = \bigcup_{i=1}^{l} \{\beta_{i,1}\} \cdot \{\beta_{i,2}\}^* \cdot \{\beta_{i,3}\} \cdot \ldots \cdot \{\beta_{i,t(i)-1}\} \cdot \{\beta_{i,t(i)}\}^* \cdot \{\beta_{i,t(i)+1}\}. \quad (4)$$

Здесь $l, t(1), \ldots, t(l) \in \mathbb{N}$ и $\beta_{1,1}, \ldots, \beta_{l,t(l)+1} \in A^*$. Применяя к (4) лемму 5 получаем, что:

$$P = \bigcup_{i=1}^{m} \{\gamma_{i,1}\} \cdot \{\gamma_{i,2}\}^* \cdot \{\gamma_{i,3}\}^* \cdot \ldots \cdot \{\gamma_{i,u(i)}\}^* \cdot \{\gamma_{i,u(i)+1}\} \quad (5)$$

для некоторых $m, u(1), \ldots, u(m) \in \mathbb{N}$ и $\gamma_{1,1}, \ldots, \gamma_{m,u(m)+1} \in A^*$. Применяя к (5) лемму 6 и используя при этом (3), имеем:

$$P = \bigcup_{i=1}^{n} \{\delta_{i,1}\} \cdot \{\delta_{i,2}\}^* \cdot \{\delta_{i,3}\} \quad (6)$$

для некоторых $n \in \mathbb{N}$, $\delta_{1,1}, \ldots, \delta_{n,3} \in A^*$. По лемме 8 все множества $\{\delta_{i,1}\} \cdot \{\delta_{i,2}\}^* \cdot \{\delta_{i,3}\}$ являются прогрессивными. Поэтому P представимо в виде конечного объединения прогрессивных множеств.

Обозначим это объединение через Σ. Если в Σ есть конечные множества, то они одноэлементны. Пусть P_1 - одно из таких множеств. Если $Sp(P_1)$ пересекается с $Sp(P_2)$ для какого-то другого $P_2 \in \Sigma$, то $P_1 \subset P_2$. Поэтому P_1 можно выкинуть из Σ. Значит можно считать, что все конечные множества из Σ спектрально независимы с остальными множествами. При этом каждое из них является общепрогрессивным. Таким образом, теперь необходимо разложить объединение бесконечных множеств из Σ в объединение спектрально независимых общепрогрессивных множеств.

Делаем это следующим образом. Если P_1 и P_2 имеют непустое пересечение, то будем писать

$$P_1 \leftrightarrow P_2.$$

Если существуют $Q_1, \ldots, Q_h \in \Sigma$, $h \geq 1$ такие, что

$$Q_1 \leftrightarrow Q_2,\ Q_2 \leftrightarrow Q_3,\ \ldots,\ Q_{h-1} \leftrightarrow Q_h,$$

то пишем

$$Q_1 \longleftrightarrow Q_2.$$

Очевидно, что это отношение эквивалентности. Оно разбивает элементы множества Σ на непересекающиеся классы эквивалентности $Z_1, \ldots, Z_n$. Из леммы 9 следует, что любые два элемента из общего класса имеют одинаковые основания. А любая пара элементов из разных классов спектрально независима, так как в противном

случае эти элементы имели бы непустое пересечение и попали бы в один класс эквивалентности. Возьмем произвольный такой класс

$$Z_i := (P_1(i), \ldots, P_{w(i)}(i)).$$

Здесь $P_1(i), \ldots, P_{w(i)}(i)$ - элементы из этого класса. Обозначаем

$$W_i = \bigcup_{j=1}^{w(i)} P_j(i),$$

где $1 \leq i \leq n$. Каждое W_i является конечным объединением прогрессивных множеств с одинаковым основанием, то есть W_i общепрогрессивно. При этом любые два разных множества W_i и W_j спектрально независимы, так как иначе нашлась бы пара

$$P_r(i) \in W_i, \quad P_s(j) \in W_j,$$

которая была бы спектрально зависимой. Но, как уже отмечалось выше, это невозможно. Таким образом, мы разбили 1-тонкое множество P в конечное объединение спектрально независимых в совокупности общепрогрессивных множеств. Вторая часть утверждения теоремы тоже доказана.

Теорема 2.2 *Имеют место следующие утверждения:*

а) любое конечное объединение попарно непересекающихся прогрессивных множеств является тонким множеством;

б) любое тонкое множество представимо в виде конечного объединения попарно непересекающихся прогрессивных множеств.

Доказательство. Докажем сначала первую часть теоремы. Пусть

$$R = \bigsqcup_{i=1}^{k} R_i,$$

где R_i - прогрессивные множества. Обозначаем алфавит, в котором рассматриваются эти множества, через A. Тогда при всех $1 \leq i \leq k$ имеем:

$$R_i = \{\alpha_i\} \cdot \{\beta_i\}^* \cdot \{\gamma_i\}$$

для некоторых $\alpha_i, \beta_i, \gamma_i \in A^*$. Очевидно, все R_i будут 1-тонкими множествами. Осталось доказать, что конечное объединение тонких множеств само является тонким множеством. Достаточно доказать это для объединения двух множеств. Пусть

$$L = L_1 \cup L_2,\ L_1 \in \mathfrak{T}_s,\ L_2 \in \mathfrak{T}_r$$

для некоторых $s, r \in \mathbb{N}$. Без ограничения общности считаем, что $s \geq r$. Тогда среди любых $2s$ слов одинаковой длины из L по принципу Дирихле найдутся s слов, которые все лежат в L_1 или все лежат в L_2. Значит среди этих s слов найдется два совпадающих. Поэтому $L \in \mathfrak{T}$. Первая часть утверждения теоремы 1 доказана.

Докажем вторую часть теоремы. Пусть $P \in \mathfrak{T}_s(A)$ для некоторого $s \in \mathbb{N}$. Множество P регулярно. Применив лемму 1, получаем, что для некоторых чисел $k, s(1), \ldots, s(k) \in \mathbb{N}$, слов $\alpha_{1,1}, \ldots, \alpha_{k,s(k)} \in A^*$ и регулярных выражений $\mathfrak{P}_{1,1}, \ldots, \mathfrak{P}_{k,s(k)-1}$

имеем

$$P = \bigcup_{i=1}^{k} |\alpha_{i,1} \cdot (\mathfrak{P}_{i,1})^* \cdot \alpha_{i,2} \cdot \ldots \cdot \alpha_{i,s(i)-1} \cdot (\mathfrak{P}_{i,s(i)-1})^* \cdot \alpha_{i,s(i)}|.$$

Поэтому

$$P = \bigcup_{i=1}^{k} \{\alpha_{i,1}\} \cdot |\mathfrak{P}_{i,1}|^* \cdot \{\alpha_{i,2}\} \cdot \ldots \cdot \{\alpha_{i,s(i)-1}\} \cdot |\mathfrak{P}_{i,s(i)-1}|^* \cdot \{\alpha_{i,s(i)}\}. \quad (1)$$

Пользуясь леммами 11 и 12 получаем существование такого числа $l_{i,j} \in \mathbb{N}$, $l_{i,j} \leq s$, для которого $|\mathfrak{P}_{i,j}|^* \in \mathfrak{T}_{l_{i,j}}(A)$. Здесь и далее i и j пробегают все значения, удовлетворяющие неравенствам

$$1 \leq i \leq k,\ 1 \leq j \leq s(i) - 1.$$

Из леммы 13 получаем

$$|\mathfrak{P}_{i,j}|^* = R_{i,j} \cup (\{\delta_{i,j}\} \cdot \{\gamma_{i,j}\}^*) \quad (2)$$

для некоторых конечных множеств $R_{i,j} \subseteq A^*$ и некоторых слов $\delta_{i,j}, \gamma_{i,j} \in A^*$. Но для любых $\alpha, \beta, \gamma \in A^*$ верно

$$(\{\alpha\} \cup \{\beta\}) \cdot \{\gamma\} = \{\alpha\gamma \cup \alpha\gamma\},\ \ \{\gamma\} \cdot (\{\alpha\} \cup \{\beta\}) = \{\gamma\alpha \cup \gamma\alpha\}. \quad (3)$$

Подставляя (2) в (1) и применяя (3), получаем

$$P = \bigcup_{i=1}^{l} \{\beta_{i,1}\} \cdot \{\beta_{i,2}\}^* \cdot \{\beta_{i,3}\} \cdot \ldots \cdot \{\beta_{i,t(i)-1}\} \cdot \{\beta_{i,t(i)}\}^* \cdot \{\beta_{i,t(i)+1}\} \quad (4)$$

для некоторых $l, t(1), \ldots, t(l) \in \mathbb{N}$ и $\beta_{1,1}, \ldots, \beta_{l,t(l)+1} \in A^*$. Применяя к (4) леммы 11,12 и 14 выводим, что

$$P = \bigcup_{i=1}^{m} \{\gamma_{i,1}\} \cdot \{\gamma_{i,2}\}^* \cdot \{\gamma_{i,3}\}^* \cdot \ldots \cdot \{\gamma_{i,u(i)}\}^* \cdot \{\gamma_{i,u(i)+1}\} \quad (5)$$

для некоторых $m, u(1), \ldots, u(m) \in \mathbb{N}$ и $\gamma_{1,1}, \ldots, \gamma_{m,u(m)+1} \in A^*$. Применяя к (5) леммы 11,12,15 и используя при этом (3), получаем

$$P = \bigcup_{i=1}^{n} \{\delta_{i,1}\} \cdot \{\delta_{i,2}\}^* \cdot \{\delta_{i,3}\} \tag{6}$$

для некоторого $n \in \mathbb{N}$ и некоторых слов $\delta_{1,1}, \ldots, \delta_{n,3} \in A^*$. Для всех $i \in \mathbb{N}$, $1 \leq i \leq n$ обозначаем через P_i множество

$$P_i := \{\delta_{i,1}\} \cdot \{\delta_{i,2}\}^* \cdot \{\delta_{i,3}\}.$$

В силу определения все P_i - прогрессивные множества. При этом

$$P = \bigcup_{i=1}^{n} P_i.$$

Доказательство второй части теоремы 2.2 завершает применение леммы 16.

4. Заключение главы 2

Обсудим вопрос о том, насколько большую по времени сложность имеет для класса $P \in \mathfrak{T}(X)$ соответствующая процедура алфавитного декодирования и можно ли ее применить на практике. Результаты этой главы говорят нам о представимости любого $P \in \mathfrak{T}(X)$ в виде конечного объединения множеств вида

$$\{\alpha_1\} \cdot \{\alpha_2\}^* \cdot \{\alpha_3\}.$$

Значит и $\tilde{f}(P)$ представимо в виде конечного объединения множеств вида

$$\{\tilde{f}(\alpha_1)\} \cdot \{\tilde{f}(\alpha_2)\}^* \cdot \{\tilde{f}(\alpha_3)\}.$$

Для каждого из таких множеств процедура декодирования тривиальна - нужно откинуть из начала и конца слова $\tilde{f}(\alpha_1)$, $\tilde{f}(\alpha_3)$ соответственно и остаток разбить на $\tilde{f}(\alpha_2)$. Значит для класса тонких языков проблем с декодированием не возникает.

В этой главе мы сделали первый шаг на пути к изучению языков из класса $RP(A)$. В следующей главе мы займемся этим вопросом вплотную. Кроме того, мы покажем, что класс тонких языков можно также определять, как класс регулярных языков с не более чем линейной функцией роста.

Глава 3

Аннотация

В этой главе рассматриваются языки из класса $RP(A)$, то есть регулярные языки с полиномиальной функцией роста в некотором произвольном конечном алфавите A. Доказывается, что их можно представить в виде конечного объединения множеств правильного линейного вида - обобщения понятия прогрессивных множеств из главы 2. Это представление далее используется в главе 5 для решения проблемы ОАД_3 - проблемы ОАД для класса $RP(A)$. Полученные результаты также позволяют продолжить исследование класса тонких языков из главы 2 и показать, что класс конечных тонких множеств в алфавите A совпадает с классом регулярных языков в алфавите A с константной функцией роста. Кроме того, доказывается результат о равенстве класса бесконечных тонких множеств в алфавите A и класса регулярных языков в алфавите A с линейной функцией роста.

1. Основные понятия и результаты.

Здесь приведены только те определения, которых еще не было до этого. Если при чтении главы какие-то термины не ясны и их нет в этом разделе, то их определения можно найти в аналогичных разделах предыдущих глав.

Пусть A - конечный алфавит, B - выходной алфавит. При всех $n \in \mathbb{N}$ обозначаем через $F_n(A, B)$ множество всех схем кодирования из $F(A, B)$, сложность которых не превосходит n.

Пусть $P \subseteq A^*$. Говорим, что P *имеет полиномиальную функцию роста* и пишем $T_P \in Pol$, если T_P ограничена сверху полиномом.

Через $RP(A)$ обозначаем множество всех не содержащих пустое слово регулярных языков в алфавите A, функция роста которых полиномиальна:

$$RP(A) := \{P \subseteq A^* \setminus \{\Lambda\} \mid T_P \in Pol\}.$$

Говорим, что регулярное выражение $\mathfrak{P}$ в алфавите A имеет *линейный вид,* если

$$\mathfrak{P} = \alpha_1 \cdot (\beta_1)^* \cdot \alpha_2 \ldots \alpha_s \cdot (\beta_s)^* \cdot \alpha_{s+1}$$

для некоторых $s \in \mathbb{N}_0$, $\alpha_1, \ldots, \alpha_{s+1} \in A^*$, $\beta_1, \ldots, \beta_s \in A^* \setminus \{\lambda\}$.

Говорим, что регулярное выражение $\mathfrak{P}$ в алфавите A имеет *правильный линейный вид,* если

$$\mathfrak{P} = \alpha_1 \cdot (\beta_1)^* \cdot \alpha_2 \ldots \alpha_s \cdot (\beta_s)^* \cdot \alpha_{s+1}$$

для некоторых $s \in \mathbb{N}_0$ и $\alpha_1, \alpha_{s+1} \in A^*$, $\alpha_2, \ldots, \alpha_s \in A^* \setminus \{\Lambda\}$, $\beta_1, \ldots, \beta_s \in A^* \setminus \{\Lambda\}$ таких, что первые буквы (если они есть, а для α_{s+1} это не всегда так) слов β_i, α_{i+1} при всех $1 \leq i \leq s$ различны. Через $L(\mathfrak{P})$ обозначаем количество всех букв алфавита A (с учетом повторений) в $\mathfrak{P}$ и называем эту величину *сложностью* $\mathfrak{P}$.

Говорим, что множество $P \subseteq A^*$ *линейного вида,* если оно может быть задано регулярным выражением линейного вида. Говорим, что множество $P \subseteq A^*$ *правильного линейного вида,* если оно может быть задано регулярным выражением правильного линейного вида.

Операцией перекидывания называем следующее преобразование:

$$(a \cdot \alpha)^* \cdot a = a \cdot (\alpha \cdot a)^*.$$

При этом говорим, что *буква a перекидывается через итерацию справа налево.*

Называем подмножество натурального ряда $T \subseteq \mathbb{N}$ *периодическим,* если существуют числа $n_0, d \in \mathbb{N}$, для которых при любом натуральном $t \geq n_0$ из $t \in T$ следует $t + d \in T$. Число d называется *длиной периода для T*. Множество

$$\{t \in T \mid t < n_0\}$$

называется *предпериодом для T*. Множество

$$\{t \in T \mid n_0 \leq t < n_0 + d\}$$

называется *периодом для T*.

Проблемой проверки однозначности алфавитного декодирования в классе регулярных языков в алфавите A с полиномиальной функцией роста (или сокращенно - проблемой $ОАД_3$) называем проверку свойства

$$P \in I(f)$$

для произвольных $f \in F(A, B)$ и $P \in RP(A)$.

Теорема 3.1 *Пусть A - конечный алфавит. Любое множество $P \in RP(A)$ может быть представлено в виде конечного объеди-*

нения множеств правильного линейного вида.

Теорема 3.2 *Имеют место следующие утверждения:*
а) класс конечных тонких множеств совпадает с классом регулярных языков с константной функцией роста;
б) класс бесконечных тонких множеств совпадает с классом регулярных языков с линейной функцией роста.

2. Доказательство вспомогательных утверждений

Лемма 1. *Пусть A - конечный алфавит, $P \subseteq A^*$, $Q \subseteq A^*$ и $T_{P\cup Q} \in POL$. Тогда $T_P \in POL$ и $T_Q \in POL$.*

Доказательство. Утверждение тривиально следует из того факта, что при всех $n \in \mathbb{N}$ выполнено

$$T_P(n) \leq T_{P\cup Q}(n) \quad \text{и} \quad T_Q(n) \leq T_{P\cup Q}(n).$$

Утверждение леммы 1 доказано.

Лемма 2. *Пусть A - конечный алфавит, $P, Q \subseteq A^*$, $P \neq \emptyset$, $Q \neq \emptyset$ и $T_{P\cdot Q} \in POL$. Тогда $T_P \in POL$ и $T_Q \in POL$.*

Доказательство. Так как $Q \neq \emptyset$, то существует $\alpha \in Q$. Из леммы 1 следует, что

$$T_{P\cdot\{\alpha\}} \in POL.$$

Но при всех $n \in \mathbb{N}$ имеем

$$T_P(n) = T_{P\cdot\{\alpha\}}(n + |\alpha|).$$

Значит $T_P \in POL$. Аналогично доказывается, что $T_Q \in POL$. Утверждение леммы 2 доказано.

Лемма 3. *Пусть A - конечный алфавит и $P \subseteq A^*$ - не измеримое множество. Тогда $T_{P^*} \notin Pol$.*

Доказательство. Так как P не измеримо, то существуют несоизмеримые слова $\alpha, \beta \in P$. Обозначаем

$$\gamma_1 := \alpha^{|\beta|}, \quad \gamma_2 := \beta^{|\alpha|}.$$

Ясно, что $|\gamma_1| = |\gamma_2|$. Обозначим эту величину через l. Из леммы 2 главы 2 следует, что $\gamma_1 \neq \gamma_2$. Пусть $n \in \mathbb{N}$ и

$$\hat{c} := c(1) \dots c(n)$$

- произвольная последовательность из 1 и 2. Всего таких последовательностей 2^n. Обозначаем

$$\gamma(\hat{c}) := \gamma_{c(1)} \cdots \gamma_{c(n)}.$$

Длина всех $\gamma(\hat{c})$ равна $l \cdot n$. Ясно, что если $\hat{c}_1 \neq \hat{c}_2$, то $\gamma(\hat{c}_1) \neq \gamma(\hat{c}_2)$. Кроме того, для всех $\hat{c}$ имеем $\gamma(\hat{c}) \in P^*$. Поэтому

$$T_{P^*}(l \cdot n) \geq 2^n.$$

Значит T_{P^*} не полиномиальна. Утверждение леммы 3 доказано.

Лемма 4. *Пусть A - конечный алфавит и слова $\alpha, \beta \in A^*$ соизмеримы. Тогда множество $\alpha^* \cdot \beta^*$ представимо в виде конечного объединения $\bigcup\limits_{i=1}^{s} \gamma_i \cdot \delta_i^*$ для некоторых $s \in \mathbb{N}$, $\gamma_i, \delta_i \in A^*$.*

Доказательство. Так как α и β соизмеримы, то для некоторых $\nu \in A^*$, $a, b \in \mathbb{N}$ имеем

$$\alpha = \nu^a, \ \beta = \nu^b.$$

Тогда

$$\alpha^* \cdot \beta^* = (\nu^a)^* \cdot (\nu^b)^* = \{\nu^{a \cdot x + b \cdot y} \mid x, y \in \mathbb{N}_0\}.$$

Пусть

$$r := \text{НОД}(a, b) \ \text{ и } \ H := \{a \cdot x + b \cdot y \mid x, y \in \mathbb{N}_0\}.$$

По лемме 3 из главы 2 существует $n_0 \in \mathbb{N}$, для которого

$$Z_{n_0,r} \subseteq H.$$

Поэтому

$$\{\nu^{a \cdot x + b \cdot y} \mid x, y \in \mathbb{N}_0\} = \{\nu^x \mid x \in H\} =$$

$$= \{\nu^x \mid x \in H \setminus Z_{n_0,r}\} \cup \{\nu^x \mid x \in Z_{n_0,r}\}.$$

Так как все элементы из H делятся нацело на r, то

$$|\{\nu^x \mid x \in H \setminus Z_{n_0,r}\}| < \infty. \tag{1}$$

Кроме того,

$$\{\nu^x \mid x \in Z_{n_0,r}\} = \nu^{n_0} \cdot (\nu^r)^*.$$

Осталось заметить, что каждый элемент конечного множества из (1) представим в виде $\gamma \cdot \delta^*$ для $\delta = \Lambda$. Утверждение леммы 4 доказано.

Лемма 5. *Пусть A - конечный алфавит и слова $\alpha, \beta \in A^*$ не соизмеримы. Тогда существует натуральное число $n \leq |\alpha|$ такое, что слово β^n не является префиксом сверхслова α^∞.*

Доказательство. Обозначаем

$$n := |\alpha|.$$

Замечаем, что

$$|\beta^n| = |\beta| \cdot |\alpha| = \left|\alpha^{|\beta|}\right|.$$

Если бы слово β^n было префиксом сверхслова α^∞, то оно было бы равно $\alpha^{|\beta|}$. Но из леммы 2 главы 2 тогда бы следовало, что слова α и β соизмеримы. Полученное противоречие завершает доказательство леммы 5.

Лемма 6. *Пусть A - конечный алфавит и P - множество линейного вида в алфавите A. Тогда оно представимо в виде конечного объединения множеств P_i правильного линейного вида в алфавите A.*

Доказательство. Из определения множеств линейного вида следует, что множество P представимо регулярным выражением ли-

нейного вида

$$\mathfrak{P} = \alpha_1 \cdot (\beta_1)^* \cdot \alpha_2 \ldots \alpha_s \cdot (\beta_s)^* \cdot \alpha_{s+1} \qquad (1)$$

для некоторых $s \in \mathbb{N}_0$ и $\alpha_1, \ldots, \alpha_{s+1} \in A^*$, $\beta_1, \ldots, \beta_s \in A^* \setminus \{\Lambda\}$. Можно сразу считать, что первые буквы (если они есть) слов β_i, α_{i+1} различны при всех $1 \leq i \leq s$. В самом деле, если это не так, то где-нибудь в выражении можно применить операцию перекидывания:

$$\ldots (a\alpha)^* \cdot (a\beta) \ldots = \ldots a \cdot (\alpha a)^* \cdot \beta \ldots .$$

Ясно, во-первых, что такое преобразование можно применить к выражению последовательно лишь конечное количество раз, так как через итерации можно перекидывать только буквы из слов α_i и делается это всегда справа налево. Во-вторых, что более интересно, окончательный результат $\mathfrak{P}'$ таких преобразований не зависит от того, в каком именно порядке перекидываются буквы через итерации. Будем при этом говорить, что $\mathfrak{P}'$ получен применением *процедуры перекидывания* к выражению (1). Пусть:

$$\mathfrak{P}' := \alpha'_1 \cdot (\beta'_1)^* \cdot \alpha'_2 \ldots \alpha'_s \cdot (\beta'_s)^* \cdot \alpha'_{s+1}.$$

Допустим, что хотя бы одно из слов $\alpha'_2, \ldots, \alpha'_s$ равно Λ. Пусть это, например, α'_2. Тогда в $\mathfrak{P}'$ рядом находятся две итерации - $(\beta'_1)^*$ и $(\beta'_2)^*$.

Наряду с процедурой перекидывания нам потребуется еще одна процедура, которую мы будем называть *процедурой расщепления.* Эта процедура применяется к выражению вида

$$(\alpha)^* \cdot (\beta)^* \tag{2}$$

и устроена следующим образом. Если слова α и β соизмеримы, то применяется лемма 4 и выражение (2) заменяется на дизъюнкцию выражений вида $\gamma \cdot \delta^*$. В каждом из этих выражений ровно одна итерация. При этом под итерацией может быть и пустой символ λ; называем такие итерации *пустыми.* Если же слова α и β не соизмеримы, то (2) заменяется на выражение

$$(\alpha)^* \cdot (\lambda)^* \vee (\alpha)^* \cdot \beta \cdot (\lambda)^* \vee (\alpha)^* \cdot \beta^2 \cdot (\lambda)^* \vee \ldots \vee$$

$$\vee (\alpha)^* \cdot \beta^{|\alpha|-1} \cdot (\lambda)^* \vee (\alpha)^* \cdot \beta^{|\alpha|} \cdot (\beta)^*.$$

После этого к выражению

$$(\alpha)^* \cdot \beta^{|\alpha|} \cdot (\beta)^*$$

применяется процедура перекидывания. Из доказательства леммы 5 следует, что результат этого преобразования будет иметь правильный линейный вид. Каждый из дизъюнктов нового выражения содержит ровно две итерации; при этом либо одна из них пустая, либо дизъюнкт имеет правильный линейный вид. Описание процедуры расщепления закончено. При этом важно помнить, что результат ее применения представляет то же множество, что и (2).

Теперь мы можем применить к $\mathfrak{P}'$ процедуру расщепления для $(\beta_1')^* \cdot (\beta_2')^*$. Новое выражение распадается на дизъюнкты, в каждом из которых или есть пустая итерация, или к самой левой из итераций неприменима операция перекидывания. Удаляем из дизъюнктов все итерации пустого символа и применяем к результатам процедуру перекидывания. Таким образом, выражение $\mathfrak{P}'$ заменено нами на дизъюнкцию выражений, в каждом из которых или стало меньше итераций, или появилось больше итераций с неприменимой к ним операцией перекидывания. Для каждого из новых дизъюнктов проверяем, есть ли в них соседние пары итераций. Если такие есть, то к самой левой из пар опять применяем операцию расщепления. Рано или поздно в получаемых выражениях или кончатся итерации, или к ним нельзя будет нигде применить операцию перекидывания. В любом случае выражения будут иметь правильный линейный вид. Утверждение леммы 6 доказано.

Лемма 7. *Пусть A - конечный алфавит. Тогда $P \subseteq A^*$ измеримо если и только если существует такое слово $\alpha \in A^*$, для которого $P \subseteq \{\alpha\}^*$.*

Доказательство. Пусть $P \subseteq A^*$ измеримо. Если P пустое, то утверждение очевидно. Пусть P непустое. Возьмем произвольное $\alpha \in P$. Обозначаем через ν его минимальное измельчение. Для любого другого слова $\beta \in P \setminus \{\alpha\}$ из измеримости P следует, что

минимальное измельчение β совпадает с ν. Поэтому $P \subseteq \{\nu\}^*$.

Пусть теперь для некоторого $\alpha \in A^*$ верно $P \subseteq \{\alpha\}^*$. Обозначаем через ν минимальное измельчение слова α. Тогда для некоторого $k \in \mathbb{N}$ имеем $\alpha = \nu^k$. Возьмем любое слово $\beta \in P \setminus \{\Lambda\}$. Так как $P \subseteq \{\alpha\}^*$, то для некоторого $m \in \mathbb{N}$ имеем $\beta = \alpha^m$. Тогда $\beta = \nu^{km}$. Пусть ν' - минимальное измельчение слова β и $\beta = (\nu')^n$ для некоторого $n \in \mathbb{N}$. Тогда

$$(\nu')^n = \beta = \nu^{km}.$$

Из леммы 2 главы 2 следует существование такого слова $\nu'' \in A^*$, что

$$|\nu''| = \text{НОД}(|\nu|, |\nu'|) \quad \text{и} \quad \nu = (\nu'')^{\frac{|\nu|}{|\nu''|}}, \quad \nu' = (\nu'')^{\frac{|\nu'|}{|\nu''|}}.$$

Поэтому $|\nu''| \leq |\nu|$ и

$$\alpha = (\nu'')^{\frac{|\nu|}{|\nu''|} \cdot k}.$$

Но ν - минимальное измельчение слова α. Поэтому $\nu = \nu''$ и

$$\beta = (\nu')^n = (\nu'')^{\frac{|\nu'|}{|\nu''|} \cdot n} = \nu^{\frac{|\nu'|}{|\nu''|} \cdot n}.$$

При этом ν' - минимальное измельчение слова β. Но

$$|\nu| = |\nu''| \leq |\nu'|$$

и значит $\nu = \nu'$. Итак, минимальное измельчение любого непустого слова из P совпадает с ν. Поэтому P измеримо. Утверждение леммы 7 доказано.

Лемма 8. *Пусть A - конечный алфавит. Тогда $P \subseteq A^* \setminus \{\Lambda\}$ регулярно и измеримо если и только если существуют такое слово $\alpha \in A^*$ и такое периодическое множество $T \subseteq \mathbb{N}$, для которых выполнено*

$$P = \{\alpha^t \mid t \in T\}.$$

Доказательство. Пусть $P = \{\alpha^t \mid t \in T\}$ для некоторого $\alpha \in A^*$ и периодического множества $T \subseteq \mathbb{N}$. Из леммы 7 следует, что P измеримо. Покажем, что P регулярно.

Так как T периодическое, то существуют числа $n_0, d \in \mathbb{N}$, для которых при всех $t \geq n_0$ из $t \in T$ следует $t + d \in T$. Пусть T' - предпериод множества T и T'' - период множества T :

$$T' := \{t \in T \mid t < n_0\},$$

$$T'' := \{t \in T \mid n_0 \leq t < n_0 + d\}.$$

Обозначаем через M множество

$$M := \{\alpha^k \mid k \in T'\}.$$

Так как M конечно, то $M \in R(A)$. Для каждого $t \in T''$ обозначаем через M_t множество

$$M_t := \{\alpha^{t+id} \mid i \in \mathbb{N}_0\}.$$

Так как

$$\{\alpha^{t+id} \mid i \in \mathbb{N}_0\} = \{\alpha^t\} \cdot \{\alpha^d\}^*,$$

то для всех $t \in T''$ имеем $M_t \in R(A)$. Осталось заметить, что

$$P = M \cup \left(\bigcup_{t \in T''} M_t \right) \quad \text{и} \quad |T''| < \infty.$$

Значит $P \in R(A)$.

Пусть теперь P - регулярное измеримое множество. Из леммы 7 следует существование $\alpha \in A^*$, для которого $P \subseteq \{\alpha\}^*$. Так как $P \in R(A)$, то по теореме Клини (см. [26]) существует абстрактный инициальный конечный автомат

$$V = (A, Q, \{0, 1\}, \varphi, \psi, q),$$

распознающий по множеству $\{1\}$ множество P. Для всех $t \in \mathbb{N}$ обозначаем через $q(t)$ состояние

$$q(t) := \varphi(q, \alpha^t).$$

Так как $|Q| < \infty$, то существуют $m, n \in \mathbb{N}$ такие, что

$$q(m) = q(m + n).$$

Тогда для всех $m_0 \in \mathbb{N}$, $m_0 \geq m$ имеем

$$q(m_0 + n) = \varphi(q, \alpha^{m_0+n}) = \varphi(q, \alpha^{(m+n)+(m_0-m)}) =$$

$$= \varphi(\varphi(q, \alpha^{m+n}), \alpha^{m_0-m}) = \varphi(q(m + n), \alpha^{m_0-m}) =$$

$$= \varphi(q(m), \alpha^{m_0-m}) = \varphi(\varphi(q, \alpha^m), \alpha^{m_0-m}) =$$

$$= \varphi(q, \alpha^{m+(m_0-m)}) = \varphi(q, \alpha^{m_0}) = q(m_0).$$

Вводим обозначения:

$$T := \{t \in \mathbb{N} \mid \alpha^t \in P\},$$

$$T' := \{t \in \mathbb{N} \mid t \leq m, \alpha^t \in P\},$$

$$T'' := \{t \in \mathbb{N} \mid m < t \leq m + n, \alpha^t \in P\}.$$

Докажем, что T - периодическое множество с предпериодом T', периодом T'' и длиной периода n. В самом деле, пусть $k > m$, $k \in T$. Тогда

$$\alpha^k \in P \ \text{ и } \ \psi(q, \alpha^k) = 1.$$

Кроме того,

$$q(k + n - 1) = q(k - 1),$$

ведь $k - 1 \geq m$ и выше мы показали, что при всех $m_0 \in \mathbb{N}$, $m_0 \geq m$ верно

$$q(m_0 + n) = q(m_0).$$

Поэтому

$$\psi(q, \alpha^{k+n}) = \psi(\varphi(q, \alpha^{k+n-1}), \alpha) = \psi(q(k + n - 1), \alpha) =$$

$$= \psi(q(k - 1), \alpha) = \psi(\varphi(q, \alpha^{k-1}), \alpha) = \psi(q, \alpha^k) = 1.$$

Значит $\alpha^{k+n} \in P$ и $k + n \in T$. Поэтому T периодическое. Утверждение леммы 8 доказано.

Лемма 9. *Пусть A - конечный алфавит и для некоторых слов $\alpha, \gamma \in A^* \setminus \{\Lambda\}$, $\beta \in A^*$ и множества $P = (\alpha)^*\beta(\gamma)^*$ верно, что $T_P \in Lin$. Тогда $P \in \mathfrak{T}(A)$.*

Доказательство. Разберем сначала случай, когда $\beta = \Lambda$. Допустим, что слова α и γ несоизмеримы. Возьмем произвольное $n \in \mathbb{N}$ и оценим снизу мощность множества $P_{\leq}(n \cdot |\alpha| \cdot |\gamma|)$. В этом множестве есть слова вида $\alpha^{a \cdot |\gamma|}\gamma^{b \cdot |\alpha|}$, где $a, b \in \mathbb{N}_0$ и $a + b \leq n$. Из несоизмеримости слов α, γ и леммы 2 главы 2 следует, что все эти слова попарно различны. Всего таких слов $n+(n-1)+\ldots+1 = \frac{n(n-1)}{2}$. Это противоречит условию $T_P \in Lin$. Значит слова α и γ соизмеримы. Разбор этого случая завершает применение леммы 4.

Пусть теперь $\beta \neq \Lambda$. Опять берем произвольное $n \in \mathbb{N}$ и для него рассматриваем слова вида

$$\alpha^{a \cdot |\gamma|}\beta\gamma^{b \cdot |\alpha|}, \tag{1}$$

где $a, b \in \mathbb{N}_0$ и $a + b \leq n$. Ясно, что все они лежат в множестве $P_{\leq}(n \cdot |\alpha| \cdot |\gamma| + |\beta|)$. Поэтому из условия $T_P \in Lin$ следует существование такого $n_0 \in \mathbb{N}$, для которого некоторые из соответствующих представлений вида (1) задают одинаковые слова. Отсюда получаем, что для некоторых $k, m \in \mathbb{N}$ выполнено условие

$$\alpha^k\beta = \beta\gamma^m. \tag{2}$$

Представим β в виде $\alpha^s\delta$ для максимально возможного $s \in \mathbb{N}_0$. Тогда из (2) получаем $\alpha^{k+s}\delta = \alpha^s\delta\gamma^m$, то есть

$$\alpha^k\delta = \delta\gamma^m. \tag{3}$$

Так как α не является префиксом δ, то из (3) выводим, что δ являет-

ся префиксом α, то есть $\alpha = \delta\rho$ для некоторого $\rho \in A^*$. Условие (3) можно теперь переписать в виде $(\delta\rho)^k\delta = \delta\gamma^m$, то есть $(\rho\delta)^k = \gamma^m$. Тогда из леммы 2 главы 2 следует соизмеримость слов $\rho\delta$ и γ. Теперь замечаем, что

$$\alpha^*\beta\gamma^* = \alpha^*\alpha^s\delta\gamma^* = \alpha^s\alpha^*\delta\gamma^* = \alpha^s(\delta\rho)^*\delta\gamma^* = \alpha^s\delta(\rho\delta)^*\gamma^*.$$

Осталось применить утверждение леммы 4. Разбор случаев завершен. Утверждение леммы 9 доказано.

3. Доказательство основных утверждений

Теорема 3.1 *Пусть A - конечный алфавит. Любое множество $P \in RP(A)$ может быть представлено в виде конечного объединения множеств правильного линейного вида.*

Доказательство. Из леммы 1 главы 2 следует, что P представимо регулярным выражением вида

$$\bigvee_{i=1}^{k} \alpha_{i,1} \cdot (\mathfrak{P}_{i,1})^* \cdot \alpha_{i,2} \cdot \ldots \cdot \alpha_{i,s(i)-1} \cdot (\mathfrak{P}_{i,s(i)-1})^* \cdot \alpha_{i,s(i)},$$

где $k, s(1), \ldots, s(k)$ - натуральные числа, $\alpha_{1,1}, \ldots, \alpha_{k,s(k)}$ - слова (возможно пустые) в алфавите A, $\mathfrak{P}_{1,1}, \ldots, \mathfrak{P}_{k,s(k)-1}$ - регулярные выражения в алфавите A. Тогда

$$P = \bigcup_{i=1}^{k} \alpha_{i,1} \cdot (|\mathfrak{P}_{i,1}|)^* \cdot \alpha_{i,2} \cdot \ldots \cdot \alpha_{i,s(i)-1} \cdot (|\mathfrak{P}_{i,s(i)-1}|)^* \cdot \alpha_{i,s(i)}. \quad (1)$$

Обозначаем

$$\mathfrak{F} := \Big\{ |\mathfrak{P}_{i,j}| \Big| 1 \le i \le k,\ 1 \le j \le s(i) - 1 \Big\}.$$

Пусть F - произвольный элемент множества $\mathfrak{F}$. Известно, что $T_P \in POL$. Поэтому из лемм 1,2 получаем $T_F \in POL$. Из леммы 3 теперь заключаем, что F измеримо. При этом F еще и регулярно. Значит по лемме 8 существуют такое слово $\alpha \in A^*$ и такое периодическое множество $T \subseteq \mathbb{N}$, для которых выполнено

$$F = \{\alpha^t \mid t \in T\}.$$

Пусть множество T имеет предпериод T', период T'' и длину периода d. Тогда

$$F = \{\alpha^t \mid t \in T'\} \cup \left(\bigcup_{t \in T''} \{\alpha^{t+id} \mid i \in \mathbb{N}_0\} \right) =$$

$$= \left\{ \alpha^t \cdot (\alpha^0)^* \,\middle|\, t \in T' \right\} \cup \left(\bigcup_{t \in T''} \{\alpha^t \cdot (\alpha^d)^*\} \right).$$

Но $|T'| < \infty$ и $|T''| < \infty$. Поэтому F можно представить в виде

$$F := \bigcup_{i=1}^{s_F} \alpha_{i,F}^{a_{i,F}} \cdot \left(\alpha_{i,F}^{b_{i,F}} \right)^* \tag{2}$$

для некоторых $s_F \in \mathbb{N}$, $\alpha_{i,F} \in A^*$ и $a_{i,F}, b_{i,F} \in \mathbb{N}_0$. Подставляем в (1) вместо всех $F \in \mathfrak{F}$ значения (2). Раскрываем дистрибутивности и получаем представление P в виде конечного объединения множеств линейного вида. Осталось применить лемму 6. Утверждение теоремы 3.1 доказано.

Теорема 3.2 *Имеют место следующие утверждения:*

а) класс конечных тонких множеств совпадает с классом регулярных языков с константной функцией роста;

б) класс бесконечных тонких множеств совпадает с классом регулярных языков с линейной функцией роста.

Доказательство. Докажем сначала первую часть утверждения теоремы. Обозначаем через A алфавит рассматриваемых классов, через S_1 - класс всех конечных множеств в алфавите A, через S_2 - класс всех конечных тонких множеств в алфавите A и через S_3 - класс всех регулярных языков с константной функцией роста в алфавите A. Пусть $P \in S_1$. Тогда, во-первых, $P \in R(A)$. Во-вторых, в P существует заведомо не больше чем $|P|$ попарно различных слов одинаковой длины. Поэтому $P \in \mathfrak{T}(A)$. Значит $S_1 = S_2$. Кроме того, если $P' \in S_1$, то $P \in R(A)$ и для любого $n \in \mathbb{N}$ имеем

$$T_{P'}(n) = |P'_{\leq}(n)| \leq |P'|.$$

Поэтому $T_{P'} \in Const$ и $P' \in S_3$. Наоборот, если $P' \in S_3$, то $T_{P'} \in Const$ и значит $|P'| < \infty$. Поэтому $S_1 = S_3$. Итак, $S_2 = S_3$. Первая часть теоремы доказана.

Докажем вторую часть теоремы. Как и в первой части, обозначаем через A алфавит рассматриваемых классов. Обозначаем через U_1 класс всех бесконечных тонких множеств в алфавите A. Через U_2 обозначаем класс всех регулярных языков с линейной функцией

роста. Пусть $P \in U_1$. Тогда для некоторого $s \in \mathbb{N}$ верно, что

$$P \in \mathfrak{T}_s(A).$$

Поэтому при всех $n \in \mathbb{N}$ в множестве P не больше s слов длины n. Значит

$$T_P(n) = |P_{\leq}(n)| \leq ns.$$

Так как $|P| = \infty$ то, как уже известно из в первой части теоремы, $T_P \notin Const$. Поэтому $T_P \in Lin$. Так как P регулярно, то $P \in U_2$. Поэтому $U_1 \subseteq U_2$. Докажем теперь, что $U_2 \subseteq U_1$. Пусть $P \in U_2$. Тогда $T_P \in Lin$, то есть и $T_P \in Pol$. Так как P регулярно, то отсюда выводим $P \in RP(A)$. Из теоремы 1 следует, что P представимо в виде конечного объединения множеств линейного вида. Применяя к этим множествам, если это необходимо, лемму 9 и теорему 2.2 из главы 2, разбиваем каждое из них в конечное объединение прогрессивных множеств вида $\alpha\beta^*\gamma$. Другими словами, все имеющиеся множества линейного вида с линейной функцией роста с помощью леммы 9 трансформируются в конечное объединение множеств линейного вида, но уже не более чем с одной итерацией внутри (если итерацию пустого слова за итерацию не считать). Значит $P \in \mathfrak{T}(A)$. Осталось заметить, что P бесконечно, ведь все конечные множества, как было только что доказано, имеют константную функцию роста. Поэтому $P \in U_1$, то есть $U_2 \subseteq U_1$. Вторая часть теоремы тоже доказана.

4. Заключение главы 3

Как и в главе 2 для класса $\mathfrak{T}(X)$ обсудим аналогичный вопрос о том, насколько большую по времени сложность имеет для класса $P \in RP(X)$ соответствующая процедура алфавитного декодирования и можно ли ее применить на практике. Ясно, что если $P \in RP(X)$, то и $\tilde{f}(P) \in RP(Y)$. В этой главе доказана представимость множества $\tilde{f}(P) \in RP(Y)$ в виде конечного объединения множеств правильного линейного вида, то есть множеств, задаваемых выражениями вида

$$\mathfrak{P} = \alpha_1 \cdot (\beta_1)^* \cdot \alpha_2 \ldots \alpha_s \cdot (\beta_s)^* \cdot \alpha_{s+1}.$$

Здесь все слова $\alpha_1, \ldots, \alpha_s, \beta_1, \ldots, \beta_s$ непусты и первые буквы (если они есть, а для α_{s+1} это не всегда так) слов β_i, α_{i+1} всегда различны при $1 \leq i \leq s$. Для таких множеств процедура декодирования тривиальна - надо откинуть из начала слово α_1, потом, пока это возможно, убирать из начала слово β_1 и так далее до тех пор, пока не останется слово α_{s+1}. На основании построенного разбиения уже тривиально восстанавливается соответствующий прообраз из языка P. Значит и в классе регулярных языков с полиномиальной функцией роста процедуру декодирования несложно реализовать технически для практического использования.

Полученное в этой главе разложение множеств $P \in RP(A)$ в

конечное объединение множеств правильного линейного вида будет использовано в главе 5 для получения полиномиальных оценок на сложность алгоритма, доставляющего решение проблемы ОАД_3. Но, прежде чем переходить к этой проблеме, мы сначала решим в главе 4 проблему ОАД_2 для класса $\mathfrak{T}(A)$ тонких языков.

Глава 4

Аннотация

В этой главе рассматривается решение проблемы ОАД$_2$ для класса $\mathfrak{T}(A)$ тонких языков в некотором произвольном алфавите A. При этом используется результат из главы 1 о решении проблемы ОАД$_1$ для класса $R(A)$ и результат из главы 2 о каноническом представлении языков из класса $\mathfrak{T}(A)$. Полученные результаты обобщаются затем в главе 5 для решения проблемы ОАД$_3$ для класса $RP(A)$ регулярных языков в алфавите A.

1. Основные понятия и результаты

Здесь приведены только те определения, которых еще не было до этого. Если при чтении главы какие-то термины не ясны и их нет в этом разделе, то их определения можно найти в аналогичных разделах предыдущих глав.

Пусть A - конечное множество. Если выражение $\mathfrak{P}$ в алфавите A является конечной дизъюнкцией выражений вида

$$\alpha \cdot (\beta)^* \cdot \gamma, \quad \text{где} \quad \alpha, \beta, \gamma \in A^*, \tag{1}$$

то его *сложностью* называем максимальную из сложностей этих выражений. Обозначаем это число через $L(\mathfrak{P})$. Класс таких дизъюнктных выражений обозначаем через $\mathfrak{X}(A)$.

Класс множеств $P \subseteq A^*$, которые можно представить в виде конечной дизъюнкции прогрессивных множеств в алфавите A, обозначаем через $U(A)$. Если $P \in U(A)$, то его *прогрессивной сложностью* называем минимальную из сложностей выражений из $\mathfrak{X}(A)$, которые задают P. Обозначаем это число через $L_p(P)$:

$$L_p(P) := \min_{\mathfrak{P} \in \mathfrak{X}(A), |\mathfrak{P}| = P} L(\mathfrak{P}). \qquad (2)$$

Замечание. Обсудим корректность этого определения. Так как P представляется в виде конечной дизъюнкции прогрессивных множеств в алфавите A, то его можно задать конечной дизъюнкцией регулярных выражений в алфавите A вида (1), то есть каким-то выражением из класса $\mathfrak{X}(A)$. Значит минимум в (2) существует.

Для произвольного натурального $n \in \mathbb{N}$ через $U^n(A)$ обозначаем класс

$$U^n(A) := \{P \in U(A) | L_p(P) \leq n\}.$$

Теорема 4.1 *Пусть A и B - конечные непустые алфавиты, $f \in F(A, B)$ и для некоторого $n \in \mathbb{N}$ верно, что $P \in U^n(A)$. Тогда $P \in I(f)$ если и только если $P_{\leq}((n+2)^2 + 4n^2 l_f^2 + l_f) \in I(f)$.*

2. Доказательство вспомогательных утверждений

Лемма 1. *Пусть A - конечный алфавит. Тогда $U(A) = \mathfrak{T}(A)$.*

Доказательство. Пусть $P \in U(A)$. Тогда P представимо в виде конечного объединения прогрессивных множеств в алфавите A. Очевидно, что прогрессивные множества в алфавите A являются тонкими в том же алфавите. Но при доказательстве теоремы 2.2 из главы 2 мы показали, что конечное объединение тонких множеств в алфавите A тоже является тонким. Значит $P \in \mathfrak{T}(A)$.

Обратно, пусть $P \in \mathfrak{T}(A)$. Из теоремы 2.2 следует, что P можно представить в виде конечного объединения прогрессивных множеств в алфавите A. Значит $P \in U(A)$. Утверждение леммы 1 доказано.

Лемма 2. *Пусть A - конечный алфавит и множество $P \subseteq A^*$ представимо регулярным выражением вида*

$$\alpha \cdot (\beta)^* \cdot \gamma,$$

где $\alpha, \beta, \gamma \in A^*$ и $|\alpha| + |\beta| + |\gamma| = n$. Тогда существует автомат

$$V \in K_{\leq}(A, E_2, n+2)$$

такой, что $1(V) = P$.

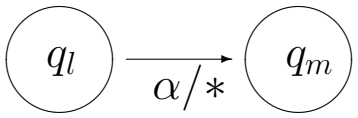
Доказательство. Если выражение

$$\alpha \cdot (\beta)^* \cdot \gamma \tag{1}$$

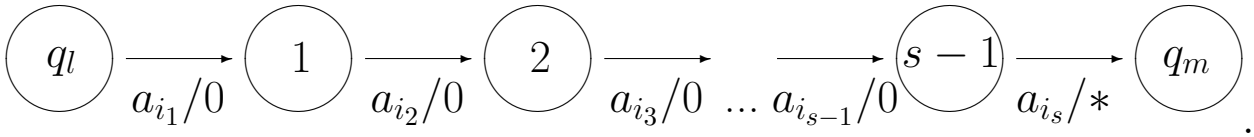
не правильного линейного вида, то или $\beta = \Lambda$, или $[_1\beta = [_1\gamma$. В первом случае

$$P = |\alpha \cdot \gamma|.$$

Составим диаграмму для автомата V. Здесь и далее во всех диаграммах звездочкой над состоянием помечено начальное состояние и для любых состояний q_l, q_m и слова $\alpha = a_{i_1} \dots a_{i_s} \in A^*$ под



подразумевается



.

Если слово α равно Λ, то все состояния $q_l, 1, 2, \dots, s-1, q_m$ отождествляются и соответствующие переходы не проводятся.

Итак, диаграмма для V имеет вид

.

Здесь, как и во всех последующих диаграммах, есть еще одно дополнительное состояние, выполняющее роль "тупика". В него отправляются все недостающие стрелки и на выходе у этих стрелок стоит символ 0. Поэтому в приведенной диаграмме всего

$$3 + (|\alpha| - 1) + (|\gamma| - 1) + 1 = n + 2$$

состояний.

Разберем теперь второй случай, когда $[_1\beta = [_1\gamma$, то есть

$$\beta = a\beta',\ \gamma = a\gamma'$$

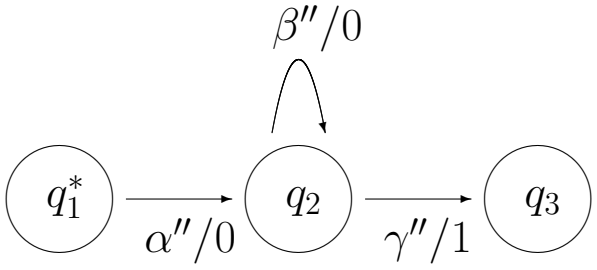
для некоторых $a \in A$ и $\beta', \gamma' \in A^*$. Применяем к выражению (1) операцию перекидывания

$$\alpha(a\beta')^* \cdot (a\gamma') = \alpha a \cdot (\beta' a)^* \cdot \gamma'$$

до тех пор, пока оно не примет правильный линейный вид

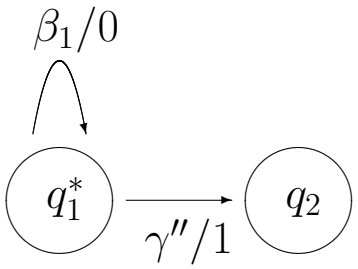
$$\alpha'' \cdot (\beta'')^* \cdot \gamma''. \tag{2}$$

При этом сложность выражения (2) по прежнему будет равна сложности выражения (1), то есть числу n. Если выражение (1) сразу было правильного линейного вида, то оно и равно выражению (2). Теперь мы можем нарисовать диаграммы для автомата V. Если $\alpha'' \neq \Lambda$ и $\gamma'' \neq \Lambda$, то диаграмма имеет вид



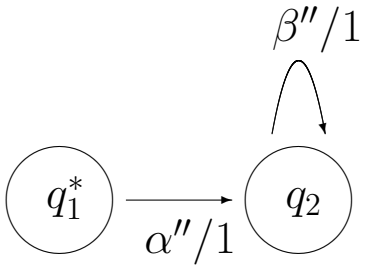
.

Если $\alpha'' = \lambda,\ \gamma'' \neq \lambda$, то диаграмма имеет вид



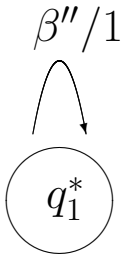
.

Если $\alpha'' \neq \lambda,\ \gamma'' = \lambda$, то диаграмма имеет вид



.

Наконец, если $\alpha'' = \gamma'' = \Lambda$, то диаграмма имеет вид



.

В каждой из этих приведенных диаграмм не более чем

$$3 + (|\alpha''| - 1) + (|\beta''| - 1) + (|\gamma''| - 1) + 1 = n + 1$$

состояний. Утверждение леммы 2 доказано.

Лемма 3. *Пусть A - конечный алфавит, множество $P \subseteq A^*$ представимо регулярным выражением вида*

$$\alpha \cdot (\beta)^* \cdot \gamma,$$

где $\alpha, \beta, \gamma \in A^*$ и $|\alpha| + |\beta| + |\gamma| = n$. И пусть $V \in K_{\leq}(A, E_2, n+2)$ - автомат из леммы 2 для P. Тогда для всех $V_1 \in [V]$ выполнено одно из двух условий:

1) $1(V_1) = \emptyset$,

2) существуют слова $\alpha'(V_1), \beta'(V_1), \gamma'(V_1) \in A^*$ такие, что

$$1(V_1) = \alpha'(V_1) \cdot (\beta'(V_1))^* \cdot \gamma'(V_1) \quad \text{и} \quad |\alpha'(V_1)| + |\beta'(V_1)| + |\gamma'(V_1)| \leq 2n.$$

Доказательство. Если $\beta = \Lambda$, то утверждение очевидно. Поэтому будем доказывать утверждение леммы только для основного случая, когда в диаграмме автомата V есть петля β''. Если в качестве начального состояния автомата V_1 берется "тупиковое" состояние автомата V, то $1(V_1) = \emptyset$. Опять же, если это состояние, находящееся на конце у слова γ'', то $1(V_1) = \emptyset$. Пусть теперь это какое-то другое состояние. Возможны три случая.

Случай 1. Это состояние находится внутри или по краям слова α''. Тогда

$$1(V_1) = \alpha''' \cdot (\beta'')^* \cdot \gamma''$$

для некоторого постфикса (возможно, пустого) α''' слова α''. Осталось положить

$$\alpha'(V_1) := \alpha''',\ \beta'(V_1) := \beta'',\ \gamma'(V_1) := \gamma''.$$

Тогда

$$|\alpha'(V_1)| + |\beta'(V_1)| + |\gamma'(V_1)| \leq |\alpha''| + |\beta''| + |\gamma''| = n.$$

Случай 2. Это состояние находится внутри петли β''. Тогда

$$1(V_1) = \beta''' \cdot (\beta'')^* \cdot \gamma''$$

для некоторого постфикса β''' слова β''. Осталось положить

$$\alpha'(V_1) := \beta''',\ \beta'(V_1) := \beta'',\ \gamma'(V_1) := \gamma''.$$

Тогда

$$|\alpha'(V_1)| + |\beta'(V_1)| + |\gamma'(V_1)| \leq 2|\beta''| + |\gamma''| \leq 2n.$$

Случай 3. Это состояние находится внутри γ''. Тогда

$$1(V_1) = \gamma'''$$

для некоторого постфикса γ''' слова γ''. Осталось положить

$$\alpha'(V_1) := \Lambda,\ \beta'(V_1) := \Lambda,\ \gamma'(V_1) := \gamma'''.$$

Тогда

$$|\alpha'(V_1)| + |\beta'(V_1)| + |\gamma'(V_1)| \leq |\gamma''| \leq n.$$

Утверждение леммы 3 доказано.

3. Доказательство основных утверждений

Теорема 4.1 *Пусть A и B - конечные непустые алфавиты, $f \in F(A, B)$ и для некоторого $n \in \mathbb{N}$ верно, что $P \in U^n(A)$. Тогда $P \in I(f)$ если и только если $P_{\leq}((n+2)^2 + 4n^2 l_f^2 + l_f) \in I(f)$.*

Доказательство. В одну сторону утверждение очевидно - из $P \in I(f)$ тривиально следует, что $P_{\leq}((n+2)^2 + 4n^2 l_f^2 + l_f) \in I(f)$. Докажем, что верно и обратное. Пусть

$$P_{\leq}((n+2)^2 + 4n^2 l_f^2 + l_f) \in I(f).$$

Так как $P \in U^n(A)$, то $L_p(P) \leq n$ и значит существует $\mathfrak{P} \in \mathfrak{X}(A)$, для которого $|\mathfrak{P}| = P$ и $L(\mathfrak{P}) \leq n$. Другими словами, P задается

выражением, являющимся конечной дизъюнкцией выражений вида

$$\alpha \cdot (\beta)^* \cdot \gamma,$$

где $\alpha, \beta, \gamma \in A^*$ и $|\alpha| + |\beta| + |\gamma| \leq n$. То есть

$$P = \bigcup_{i=1}^{s} \alpha_i \cdot (\beta_i)^* \cdot \gamma_i,$$

где $s \in \mathbb{N}$ и при $1 \leq i \leq s$ верно $\alpha_i, \beta_i, \gamma_i \in A^*$ и $|\alpha_i| + |\beta_i| + |\gamma_i| \leq n$. Обозначаем

$$P_i := \alpha_i \cdot (\beta_i)^* \cdot \gamma_i.$$

По лемме 2 при каждом значении $1 \leq i \leq s$ существует автомат

$$V_i \in K_{\leq}(A, E_2, n + 2),$$

для которого $1(V_i) = P_i$. Кроме того, для любого $V \in [V_i]$ из леммы 3 получаем, что или $1(V) = \emptyset$, или же существуют слова

$$\alpha_i'(V), \beta_i'(V), \gamma_i'(V) \in A^*,$$

для которых выполнено

$$1(V) = \alpha_i'(V) \cdot (\beta_i'(V))^* \cdot \gamma_i'(V),$$

$$|\alpha_i'(V)| + |\beta_i'(V)| + |\gamma_i'(V)| \leq 2n.$$

В первом случае $\tilde{f}(1(V)) = \emptyset$. Тогда очевидно, что существует автомат

$$W \in K_{\leq}(B, E_2, 2nl_f + 2),$$

для которого $1(W) = \tilde{f}(1(V))$. Во втором случае имеем

$$\tilde{f}(1(V)) = \tilde{f}(\alpha_i'(V)) \cdot (\tilde{f}(\beta_i'(V)))^* \cdot \tilde{f}(\gamma_i'(V)), \qquad (1)$$

$$|\tilde{f}(\alpha_i'(V))| + |\tilde{f}(\beta_i'(V))| + |\tilde{f}(\gamma_i'(V))| \leq$$

$$\leq l_f \cdot (|\alpha_i'(V)| + |\beta_i'(V)| + |\gamma_i'(V)|) \leq 2nl_f.$$

Далее, из леммы 2, примененной к выражению (1) в алфавите B, следует существование автомата

$$W \in K_{\leq}(B, E_2, 2nl_f + 2)$$

такого, что $1(W) = \tilde{f}(1(V))$. Все условия теоремы 1.2 из главы 1 выполнены. Значит $P \in I(f)$. Утверждение теоремы доказано.

Замечание. Оценку $(n+2)^2+(2nl_f+2)^2+l_f$ из формулировки теоремы, вообще говоря, можно понизить. Для этого надо углубиться в доказательство теоремы 1.2 и заметить, что далеко не все пары состояний, например, автоматов V_i могут встретиться в общем образе двух скленных слов из множества P. Это связано с тем, что один раз попав в "тупиковое"состояние из него уже нельзя выйти. Значит пар, в которых такое состояние есть, не будет. Но такие детали нам не важны, поэтому формально доказывать более точные оценки мы не будем.

4. Заключение главы 4

В этой главе нами была получена квадратичная по n и l_f оценка на длину перебираемых слов в алгоритме, доставляющем решение проблемы ОАД_2. Отсюда получаем тривиальное следствие, что проблема ОАД_2 имеет полиномиальное решение, то есть лежит в классе P. (о классах P и NP можно прочитать, например, в [35].) Это говорит в пользу того, чтобы в рамках нашей модели брать язык P из класса $\mathfrak{T}(X)$. В следующей главе эти результаты будут перенесены на более общий и потому более трудоемкий (но и более интересный) для исследования случай решения проблемы ОАД_3.

Глава 5

Аннотация

В этой главе приводится решение проблемы ОАД$_3$ для класса $RP(A)$ регулярных языков с полиномиальной функцией роста в некотором произвольном алфавите A. Это решение основывается на теореме 1.3 из главы 1, результатах из главы 3 о разложении языков из класса $RP(A)$ в конечное объединение множеств правильного линейного вида и на технике, использованной в главе 4.

1. Основные понятия и результаты

Здесь приведены только те определения, которых еще не было до этого. Если при чтении главы какие-то термины не ясны и их нет в этом разделе, то их определения можно найти в аналогичных разделах предыдущих глав.

Пусть A - конечный алфавит. Обозначаем через $RP^1(A)$ множество всех множеств линейного вида в алфавите A, а через $WRP^1(A)$ - множество всех множеств правильного линейного вида в алфавите A. При всех $n \in \mathbb{N}$ обозначаем через $RP^1_n(A)$ множество всех $P \in RP^1(A)$, которые представимы регулярными выражениями линейного вида сложности не выше n. Через $RP_n(A)$ обозначаем множество всех $P \in RP(A)$, которые могут быть получены конечным объединением множеств из $RP^1_n(A)$. Для произвольного $n \in \mathbb{N}$ обозначаем через $WRP^1_n(A)$ множество всех $P \in WRP^1(A)$,

которые представимы регулярными выражениями правильного линейного вида сложности не выше n. Через $WRP_n(A)$ обозначаем множество всех $P \in RP(A)$, которые можно получить конечным объединением множеств из $WRP_n^1(A)$.

Теорема 5.1 *Пусть A и B - конечные непустые алфавиты, $P \in WRP_n(A)$ и $f \in F(A,B)$. Тогда $P \in I(f)$ если и только если $P_{\leq}((n+2)^2 + (4n^2 l_f^2 + 2)^2 + l_f) \in I(f)$.*

Теорема 5.2 *Пусть A и B - конечные непустые алфавиты, $P \in RP_n(A)$ и $f \in F(A,B)$. Тогда $P \in I(f)$ если и только если $P_{\leq}((n^2+2)^2 + (4n^4 l_f^2 + 2)^2 + l_f) \in I(f)$.*

2. Доказательство вспомогательных утверждений

Лемма 1. *Пусть A - конечный алфавит и $P \in WRP_n^1(A)$ для некоторого $n \in \mathbb{N}$. Тогда существует $V \in K_{\leq}(A, E_2, n+2)$ такой, что $1(V) = P$.*

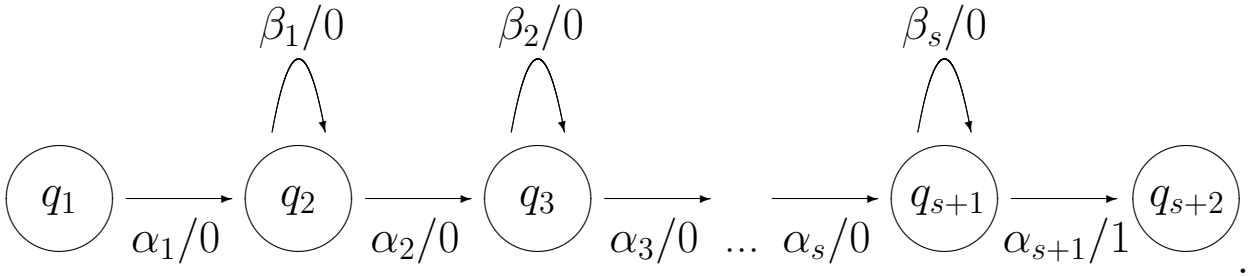
Доказательство. Эта лемма является обобщением леммы 2 из предыдущей главы. Все обозначения для диаграмм автоматов взяты оттуда.

Так как $P \in WRP_n^1(A)$, то P представимо выражением правильного линейного вида сложности не выше n :

$$P = |\alpha_1 \cdot (\beta_1)^* \cdot \alpha_2 \cdot (\beta_2)^* \cdot \ldots \cdot \alpha_s \cdot (\beta_s)^* \cdot \alpha_{s+1}| \,.$$

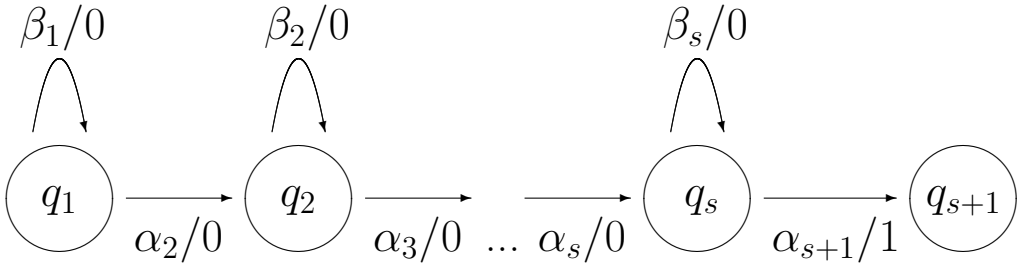
Построим диаграмму Мура для V. Возможно 4 случая.

Случай 1. $\alpha_1 \neq \Lambda,\ \alpha_{s+1} \neq \Lambda$. Тогда соответствующая диаграмма имеет вид



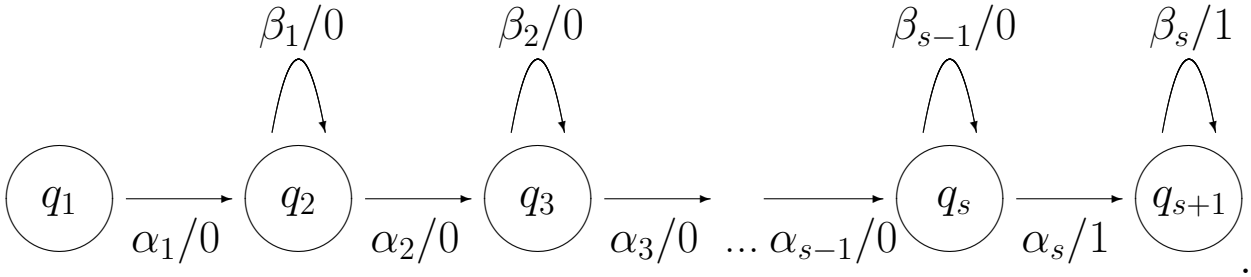
.

Случай 2. $\alpha_1 = \Lambda,\ \alpha_{s+1} \neq \Lambda$. Тогда диаграмма имеет вид



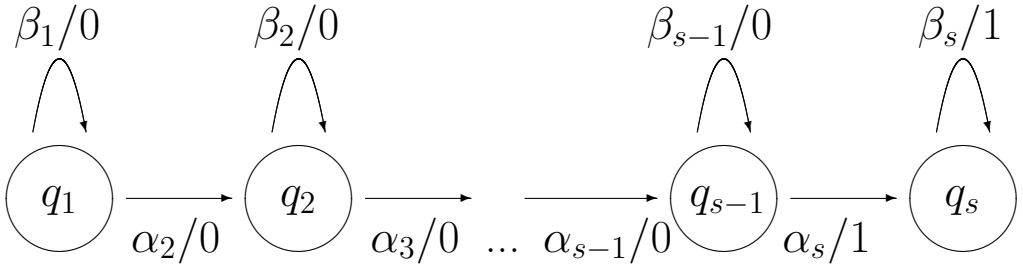
.

Случай 3. $\alpha_1 \neq \Lambda,\ \alpha_{s+1} = \Lambda$. Тогда диаграмма имеет вид



.

Случай 4. $\alpha_1 = \alpha_{s+1} = \Lambda$. Тогда диаграмма имеет вид



.

Во всех этих диаграммах есть еще одно дополнительное состоя-

ние, выполняющее роль "тупика". В него отправляются все недостающие стрелки и на выходе у этих стрелок символ 0. В каждой из этих диаграмм не более чем

$$(s+2)+(l(\alpha_1)-1)+(l(\beta_1)-1)+\ldots+(l(\beta_s)-1)+(l(\alpha_{s+1})-1)+1=$$

$$=l(\alpha_1)+l(\beta_1)+\ldots+l(\beta_s)+l(\alpha_{s+1})+(s+2)-(2s+1)+1\leq$$

$$\leq l(\alpha_1)+l(\beta_1)+\ldots+l(\beta_s)+l(\alpha_{s+1})+2\leq n+2$$

состояний. Утверждение леммы 1 доказано.

Лемма 2. *Пусть A - конечный алфавит, $P\in WRP_n^1(A)$ для некоторого $n\in\mathbb{N}$ и V - автомат из леммы 1 для множества P. Тогда для всех $V_1\in[V]$ выполнено одно из двух условий:*

1) $1(V_1)=\emptyset$;

2) $1(V_1)\in WRP_{2n}^1(A)$.

Доказательство. Эта лемма является обобщением леммы 3 из предыдущей главы. Так как $P\in WRP_n^1(A)$, то P представимо выражением правильного линейного вида сложности не выше n :

$$P=|\alpha_1\cdot(\beta_1)^*\cdot\alpha_2\cdot(\beta_2)^*\cdot\ldots\cdot\alpha_s\cdot(\beta_s)^*\cdot\alpha_{s+1}|\,. \qquad (1)$$

Разберем, для примера, случай, когда $\alpha_1\neq\Lambda$ и $\alpha_{s+1}\neq\Lambda$. Остальные случаи разбираются аналогично. Если в качестве начального состояния автомата V_1 берется "тупиковое" состояние автомата V, то $1(V_1)=\emptyset$. Опять же, если это состояние, находящееся на конце

у слова α_{s+1}, то $1(V_1) = \emptyset$. Пусть теперь это какое-то другое состояние. Возможны три принципиально разных случая.

Случай 1. Это состояние находится внутри или по краям одного из слов α_i, где $i \leq s$. Тогда

$$1(V_1) = |\alpha' \cdot (\beta_i)^* \cdot \ldots (\beta_s)^* \cdot \alpha_{s+1}| \tag{2}$$

для некоторого постфикса (возможно, пустого) α' слова α_i. Так как выражение из (1) имеет правильный линейный вид, то и множество (2) имеет правильный линейный вид. Кроме того,

$$L(\alpha' \cdot (\beta_i)^* \cdot \ldots (\beta_s)^* \cdot \alpha_{s+1}) \leq$$

$$\leq L(\alpha_1 \cdot (\beta_1)^* \cdot \alpha_2 \cdot (\beta_2)^* \cdot \ldots \cdot \alpha_s \cdot (\beta_s)^* \cdot \alpha_{s+1}) \leq n.$$

Значит $1(V_1) \in WRP_n^1(A)$.

Случай 2. Это состояние находится внутри какой-то из петель β_i, где $i \leq s$. Тогда

$$1(V_1) = |\beta' \cdot (\beta_i)^* \cdot \ldots (\beta_s)^* \cdot \alpha_{s+1}| \tag{3}$$

для некоторого постфикса β' слова β_i. Так как выражение из (1) имеет правильный линейный вид, то и множество (3) имеет правильный линейный вид. Кроме того,

$$L(\beta' \cdot (\beta_i)^* \cdot \ldots (\beta_s)^* \cdot \alpha_{s+1}) \leq$$

$$\leq 2L(\alpha_1 \cdot (\beta_1)^* \cdot \alpha_2 \cdot (\beta_2)^* \cdot \ldots \cdot \alpha_s \cdot (\beta_s)^* \cdot \alpha_{s+1}) \leq 2n.$$

Значит $1(V_1) \in WRP^1_{2n}(A)$.

Случай 3. Это состояние находится внутри α_{s+1}. Тогда

$$1(V_1) = |\gamma'| \tag{4}$$

для некоторого постфикса γ' слова α_{s+1}. Очевидно, что множество (4) имеет правильный линейный вид и

$$L(\gamma') \leq L(\alpha_1 \cdot (\beta_1)^* \cdot \alpha_2 \cdot (\beta_2)^* \cdot \ldots \cdot \alpha_s \cdot (\beta_s)^* \cdot \alpha_{s+1}) \leq n.$$

Значит $1(V_1) \in WRP^1_n(A)$.

Все случаи разобраны. Утверждение леммы 2 доказано.

Лемма 3. *Пусть A - конечный алфавит и $P \in RP_n(A)$ для некоторого $n \in \mathbb{N}$. Тогда $P \in WRP_{n^2}(A)$.*

Доказательство. Доказательство леммы основано на процедуре, описанной в лемме 6 главы 3. Рекомендуется предварительно вспомнить используемую там технику.

Очевидно, что достаточно доказать утверждение лишь для случая, когда $P \in RP^1_n(A)$. Пусть P представимо регулярным выражением линейного вида

$$\mathfrak{P} = \alpha_1 \cdot (\beta_1)^* \cdot \alpha_2 \ldots \alpha_s \cdot (\beta_s)^* \cdot \alpha_{s+1}, \tag{1}$$

где $s \in \mathbb{N}_0$, $\alpha_1, \ldots, \alpha_{s+1} \in A^*$, $\beta_1, \ldots, \beta_s \in A^* \setminus \{\Lambda\}$. Рассмотрим более подробно процедуру преобразования этого выражения в конечную дизъюнкцию выражений правильного линейного вида,

которая описана в лемме 6 главы 3. При выполнении процедуры "перекидывания" сложность рассматриваемых линейных выражений остается неизменной. А процедур "расщепления" всего будет проведено не более $s-1$ - по числу соседних пар итераций. Для каждой из этих процедур возможны два случая.

Случай 1. Под расщепляемыми итерациями находятся соизмеримые слова. Без ограничения общности, рассмотрим случай, когда это β_1^* и β_2^*. Для некоторых $\nu \in A^*$, $a, b \in \mathbb{N}$ имеем $\beta_1 = \nu^a$, $\beta_2 = \nu^b$. Получаем выражение

$$(\nu^a)^* \cdot (\nu^b)^*. \tag{2}$$

Напомним, что в этом случае к выражению (2) применяется лемма 4 главы 3. Более подробно, пусть

$$r := \text{НОД}(a, b) \quad \text{и} \quad H := \{a \cdot x + b \cdot y \mid x, y \in \mathbb{N}_0\}.$$

Из леммы 3 главы 2 берется $n_0 \in \mathbb{N}$, для которого

$$Z_{n_0,r} \subseteq H.$$

Ясно, что в качестве n_0 можно взять некоторое число, не превосходящее $a \cdot b$. (Этот факт легко доказать, рассмотрев остатки по модулю b среди чисел $0, a, 2a, \ldots, (b-1)a$.) Далее выражение (2) заменяется на дизъюнкцию выражения $\nu^{n_0} \cdot (\nu^r)^*$ и по отдельности выражений для всех элементов множества $\{\nu^x \mid x \in H \setminus Z_{n_0,r}\}$. Посчитаем, насколько при такой замене могла увеличиться по срав-

нению с исходной сложностью сложность новых линейных выражений. Эта разница не больше чем

$$|\nu^{n_0}| + \left|\nu^{\text{НОД(a,b)}}\right| - |\nu^a| - |\nu^b| = |\nu|(n_0 + \text{НОД(a,b)} - a - b) \leq$$

$$\leq |\nu| \cdot a \cdot b \leq (|\nu| \cdot a) \cdot (|\nu| \cdot b) = |\beta_1| \cdot |\beta_2|.$$

Случай 2. Под расщепляемыми итерациями находятся несоизмеримые слова. Опять, без ограничения общности, считаем, что это β_1^* и β_2^*. Тогда выражение

$$(\beta_1)^* \cdot (\beta_2)^* \tag{3}$$

заменяется на выражение

$$(\beta_1)^* \cdot (\lambda)^* \vee (\beta_1)^* \cdot \beta_2 \cdot (\lambda)^* \vee (\beta_1)^* \cdot \beta_2^2 \cdot (\lambda)^* \vee \ldots \vee$$

$$\vee (\beta_1)^* \cdot \beta_2^{|\beta_1|-1} \cdot (\lambda)^* \vee (\beta_1)^* \cdot \beta_2^{|\beta_1|} \cdot (\beta_2)^*.$$

Посчитаем, насколько при такой замене могла увеличиться по сравнению со сложностью для (3) сложность новых линейных выражений. Эта разница не больше чем

$$\left(|\beta_1| + \left|\beta_2^{|\beta_1|}\right| + |\beta_2|\right) - (|\beta_1| + |\beta_2|) = \left|\beta_2^{|\beta_1|}\right| = |\beta_1| \cdot |\beta_2|.$$

Таким образом, в каждом из обоих случаев сложность новых линейных выражений увеличивается не более чем на $|\beta_1| \cdot |\beta_2|$. Осталось заметить, что в новых выражениях длина слов под итерациями может только уменьшиться: в первом случае происходит замена $\beta_1 = \nu^a$ и $\beta_2 = \nu^b$ на $\nu^{\text{НОД(a,b)}}$ или λ, во втором - замена

β_2 на β_2 или λ. Таким образом, после применения всех процедур "расщепления" сложность новых линейных выражений увеличивается не более чем на

$$|\beta_1| \cdot |\beta_2| + |\beta_2| \cdot |\beta_3| + \ldots + |\beta_{s-1}| \cdot |\beta_s|.$$

Сложность исходного линейного выражения (1) равна

$$|\alpha_1| + \ldots + |\alpha_{s+1}| + |\beta_1| + \ldots + |\beta_s|$$

и, по условию леммы, не превосходит n. Отсюда получаем следующую оценку сверху на сложность окончательных выражений правильного линейного вида:

$$(|\alpha_1| + \ldots + |\alpha_{s+1}| + |\beta_1| + \ldots + |\beta_s|) + |\beta_1| \cdot |\beta_2| +$$

$$+|\beta_2| \cdot |\beta_3| + \ldots + |\beta_{s-1}| \cdot |\beta_s| \leq |\alpha_1|^* + \ldots + |\alpha_{s+1}|^* + |\beta_1|^* + \ldots + |\beta_s|^* +$$

$$+2|\beta_1| \cdot |\beta_2| + 2|\beta_2| \cdot |\beta_3| + \ldots + 2|\beta_{s-1}| \cdot |\beta_s| \leq$$

$$\leq (|\alpha_1| + \ldots + |\alpha_{s+1}| + |\beta_1| + \ldots + |\beta_s|)^2 = n^2.$$

Утверждение леммы 3 доказано.

3. Доказательство основных утверждений

Теорема 5.1 *Пусть A и B - конечные непустые алфавиты, $P \in WRP_n(A)$ и $f \in F(A, B)$. Тогда $P \in I(f)$ если и только если $P_{\leq}((n+2)^2 + (4n^2 l_f^2 + 2)^2 + l_f) \in I(f)$.*

Доказательство. Для доказательства теоремы мы используем теорему 1.3 из главы 1 при

$$k := n + 2 \quad \text{и} \quad m := 4n^2 l_f^2 + 2.$$

Покажем, что все ее условия выполнены, то есть что P представимо в виде конечного объединения множеств P_i, для которых при всех i имеем:

1) существует $V_i \in K_{\leq}(A, E_2, n+2)$, для которого $1(V_i) = P_i$;

2) для каждого $V \in [V_i]$ существует конечное множество автоматов

$$W_i \in K_{\leq}(B, E_2, 4n^2 l_f^2 + 2),\ 1 \leq i \leq s,\ s \in \mathbb{N} \tag{1}$$

таких, что

$$\bigcup_{i=1}^{s} 1(W_i) = \tilde{f}(1(V)).$$

Так как $P \in WRP_n(A)$, то P можно представить в виде конечного объединения множеств из $WRP_n^1(A)$, то есть множеств правильного линейного вида, которые представимы выражениями правильного линейного вида в алфавите A длины не выше n. Эти множества и берем за P_i. Соответствующие им регулярные выражения обозначаем через $\mathfrak{P}_i$. Выполнение условия 1) теперь следует из леммы 1. Покажем, что и условие 2) выполнено. Пусть для некоторого i имеем $V \in [V_i]$. Из леммы 2 следует, что или $1(V) = \emptyset$, или $1(V) \in WRP_{2n}^1(A)$. В первом случае свойство 2), очевидно, выпол-

нено. Во втором случае рассмотрим множество

$$X := \tilde{f}(1(V)).$$

Ясно, что это множество линейного вида в алфавите B. Так как

$$1(V) \in WRP^1_{2n}(A),$$

то множество X можно задать выражением линейного вида в алфавите B со сложностью, не превосходящей $2n \cdot l_f$. То есть

$$X \in RP^1_{2n \cdot l_f}(B).$$

Применяем теперь к множеству X процедуру из леммы 6 главы 3. Эта процедура разлагает X в конечное объединение множеств X_i правильного линейного вида в алфавите B :

$$X := \bigcup_{i=1}^{s} X_i.$$

Число s здесь будет совпадать с s из (1), поэтому дальнейшей путаницы в обозначениях не возникнет. Из доказательства леммы 3 следует, что для всех $1 \leq i \leq s$ имеем

$$X_i \in WRP^1_{4n^2 l_f^2}(B).$$

Осталось заметить, что тогда при всех $1 \leq i \leq s$ из леммы 1 следует существование автоматов

$$V_i' \in K_{\leq}(B, E_2, 4n^2 l_f^2 + 2)$$

таких, что $1(V_i') = X_i$. Эти-то V_i' мы и берем за W_i из (1). Все условия теоремы 1.3 выполнены. Это замечание завершает доказательство теоремы.

Теорема 5.2 *Пусть A и B - конечные непустые алфавиты, $P \in RP_n(A)$ и $f \in F(A,B)$. Тогда $P \in I(f)$ если и только если $P_{\leq}((n^2+2)^2 + (4n^4 l_f^2 + 2)^2 + l_f) \in I(f)$.*

Доказательство. Утверждение теоремы является простым следствием теоремы 5.1 и леммы 3.

4. Заключение главы 5

В этой главе нами была получена полиномиальная по n и l_f оценка на длину перебираемых слов в алгоритме, доставляющем решение проблемы ОАД_3. При этом, если за n берется максимальная длина выражений правильного линейного вида в дизъюнктивном представлении множества P, то ограничивающий длину полином имеет четвертую степень как по n, так и по l_f. Если же в дизъюнктивном представлении можно использовать любые выражения линейного вида (не обязательно правильные), то степень полинома по n поднимается до 8. Но само n при этом может стать значительно меньше. Отсюда получаем тривиальное следствие, что проблема ОАД_3 имеет полиномиальное решение, то есть лежит в классе P.

Этот результат говорит в пользу того, чтобы в рамках нашей модели брать язык P из класса $RP(X)$. Итак, у нас возникает два варианта для выбора P - класс $\mathfrak{T}(X)$ и класс $RP(X)$.

Для каждого из этих случаев поговорим о безопасности нашей модели. Во-первых, система защищена тем, что параметры процедуры вложения исходных сообщений в язык P скрыты от сторонних пользователей и конкурентов. Во-вторых, скрыты и параметры кодирования сообщения $\tilde{f}(\alpha)$, которое нужно для борьбы с помехами при передаче по каналу. В третьих, даже если все эти параметры будут взломаны, еще остаются неизвестными P и f. И если f еще можно каким-то образом вскрыть, исследуя частотные характеристики сообщения $\tilde{f}(\alpha)$, то найти P крайне затруднительно. Для класса $\mathfrak{T}(X)$ это еще теоретически возможно, но в классе $RP(X)$ каких-либо путей для идентификации P, кроме разве что утечки информации, не предвидится. А не зная P, сторонний конкурент не сможет изготовить продукт, аналогичный нашему по своим свойствам. При выборе между классами $\mathfrak{T}(X)$ и $RP(X)$ следует руководствоваться в первую очередь тем, что для нас важнее - безопасность или простота реализации. В первом случае лучше брать класс $RP(X)$, а во втором - $\mathfrak{T}(X)$.

В следующей главе мы перейдем к вопросу о том, как в рамках нашей модели сравнивать между собой эффективность использо-

вания соответствующих языков и функций алфавитного кодирования. Для этого мы введем понятие полезности для регулярных множеств и функций алфавитного кодирования и научимся алгоритмически сравнивать их между собой (множества - с множествами, а функции - с функциями).

Глава 6

Аннотация

В главе изучаются две проблемы вложения: проблема вложения классов допустимых регулярных языков, задаваемых функциями алфавитного кодирования (сокращенно - проблема ВКД$_1$) и проблема вложения классов допустимых функций алфавитного кодирования, задаваемых регулярными языками (сокращенно - проблема ВКД$_2$).

Для доказательства алгоритмической разрешимости проблемы ВКД$_1$ используется тот факт, что отношение синонимии, возникающее на языке при алфавитном кодировании, регулярно в его обобщенном алфавите. Это утверждение принадлежит Ал. А. Маркову и его можно найти в [22]. Далее идет ссылка на то, что проблема проверки вложимости двух произвольных регулярных множеств в общем алфавите алгоритмически разрешима. Доказательство этого факта можно найти в [26]. Из этих двух фактов уже несложно выводится требуемое утверждение.

Далее в главе приводится результат об алгоритмической разрешимости проблемы ВКД$_2$ для случая, когда мощность входного алфавита равна двум. При этом показывается, что классы допустимых функций кодирования можно (в некотором смысле) представлять подмножествами рациональных чисел. Это позволяет свести исследуемую проблему к проверке вложения некоторых подмножеств рациональных чисел. И уже для этой проблемы успешно разрешается вопрос об ее алгоритмической разрешимости. При этом используются следующие общеизвестные утверждения из теории контекстно-свободных языков:

1) множество контекстно-свободных языков совпадает с множеством языков, распознаваемых автоматами с магазинной памятью;

2) пересечение и разность контекстно-свободного и регулярного языков - контекстно-свободный язык;

3) проблема проверки пустоты контекстно-свободного языка, заданного автоматом с магазинной памятью, алгоритмически разрешима.

Подробно об этих результатах можно прочитать, например, в [36-37].

1. Основные понятия и результаты

Здесь приведены только те определения, которых еще не было до этого. Если при чтении главы какие-то термины не ясны и их нет в этом разделе, то их определения можно найти в аналогичных разделах предыдущих глав.

Пусть A, B - некоторые конечные алфавиты и τ - конечный набор строк вида $\alpha \to \beta$, где $\alpha, \beta \in (A \cup B)^*$ и $\alpha \notin A^*$. Пусть, кроме того, b - некоторая буква алфавита B. Тогда *контекстно-свободной грамматикой* называем набор

$$X := (A, B, \tau, b).$$

Выводом в этой грамматике называем произвольную конечную последовательность слов

$$\alpha_1, \alpha_2, \ldots, \alpha_s \in (A \cup B)^*,$$

в которой $\alpha_1 = b$, $\alpha_s \in A^*$ и каждое следующее слово α_{s+1} получено из предыдущего слова α_s заменой некоторого его подслова β_1 на слово β_2, причем в τ есть строка вида $\beta_1 \to \beta_2$. Слово α_s называем *результатом этого вывода.* Множество всевозможных результатов вывода в контекстно-свободной грамматике X обозначаем через $|X|$.

Пусть A - конечный алфавит и $P \subseteq A^*$. Говорим, что P - *контекстно-свободный язык в алфавите* A, если для некоторых B, τ, b существует контекстно-свободная грамматика

$$X := (A, B, \tau, b),$$

для которой $|X| = P$. Множество всех контекстно-свободных языков в алфавите A обозначаем через $\mathbb{CF}(A)$.

Инициальным конечным автоматом с магазинной памятью называется набор

$$V = (A, Q, B, \varphi, \psi, \lambda, q_0),$$

где A, Q, B - конечные множества, φ - функция, определенная на множестве $A \times Q \times B$ и принимающая значения из Q, ψ - функция, определенная на множестве $A \times Q \times B$ и принимающая значения из $\{d\} \cup B$, λ - специальная буква алфавита B. Здесь d - служебный символ, не входящий в алфавит B. При этом на функцию ψ накладывается дополнительное ограничение: если третий элемент из тройки в $A \times Q \times B$ равен λ, то под действием ψ она не должна переходить в d. Множества A, Q, B называются соответственно *входным алфавитом*, *алфавитом состояний* и *алфавитом стека.* Функция φ называется *функцией переходов*, а функция ψ - *функцией выходов*. Буква λ называется *пустым символом*, а символ d - *символом стирания*. Множество всех инициальных конечных автоматов с магазинной памятью с входным алфавитом A обозначаем через $\mathbb{PDA}(A)$.

Пусть $V = (A, Q, B, \varphi, \psi, \lambda, q_0)$ - инициальный конечный автомат с магазинной памятью и α - слово в алфавите A. Определим результат применения автомата V к слову α. На автомат посту-

пает первая буква a_{i_1} слова α. В этот момент времени автомат имеет начальное состояние q_0 и в его магазинной памяти хранится пустое слово в алфавите B. По функции φ автомат вычисляет состояние $q_1 := \varphi(a_{i_1}, q_0, \lambda)$ в следующий момент времени. После этого автомат по функции ψ вычисляет $\psi(a_{i_1}, q_0, \lambda)$. Так как это не символ стирания, то это какая-то буква алфавита B. Если это λ, то автомат не меняет слово из магазинной памяти. Если же это любая другая буква, то автомат дописывает ее с конца к слову из магазинной памяти. В следующий момент времени на автомат поступает вторая буква a_{i_2} слова α. Через b_1 обозначаем последнюю букву слова из магазинной памяти. Если в памяти хранится пустое слово, то $b_1 := \lambda$. По функции φ автомат вычисляет состояние $q_2 := \varphi(a_{i_2}, q_1, b_1)$ в следующий момент времени. После этого автомат по функции ψ вычисляет $\psi(a_{i_2}, q_1, b_1)$. Если это символ стирания, то в слове из магазинной памяти стирается последняя буква b_1. Если это пустой символ, то слово из магазинной памяти не меняется. Если это любая другая буква, то автомат дописывает ее с конца к слову из магазинной памяти. И так далее. В тот момент времени, когда на автомат подается последняя буква a_{i_s} слова α, он вычисляет $\psi(a_{i_s}, q_{s-1}, b_{s-1})$, выполняет соответствующие преобразования со словом из магазинной памяти и заканчивает работу. *Результатом применения автомата V к слову $\alpha \in A^*$* называем

слово из магазинной памяти автомата после окончания его работы. Говорим, что *слово* $\alpha \in A^*$ *распознается автоматом* V, если результат применения автомата V к слову α равен пустому слову. Множество всех принимаемых автоматом V слов в алфавите A обозначаем через $|V|$ и называем *множеством, распознаваемым автоматом* V. Также говорим, что *автомат* V *задает множество* $|V|$.

Введем понятие обобщенного источника. *Обобщенным источником в алфавите* A называем конечный ориентированный граф $G = (V, E)$, у которого выделены начальная и финальная вершины $v, w \in V$ и каждому ребру $\rho \in E$ приписано либо пустое слово Λ, либо буква алфавита A. Допускается наличие в графе петель и кратных ребер. *Путем* в обобщенном источнике $G = (V, E)$ будем называть последовательность

$$\pi = (v_1, \rho_1, v_2, \rho_2, \ldots, \rho_n, v_{n+1}),$$

где $n \in \mathbb{N}$, $v_1, v_2, \ldots, v_{n+1} \in V$ и при всех $i = 1, \ldots, n$ ребро $\rho_i \in E$ ведет из вершины v_i в вершину v_{i+1}. Пути π сопоставляем слово

$$[\pi] := a_1 \ldots a_n,$$

где при всех $i = 1, \ldots n$ через a_i обозначен символ, приписанный ребру ρ_i. Говорим, что путь π *ведет от вершины* v_1 *к вершине* v_{n+1} и пишем $\pi(v_1 \to v_{n+1})$. Пусть $u \in V$ и $\alpha \in A^* \setminus \{\Lambda\}$. Тогда через

$\theta(u, \alpha)$ обозначаем множество

$$\theta(u, \alpha) := \{u' \in E| \text{ существует путь } \pi(u \to u') \text{ такой, что } [\pi] = \alpha.\}$$

Говорим, что обощенный источник G с начальной вершиной v и финальной вершиной w *задает событие* $|G|$

$$|G| := \{\alpha \in A^* \setminus \{\Lambda\} | w \in \theta(v, \alpha)\}.$$

Пусть A, B - конечные алфавиты и $f \in F(A, B)$. Обозначаем через $\mathbb{R}(f)$ множество

$$\mathbb{R}(f) := \left\{ P \in R(A) \middle| (\tilde{f})_P - \text{инъекция} \right\},$$

называемое *классом допустимых регулярных языков для схемы* f. Пусть $f_1, f_2 \in F(A, B)$. Говорим, что f_1 *вкладывается в* f_2 и пишем $f_1 \leq f_2$, если

$$\mathbb{R}(f_1) \subseteq \mathbb{R}(f_2).$$

Проблемой ВКД$_1$ *в алфавитах* A, B называется проблема проверки свойства

$$f_1 \leq f_2$$

для произвольных $f_1, f_2 \in F(A, B)$.

Пусть A, B - конечные алфавиты и $P \in R(A)$. Обозначаем через $\mathbb{F}(P)$ множество

$$\mathbb{F}(P) := \left\{ f \in F(A, B) \middle| (\tilde{f})_P - \text{инъекция} \right\},$$

называемое *классом допустимых схем кодирования для языка* P. Для произвольных $P_1, P_2 \in R(A)$ говорим, что P_1 *вкладывается в* P_2 и пишем $P_1 \leq P_2$, если

$$\mathbb{F}(P_1) \subseteq \mathbb{F}(P_2).$$

Проблемой $ВКД_2$ *в алфавитах* A, B называется проблема проверки свойства

$$P_1 \leq P_2$$

для произвольных $P_1, P_2 \in R(A)$.

Называем *обобщенным алфавитом* $\tilde{A}$ множество

$$\tilde{A} := A \times \{\Lambda\} \cup \{\Lambda\} \times A.$$

Каждому слову в обобщенном алфавите $\tilde{A}$ можно естественным образом сопоставить упорядоченную пару слов в алфавите A в силу того факта, что $\tilde{A} \subset (A \cup \Lambda) \times (A \cup \Lambda)$ и $(A \cup \Lambda)^* = A^*$.

Пусть $f \in F(A, B)$. Называем *отношением синонимии на* f множество $S(f)$:

$$S(f) := \{(\alpha, \beta) \in A^* \times A^* \mid \tilde{f}(\alpha) = \tilde{f}(\beta)\}.$$

Через $\mathbf{K}(A)$ обозначаем множество

$$\mathbf{K}(A) := \{P \subseteq (A \times A)^* \mid P - \text{контекстно-свободно}\}.$$

Пусть A, B - конечные алфавиты и $A = \{a_1, a_2\}$, $\alpha \in A^*$. Через $n_1(\alpha)$ и $n_2(\alpha)$ обозначаем количество букв a_1 и a_2 в слове α

соответственно. Также через $\mathbf{F}_1(A)$ и $\mathbf{F}_2(A)$ обозначаем множества

$$\mathbf{F}_1(A,B) := \{f \in F(A,B) \mid f(a_1), f(a_2) \text{ соизмеримы}\},$$

$$\mathbf{F}_2(A,B) := \{f \in F(A,B) \mid f(a_1), f(a_2) \text{ несоизмеримы}\}.$$

Называем множество $P \subseteq \{a_1, a_2\}^*$ *примитивным*, если для любых слов $\alpha, \beta \in P$ из одновременного выполнения равенств

$$n_1(\alpha) = n_1(\beta),\ n_2(\alpha) = n_2(\beta)$$

следует, что $\alpha = \beta$. Через $T(P)$ обозначаем множество

$$T(P) := \{(n_1(\alpha), n_2(\alpha)) \mid \alpha \in P\}.$$

Пусть $k \in \mathbb{N}$ и $\hat{v}, \hat{v}_1, \ldots, \hat{v}_k \in \mathbb{N}_0 \times \mathbb{N}_0$ - двумерные вектора с неотрицательными целочисленными координатами. Тогда множество

$$L(\hat{v}, V) := \left\{ \hat{v} + \sum_{i=1}^{k} c_i \hat{v}_i \,\middle|\, c_i \in \mathbb{N}_0 \right\}$$

называем *пучком с началом в* $\hat{v}$ *и базисом* $V := \{\hat{v}_1, \ldots, \hat{v}_k\}$.

Пусть $C \subseteq \mathbb{N}_0 \times \mathbb{N}_0$. Через $\mathbb{Q}$ обозначаем множество рациональных чисел. Тогда

$$\mathbb{H}(C) := \{x \in \mathbb{Q} \mid x < 0\} \cap$$

$$\cap \left\{ \frac{u_1 - w_1}{u_2 - w_2} \,\middle|\, (u_1, u_2), (w_1, w_2) \in C, u_2 \neq w_2 \right\}.$$

Называем два вектора $\hat{v}_1, \hat{v}_2 \in \mathbb{N}_0 \times \mathbb{N}_0$ *коллинеарными*, если существует $\alpha \in \mathbb{R}$, для которого $\hat{v}_1 = \alpha \cdot \hat{v}_2$. Если такого α не существует, то называем вектора *неколлинеарными.*

Для $x \in \mathbb{R}$ через $]x[$ обозначаем наименьшее целое число, больше или равное x.

Теорема 6.1 Пусть A, B - конечные алфавиты. Тогда проблема ВКД_1 в алфавитах A, B алгоритмически разрешима.

Теорема 6.2 Пусть A, B - конечные алфавиты и $A = \{a_1, a_2\}$. Тогда проблема ВКД_2 в алфавитах A, B алгоритмически разрешима.

2. Доказательство вспомогательных утверждений

Лемма 1. *Пусть A, B конечные алфавиты и $f \in F(A, B)$. Тогда существует обобщенный источник в обобщенном алфавите $\tilde{A}$, задающий событие $S(f)$.*

Доказательство этого факта приведено в [22].

Лемма 2. *Пусть A - конечный алфавит. Тогда проблема проверки вложения $P_1 \subseteq P_2$ по произвольным $P_1, P_2 \in R(A)$ алгоритмически разрешима.*

Доказательство этого факта приведено в [26].

Лемма 3. *Пусть A - конечный алфавит. Тогда множество*

$\mathbb{CF}(A)$ *контекстно -свободных языков в алфавите* A *совпадает с множеством языков, распознаваемых автоматами из* $\mathbb{PDA}(A)$.

Доказательство этого факта приведено в [36-37].

Лемма 4. *Пусть* $P_1 \in \mathbb{CF}(A)$, $P_2 \in R(A)$ *для некоторого конечного алфавита* A. *Тогда* $P_1 \cap P_2 \in \mathbb{CF}(A)$.

Доказательство этого факта приведено в [36-37].

Лемма 5. *Пусть* $P_1 \in \mathbb{CF}(A)$, $P_2 \in R(A)$ *для некоторого конечного алфавита* A. *Тогда* $P_1 \setminus P_2 \in \mathbb{CF}(A)$.

Доказательство этого факта приведено в [36-37].

Лемма 6. *Пусть* A *- конечный алфавит. Тогда проблема проверки на пустоту для произвольного* $P \in \mathbb{CF}(A)$ *алгоритмически разрешима.*

Доказательство этого факта приведено в [36-37].

Лемма 7. *Пусть* A, B *- конечные алфавиты и даны схемы* $f_1, f_2 \in F(A, B)$. *Тогда* $f_1 \leq f_2$ *если и только если* $S(f_2) \subseteq S(f_1)$.

Доказательство. Пусть $f_1 \leq f_2$. Из определения получаем

$$\mathbb{R}(f_1) \subseteq \mathbb{R}(f_2). \tag{1}$$

Рассмотрим произвольную пару слов $(\alpha, \beta) \in S(f_2)$. Предположим, что $(\alpha, \beta) \notin S(f_1)$. Тогда

$$f_1(\alpha) \neq f_1(\beta), \tag{2}$$

$$f_2(\alpha) = f_2(\beta). \tag{3}$$

Обозначаем через P множество $\{\alpha, \beta\}$. Из (2) и (3) получаем, что $P \in \mathbb{R}(f_1)$ и $P \notin \mathbb{R}(f_2)$. Тогда (1) неверно. Значит $S(f_2) \subseteq S(f_1)$.

Пусть теперь $S(f_2) \subseteq S(f_1)$ и при этом неверно, что $f_1 \leq f_2$, то есть свойство (1) нарушается. Тогда существует $P \in R(A)$, для которого выполнено

$$P \in \mathbb{R}(f_1), \quad P \notin \mathbb{R}(f_2).$$

Так как $P \notin \mathbb{R}(f_2)$, то $(\tilde{f_2})_P$ - не инъекция. Поэтому существуют $\alpha, \beta \in P, \alpha \neq \beta$ такие, что $\tilde{f_2}(\alpha) = \tilde{f_2}(\beta)$. Значит $(\alpha, \beta) \in S(f_2)$. Так как $S(f_2) \subseteq S(f_1)$, то и $(\alpha, \beta) \in S(f_1)$. Итак, $\tilde{f_1}(\alpha) = \tilde{f_1}(\beta)$. Значит $(\tilde{f_1})_P$ - не инъекция, то есть $P \notin \mathbb{R}(f_1)$. Полученное противоречие завершает доказательство леммы.

Лемма 8. *Пусть A - конечный алфавит. Тогда проблема проверки на измеримость по произвольному $P \in R(A)$ алгоритмически разрешима.*

Доказательство. Если множество P пусто, то оно измеримо. Пусть теперь P непусто. Возьмем произвольное $\alpha \in P$. Обозначим через ν минимальное измельчение слова α. Очевидно, что его можно найти. Из доказательства леммы 7 главы 3 следует, что множество P измеримо если и только если

$$P \subseteq \{\nu\}^*. \tag{1}$$

Так как $\{\nu\}^* \in R(A)$, то из леммы 2 следует алгоритмическая разрешимость проверки свойства (1). Утверждение леммы 8 доказано.

Лемма 9. *Пусть A, B - конечные алфавиты, $A = \{a_1, a_2\}$, $f \in F(A, B)$ и γ, δ - непустые попарно различные слова из A^*, для которых $\tilde{f}(\gamma) = \tilde{f}(\delta)$. Тогда $f \in \mathbf{F}_1(A, B)$.*

Доказательство. Будем доказывать утверждение индукцией по длине

$$n := |\tilde{f}(\gamma)| = |\tilde{f}(\delta)|.$$

База индукции. $n = 1$:

Тогда $|\gamma| = |\delta| = 1$. Так как $\gamma \neq \delta$, то без ограничения общности считаем, что $\gamma = a_1$ и $\delta = a_2$. Тогда

$$f(a_1) = \tilde{f}(\gamma) = \tilde{f}(\delta) = f(a_2).$$

Поэтому $f(a_1)$, $f(a_2)$ соизмеримы и $f \in \mathbf{F}_1(A, B)$.

Переход индукции. $n \to n + 1$:

Пусть $|\tilde{f}(\gamma)| = |\tilde{f}(\delta)| = n + 1$. Разбираем два случая.

Случай 1.

Первые буквы слов γ и δ совпадают и без ограничения общности равны a_1. Обозначаем через γ', δ' слова, получающиеся из γ, δ отбрасыванием первой буквы:

$$\gamma := a_1\gamma', \quad \delta := a_1\delta'.$$

Предположим, что слово γ' пустое. Тогда

$$f(a_1) = \tilde{f}(\gamma) = \tilde{f}(\delta) = f(a_1)\tilde{f}(\delta').$$

Значит $\delta' = \Lambda$ и $\delta = a_1$. Это противоречит тому, что $\gamma \neq \delta$. Значит слово γ' не пустое. Аналогично доказывается, что слово δ' не пустое. Так как $\gamma \neq \delta$, то и $\gamma' \neq \delta'$. Слова $\tilde{f}(\gamma'), \tilde{f}(\delta')$ равны, так как получаются из одинаковых слов $\tilde{f}(\gamma)$, $\tilde{f}(\delta)$ откидыванием общего префикса $f(a_1)$. Кроме того,

$$|\tilde{f}(\gamma')| = |\tilde{f}(\delta')| < |\tilde{f}(\gamma)| = |\tilde{f}(\delta)| = n + 1.$$

По предположению индукции, примененному к словам γ', δ', получаем $f \in \mathbf{F}_1(A, B)$. В этом случае переход индукции доказан.

Случай 2.

Первые буквы слов γ и δ различны. Без ограничения общности считаем, что γ начинается с a_1, а δ начинается с a_2. Пусть

$$|f(a_1)| = |f(a_2)|.$$

Так как слова $f(a_1), f(a_2)$ - префиксы слова $\tilde{f}(\gamma) = \tilde{f}(\delta)$, то тогда

$$f(a_1) = f(a_2).$$

Значит $f \in \mathbf{F}_1(A, B)$ и утверждение доказано. Поэтому далее без ограничения общности считаем, что

$$|f(a_1)| > |f(a_2)|.$$

Так как $|\tilde{f}(\gamma) = |\tilde{f}(\delta)|$, то $|\delta| > 1$. Для некоторого $\rho \in B^* \setminus \{\Lambda\}$ имеем

$$f(a_1) := f(a_2)\rho.$$

Определяем новую функцию $\tilde{f}' \in F(A, B)$:

$$\tilde{f}'(a_1) := \rho, \quad \tilde{f}'(a_2) := f(a_2).$$

Заменим в словах γ, δ каждую букву a_1 на слово a_2a_1. Полученные в результате слова обозначаем через γ', δ'. Слово γ' начинается с a_2a_1. А слово δ' начинается с a_2a_2, так как $|\delta| > 1$. Значит $\gamma' \neq \delta'$. Кроме того,

$$\tilde{f}'(\gamma') = \tilde{f}(\gamma) = \tilde{f}(\delta) = \tilde{f}'(\delta'),$$

так как

$$f(a_1) = f(a_2)\rho = \tilde{f}'(a_2)\tilde{f}'(a_1) = \tilde{f}'(a_2a_1) \quad \text{и} \quad f(a_2) = \tilde{f}'(a_2).$$

Теперь убираем из слов γ', δ' общую первую букву a_2. Новые слова обозначаем через γ'', δ''. Они непусты и для них по-прежнему верно, что $\gamma'' \neq \delta''$ и $\tilde{f}'(\gamma'') = \tilde{f}'(\delta'')$. При этом

$$|\tilde{f}'(\gamma'')| = |\tilde{f}'(\delta'')| < |\tilde{f}'(\delta')| = |\tilde{f}(\delta)| = n + 1.$$

Значит к паре слов γ'', δ'' можно применить предположение индукции. Получаем, что $\tilde{f}' \in \mathbf{F}_1(A, B)$. Значит слова $\tilde{f}'(a_1) = \rho$ и $\tilde{f}'(a_2) = f(a_2)$ соизмеримы. Тогда и слова $f(a_2)$, $f(a_2)\rho$ соизмеримы, то есть $f \in \mathbf{F}_1(A, B)$. Переход индукции и утверждение леммы 9 доказаны.

Лемма 10. *Пусть A, B - конечные алфавиты, $A = \{a_1, a_2\}$, $f \in \mathbf{F}_1(A, B)$ и множество $P \in R(A)$ представимо в виде*

$$P = \bigcup_{i=1}^{k} C_{i,1} \cdot (P_{i,1})^* \cdot C_{i,2} \cdot \ldots \cdot C_{i,s(i)-1} \cdot (P_{i,s(i)-1})^* \cdot C_{i,s(i)},$$

где $k, s(1), \ldots, s(k)$ - натуральные числа, все $C_{i,j}, P_{i,j}$ - непустые множества слов в алфавите A и для некоторых

$$1 \leq i_0 \leq k,\ 1 \leq j_0 < s(i_0)$$

множество P_{i_0,j_0} не измеримо. Тогда $f \notin \mathbb{F}(P)$.

Доказательство. Так как $f \in \mathbf{F}_1(A, B)$, то для некоторых чисел $r, m \in \mathbb{N}$ и слова $\nu \in B^* \setminus \{\Lambda\}$ имеем

$$f(a_1) = \nu^r, f(a_2) = \nu^m. \tag{1}$$

Так как множество P_{i_0,j_0} не измеримо, то в нем существует пара различных непустых несоизмеримых слов α, β. Из (1) для них следует, что

$$\tilde{f}(\alpha\beta) = \tilde{f}(\alpha)\tilde{f}(\beta) =$$

$$= (f(a_1))^{n_1(\alpha)}(f(a_2))^{n_2(\alpha)}(f(a_1))^{n_1(\beta)}(f(a_2))^{n_2(\beta)} =$$

$$= (\nu^r)^{n_1(\alpha)}(\nu^m)^{n_2(\alpha)}(\nu^r)^{n_1(\beta)}(\nu^m)^{n_2(\beta)} =$$

$$= \nu^{r\cdot n_1(\alpha)+m\cdot n_2(\alpha)}\nu^{r\cdot n_1(\beta)+m\cdot n_2(\beta)} =$$

$$= \nu^{r\cdot(n_1(\alpha)+n_1(\beta))+m\cdot(n_2(\alpha)+n_2(\beta))} = \nu^{r\cdot n_1(\beta)+m\cdot n_2(\beta)}\nu^{r\cdot n_1(\alpha)+m\cdot n_2(\alpha)} =$$

$$= (f(a_1))^{n_1(\beta)}(f(a_2))^{n_2(\beta)}(f(a_1))^{n_1(\alpha)}(f(a_2))^{n_2(\alpha)} =$$

$$= \tilde{f}(\beta)\tilde{f}(\alpha) = \tilde{f}(\beta\alpha).$$

То есть

$$\tilde{f}(\alpha\beta) = \tilde{f}(\beta\alpha). \tag{2}$$

Докажем, что $\alpha\beta \neq \beta\alpha$. Если бы это было не так, то из леммы 8 при

$$\gamma := a_1a_2,\ \delta := a_2a_1,\ f(a_1) := \alpha,\ f(a_2) := \beta$$

следовало бы, что $f \in \mathbf{F}_1(A)$. Это вступает в противоречие с несоизмеримостью слов α, β. Итак,

$$\alpha\beta \neq \beta\alpha. \tag{3}$$

Для каждого $1 \leq j \leq s(i_0)$ берем по одному произвольному элементу α_j из $C_{i_0,j}$. И для каждого $1 \leq j < s(i_0)$, $j \neq j_0$ берем по одному произвольному элементу β_j из $P_{i_0,j}$. Обозначаем через α', β' слова

$$\alpha' := \alpha_1\beta_1 \ldots \alpha_{j_0-1}\beta_{j_0-1}\alpha_{j_0}\alpha\beta\alpha_{j_0+1}\beta_{j_0+1} \ldots \alpha_{s(i_0)-1}\beta_{s(i_0)}\alpha_{s(i_0)},$$

$$\beta' := \alpha_1\beta_1 \ldots \alpha_{j_0-1}\beta_{j_0-1}\alpha_{j_0}\beta\alpha\alpha_{j_0+1}\beta_{j_0+1} \ldots \alpha_{s(i_0)-1}\beta_{s(i_0)}\alpha_{s(i_0)}.$$

Ясно, что $\alpha', \beta' \in P$. Из (2) следует, что $\tilde{f}(\alpha') = \tilde{f}(\beta')$. Наконец, из (3) получаем $\alpha' \neq \beta'$. Значит $f \notin \mathbb{F}(P)$. Утверждение леммы 10 доказано.

Лемма 11. *Пусть A, B - конечные алфавиты, $A = \{a_1, a_2\}$, $P \in R(A)$ - не примитивный язык, $f \in F(A, B)$ и $f \in \mathbb{F}(P)$. Тогда $f \in \mathbf{F}_2(A, B)$.*

Доказательство. Предположим, что $f \notin \mathbf{F}_2(A, B)$. Тогда для некоторых чисел $k, m \in \mathbb{N}$ и слова $\nu \in B^* \setminus \{\Lambda\}$ получаем

$$f(a_1) = \nu^k,\ f(a_2) = \nu^m.$$

Так как $P \in R(A)$ - не примитивный язык, то существуют слова $\alpha, \beta \in P$, для которых $n_1(\alpha) = n_1(\beta)$, $n_2(\alpha) = n_2(\beta)$ и $\alpha \neq \beta$. Тогда

$$\tilde{f}(\alpha) = (f(a_1))^{n_1(\alpha)}(f(a_2))^{n_2(\alpha)} = (f(a_1))^{n_1(\beta)}(f(a_2))^{n_2(\beta)} = \tilde{f}(\beta).$$

Значит $f \notin \mathbb{F}(P)$. Полученное противоречие завершает доказательство леммы.

Лемма 12. *Пусть A - конечный алфавит, $A = \{a_1, a_2\}$. Тогда для множества*

$$\mathbf{C} := \{(\alpha, \beta) \in A^* \times A^* \mid n_1(\alpha) = n_1(\beta), n_2(\alpha) = n_2(\beta)\}$$

верно, что $\mathbf{C} \in \mathbf{K}(A)$.

Доказательство. Из леммы 3 следует, что для доказательства леммы достаточно построить автомат $V \in \mathbb{PDA}(A \times A)$, задающий множество $\mathbf{C}$. Входным алфавитом, алфавитом состояний и алфавитом стека в V будут соответственно множества $A \times A$, $\{q_1, q_2, q_3\}$, $\{1, \lambda\}$. Состояние q_1 соответствует тому, что в уже обработанном начале входного слова одинаковое количество букв a_1 и a_2 в первой и второй позициях. Состояние q_2 соответствует тому, что в уже обработанном начале входного слова в первой позиции больше

букв a_1, чем во второй позиции. Состояние q_3 соответствует тому, что в уже обработанном начале входного слова в первой позиции больше букв a_2, чем во второй позиции. В начальный момент времени автомат находится в состоянии q_1 и стек пуст. Опишем функционирование автомата.

текущее состояние ленты	текущий символ ленты	последний символ в стеке	новое состояние	действие над стеком
q_1, q_2 или q_3	(a_1, a_1)	пусто	q_1	ничего
q_1, q_2 или q_3	(a_1, a_2)	пусто	q_2	пишем 1
q_1, q_2 или q_3	(a_2, a_1)	пусто	q_3	пишем 1
q_1, q_2 или q_3	(a_2, a_2)	пусто	q_1	ничего
q_2	(a_1, a_1)	1	q_2	ничего
q_2	(a_1, a_2)	1	q_2	пишем 1
q_2	(a_2, a_1)	1	q_2	стираем 1
q_2	(a_2, a_2)	1	q_2	ничего
q_3	(a_1, a_1)	1	q_3	ничего
q_3	(a_1, a_2)	1	q_3	стираем 1
q_3	(a_2, a_1)	1	q_3	пишем 1
q_3	(a_2, a_2)	1	q_3	ничего

Автомат следит за разницей для уже обработанной части входного

слова между количеством букв a_1 в первой и второй позициях и хранит эту разницу в стеке. Ясно, что стек автомата пуст тогда и только тогда, когда разница равна 0. Значит автомат функционирует правильно и распознает те и только те пары слов, в которых одинаковое количество букв a_1 и одинаковое количество букв a_2. Утверждение леммы 12 доказано.

Лемма 13. *Пусть A, B - конечные алфавиты, $A = \{a_1, a_2\}$, есть схема $f \in \mathbf{F}_1(A, B)$, $P \in \mathbb{R}(A)$ - примитивный язык и пусть для некоторых чисел $k, m \in \mathbb{N}$ и слова $\nu \in B^* \setminus \{\Lambda\}$ верно, что $f(a_1) = \nu^k$, $f(a_2) = \nu^m$. Тогда $f \notin \mathbb{F}(P)$ если и только если существует такая пара слов $\alpha, \beta \in P$, для которой выполнено*

$$\frac{n_1(\alpha) - n_1(\beta)}{n_2(\alpha) - n_2(\beta)} = -\frac{m}{k}.$$

Доказательство. Пусть для некоторых слов $\alpha, \beta \in P$ выполнено равенство

$$\frac{n_1(\alpha) - n_1(\beta)}{n_2(\alpha) - n_2(\beta)} = -\frac{m}{k}. \tag{1}$$

Отсюда следут, что $\alpha \neq \beta$ и

$$n_1(\alpha)k + n_2(\alpha)m = n_1(\beta)k + n_2(\beta)m. \tag{2}$$

Так как $f(a_1) = \nu^k$, $f(a_2) = \nu^m$, то из (2) получаем

$$\tilde{f}(\alpha) = (f(a_1))^{n_1(\alpha)}(f(a_2))^{n_2(\alpha)} = (\nu^k)^{n_1(\alpha)}(\nu^m)^{n_2(\alpha)} =$$

$$= \nu^{n_1(\alpha)k+n_2(\alpha)m} = \nu^{n_1(\beta)k+n_2(\beta)m} = (\nu^k)^{n_1(\beta)}(\nu^m)^{n_2(\beta)} =$$

$$= (f(a_1))^{n_1(\beta)}(f(a_2))^{n_2(\beta)} = \tilde{f}(\beta).$$

Но $\alpha \neq \beta$. Значит $f \notin \mathbb{F}(P)$.

Пусть теперь $f \notin \mathbb{F}(P)$. Тогда существуют слова $\alpha, \beta \in P$, $\alpha \neq \beta$, для которых $\tilde{f}(\alpha) = \tilde{f}(\beta)$. Значит

$$\nu^{n_1(\alpha)k+n_2(\alpha)m} = (\nu^k)^{n_1(\alpha)}(\nu^m)^{n_2(\alpha)} = (f(a_1))^{n_1(\alpha)}(f(a_2))^{n_2(\alpha)} =$$

$$= \tilde{f}(\alpha) = \tilde{f}(\beta) = (f(a_1))^{n_1(\beta)}(f(a_2))^{n_2(\beta)} =$$

$$= (\nu^k)^{n_1(\beta)}(\nu^m)^{n_2(\beta)} = \nu^{n_1(\beta)k+n_2(\beta)m}.$$

Поэтому

$$(n_1(\alpha) - n_1(\beta))k = -(n_2(\alpha) - n_2(\beta))m.$$

Если $n_2(\alpha) - n_2(\beta) = 0$, то и $n_1(\alpha) - n_1(\beta) = 0$. Это противоречит примитивности языка P. Значит $n_2(\alpha) - n_2(\beta) \neq 0$ и

$$\frac{n_1(\alpha) - n_1(\beta)}{n_2(\alpha) - n_2(\beta)} = -\frac{m}{k}.$$

Утверждение леммы 13 доказано.

Лемма 14. *Пусть A, B - конечные алфавиты, $A = \{a_1, a_2\}$, $f \in \mathbf{F}_2(A, B)$ и $P \in R(A)$. Тогда $f \in \mathbb{F}(P)$.*

Доказательство. Допустим, что $f \notin \mathbb{F}(P)$. Тогда для некоторых слов попарно различных слов $\alpha, \beta \in P$ верно $\tilde{f}(\alpha) = \tilde{f}(\beta)$. Тогда по лемме 9 получаем $f \in \mathbf{F}_1(A, B)$. Но это невозможно. Полученное противоречие завершает доказательство леммы.

Лемма 15. *Пусть A - конечный алфавит. Тогда проблема проверки на примитивность по произвольному $P \in R(A)$ алгоритмически разрешима.*

Доказательство. Пусть $P \in R(A)$. Вводим обозначения:

$$\mathbf{C} := \{(\alpha, \beta) \in A^* \times A^* \mid n_1(\alpha) = n_1(\beta), n_2(\alpha) = n_2(\beta)\},$$

$$\mathbf{I} := \{(\alpha, \alpha) \mid \alpha \in A^*\}.$$

Так как

$$(P \times P) \cap \mathbf{C} = \{(\alpha, \beta) \in P \times P \mid n_1(\alpha) = n_1(\beta), n_2(\alpha) = n_2(\beta)\},$$

то P примитивно если и только если

$$(P \times P) \cap \mathbf{C} \subseteq \mathbf{I}. \tag{1}$$

Поэтому исходная проблема может быть сведена к проверке на пустоту множества (1). Из леммы 12 следует, что $\mathbf{C} \in \mathbf{K}(A)$. Далее, так как выполнено $P \times P \in R(A \times A)$, то из леммы 4 получаем $(P \times P) \cap \mathbf{C} \in \mathbf{K}(A)$. Но $\mathbf{I} \in R(A \times A)$. Отсюда и из леммы 5 заключаем $((P \times P) \cap \mathbf{C}) \setminus \mathbf{I} \in \mathbf{K}(A)$. Для завершения доказательства осталось воспользоваться леммой 6. Утверждение леммы 15 доказано.

Лемма 16. *Пусть L - пучок с началом $\hat{v}$ и базисом V. Пусть $\hat{v}_1, \hat{v}_2 \in V$ - пара неколлинеарных векторов. Тогда*

$$\mathbb{H}(L) = \{x \in \mathbb{Q} \mid x < 0\}.$$

Доказательство. Очевидно, что $\mathbb{H}(L) \subseteq \{x \in \mathbb{Q} \mid x < 0\}$. Покажем теперь, что для любых $k, l \in \mathbb{N}$ верно

$$-\frac{k}{l} \in \mathbb{H}(L). \tag{1}$$

Вводим обозначения:

$$(x_0, y_0) := \hat{v},\ (x_1, y_1) := \hat{v}_1,\ (x_2, y_2) := \hat{v}_2,$$

$$(u_1, u_2) = \hat{u} := \hat{v} + (ky_2 + lx_2)\hat{v}_1,$$

$$(w_1, w_2) = \hat{w} := \hat{v} + (lx_1 + ky_1)\hat{v}_2.$$

Очевидно, что $\hat{u}, \hat{w} \in L$. Покажем, что $u_2 \neq w_2$.

$$u_2 - w_2 = (y_0 + ky_2y_1 + lx_2y_1) - (y_0 + lx_1y_2 + ky_1y_2) = l(x_2y_1 - x_1y_2).$$

Рассмотрим семь случаев.

1. $x_1 = x_2 = 0$. Тогда вектора $\hat{v}_1, \hat{v}_2 \in V$ коллинеарны, а это невозможно.

2. $x_1 = y_1 = 0$. Тогда вектора $\hat{v}_1, \hat{v}_2 \in V$ коллинеарны, а это невозможно.

3. $x_1 = 0$, $y_1 \neq 0$, $x_2 \neq 0$. Тогда $l(x_2y_1 - x_1y_2) \neq 0$.

4. $x_1 \neq 0$, $y_1 = y_2 = 0$. Тогда вектора $\hat{v}_1, \hat{v}_2 \in V$ коллинеарны, а это невозможно.

5. $x_1 \neq 0$, $y_1 = 0$, $y_2 \neq 0$, $x_2 = 0$. Тогда $l(x_2y_1 - x_1y_2) \neq 0$.

6. $x_1 \neq 0$, $y_1 = 0$, $y_2 \neq 0$, $x_2 \neq 0$. Так как вектора $\hat{v}_1, \hat{v}_2 \in V$ не

коллинеарны, то $\frac{x_1}{x_2} \neq \frac{y_1}{y_2}$. Значит $l(x_2y_1 - x_1y_2) \neq 0$.

7. $x_1 \neq 0$, $y_1 \neq 0$. Так как вектора $\hat{v}_1, \hat{v}_2 \in V$ не коллинеарны, то $\frac{x_2}{x_1} \neq \frac{y_2}{y_1}$. Значит $l(x_2y_1 - x_1y_2) \neq 0$.

Итак, мы показали, что $u_2 \neq w_2$. Далее

$$u_1 - w_1 = (x_0 + ky_2x_1 + lx_2x_1) - (x_0 + lx_1x_2 + ky_1x_2) = k(y_2x_1 - x_2y_1).$$

Значит

$$\frac{u_1 - w_1}{u_2 - w_2} = \frac{k(y_2x_1 - x_2y_1)}{l(x_2y_1 - x_1y_2)} = -\frac{k}{l}.$$

Свойство (1) выполнено. Утверждение леммы 16 доказано.

Лемма 17. *Пусть* $k \in \mathbb{N}$ *и* $\hat{v}_1, \ldots, \hat{v}_k \in \mathbb{N}_0 \times \mathbb{N}_0$ - *семейство попарно коллинеарных векторов. Тогда существуют* $\hat{v} \in \mathbb{N}_0 \times \mathbb{N}_0$ *и* $n_1, \ldots, n_k \in \mathbb{N}_0$ *такие, что*

$$\hat{v}_i = n_i\hat{v} \qquad \textit{при} \;\; 1 \leq i \leq k.$$

Доказательство. Если в семействе $\hat{v}_1, \ldots, \hat{v}_k$ есть нулевые векторы $\hat{v}_i = (0, 0)$, то для них можно положить $n_i := 0$. Считаем теперь, что все $\hat{v}_i$ не равны $(0, 0)$. Здесь и далее везде, где это не обговорено отдельно, подразумеваем, что приводимые равенства выполнены при всех $1 \leq i \leq k$. Обозначаем через x_i, y_i координаты вектора v_i. Возможны три случая.

Случай 1. $x_1 = 0$. Тогда в силу попарной коллинеарности векторов

v_i получаем $x_i = 0$. Положим

$$\hat{v} := (0, 1),$$

$$n_i := y_i.$$

Очевидно, что

$$\hat{v}_i = (0, y_i) = (0, n_i) = n_i \hat{v}.$$

Случай 2. $y_1 = 0$. Этот случай сводится к случаю 1 перестановкой координат.

Случай 3. $x_1 \neq 0, y_1 \neq 0$. Так как вектора v_i попарно коллинеарны, то для некоторых чисел $c_i \in \mathbb{R}$ верно

$$(x_i, y_i) = (c_i x_1, c_i y_1). \tag{1}$$

Так как $\hat{v}_i \neq (0, 0)$, то $c_i \neq 0$. Тогда $c_i = \frac{x_i}{x_1} \in \mathbb{Q}$ и существуют $m_i, l_i \in \mathbb{N}$, для которых

$$c_i = \frac{m_i}{l_i}, \quad \text{НОД}(m_i, l_i) = 1. \tag{2}$$

Из (1) и (2) следует, что числа x_1 и y_1 делятся нацело на l_i. Значит и число НОД(x_1, y_1) делится нацело на l_i. Положим

$$\hat{v} := \left(\frac{x_1}{\text{НОД}(x_1, y_1)}, \frac{y_1}{\text{НОД}(x_1, y_1)} \right)$$

и

$$n_i := \frac{m_i \cdot \ \text{НОД}(x_1, y_1)}{l_i}.$$

Как было показано выше, $n_i \in \mathbb{N}$. Осталось заметить, что

$$\hat{v}_i = (x_i, y_i) = (\frac{m_i}{l_i} x_1, \frac{m_i}{l_i} y_1) =$$

$$= \left(\frac{m_i \cdot \ \text{НОД}(x_1, y_1) \cdot x_1}{l_i \cdot \text{НОД}(x_1, y_1)}, \frac{m_i \cdot \ \text{НОД}(x_1, y_1) \cdot y_1}{l_i \cdot \text{НОД}(x_1, y_1)} \right) = n_i \hat{v}.$$

Разбор случаев завершен. Утверждение леммы 17 доказано.

Лемма 18. *Пусть* $k \in \mathbb{N}$ *и* $\hat{v}_1, \ldots, \hat{v}_k \in \mathbb{N}_0 \times \mathbb{N}_0$ - *семейство попарно коллинеарных векторов, ни один из которых не равен* $(0,0)$. *Тогда существуют такие* $c_0 \in \mathbb{R}$, $\hat{u} \in (\mathbb{N}_0 \times \mathbb{N}_0) \setminus (0,0)$ *и* $D \subseteq \{i \in \mathbb{N} \mid 1 \leq i < c_0\}$, *для которых выполнено*

$$\left\{ \sum_{i=1}^{k} c_i \hat{v}_i \,\middle|\, c_i \in \mathbb{N}_0 \right\} = \{n\hat{u} \mid n \in D\} \cup \{n\hat{u} \mid n \geq c_0\}.$$

Доказательство. Ниже, как и при доказательстве леммы 17, считаем, что приводимые равенства выполнены при $1 \leq i \leq k$. Из леммы 17 следует существование векторов $\hat{v} \in \mathbb{N}_0 \times \mathbb{N}_0$ и чисел $n_i \in \mathbb{N}_0$, для которых

$$\hat{v}_i = n_i \hat{v}. \tag{1}$$

Так как $\hat{v}_i \neq (0,0)$, то $n_i \neq 0$.

Вводим обозначения:

$$L := \left\{ \sum_{i=1}^{k} c_i \hat{v}_i \,\middle|\, c_i \in \mathbb{N}_0 \right\},$$

$$H := \{c_1 \cdot n_1 + \ldots + c_k \cdot n_k \mid c_1, \ldots, c_k \in \mathbb{N}_0\},$$

$$r := \text{НОД}(n_1, \ldots, n_k).$$

Так как выполнено (1), то

$$L = \left\{ \sum_{i=1}^{k} c_i n_i \hat{v} \,\middle|\, c_i \in \mathbb{N}_0 \right\} = \{h\hat{v} \mid h \in H\}. \tag{2}$$

Из леммы 3 главы 2 следует существование числа $n_0 \in \mathbb{N}$ такого, что

$$Z_{n_0,r} \subseteq H.$$

С другой стороны, если $h \in H$, то h делится нацело на r. Поэтому

$$\{h \in H \mid h \geq n_0\} = Z_{n_0,r} \tag{3}$$

и

$$\left\{ \frac{h}{r} \middle| h \in H \right\} \in \mathbb{N}. \tag{4}$$

Теперь мы готовы к тому, чтобы задать искомые $c_0, \hat{u}, D$ из формулировки леммы:

$$c_0 := \frac{n_0}{r}, \quad \hat{u} := r\hat{v}, \quad D := \left\{ \frac{h}{r} \middle| h \in H, h < n_0 \right\}.$$

Из (4) получаем, что

$$D \subseteq \{i \in \mathbb{N} \mid 1 \leq i < c_0\}.$$

Так как выполнено (3), то

$$L = \{h\hat{v} \mid h \in H\} = \{h\hat{v} \mid h \in H, h < n_0\} \cup \{h\hat{v} \mid h \in H, h \geq n_0\} =$$

$$= \left\{ \frac{h}{r} r\hat{v} \mid h \in H, h < n_0 \right\} \cup \{nr\hat{v} \mid nr \geq n_0\} =$$

$$= \{nr\hat{v} \mid n \in D\} \cup \left\{ nr\hat{v} \mid n \geq \frac{n_0}{r} \right\} = \{n\hat{u} \mid n \in D\} \cup \{n\hat{u} \mid n \geq c_0\}.$$

Утверждение леммы 18 доказано.

Лемма 19. *Пусть L - пучок с началом $\hat{v}$ и базисом V, в котором любая пара векторов $\hat{v}_1, \hat{v}_2 \in V$ коллинеарна. Тогда L представим*

в виде конечного объединения

$$L = \bigcup_{i=1}^{k} L_i,$$

где $k \in \mathbb{N}$ *и все* L_i *- пучки с одноэлементными базисами.*

Доказательство. Если базис V одноэлементен, то утверждение очевидно. Теперь разберем случай, когда $|V| \geq 2$. Если в базисе V есть нулевой вектор $(0,0)$, то его можно оттуда выкинуть, получив при этом тот же самый пучок L. Поэтому будем далее считать, что $(0,0) \notin V$. В силу леммы 18 существуют такие число $c_0 \in \mathbb{R}$, вектор $\hat{u} \in (\mathbb{N}_0 \times \mathbb{N}_0) \setminus (0,0)$ и множество $D \subseteq \{i \in \mathbb{N} \mid 1 \leq i < c_0\}$, для которых выполнено

$$\left\{ \sum_{i=1}^{k} c_i \hat{v}_i \,\middle|\, c_i \in \mathbb{N}_0 \right\} = \{n\hat{u} \mid n \in D\} \cup \{n\hat{u} \mid n \geq c_0\}.$$

Поэтому

$$L = \{\hat{v} + n\hat{u} \mid n \in D\} \cup \{\hat{v} + n\hat{u} \mid n \geq c_0\}. \tag{1}$$

Ясно, что множество

$$L_0 := \{\hat{v} + n\hat{u} \mid n \geq c_0\}$$

является пучком с началом $\hat{v}+]c_0[\hat{u}$ и одноэлементным базисом $\{\hat{u}\}$. Обозначаем через m мощность множества D. Если $m = 0$, то $L = L_0$, $k = 1$ и утверждение доказано. Пусть теперь $m \in \mathbb{N}$. Упорядочиваем элементы множества D:

$$D = \{d_1, d_2, \ldots, d_m\}.$$

Для всех $1 \leq i \leq m$ через L_i обозначаем пучок с началом $\hat{v} + d_i\hat{u}$ и одноэлементным базисом $\{(0,0)\}$. Тогда из (1) получаем

$$L = \bigcup_{i=1}^{m} L_i \cup L_0.$$

Утверждение леммы 19 доказано.

Лемма 20. *Пусть L_1, L_2 - бесконечные пучки с одноэлементными базисами $\{v_1\}$, $\{v_2\}$ соответственно, векторы v_1, v_2 коллинеарны и $L_1 \cap L_2 \neq \emptyset$. Тогда $L_1 \cup L_2$ представимо в виде конечного объединения*

$$\bigsqcup_{i=1}^{k} L_i$$

попарно непересекающихся пучков с одноэлементными базисами.

Доказательство. Пусть $(a_1, b_1), (c_1, d_1)$ - соответственно начала пучков L_1, L_2 и

$$v_1 := (a_2, b_2),\ v_2 := (c_2, d_2). \tag{1}$$

Так как пучки бесконечные, то

$$(a_2, b_2) \neq (0,0),\ (c_2, d_2) \neq (0,0).$$

Без ограничения общности считаем $a_2 \neq 0$. Из коллинерности векторов в (1) получаем, что и $c_2 \neq 0$. Пусть (x, y) - такая точка из $L_1 \cap L_2$, для которой значение $x + y$ минимально. Опять же, из коллинеарности векторов в (1) имеем

$$\frac{\text{НОК}(a_2, c_2)}{a_2}(a_2, b_2) = \left(\text{НОК}(a_2, c_2), \text{НОК}(a_2, c_2)\frac{b_2}{a_2}\right) =$$

$$= \left(\text{НОК}(a_2, c_2), \text{НОК}(a_2, c_2)\frac{d_2}{c_2}\right) = \frac{\text{НОК}(a_2, c_2)}{c_2}(c_2, d_2)$$

и

$$L_1 \cap L_2 = \left\{ \left(x + k \cdot \text{НОК}(a_2, c_2), y + k \cdot \text{НОК}(a_2, c_2)\frac{b_2}{a_2} \middle| \, k \in \mathbb{N}_0 \right) \right\}.$$

Значит $L_1 \cap L_2$ - пучок с началом (x, y) и одноэлементным базисом $\{v\}$, где

$$v := \left(\text{НОК}(a_2, c_2), \text{НОК}(a_2, c_2)\frac{b_2}{a_2}\right).$$

Далее множество $L_1 \backslash L_2$ разлагаем в конечное объединение попарно непересекающихся множеств вида

$$\left\{ (x + a_2 i, y + b_2 i) + k \cdot \text{НОК}(a_2, c_2) \left(1, \frac{b_2}{a_2}\right) \middle| \, k \in \mathbb{N}_0 \right\},$$

где $1 \leq i < \frac{\text{НОК}(a_2, c_2)}{a_2}$, и еще, возможно, одноэлементных множеств вида

$$\{(x_0, y_0) \in L_1 | x_0 < x, y_0 < y\}.$$

Очевидно, что все эти множества - пучки с одноэлементным базисом (для одноэлементных множеств этим базисным элементом будет $(0, 0)$). Аналогично поступаем и с $L_2 \setminus L_1$. Утверждение леммы 20 доказано.

Лемма 21. *Пусть* $a, b \in \mathbb{N}$, $a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2 \in \mathbb{N}_0$, $(a_2, b_2) \neq (0, 0)$, $(c_2, d_2) \neq (0, 0)$, L_1 - *пучок с началом* (a_1, b_1) *и базисом* $\{(a_2, b_2)\}$, L_2 - *пучок с началом* (c_1, d_1) *и базисом* $\{(c_2, d_2)\}$

и $L_1 \cap L_2 = \emptyset$. *Тогда* $-\frac{a}{b} \in \mathbb{H}(L_1 \cup L_2)$ *если и только если* $bc_1 - ba_1 + ad_1 - ab_1$ *делится нацело на НОД*$(ab_2 + ba_2, ad_2 + bc_2)$.

Доказательство. Пусть для некоторых $a, b \in \mathbb{N}$ выполнено

$$-\frac{a}{b} \in \mathbb{H}(L_1 \cup L_2).$$

Тогда существуют векторы

$$(s_1, s_2), (s_3, s_4) \in L_1 \cup L_2$$

такие, что $s_2 \neq s_4$ и

$$\frac{s_1 - s_3}{s_2 - s_4} = -\frac{a}{b}. \tag{1}$$

Разбираем три случая.

Случай 1. $(s_1, s_2), (s_3, s_4) \in L_1$. Тогда существуют $m, n \in \mathbb{N}_0$, для которых выполнено

$$(s_1, s_2) = (a_1 + ma_2, b_1 + mb_2), \tag{2}$$

$$(s_3, s_4) = (a_1 + na_2, b_1 + nb_2). \tag{3}$$

Подставляем (2) и (3) в (1):

$$-\frac{a}{b} = \frac{s_1 - s_3}{s_2 - s_4} = \frac{(a_1 + ma_2) - (a_1 + na_2)}{(b_1 + mb_2) - (b_1 + nb_2)} = \frac{(m-n)a_2}{(m-n)b_2} = \frac{a_2}{b_2} \geq 0.$$

Поэтому этот случай невозможен.

Случай 2. $(s_1, s_2), (s_3, s_4) \in L_2$. Этот случай полностью аналогичен предыдущему.

Случай 3. $(s_1, s_2) \in L_1, (s_3, s_4) \in L_2$. Тогда существуют $m, n \in \mathbb{N}_0$ такие, что

$$(s_1, s_2) = (a_1 + ma_2, b_1 + mb_2), \tag{4}$$

$$(s_3, s_4) = (c_1 + nc_2, d_1 + nd_2). \tag{5}$$

Подставляем (4) и (5) в (1):

$$-\frac{a}{b} = \frac{s_1 - s_3}{s_2 - s_4} = \frac{(a_1 + ma_2) - (c_1 + nc_2)}{(b_1 + mb_2) - (d_1 + nd_2)}.$$

Тогда

$$-a(b_1 - d_1 - nd_2 + mb_2) = b(a_1 - c_1 - nc_2 + ma_2),$$

то есть

$$m(ab_2 + ba_2) - n(ad_2 + bc_2) = bc_1 - ba_1 + ad_1 - ab_1.$$

Отсюда выводим, что $bc_1 - ba_1 + ad_1 - ab_1$ делится нацело на НОД$(ab_2 + ba_2, ad_2 + bc_2)$.

Докажем теперь обратное утверждение. Пусть для некоторых $a, b \in \mathbb{N}$ верно, что число $bc_1 - ba_1 + ad_1 - ab_1$ делится нацело на НОД$(ab_2 + ba_2, ad_2 + bc_2)$, то есть

$$bc_1 - ba_1 + ad_1 - ab_1 = k \cdot \text{НОД}(ab_2 + ba_2, ad_2 + bc_2) \tag{6}$$

для некоторого $k \in \mathbb{Z}$. Так как

$$(a_2, b_2) \neq (0, 0),\ (c_2, d_2) \neq (0, 0),$$

то $ab_2 + ba_2 \in \mathbb{N}$, $ad_2 + bc_2 \in \mathbb{N}$. Из (6) и расширенного алгоритма Евклида следует существование $m, n \in \mathbb{N}_0$ таких, что

$$m(ab_2 + ba_2) - n(ad_2 + bc_2) = bc_1 - ba_1 + ad_1 - ab_1.$$

Отсюда получаем

$$-a(b_1 - d_1 - nd_2 + mb_2) = b(a_1 - c_1 - nc_2 + ma_2). \qquad (7)$$

Если $b_1 - d_1 - nd_2 + mb_2 = 0$, то выполнено $a_1 - c_1 - nc_2 + ma_2 = 0$ и $b_1 + mb_2 = d_1 + nd_2$, $a_1 + ma_2 = c_1 + nc_2$. Но

$$(a_1 + ma_2, b_1 + mb_2) \in L_1,\ (c_1 + nc_2, d_1 + nd_2) \in L_2. \qquad (8)$$

Это противоречит тому, что $L_1 \cap L_2 = \emptyset$. Значит на самом деле $b_1 - d_1 - nd_2 + mb_2 \neq 0$ и равенство (7) можно записать иначе:

$$-\frac{a}{b} = \frac{a_1 - c_1 - nc_2 + ma_2}{b_1 - d_1 - nd_2 + mb_2} = \frac{(a_1 + ma_2) - (c_1 + nc_2)}{(b_1 + mb_2) - (d_1 + nd_2)}.$$

Из (8) теперь следует, что $-\frac{a}{b} \in \mathbb{H}(L_1 \cup L_2)$. Утверждение леммы 21 доказано.

Лемма 22. *Пусть* $a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2 \in \mathbb{N}_0$, $(a_2, b_2) \neq (0, 0)$ *и* $(c_2, d_2) \neq (0, 0)$. *Тогда существуют числа* $k \in \mathbb{N}$, $s \in \mathbb{N}_0$, $e_1, \ldots, e_k \in \mathbb{N}_0$, $f_1 \ldots, f_k \in \mathbb{N}_0$, $q_1, q_2 \in \{0, 1\}$ *такие, что для любой пары взаимно простых* $a, b \in \mathbb{N}$ *число* $b(c_1 - a_1) + a(d_1 - b_1)$ *делится нацело на число НОД*$(ab_2 + ba_2, ad_2 + bc_2)$ *в том и только в том случае, когда*

$$(a, b) \in \bigcup_{i=1}^{k} \left\{ (msq_1 + e_i, nsq_2 + f_i) \,\middle|\, m, n \in \mathbb{N}_0 \right\}.$$

Доказательство. Во всех приводимых в доказательстве выражениях с n и m считаем, что это произвольные взаимно простые

числа из $\mathbb{N}$. Случай, когда $b_2 d_2 = 0$, разберем в самом конце. Пока же будем исходить из того, что $b_2 \neq 0$ и $d_2 \neq 0$. Тогда к паре чисел b_2, d_2 можно применить алгоритм Евклида. В результате получаем, что для некоторых $\hat{a}_2, \hat{c}_2 \in \mathbb{Z}$ выполнено равенство

$$\text{НОД}(ab_2 + ba_2, ad_2 + bc_2) = \text{НОД}(a\hat{b}_2 + b\hat{a}_2, a\hat{b}_2 + b\hat{c}_2), \qquad (1)$$

где $\hat{b}_2 := \text{НОД}(b_2, d_2)$. Возможны два случая.

Случай 1. $\hat{a}_2 = \hat{c}_2$. Тогда из (1) получаем:

$$\text{НОД}(ab_2 + ba_2, ad_2 + bc_2) = a\hat{b}_2 + b\hat{a}_2. \qquad (2)$$

Так как $a_2, b_2 \in \mathbb{N}_0$, $\hat{b}_2 = \text{НОД}(b_2, d_2)$ и (2) верно для любых чисел $a, b \in \mathbb{N}$, то

$$\hat{a}_2 \in \mathbb{N}_0, \hat{b}_2 \in \mathbb{N}. \qquad (3)$$

Рассматриваем два подслучая.

Случай 1.1. $d_1 - b_1 = k\hat{b}_2$, $c_1 - a_1 = k\hat{a}_2$ для некоторого $k \in \mathbb{Z}$. Тогда число

$$b(c_1 - a_1) + a(d_1 - b_1) \qquad (4)$$

делится нацело на (2) при любых натуральных a, b и в утверждении леммы можно положить

$$q_1 := 1,\ q_2 := 1,\ s := 1,\ k := 1,\ e_1 := 0,\ f_1 := 0.$$

Случай 1.2. Не существует такого числа $k \in \mathbb{Z}$, для которого одновременно выполнено $d_1 - b_1 = k\hat{b}_2$ и $c_1 - a_1 = k\hat{a}_2$. Тогда (4) делится

нацело на (2) если и только если существует $s \in \mathbb{Z}$, для которого имеем

$$b(c_1 - a_1) + a(d_1 - b_1) = s(a\hat{b}_2 + b\hat{a}_2),$$

то есть

$$a(d_1 - b_1 - s\hat{b}_2) + b(c_1 - a_1 - s\hat{a}_2) = 0. \tag{5}$$

Если $d_1 - b_1 - s\hat{b}_2 = 0$, то из условия разбираемого подслучая следует, что $c_1 - a_1 \neq s\hat{a}_2$, то есть (5) не выполнено ни для каких $a, b \in \mathbb{N}$. Тогда в утверждении леммы можно положить

$$q_1 := 0,\ q_2 := 0,\ s := 0,\ k := 1,\ e_1 := 0,\ f_1 := 0.$$

Если же $d_1 - b_1 - s\hat{b}_2 \neq 0$, то (5) можно переписать в виде

$$\frac{a}{b} = \frac{s\hat{a}_2 - c_1 + a_1}{d_1 - b_1 - s\hat{b}_2}. \tag{6}$$

Если $\hat{a}_2 \neq 0$, то из (3) получаем, что неравенсво $\frac{s\hat{a}_2 - c_1 + a_1}{d_1 - b_1 - s\hat{b}_2} > 0$ выполнено лишь для конечного количества значений $s \in \mathbb{Z}$ и в утверждении леммы нужно положить

$$q_1 := 0,\ q_2 := 0.$$

Остальные константы подбираются тривиально.

Если же $\hat{a}_2 = 0$, то (6) принимает вид

$$\frac{a}{b} = \frac{a_1 - c_1}{d_1 - b_1 - s\hat{b}_2}$$

и в утверждении леммы нужно положить

$$q_1 := 0,\ q_2 := 1.$$

Остальные константы подбираются тривиально.

Случай 2. $\hat{a}_2 \neq \hat{c}_2$. Тогда из (1) получаем:

$$\text{НОД}(ab_2 + ba_2, ad_2 + bc_2) = \text{НОД}(b(\hat{a}_2 - \hat{c}_2), a\hat{b}_2 + b\hat{c}_2). \qquad (7)$$

Пусть для некоторых $a, b \in \mathbb{N}$ верно, что $b(c_1 - a_1) + a(d_1 - b_1)$ делится нацело на (7). Через m обозначаем число $\text{НОД}(b, \hat{b}_2)$. Пусть

$$b := m \cdot m', \ \hat{b}_2 = m \cdot m''.$$

Так как $\text{НОД}(m', m'') = 1$, то

$$\text{НОД}(b(\hat{a}_2 - \hat{c}_2), a\hat{b}_2 + b\hat{c}_2) = \text{НОД}(mm'(\hat{a}_2 - \hat{c}_2), amm'' + mm'\hat{c}_2) =$$

$$= m\text{НОД}(m'(\hat{a}_2 - \hat{c}_2), am'' + m'\hat{c}_2) =$$

$$= m\text{НОД}(\hat{a}_2 - \hat{c}_2, am'' + m'\hat{c}_2) \qquad (8)$$

и значение выражения из (7) равно значению выражения из (8). Пусть теперь $x \in \mathbb{N}$ - произвольное натуральное число и

$$\hat{b} := b + m\hat{b}_2(\hat{a}_2 - \hat{c}_2)x.$$

Тогда

$$\hat{b} := m \cdot (m' + \hat{b}_2(\hat{a}_2 - \hat{c}_2)x), \ \hat{b}_2 = m \cdot m'',$$

$$\text{НОД}(\hat{b}, \hat{b}_2) = \text{НОД}(b + m\hat{b}_2(\hat{a}_2 - \hat{c}_2)x, \hat{b}_2) = \text{НОД}(b, \hat{b}_2) = m.$$

Значит $\text{НОД}(m' + \hat{b}_2(\hat{a}_2 - \hat{c}_2)x, m'') = 1$ и, подставляя в (8) $\hat{b}$ вместо b , получаем

$$\text{НОД}(\hat{b}(\hat{a}_2 - \hat{c}_2), a\hat{b}_2 + \hat{b}\hat{c}_2) =$$

$$= m\text{НОД}(\hat{a}_2 - \hat{c}_2, am'' + (m' + \hat{b}_2(\hat{a}_2 - \hat{c}_2)x)\hat{c}_2) =$$

$$= m\text{НОД}(\hat{a}_2 - \hat{c}_2, am'' + m'\hat{c}_2).$$

Далее, так как

$$\hat{b}(c_1 - a_1) + a(d_1 - b_1) =$$

$$= (b + m\hat{b}_2(\hat{a}_2 - \hat{c}_2)x)(c_1 - a_1) + a(d_1 - b_1) =$$

$$= b(c_1 - a_1) + a(d_1 - b_1) + m\hat{b}_2(\hat{a}_2 - \hat{c}_2)x(c_1 - a_1)$$

и число $b(c_1 - a_1) + a(d_1 - b_1)$ делится нацело на (7), то и число $\hat{b}(c_1 - a_1) + a(d_1 - b_1)$ делится нацело на (7), а значит и на (8). Поэтому пара $(a, \hat{b})$ удовлетворяет условию леммы. Аналогично, пусть

$$\hat{a} := a + m\hat{b}_2(\hat{a}_2 - \hat{c}_2)x.$$

Подставляем в (8) $\hat{a}$ вместо a и получаем

$$\text{НОД}(b(\hat{a}_2 - \hat{c}_2), \hat{a}\hat{b}_2 + b\hat{c}_2) =$$

$$= m\text{НОД}(\hat{a}_2 - \hat{c}_2, (a + m\hat{b}_2(\hat{a}_2 - \hat{c}_2)x)m'' + m'\hat{c}_2) =$$

$$= m\text{НОД}(\hat{a}_2 - \hat{c}_2, am'' + m'\hat{c}_2).$$

При этом

$$b(c_1 - a_1) + \hat{a}(d_1 - b_1) =$$

$$= b(c_1 - a_1) + (a + m\hat{b}_2(\hat{a}_2 - \hat{c}_2)x)(d_1 - b_1) =$$

$$= b(c_1 - a_1) + a(d_1 - b_1) + m\hat{b}_2(\hat{a}_2 - \hat{c}_2)x(d_1 - b_1)$$

и так как число $b(c_1 - a_1) + a(d_1 - b_1)$ делится нацело на (7), то и число $b(c_1 - a_1) + \hat{a}(d_1 - b_1)$ делится нацело на (7), а значит и на (8). Поэтому пара $(\hat{a}, b)$ тоже удовлетворяет условию леммы. Значит все удовлетворяющие условию леммы пары (a, b) распадаются на серии

$$a = rx + r_1,$$

$$b = rx + r_2$$

для некоторых $0 \leq r_1 < r$, $0 \leq r_2 < r$. Количество этих серий конечно и они совпадают с множествами из формулировки леммы при $q_1 = q_2 = 1$.

Пусть теперь хотя бы одно из чисел b_2, d_2 равно 0. Если $b_2 = 0$ и $d_2 \neq 0$, то

$$\text{НОД}(ab_2 + ba_2, ad_2 + bc_2) = \text{НОД}(ba_2, ad_2 + bc_2)$$

и этот вариант разбирается точно так же, как и случай 2. Если $b_2 \neq 0$ и $d_2 = 0$, то изменением порядка переменных в (a, b) этот случай сводится к предыдущему. Осталось разобрать последний вариант - когда $b_2 = d_2 = 0$. Тогда

$$\text{НОД}(ab_2 + ba_2, ad_2 + bc_2) = \text{НОД}(ba_2, bc_2) = b \cdot \text{НОД}(a_2, c_2).$$

Пусть для некоторых взаимно простых $a, b \in \mathbb{N}$ выполнено условие леммы, то есть число $b(c_1 - a_1) + a(d_1 - b_1)$ делится нацело на $b \cdot \text{НОД}(a_2, c_2)$. Так как $\text{НОД}(a, b) = 1$, то $d_1 - b_1$ делится нацело на

b. Значит b может принимать лишь конечное количество значений. Кроме того, для

$$\hat{a} := a + \text{НОД}(a_2, c_2)$$

число $b(c_1 - a_1) + \hat{a}(d_1 - b_1)$ тоже делится нацело на $b \cdot \text{НОД}(a_2, c_2)$, так как

$$(b(c_1-a_1)+\hat{a}(d_1-b_1))-(b(c_1-a_1)+a(d_1-b_1)) = (d_1-b_1)\cdot\text{НОД}(a_2, c_2).$$

Значит все удовлетворяющие условию леммы пары (a, b) распадаются на конечное количество серий

$$a = rx + r_1,$$

$$b = r_2$$

и в формулировке леммы можно положить

$$q_1 := 1,\ q_2 := 0.$$

Остальные константы подбираются тривиально. Утверждение леммы 22 доказано.

Лемма 23. *Пусть* $a_1, b_1, c_1, c_2, d_1, d_2 \in \mathbb{N}_0$, L_1 *- пучок с началом* (a_1, b_1) *и базисом* $\{(0, 0)\}$, L_2 *- пучок с началом* (c_1, d_1) *и базисом* $\{(c_2, d_2)\}$, $L_1 \cap L_2 = \emptyset$ *и* $\mathbb{H}(L_1 \cup L_2) \neq \emptyset$. *Тогда существуют числа* $q_1, q_2 \in \{0, 1\}$, $k, s_1, \ldots, s_k \in \mathbb{N}$, $e_1, \ldots, e_k, f_1 \ldots, f_k \in \mathbb{N}_0$ *такие, что для любой пары взаимно простых чисел* $a, b \in \mathbb{N}$ *выполнено*

$-\frac{a}{b} \in \mathbb{H}(L_1 \cup L_2)$ *если и только если*

$$(a,b) \in \bigcup_{i=1}^{k} \left\{ (ms_i q_1 + e_i), (ns_i q_2 + f_i) \,\middle|\, m,n \in \mathbb{N}_0 \right\}.$$

Доказательство. Разбираем случаи.

Случай 1. $c_2 = d_2 = 0$. Тогда из непустоты множества $\mathbb{H}(L_1 \cup L_2)$ получаем, что $b_1 \neq d_1$ и

$$\mathbb{H}(L_1 \cup L_2) = \left\{ -\frac{a_1 - c_1}{b_1 - d_1} \right\},$$

причем $\frac{a_1 - c_1}{b_1 - d_1} > 0$. Осталось положить

$$q_1 := 0,\ q_2 := 0,\ k := 1,\ s_1 := 1,\ e_1 = \frac{|a_1 - c_1|}{c},\ f_1 = \frac{|b_1 - d_1|}{c},$$

где $c = \text{НОД}(|a_1 - c_1|, |b_1 - d_1|)$.

Случай 2. $c_2 = 0$, $d_2 \neq 0$, $a_1 < c_1$, $b_1 \leq d_1$. Обозначем для краткости множество $\{x \in \mathbb{Q} \mid x < 0\}$ через $\hat{\mathbb{Q}}$. Тогда

$$\mathbb{H}(L_1 \cup L_2) =$$

$$= \hat{\mathbb{Q}} \cap \left\{ \frac{u_1 - w_1}{u_2 - w_2} \,\middle|\, (u_1,u_2),(w_1,w_2) \in L_1 \cup L_2, u_2 \neq w_2 \right\} =$$

$$= \hat{\mathbb{Q}} \cap \left\{ \frac{u_1 - w_1}{u_2 - w_2} \,\middle|\, (u_1,u_2) \in L_1, (w_1,w_2) \in L_2, u_2 \neq w_2 \right\} =$$

$$= \hat{\mathbb{Q}} \cap \left\{ \frac{a_1 - c_1}{b_1 - d_1 - nd_2} \,\middle|\, n \in \mathbb{N}_0, b_1 - d_1 - nd_2 \neq 0 \right\} = \emptyset,$$

так как $a_1 - c_1 < 0$, $b_1 - d_1 - nd_2 \leq 0$ при любом $n \in \mathbb{N}_0$. Этот случай невозможен.

Случай 3. $c_2 = 0$, $d_2 \neq 0$, $a_1 < c_1$, $b_1 > d_1$. Тогда

$$\mathbb{H}(L_1 \cup L_2) = \hat{\mathbb{Q}} \cap \left\{ \frac{a_1 - c_1}{b_1 - d_1 - nd_2} \middle| n \in \mathbb{N}_0, b_1 - d_1 - nd_2 \neq 0 \right\} =$$

$$= \left\{ \frac{a_1 - c_1}{b_1 - d_1 - nd_2} \middle| n \in \mathbb{N}_0, n < \frac{b_1 - d_1}{d_2} \right\}.$$

Поэтому множество $\mathbb{H}(L_1 \cup L_2)$ конечно и непусто. Пусть n_0 - максимальное целое неотрицательное число, для которого выполнено $n_0 < \frac{b_1 - d_1}{d_2}$. Осталось положить

$$k := n_0 + 1,\ s_1 := 1,\ \ldots,\ s_k := 1,\ q_1 := 0,\ q_2 := 0,$$

$$e_i := \frac{a_1 - c_1}{c_i},\ f_i := \frac{b_1 - d_1 - (i-1)d_2}{c_i},$$

где $c_i = \text{НОД}(a_1 - c_1, b_1 - d_1 - (i-1)d_2)$ при $1 \leq i \leq k$.

Случай 4. $c_2 = 0$, $d_2 \neq 0$, $a_1 = c_1$. Тогда

$$\mathbb{H}(L_1 \cup L_2) = \hat{\mathbb{Q}} \cap \left\{ \frac{a_1 - c_1}{b_1 - d_1 - nd_2} \middle| n \in \mathbb{N}_0, b_1 - d_1 - nd_2 \neq 0 \right\} = \emptyset.$$

Этот случай невозможен.

Случай 5. $c_2 = 0$, $d_2 \neq 0$, $a_1 > c_1$, $b_1 < d_1$. Тогда

$$\mathbb{H}(L_1 \cup L_2) = \hat{\mathbb{Q}} \cap \left\{ \frac{a_1 - c_1}{b_1 - d_1 - nd_2} \middle| n \in \mathbb{N}_0, b_1 - d_1 - nd_2 \neq 0 \right\} =$$

$$= \left\{ \frac{c_1 - a_1}{nd_2 + d_1 - b_1} \middle| n \in \mathbb{N}_0, b_1 - d_1 - nd_2 \neq 0 \right\} =$$

$$= \bigcup_{i=0}^{c_1 - a_1 - 1} \left\{ \frac{c_1 - a_1}{n(c_1 - a_1)d_2 + id_2 + d_1 - b_1} \middle| n \in \mathbb{N}_0 \right\}.$$

При $0 \leq i < c_1 - a_1$ вводим обозначения

$$c_i := \text{НОД}(c_1 - a_1, id_2 + d_1 - b_1),$$

$$k_i := \frac{c_1 - a_1}{c_i},\ l_i := \frac{id_2 + d_1 - b_1}{c_i}.$$

Имеем:

$$\bigcup_{i=0}^{c_1-a_1-1} \left\{ \frac{c_1 - a_1}{n(c_1 - a_1)d_2 + id_2 + d_1 - b_1} \middle| n \in \mathbb{N}_0 \right\} =$$

$$= \bigcup_{i=0}^{c_1-a_1-1} \left\{ \frac{k_i}{nk_id_2 + l_i} \middle| n \in \mathbb{N}_0 \right\}.$$

Так как при всех $0 \leq i < c_1 - a_1$ верно, что

$$\text{НОД}(k_i, nk_id_2 + l_i) = \text{НОД}(k_i, l_i) = 1,$$

то это искомое представление. Осталось положить

$$q_1 := 0,\ q_2 := 1,\ k := c_1 - a_1,$$

$$s_i := k_{i-1}d_2,\ e_i := k_{i-1},\ f_i := l_{i-1} \text{ при } 1 \leq i \leq k.$$

Случай 6. $c_2 = 0$, $d_2 \neq 0$, $a_1 > c_1$, $b_1 \geq d_1$. Тогда

$$\mathbb{H}(L_1 \cup L_2) = \hat{\mathbb{Q}} \cap \left\{ \frac{a_1 - c_1}{b_1 - d_1 - nd_2} \middle| n \in \mathbb{N}_0, b_1 - d_1 - nd_2 \neq 0 \right\} =$$

$$= \left\{ \frac{c_1 - a_1}{nd_2 + n_0d_2 + d_1 - b_1} \middle| n \in \mathbb{N}_0 \right\},$$

где n_0 - минимальное целое неотрицательное число, для которого выполнено неравенсво $n_0 > \frac{b_1 - d_1}{d_2}$. Для всех $0 \leq i < c_1 - a_1$ вводим обозначения

$$c_i := \text{НОД}(c_1 - a_1, id_2 + n_0d_2 + d_1 - b_1),$$

$$k_i := \frac{c_1 - a_1}{c_i},\ l_i := \frac{id_2 + n_0d_2 + d_1 - b_1}{c_i}.$$

Тогда

$$\left\{ \frac{c_1 - a_1}{nd_2 + n_0 d_2 + d_1 - b_1} \middle| \, n \in \mathbb{N}_0 \right\} =$$

$$= \bigcup_{i=0}^{c_1 - a_1 - 1} \left\{ \frac{c_1 - a_1}{n(c_1 - a_1)d_2 + n_0 d_2 + id_2 + d_1 - b_1} \middle| \, n \in \mathbb{N}_0 \right\} =$$

$$= \bigcup_{i=0}^{c_1 - a_1 - 1} \left\{ \frac{k_i}{nk_i d_2 + l_i} \middle| \, n \in \mathbb{N}_0 \right\}.$$

Так как при всех $0 \leq i < c_1 - a_1$ верно, что

$$\text{НОД}(k_i, nk_i d_2 + l_i) = \text{НОД}(k_i, l_i) = 1,$$

то это искомое представление. Осталось положить

$$q_1 := 0, \ q_2 := 1, \ k := c_1 - a_1,$$

$$s_i := k_{i-1} d_2, \ e_i := k_{i-1}, \ f_i := l_{i-1} \text{ при } 1 \leq i \leq k.$$

Случай 7. $c_2 \neq 0$, $d_2 = 0$. Заменой координат и изменением порядка переменных в (a, b) этот случай сводится к предыдущим.

Случай 8. $c_2 \neq 0$, $d_2 \neq 0$. Тогда

$$\mathbb{H}(L_1 \cup L_2) = \hat{\mathbb{Q}} \cap \left\{ \frac{a_1 - c_1 - mc_2}{b_1 - d_1 - nd_2} \middle| \, n \in \mathbb{N}_0, b_1 - d_1 - nd_2 \neq 0 \right\}.$$

Так как $c_2 \neq 0$, $d_2 \neq 0$, то существует такое $n_0 \in \mathbb{N}_0$, для которого при всех $n \geq n_0$ имеем

$$a_1 - c_1 - mc_2 < 0, \ b_1 - d_1 - nd_2 < 0.$$

Значит $|\mathbb{H}(L_1 \cup L_2)| < \infty$. Далее доказательство повторяет доказательство для случая 3.

Разбор случаев закончен. Утверждение леммы 23 доказано.

Лемма 24. *Пусть* $w_1, w_2 \in \{0, 1\}$, $u \in \mathbb{N}$, $a, b \in \mathbb{N}_0$. *Тогда существует алгоритм, проверяющий существование таких* $m, n \in \mathbb{N}_0$, *для которых*

$$uw_1n + a > 0,$$

$$uw_2m + b > 0,$$

$$НОД(uw_1n + a, uw_2m + b) = 1.$$

Доказательство. Разбираем случаи.

Случай 1. $w_1 = w_2 = 0$. Тогда при всех $m, n \in \mathbb{N}_0$ верно, что

$$\text{НОД}(uw_1n + a, uw_2m + b) = \text{НОД}(a, b).$$

В этом случае утверждение очевидно.

Случай 2. $w_1 = 0$, $w_2 = 1$, $a = 0$. Здесь искомых $m, n \in \mathbb{N}_0$, очевидно, не существует.

Случай 3. $w_1 = 0$, $w_2 = 1$, $a > 0$, $b > 0$. Проверяем, существует ли $k \in \mathbb{N}_0$, $0 \leq k < a$, для которого $(a, b+ku) = 1$. Если да, то искомые $m, n \in \mathbb{N}_0$ найдены. Если нет, то таких $m, n \in \mathbb{N}_0$ не существует, так как при $k \in \mathbb{N}$, $k \geq a$ верно, что

$$\text{НОД}(a, b + ku) = \text{НОД}(a, b + (k - a)u).$$

Случай 4. $w_1 = 0$, $w_2 = 1$, $a > 0$, $b = 0$. Случай сводится к предыдущему заменой b на $b + u$.

Случай 5. $w_1 = 1$, $w_2 = 0$. Случай сводится к предыдущим перестановкой w_1 и w_2.

Случай 6. $w_1 = 1$, $w_2 = 1$, $a = b = 0$. Если $u = 1$, то в качестве искомых $m, n \in \mathbb{N}_0$ можно взять, например, 1. Если $u > 1$, то таких $m, n \in \mathbb{N}_0$, очевидно, не существует.

Случай 7. $w_1 = 1$, $w_2 = 1$, $a = 0$, $b > 0$. Проверяем, верно ли, что

$$\text{НОД}(u, b) > 1.$$

Если не верно, то в качестве искомых $m, n \in \mathbb{N}_0$ можно взять $n = 1$ и $m = 0$. А если верно, то при всех $k \in \mathbb{N}_0$ имеем

$$\text{НОД}(u, b + uk) = \text{НОД}(u, b) > 1$$

и искомых $m, n \in \mathbb{N}_0$ не существует.

Случай 8. $w_1 = 1$, $w_2 = 1$, $a > 0$, $b = 0$. Случай сводится к предыдущему перестановкой w_1 и w_2.

Случай 9. $w_1 = 1$, $w_2 = 1$, $a = 1$, $b > 0$. Тогда в качестве искомых $m, n \in \mathbb{N}_0$ можно взять, например, 0.

Случай 10. $w_1 = 1$, $w_2 = 1$, $a > 0$, $b = 1$. Случай сводится к предыдущему перестановкой w_1 и w_2.

Случай 11. $w_1 = 1$, $w_2 = 1$, $a = b > 1$. Проверяем, верно ли, что

$$\text{НОД}(a, a + u) > 1.$$

Если не верно, то в качестве искомых $m, n \in \mathbb{N}_0$ можно взять $n = 0$

и $m = 1$. А если верно, то при всех $k \in \mathbb{N}_0$ имеем

$$\text{НОД}(a, a + uk) > 1$$

и искомых $m, n \in \mathbb{N}_0$ не существует.

Случай 12. $w_1 = 1$, $w_2 = 1$, $a > b > 1$. Вводим обозначения:

$$c_1 := \text{НОД}(a, u),\ c_2 := \text{НОД}(b, u).$$

Если $\text{НОД}(c_1, c_2) > 1$, то при всех $k, l \in \mathbb{N}_0$ получаем

$$\text{НОД}(a + ul, b + uk) \geq \text{НОД}(b, u) > 1$$

и искомых $m, n \in \mathbb{N}_0$ не существует. Если же $\text{НОД}(c_1, c_2) = 1$, то по теореме Дирихле о простых числах в арифметических прогрессиях (см., например, [38-40]) существует $n_0 \in \mathbb{N}_0$, для которого

$$a + un_0 = c_1 p_1,$$

где p_1 - простое число, большее c_2. По этой же теореме существует число $m_0 \in \mathbb{N}_0$ такое, что

$$b + um_0 = c_2 p_2,$$

где p_2 - простое число, большее p_1 и c_1. Тогда

$$\text{НОД}(a + un_0, b + um_0) = \text{НОД}(c_1 p_1, c_2 p_2) = 1$$

и в качестве искомых $m, n \in \mathbb{N}_0$ можно взять $n = n_0$ и $m = m_0$.

Разбор случаев закончен. Утверждение леммы 24 доказано.

3. Доказательство основных утверждений

Теорема 6.1 Пусть A, B - конечные алфавиты. Тогда проблема ВКД_1 в алфавитах A, B алгоритмически разрешима.

Доказательство. Пусть $f_1, f_2 \in F(A, B)$. В силу леммы 7 для решения проблемы достаточно проверить верно ли вложение

$$S(f_2) \subseteq S(f_1).$$

Из леммы 1 следует, что множества $S(f_1), S(f_2)$ регулярны в обобщенном алфавите $\tilde{A}$. Здесь мы воспользовались классическим результатом о том, что обобщенные источники задают регулярные события в своем алфавите (см., например, [26]). Доказательство теоремы завершает применение леммы 2.

Замечание. Можно заметить, что в доставляющем решение алгоритме не используется регулярность допустимых языков. Поэтому этот алгоритм будет правильно работать и в том случае, когда на допустимые языки не накладывается ограничение регулярности. Акцентирование внимания именно на регулярных языках объясняется тем фактом, что для них проблема распознавания однозначности алфавитного декодирования алгоритмически разрешима.

Теорема 6.2 Пусть A, B - конечные алфавиты и $A = \{a_1, a_2\}$. Тогда проблема ВКД_2 в алфавитах A, B алгоритмически разрешима.

Доказательство. Пусть $P_1, P_2 \in R(A) \setminus \{\Lambda\}$ - произвольные регулярные языки в алфавите A, не содержащие пустое слово. Из леммы 15 следует, что мы можем проверить P_1 на примитивность.

Пусть множество P_1 не примитивно. Тогда возьмем произвольную схему $f \in \mathbb{F}(P_1)$. Из леммы 11 мы знаем, что $f \in \mathbf{F}_2(A, B)$. Тогда из леммы 14 получаем $f \in \mathbb{F}(P_2)$. Значит $\mathbb{F}(P_1) \subseteq \mathbb{F}(P_2)$ и алгоритм завершен.

Пусть теперь множество P_1 примитивно. Используем лемму 1 главы 2. Из нее следует, что P_1 представимо регулярным выражением вида

$$\bigvee_{i=1}^{k} \alpha_{i,1} \cdot (\mathfrak{P}_{i,1})^* \cdot \alpha_{i,2} \cdot \ldots \cdot \alpha_{i,s(i)-1} \cdot (\mathfrak{P}_{i,s(i)-1})^* \cdot \alpha_{i,s(i)},$$

где $k, s(1), \ldots, s(k)$ - натуральные числа, $\alpha_{1,1}, \ldots, \alpha_{k,s(k)}$ - слова (возможно пустые) в алфавите A, $\mathfrak{P}_{1,1}, \ldots, \mathfrak{P}_{k,s(k)-1}$ - регулярные выражения в алфавите A. Поэтому

$$P_1 = \bigcup_{i=1}^{k} \{\alpha_{i,1}\} \cdot |\mathfrak{P}_{i,1}|^* \cdot \{\alpha_{i,2}\} \cdot \ldots \cdot \{\alpha_{i,s(i)-1}\} \cdot |\mathfrak{P}_{i,s(i)-1}|^* \cdot \{\alpha_{i,s(i)}\}. \quad (1)$$

По лемме 8 мы умеем проверять множества $|\mathfrak{P}_{i_0,j_0}|$ при (i_0, j_0) $1 \leq i_0 \leq k$, $1 \leq j_0 < s(i_0)$ на измеримость. Если для некоторой пары множество $|\mathfrak{P}_{i_0,j_0}|$ не измеримо, то по лемме 10 для произвольного $f \in \mathbb{F}(P_1)$ верно $f \notin \mathbf{F}_1(A, B)$, то есть $f \in \mathbf{F}_2(A, B)$. Тогда в силу леммы 14 $f \in \mathbb{F}(P_2)$, а значит $\mathbb{F}(P_1) \subseteq \mathbb{F}(P_2)$ и алгоритм завершен. Поэтому считаем теперь, что все множества $|\mathfrak{P}_{i_0,j_0}|$

измеримы. Тогда и множества $|\mathfrak{P}_{i_0,j_0}|^*$ измеримы. Из леммы 8 главы 3 получаем существование слов $\alpha_{i_0,j_0} \in A^*$ и периодических множеств $T_{i_0,j_0} \subseteq \mathbb{N}$, для которых выполнено

$$|\mathfrak{P}_{i_0,j_0}|^* = \{\alpha_{i_0,j_0}^t \mid t \in T_{i_0,j_0}\}.$$

Из доказательства этой леммы также следует, что

$$|\mathfrak{P}_{i_0,j_0}|^* = \left\{\alpha_{i_0,j_0}^k \mid k \in T'_{i_0,j_0}\right\} \bigcup \left(\bigcup_{t \in T''_{i_0,j_0}} \left\{\alpha^{t+id_{i_0,j_0}} \mid i \in \mathbb{N}_0\right\} \right),$$

где $T'_{i_0,j_0}, T''_{i_0,j_0}, d_{i_0,j_0}$ - предпериод, период и длина периода множества $|\mathfrak{P}_{i_0,j_0}|^*$ соответственно. Так как период и предпериод - конечные множества, то, подставляя эти выражения в формулу (1) и используя дистрибутивность, получаем

$$P_1 = \bigcup_{i=1}^{n} \{\beta_{i,1}\} \cdot \{\gamma_{i,1}\}^* \cdot \{\beta_{i,2}\} \cdot \ldots \cdot \{\beta_{i,r(i)-1}\} \cdot \{\gamma_{i,r(i)-1}\}^* \cdot \{\beta_{i,r(i)}\}$$

для некоторых натуральных чисел $n, r(1), \ldots, r(n)$, некоторых слов (возможно, пустых) $\beta_{1,1}, \ldots, \beta_{n,r(n)} \in A^*$ и некоторых слов $\gamma_{1,1}, \ldots, \gamma_{n,r(n)-1} \in A^* \setminus \{\Lambda\}$. Если $f \in \mathbb{F}(P_1) \cap \mathbf{F}_2(A,B)$, то из леммы 14 следует, что $f \in \mathbb{F}(P_2)$. Поэтому нам достаточно проверить, для любого ли $f \in \mathbb{F}(P_1)$ выполение условия $f \in \mathbf{F}_1$ влечет за собой выполение условия $f \in \mathbb{F}(P_2)$. Отсюда и из леммы 13 получаем, что при сделанных выше предположениях $\mathbb{F}(P_1) \subseteq \mathbb{F}(P_2)$ если и только если $\mathbb{H}(T(P_2)) \subseteq \mathbb{H}(T(P_1))$. Замечаем, что

$$T(P_1) = T\left(\bigcup_{i=1}^{n} \{\beta_{i,1}\} \cdot \{\gamma_{i,1}\}^* \cdot \ldots \cdot \{\gamma_{i,r(i)-1}\}^* \cdot \{\beta_{i,r(i)}\} \right) =$$

$$= \left\{ (n_1(\alpha), n_2(\alpha)) \,\middle|\, \alpha \in \bigcup_{i=1}^{n} \{\beta_{i,1}\} \cdot \{\gamma_{i,1}\}^* \cdot \ldots \cdot \{\beta_{i,r(i)}\} \right\} =$$

$$= \bigcup_{i=1}^{n} \left\{ \sum_{j=1}^{r(i)} \Big(n_1(\beta_{i,j}), n_2(\beta_{i,j}) \Big) + \right.$$

$$\left. + \sum_{j=1}^{r(i)-1} m_j \Big(n_1(\gamma_{i,j}), n_2(\gamma_{i,j}) \Big) \,\middle|\, m_j \in \mathbb{N}_0 \right\} =$$

$$= \bigcup_{i=1}^{n} L \left(\left(\sum_{j=1}^{r(i)} n_1(\beta_{i,j}), \sum_{j=1}^{r(i)} n_2(\beta_{i,j}) \right), \bigcup_{j=1}^{r(i)-1} \{(n_1(\gamma_{i,j}), n_2(\gamma_{i,j}))\} \right).$$

Таким образом, задача свелась к тому, чтобы по двум конечным семействам пучков X_1, X_2 проверить, верно ли, что

$$\mathbb{H}(X_1) \subseteq \mathbb{H}(X_2).$$

Если в каком-то из пучков $L \in X_2$ в базисе есть неколлинеарные векторы, то из леммы 16 получаем

$$\mathbb{H}(L) = \{x \in \mathbb{Q} \mid x < 0\}.$$

Значит

$$\mathbb{H}(X_1) \subseteq \{x \in \mathbb{Q} \mid x < 0\} = \mathbb{H}(X_2).$$

В этом случае алгоритм завершен. Аналогично, если в каком-то из пучков $L \in X_1$ в базисе есть неколлинеарные векторы, то

$$\mathbb{H}(X_1) = \{x \in \mathbb{Q} \mid x < 0\}.$$

В этом случае $\mathbb{H}(X_1) \subseteq \mathbb{H}(X_2)$ выполнено если и только если $\mathbb{H}(X_2) = \{x \in \mathbb{Q} \mid x < 0\}$. Тогда алгоритм работает тривиально по

общей схеме, приводимой нами ниже в конце этого доказательства. Считаем теперь, что в базисах пучков из X_1, X_2 нет неколлинеарных векторов. Тогда в силу леммы 19 можно считать базисы всех пучков одноэлементными. Далее разбиваем пучки из X_1 на классы эквивалентности: в один и тот же класс попадают пучки, базисные векторы которых коллинеарны. Разбираем два случая.

Случай 1. Какие-то два пучка L_1, L_2 из разных классов эквивалентности пересекаются. Тогда X_1 содержит в себе пучок, у которого есть два базисных неколлинеарных вектора. Применение леммы 16 завершает исследование этого случая.

Случай 2. Любые два пучка L_1, L_2 из разных классов не пересекаются. С каждым из классов эквивалентности теперь можно разбираться отдельно. Все пучки делятся на два типа: одноэлементные (базисный вектор равен $(0, 0)$) и бесконечные (базисный вектор не равен $(0, 0)$). Если одноэлементный пучок L_1 пересекается с каким-то другим пучком L_2, то $L_1 \subseteq L_2$. Значит L_1 можно просто выкинуть. Поэтому далее считаем, что все одноэлементные пучки изолированы от остальных. По лемме 20 объединение пучков из общего класса эквивалентности (а таких пучков, как мы помним, конечное количество) можно разбить в конечное объединение непересекающихся пучков с одноэлементным базисом.

Теперь считаем, что все пучки из X_1 попарно не пересекают-

ся. Применяя, если необходимо, тот же процесс для X_2, считаем, что и все пучки из X_2 попарно не пересекаются. Теперь можно использовать результаты лемм 21, 22, 23. Из них получаем разбиение множеств $\mathbb{H}(X_1)$ и $\mathbb{H}(X_2)$ в конечное объединение по $i = 1, \ldots, k$ серий вида

$$(a, b) \in \{(ms_iq_1 + e_i), (ns_iq_2 + f_i) | m, n \in \mathbb{N}_0\}, \quad \text{НОД}(a, b) = 1. \quad (2)$$

Здесь $q_1, q_2 \in \{0, 1\}$, $k, s_1, \ldots, s_k \in \mathbb{N}$, $e_1, f_1 \ldots e_k, f_k \in \mathbb{N}_0$. Ясно, что разность любых двух серий вида (2) можно в свою очередь разложить в конечное объединение непересекающихся серий того же вида. Значит и разность $\mathbb{H}(X_1) \setminus \mathbb{H}(X_2)$ разбивается в конечное объединение непересекающихся серий. Доказательство теоремы завершает применение для каждой из них леммы 24 о проверке на пустоту.

В завершение приводим краткое описание доставляющего решение алгоритма.

Шаг 1. Проверяем P_1 на примитивность. Если P_1 не примитивно, то

$$\mathbb{F}(P_1) \subseteq \mathbb{F}(P_2)$$

и алгоритм завершен. Иначе шаг 2.

Шаг 2. Представляем P_1 в виде

$$P_1 = \bigcup_{i=1}^{k} \{\alpha_{i,1}\} \cdot |\mathfrak{P}_{i,1}|^* \cdot \{\alpha_{i,2}\} \cdot \ldots \cdot \{\alpha_{i,s(i)-1}\} \cdot |\mathfrak{P}_{i,s(i)-1}|^* \cdot \{\alpha_{i,s(i)}\}.$$

Далее шаг 3.

Шаг 3. Проверяем множества

$$|\mathfrak{P}_{i_0,j_0}|,\ 1 \le i_0 \le k,\ 1 \le j_0 < s(i_0)$$

на измеримость. Если хоть одно из этих множеств не измеримо, то $\mathbb{F}(P_1) \subseteq \mathbb{F}(P_2)$ и алгоритм завершен. Иначе шаг 4.

Шаг 4. Представляем P_1 и P_2 в виде

$$\bigcup_{i=1}^{n} \{\beta_{i,1}\} \cdot \{\gamma_{i,1}\}^* \cdot \{\beta_{i,2}\} \cdot \ldots \cdot \{\beta_{i,r(i)-1}\} \cdot \{\gamma_{i,r(i)-1}\}^* \cdot \{\beta_{i,r(i)}\},$$

где $n, r(1), \ldots, r(n)$ - натуральные числа, $\beta_{1,1}, \ldots, \beta_{n,r(n)}$ - слова (возможно пустые) в алфавите A, $\gamma_{1,1}, \ldots, \gamma_{n,r(n)-1}$ - непустые слова в алфавите A. Далее шаг 5.

Шаг 5. Составляем для P_1 и P_2 пучки X_2 и X_1 соответственно по формулам

$$\bigcup_{i=1}^{n} L\left(\left(\sum_{j=1}^{r(i)} n_1(\beta_{i,j}), \sum_{j=1}^{r(i)} n_2(\beta_{i,j})\right), \bigcup_{j=1}^{r(i)-1} \{(n_1(\gamma_{i,j}), n_2(\gamma_{i,j}))\}\right).$$

Далее шаг 6.

Шаг 6. Если в каком-то из пучков $L \in X_2$ в базисе есть неколлинеарные векторы, то $\mathbb{F}(P_1) \subseteq \mathbb{F}(P_2)$ и алгоритм завершен. Иначе шаг 7.

Шаг 7. Если в каком-то из пучков $L \in X_1$ в базисе есть неколлинеарные векторы, то $\mathbb{H}(X_1) = \{x \in \mathbb{Q} \mid x < 0\}$ и $\mathbb{H}(X_1)$ задается тривиальной серией

$$(a, b) \in \{(m+1, n+1) | m, n \in \mathbb{N}_0\}, \quad \text{НОД}(a, b) = 1.$$

Далее шаг 11. Иначе шаг 8.

Шаг 8. Если в X_2 есть пересекающиеся пучки с неколлинеарными базисными векторами, то $\mathbb{F}(P_1) \subseteq \mathbb{F}(P_2)$ и алгоритм завершен. Иначе шаг 9.

Шаг 9. Если в X_1 есть пересекающиеся пучки с неколлинеарными базисными векторами, то $\mathbb{H}(X_1)$ задается тривиальной серией

$$(a, b) \in \{(m+1, n+1) | m, n \in \mathbb{N}_0\}, \quad \text{НОД}(a, b) = 1.$$

Далее шаг 11. Иначе шаг 10.

Шаг 10. Выкидываем все дублирующие одноэлементные пучки. Оставшиеся пучки разбиваем на непересекающиеся. Далее шаг 11.

Шаг 11. Разбиваем множество

$$\mathbb{H}(X_1) \setminus \mathbb{H}(X_2)$$

в конечное объединение непересекающихся серий. Далее шаг 12.

Шаг 12. Проверяем серии на пустоту. Если хоть одна непуста, то

$$\mathbb{F}(P_1) \not\subseteq \mathbb{F}(P_2)$$

и алгоритм завершен. Иначе

$$\mathbb{F}(P_1) \subseteq \mathbb{F}(P_2)$$

и алгоритм опять завершен.

Утверждение теоремы доказано.

Замечание. Приводимый здесь алгоритм не претендует на оптимальность и в дальнейшем будет улучшаться. Кроме того, боль-

шой интерес представляет решение проблемы ВКД_2 для случая, когда $|A| > 2$. В этом направлении тоже ведутся соответствующие изыскания.

4. Заключение главы 6

В этой главе мы научились алгоритмически сравнивать между собой произвольные функции алфавитного кодирования в терминах их полезности на классах регулярных языков. Кроме того, была решена аналогичная проблема сравнения полезности регулярных языков в алфавите мощности 2. Эти результаты являются существенными при реализации нашей модели. Действительно, чем больше класс $\mathbb{R}(f)$ допустимых регулярных языков для f, тем эта функция нам полезнее. Аналогично, чем больше класс $\mathbb{F}(P)$ допустимых схем кодирования для P, тем этот язык для нас лучше. Результаты этой главы позволяют нам в этом смысле сравнивать между собой разные $f \in F(X, Y)$ и, тоже между собой, разные $P \in X^*$. Значит у нас есть инструмент, позволяющий, например, по уже выбранному языку P выбрать наиболее перспективную для нас функцию $f \in \mathbb{F}(P)$. Или, наоборот, для уже имеющейся функции f выбрать богатый язык $P \in \mathbb{R}(f)$. Это может быть особенно полезно, если один объект из пары P, f стал известен конкурентам, а второй еще нет. Тогда, не меняя второй объект, можно из соот-

ветствующего ему широкого класса допустимых первых объектов выбрать новый. Это позволяет нам не только экономить время на перестройку системы, но и еще и минимизировать соответствующие производственные издержки.

Заключение

В диссертации получены следующие основные результаты.

- Получены верхние полиномиальные оценки на сложность решения проблемы однозначности алфавитного кодирования в классах $\mathfrak{T}(A)$ и $RP(A)$.
- Построена полная неизбыточная классификация внутренней структуры элементов из класса $\mathfrak{T}(A)$ в терминах конечных объединений прогрессивных множеств.
- Построена полная неизбыточная классификация внутренней структуры элементов из класса $RP(A)$ в терминах конечных объединений множеств правильного линейного вида.
- Найдена квадратичная зависимость между сложностью представления множеств в прогрессивном и правильном линейном видах.
- Доказана алгоритмическая разрешимость проблемы ВКД_1.
- Доказана алгоритмическая разрешимость проблемы ВКД_2 для случая, когда мощность входного алфавита равна двум.
- Предложена модель, демонстрирующая применение полученных результатов на практике.

Полученные результаты демонстрируют, что алфавитное кодирование элементов из классов $\mathfrak{T}(A)$ и $RP(A)$ может быть применено как в области сжатия информации, так и при передаче информации по каналу связи на длительные расстояния.

Краткий список обозначений

$\mathbb{N}$ - множество натуральных чисел.

$V = (A, Q, B, \varphi, \psi)$ - абстрактный конечный автомат.

V_q - инициальный абстрактный конечный автомат с начальным состоянием q.

$K(A, B, n)$ - множество всех инициальных абстрактных конечных автоматов с входным алфавитом A, выходным алфавитом B и алфавитом состояний мощности n.

$K_{\leq}(A, B, n)$ - множество всех инициальных абстрактных конечных автоматов с входным алфавитом A, выходным алфавитом B и алфавитом состояний мощности не выше n.

$[V_q]$ - множество инициальных абстрактных конечных автоматов, полученных из V_q изменением начального состояния.

$|\gamma|$ - длина слова γ.

$\gamma_{l,m}$ - подслово слова γ с l-ой до m-ой буквы.

$P_{\leq}(n)$ - множество слов из P, длина которых не превосходит n.

$B'(V_q)$ - множество слов, распознаваемых автоматом V_q по последней букве выходного слова с помощью множества B'.

E_2 - множество $\{0, 1\}$.

$V = (A, Q, B, \gamma)$ - недетерминированный конечный автомат.

$V_{Q'}$ - инициальный недетерминированный конечный автомат с начальным множеством состояний Q'.

$B'(V_{Q'})$ - множество слов, распознаваемых автоматом $V_{Q'}$ по последней букве выходного слова с помощью множества B'.

$\tilde{K}(A, B, n)$ - класс инициальных недетерминированных конечных автоматов с n состояниями.

$P_1 \bigcup P_2$ - объединение множеств P_1 и P_2.

$P_1 \cdot P_2$ - конкатенация множеств P_1 и P_2.

$(P_1)^*$ - итерация множества P_1.

$\emptyset$ - пустое множество.

$|\mathfrak{P}|$ - множество, представимое регулярным выражением $\mathfrak{P}$.

$R(A)$ - множество всех регулярных множеств в алфавите A.

λ - регулярное выражение, представляющее пустое множество.

$F(A, B)$ - множество всех схем кодирования из A в B.

L_f - длина схемы кодирования f.

l_f - сложность схемы кодирования f.

$\tilde{f}$ - алфавитное кодирование по схеме f.

$(\tilde{f})_P$ - функция, полученная из $\tilde{f}$ сужением области определения до P.

$I(f)$ - класс допустимых языков для схемы f.

ОАД$_1$ - проблема распознавания свойства однозначности алфавитного декодирования в классе регулярных языков.

$\mathbb{N}_0$ - множество $\mathbb{N} \cup \{0\}$.

$Sp(P)$ - спектр (множество длин слов) множества P.

$|P|$ - мощность множества P.

$\mathfrak{T}_s(A)$ - множество всех s-тонких языков в алфавите A.

$\mathfrak{T}(A)$ - множество всех тонких языков в алфавите A.

$T_n(P)$ - количество слов из P длины не больше n.

T_P - функция роста языка P.

ОАД$_2$ - проблема распознавания свойства однозначности алфавитного декодирования в классе тонких языков.

$F_n(A, B)$ - множество всех схем кодирования из A в B, сложность которых не превосходит n.

$RP(A)$ - множество всех не содержащих пустое слово регулярных языков в алфавите A с полиномиальной функцией роста.

ОАД$_3$ - проблема распознавания свойства однозначности алфавитного декодирования в классе регулярных языков с полиномиальной функцией роста.

$\mathfrak{X}(A)$ - множество выражений в алфавите A, являющихся конечной дизъюнкцией выражений вида $\alpha \cdot (\beta)^* \cdot \gamma$.

$L(\mathfrak{P})$ - сложность выражений из класса $\mathfrak{X}(A)$.

$U(A)$ - класс множеств, представимых выражениями из $\mathfrak{X}(A)$.

$L_p(P)$ - прогрессивная сложность множества из класса $U(A)$.

$U^n(A)$ - множество элементов из $U(A)$, прогрессивная сложность которых не превосходит n.

$RP^1(A)$ - множество, состоящее из всех множеств линейного вида

в алфавите A.

$WRP^1(A)$ - множество, состоящее из всех множеств правильного линейного вида в алфавите A.

$RP^1_n(A)$ - множество, состоящее из всех множеств линейного вида в алфавите A, которые представимы регулярными выражениями линейного вида сложности не выше n.

$WRP^1_n(A)$ - множество, состоящее из всех множеств правильного линейного вида в алфавите A, которые представимы регулярными выражениями правильного линейного вида сложности не выше n.

$RP_n(A)$ - множество, состоящее из всех конечных объединений множеств из $RP^1_n(A)$.

$WRP_n(A)$ - множество, состоящее из всех конечных объединений множеств из $WRP^1_n(A)$.

(A, B, τ, b) - контекстно-свободная грамматика.

$|X|$ - множество слов, выводимых в контекстно-свободной грамматике X.

$\mathbb{CF}(A)$ - множество всех контекстно-свободных языков в алфавите A.

$(A, Q, B, \varphi, \psi, \lambda, q_0)$ - инициальный конечный автомат с магазинной памятью.

$\mathbb{PDA}(A)$ - множество всех инициальных конечных автоматов с магазинной памятью с входным алфавитом A.

$|V|$ - множество слов, распознаваемое инициальным конечным автоматом V с магазинной памятью.

$|G|$ - событие, задаваемое обобщенным источником G.

$\mathbb{R}(f)$ - класс допустимых регулярных языков для схемы f.

$\mathbb{F}(P)$ - класс допустимых схем кодирования для языка P.

ВКД_1 - проблема распознавания свойства $f_1 \leq f_2$ для произвольной пары схем кодирования из A в B.

ВКД_2 - проблема распознавания свойства $P_1 \leq P_2$ для произвольной пары регулярных языков в алфавите A.

$\tilde{A}$ - обобщенный алфавит для алфавита A.

$S(f)$ - отношение синонимии на схеме кодирования f.

$L(\hat{v}, V)$ - пучок с началом в $\hat{v}$ и базисом V.

$\mathbb{Q}$ - множество рациональных чисел.

$]x[$ - целая часть сверху от числа x.

Список литературы

[1] C. E. Shannon. *A Mathematical Theory of Communication*, Bell System Technical Journal, 1952, № 27, pp. 379-423.

[2] К. Э. Шеннон. *Математическая теория связи.* В сб. "Работы по теории информации и кибернетики". Издательство иностранной литературы, 1963.

[3] Marcel J. E. Golay, *Notes on Digital Coding*, Proc. IRE 37: 657, 1949.

[4] L. G. Kraft. *A device for quantizing, grouping, and coding amplitude modulated pulses*, Cambridge, MA: MS Thesis, Electrical Engineering Department, Massachusetts Institute of Technology, 1949.

[5] B. McMillan. *Two inequalities implied by unique decipherability,* IEEE Trans. Information Theory 2 (4), pp. 115-116, 1956.

[6] D.A. Huffman. *A Method for the Construction of Minimum-Redundancy Codes,* Proceedings of the I.R.E., pp 1098-1102, 1952.

[7] D. E. Muller. *Application of boolean algebra to switching circuit design and to error detection*, IRE Transactions on Electronic Computers, 3:6-12, 1954.

[8] Irving S. Reed. *A class of multiple-error-correcting codes and the decoding scheme*, Transactions of the IRE Professional Group on Information Theory, 4:38-49, 1954.

[9] A. Hocquenghem. *Codes correcteurs d'erreurs,* Chiffres (in French) (Paris) 2, pp 147-156, 1959.

[10] R. C. Bose; D. K. Ray-Chaudhuri. *On A Class of Error Correcting Binary Group Codes*, Information and Control 3 (1), pp 68-79, 1960.

[11] Irving S. Reed; Gustave Solomon, *Polynomial Codes over Certain Finite Fields*, Journal of the Society for Industrial and Applied Mathematics (SIAM) 8 (2), pp 300-304, 1960.

[12] Robert G. Gallager, *Low Density Parity Check Codes,* Monograph, M.I.T. Press, 1963.

[13] A. Viterbi. *Error bounds for convolutional codes and an asymptotically optimum decoding algorithm.* IEEE Transactions on Information Theory 13 (2): 260-269, 1967.

[14] Jorma Rissanen. *Generalized Kraft Inequality and Arithmetic Coding*, IBM Journal of Research and Development 20 (3), pp 198-203, 1976.

[15] Jacob Ziv, Abraham Lempel. *A Universal Algorithm for Sequential Data Compression*, IEEE Transactions on Information Theory, 23(3), pp. 337-343, 1977.

[16] Jacob Ziv, Abraham Lempel. *Compression of Individual Sequences Via Variable-Rate Coding*, IEEE Transactions on Information Theory, 24(5), pp. 530-536, 1978.

[17] C. Berrou, A. Glavieux, P. Thitimayshima. *Near Shannon Limit Error - Correcting Coding and Decoding: Turbo-Codes*, Ecole Nationale Superieure des Telecommunications de Bretagne, France, 1993.

[18] Michael Burrows; David J. Wheeler.*A block sorting lossless data compression algorithm,* Technical Report 124, Digital Equipment Corporation, 1994.

[19] E. Arikan. *Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels,* IEEE Transactions on Information Theory, vol.55, №7, pp. 3051-3073, 2009.

[20] N. Chomsky. *Three Models for the Description of Language,* IRE Transactions on Information Theory IT-2, 113-24, 1956.

[21] John E. Hopcroft; Jeffrey D. Ullman. Introduction to Automata Theory, Languages, and Computation (1st ed.), Addison-Wesley, 1979.

[22] Ал. А. Марков. *Основания общей теории кодов*. Проблемы кибернетики, 1976, №31, с. 77-108.

[23] Ал. А. Марков. *Введение в теорию кодирования.* М: Наука, 1982.

[24] С. В. Яблонский. *Введение в дискретную математику.* М.: Наука, 1986.

[25] Л. П. Жильцова. *Современные проблемы теории кодирования.* Учебное пособие, Нижний Новгород, 2007.

[26] В. Б. Кудрявцев, С. В. Алешин, А. С. Подколзин. *Введение в теорию автоматов.* М.: Наука, 1985.

[27] B. M. Oliver. *Efficient Codiny.* BSTJ, 1952, № 3.

[28] Р. В. Хемминг. *Коды с обнаружением и исправлением ошибок.* В сб. "Теория передачи сообщений". Издательство иностранной литературы, 1956.

[29] W. W. Peterson, D. T. Brown. *Cyclic Codes for Error Delection.* PIRE, 1961, № I.

[30] В. Д. Колесник, Е. Т. Мирончиков. *Декодирование циклических кодов.* Издательство "Связь", 1968.

[31] Р. Дж. Галлагер. *Коды с малой плотностью проверок на четность.* Издательство "Мир", 1966.

[32] А. Н. Колмогоров. *Теория передачи информации.* Издательство АН СССР, 1956.

[33] E. N. Gilbert, E. F. Mooge. *Variable-Length Binary Encoding.* BSTJ, 1959, № 4.

[34] А. В. Чашкин. *Лекции по дискретной математике.* Изд. МГУ, учебное пособие, М.: 2007.

[35] В. А. Носов. *Основы теории алгоритмов и анализа их сложности.* М.: Наука, 1992.

[36] Y. Bar-Hillel, M. Perles, and E. Shamir. *On formal properties of simple phrase-structure grammars*, Z. Phonetic. Sprachwiss. Kommu-nikationsforsch. 14 (1961), pp. 143-172.

[37] S. Ginsburg. *The mathematical theory of context-free languages.* Santa Monica: McGraw-Hill Book Company, 1966.

[38] И. М. Виноградов. *Основы теории чисел.* Физматгиз, 1959.

[39] Ю. В. Линник, А.О. Гельфанд. *Элементарные методы в аналитической теории чисел.* Физматгиз, 1992.

[40] М. М. Постников. *Теорема Ферма. Введение в теорию алгебраических чисел.* М.: Наука, 1986.

Работы автора по теме диссертации

1. П. С. Дергач. *Об однозначности алфавитного декодирования.* Дискретная математика -М.: Наука, том 24, № 4, с. 80-90, 2012.

2. П. С. Дергач. *Об однозначности алфавитного декодирования*

общерегулярных сверхязыков. Дискретная математика -М.: Наука, том 26, № 1, с. 32-48, 2014.

3. P. S. Dergach. *On uniqueness of alphabetical decoding of* $\emptyset$*-regular languages*. Discrete Mathematics and Applications, издательство V S P (Netherlands), том 24, № 3, с. 139-152, 2014.

4. П. С. Дергач. *О каноническом регулярном представлении S-тонких языков*. Интеллектуальные системы, изд. МГУ, М., том 18, № 1, с. 211-242, 2014.

5. П. С. Дергач. *О проблеме вложения допустимых классов*. Интеллектуальные системы, изд. МГУ, М., том 19, № 2, с. 143-174, 2015.