ФГБОУ ВО «Московский государственный университет
имени М.В. Ломоносова»

На правах рукописи

Дергач Пётр Сергеевич

Алфавитное кодирование регулярных языков с полиномиальной функцией роста

01.01.09 — дискретная математика и математическая кибернетика

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук

Москва — 2016

Работа выполнена на кафедре математической теории интеллектуальных систем Механико-математического факультета ФГБОУ ВО «Московский государственный университет имени М.В. Ломоносова».

Научный руководитель: **Кудрявцев Валерий Борисович**, доктор физико-математических наук, академик, профессор

Официальные оппоненты: **Орлов Валентин Александрович** доктор физико-математических наук, профессор кафедры "Информационная безопасность"(Московского государственного технического университета имени Н.Э. Баумана)

Летуновский Алексей Александрович кандидат физико-математических наук, Техкомпания "Хуавей",консультант

Ведущая организация: **ФГБОУ ВО "Московский технологический университет"**

Защита диссертации состоится 21 октября 2016 г. в 16 ч. 45 м. на заседании диссертационного совета Д.501.001.84, на базе ФГБОУ ВО МГУ имени М.В. Ломоносова, по адресу: Российская Федерация, 119234, Москва, ГСП-1, Ленинские горы, д.1, ФГБОУ ВО МГУ имени М.В. Ломоносова, Механико-математический факультет, аудитория 14-08.

С диссертацией можно ознакомиться в Фундаментальной библиотеке ФГБОУ ВО МГУ имени М.В. Ломоносова, по адресу: Москва, Ломоносовский проспект, д. 27, сектор А., http://mech.math.msu.su/∼snark/index.cgi, https://istina.msu.ru/dissertations/21301300.

Автореферат разослан 21 сентября 2016 г.

Ученый секретарь
диссертационного совета
Д.501.001.84, на базе
ФГБОУ ВО МГУ имени М.В. Ломоносова,
доктор физико-математических наук,
профессор **Шафаревич Андрей Игоревич**

Общая характеристика работы

Актуальность темы исследования. *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[1]. 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[2]. 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*[3]. Своим именем оно также обязано американскому ученому и политику Броквею Мак-Миллану, который в 1956 году доказал выполнение этого неравенства для разделимых кодов[4]. Затем американский ученый Дэвид Хаффман, будучи аспирантом создает первый код оптимального сжатия, широко известный теперь как *код Хаффмана*[5]. В 1954 году американцы Ирвинг С. Рид и Дэвид Е. Мюллер придумывают *блочные коды Рида-Маллера*[6]. Далее французский математик Алексис Хоквингем[7] и независимо от него чуть позже американцы Радж Чандра Боуз и Двайджендра Камар Рей-Чоудхури[8] 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*[9]. Затем результаты в области теории кодирования идут

[1] см. C. E. Shannon. *A Mathematical Theory of Communication*, Bell System Technical Journal, № 27, pp. 379-423, 1948, см. также перевод К. Э. Шеннон. *Математическая теория связи.* В сб. "Работы по теории информации и кибернетики". Издательство иностранной литературы, 1963.

[2] см. Marcel J. E. Golay, *Notes on Digital Coding*, Proc. IRE 37: 657, 1949.

[3] см. 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.

[4] см. B. McMillan. *Two inequalities implied by unique decipherability,* IEEE Trans. Information Theory 2 (4), pp. 115-116, 1956.

[5] см. D.A. Huffman. *A Method for the Construction of Minimum-Redundancy Codes,* Proceedings of the I.R.E., pp 1098-1102, 1952.

[6] см. D. E. Muller. *Application of boolean algebra to switching circuit design and to error detection*, IRE Transactions on Electronic Computers, 3:6-12, 1954 и 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.

[7] см. A. Hocquenghem. *Codes correcteurs d'erreurs,* Chiffres (in French) (Paris) 2, pp 147-156, 1959.

[8] см. 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.

[9] см. 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.

по нарастающей. К наиболее важным из них, пожалуй, стоит отнести создание *LDPC-кодов*[10] (Роберт Галлагер, 1963), возникновение *алгоритма Витерби* по декодированию сверточных кодов[11] (Эндрю Витерби, 1967), создание *арифметического кодирования*[12] (Йорма Риссанен, 1976), появление *алгоритмов сжатия* LZ-77[13] и LZ-78[14] (Якоб Зив, Авраам Лемпел, 1977-1978), появление *турбо-кодов*[15] (Клод Берроу , Алэйн Главиукс и П.Ситимашимой, 1993), *преобразование Барроуза-Уилера*[16] (Майкл Барроуз, Дэвид Уилер, 1994), изобретение *полярных кодов*[17] (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,*[18] 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

[10] см. Robert G. Gallager, *Low Density Parity Check Codes,* Monograph, M.I.T. Press, 1963.

[11] см. A. Viterbi. *Error bounds for convolutional codes and an asymptotically optimum decoding algorithm.* IEEE Transactions on Information Theory 13 (2): 260-269, 1967.

[12] см. Jorma Rissanen. *Generalized Kraft Inequality and Arithmetic Coding*, IBM Journal of Research and Development 20 (3), pp 198-203, 1976.

[13] см. Jacob Ziv, Abraham Lempel. *A Universal Algorithm for Sequential Data Compression*, IEEE Transactions on Information Theory, 23(3), pp. 337-343, 1977.

[14] см. Jacob Ziv, Abraham Lempel. *Compression of Individual Sequences Via Variable-Rate Coding*, IEEE Transactions on Information Theory, 24(5), pp. 530-536, 1978.

[15] см. 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.

[16] см. Michael Burrows; David J. Wheeler.*A block sorting lossless data compression algorithm,* Technical Report 124, Digital Equipment Corporation, 1994.

[17] см. 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.

[18] см. N. Chomsky. *Three Models for the Description of Language,* IRE Transactions on Information Theory IT-2, 113-24, 1956.

- *абстрактными конечными автоматами*[19].

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 кодирования в России[20]. 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[21]. 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[22]. Также известно, что в классе контекстно-свободных грамматик проблема ОАД при ряде дополнительных незначительных ограничений алгоритмически неразрешима[23]. 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

[19] доказательства этих утверждений можно найти, например, в John E. Hopcroft; Jeffrey D. Ullman. Introduction to Automata Theory, Languages, and Computation (1st ed.), Addison-Wesley, 1979.

[20] см. Ал. А. Марков. *Основания общей теории кодов.* Проблемы кибернетики, 1976, №31, с. 77-108.

[21] см. Ал. А. Марков. *Введение в теорию кодирования.* М: Наука, 1982.

[22] см. С. В. Яблонский. *Введение в дискретную математику.* М.: Наука, 1986.

[23] см. Л. П. Жильцова. *Современные проблемы теории кодирования.* Учебное пособие, Нижний Новгород, 2007.

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 разрешима. 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 однозначности алфавитного кодирования регулярных языков.

- Привести для языков из класса тонких языков полное неизбыточное описание в терминах конечных объединений *прогрессивных множеств.*
- Привести для языков из класса регулярных языков с полиномиальной функцией роста полное неизбыточное описание в терминах конечных объединений *множеств правильного линейного вида.*
- Применить разработанный подход решения проблемы однозначности алфавитного декодирования к классам тонких языков и регулярных языков с полиномиальной функцией роста и, используя информацию об их структуре, улучшить его, получив соответствующие оценки на сложность решающего алгоритма.
- Найти алгоритмическое решение проблемы сравнения полезности произвольной пары функций алфавитного кодирования в общих алфавитах.
- Найти алгоритмическое решение проблемы сравнения полезности произвольной пары регулярных языков в общем алфавите.
- Показать, что процедура алфавитного декодирования в классах тонких языков и регулярных языков с полиномиальной функцией роста имеет простую сложность.

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

- Предложен новый автоматно-алгебраический подход к решению проблемы однозначности алфавитного кодирования регулярных языков.
- Приведена классификация внутренней структуры тонких множеств в терминах конечных объединений прогрессивных множеств.
- Приведена классификация внутренней структуры регулярных языков с полиномиальной функцией роста в терминах конечных объединений множеств правильного линейного вида.
- Показана полиномиальность верхних оценок на сложность решения проблемы однозначности алфавитного кодирования регулярных языков новым автоматно-алгебраическим подходом.
- Показано, что проблема сравнения полезности произвольной пары функций алфавитного кодирования в общих алфавитах алгоритмически разрешима.

- Показано, что проблема сравнения полезности произвольной пары регулярных языков в общем алфавите алгоритмически разрешима в случае, когда мощность этого алфавита равна двум.
- Предложена модель, демонстрирующая практическую значимость полученных результатов.
- Исследован вопрос о сложности процедуры алфавитного декодирования в классах тонких языков и регулярных языков с полиномиальной функцией роста.

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 тонких языков и регулярных языков с полиномиальной функцией роста. 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 каналу 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 тонких языков и регулярных языков с полиномиальной функцией роста.
2. Построение полной неизбыточной классификации внутренней структуры элементов из класса тонких языков в терминах конечных объединений прогрессивных множеств.
3. Построение полной неизбыточной классификации внутренней структуры элементов из класса регулярных языков с полиномиальной функцией роста в терминах конечных объединений множеств правильного линейного вида.

Степень достоверности и апробация результатов. Результаты прошли апробацию на международных и всероссийских научных конференциях, а также на авторитетных научных семинарах.

Конференции.

- Международная конференция студентов, аспирантов и молодых ученых "Ломоносовские чтения" (7-15 апреля 2011 года, 2-9 апреля 2012 года, 15-26 апреля 2013 года, 14-23 апреля 2014 года, 18-27 апреля 2016 года, Москва, МГУ).
- XI Международный семинар "Дискретная математика и ее приложения", посвященный 80-летию со дня рождения О.Б. Лупанова (18-23 июня 2012, Москва, МГУ).

Семинары.

- Семинар "Теория автоматов" под руководством академика, профессора, д.ф.-м.н. В. Б. Кудрявцева, механико-математический факультет МГУ им. М.В. Ломоносова (2008 - 2016 г.г.).
- Семинар "Кибернетика и информатика" под руководством академика, профессора, д.ф.-м.н. В. Б. Кудрявцева, механико-математический факультет МГУ им. М. В. Ломоносова (2008 - 2016 г.г.).
- Семинар "Вопросы сложности алгоритмов поиска" под руководством академика АТН РФ, профессора, д.ф.-м.н. Э. Э. Гасанова, механико-математический факультет МГУ им. М. В. Ломоносова (2013 - 2016 г.г.).
- Семинар "Дискретный анализ" под руководством член-корр. АТН РФ, профессора, д.ф.-м.н. С. В. Алешина, механико-математический факультет МГУ им. М. В. Ломоносова (2013 - 2016 г.г.).
- Семинар "Теория графов и синтез БИС" под руководством доцента, к.ф.-м.н. А. А. Часовских, механико-математический факультет МГУ им. М. В. Ломоносова (2013 - 2016 г.г.).

Публикации. Материалы диссертации опубликованы в 5 печатных работах, из них 5 статей в списке ВАК.

Структура и объем диссертации. Диссертация состоит из введения, раздела благодарностей, 6 глав, заключения, краткого списка обозначений и библиографии. Общий объем диссертации - 213 страниц, из них 200 страниц текста. Библиография включает в себя 40 наименований на 5 страницах.

Краткое содержание работы.

Во введении кратко изложена история вопроса, поставлены основные цели и задачи диссертации, обоснованы научная новизна и практическая значимость работы, сформулированы методология и выносимые на защиту положения, описана структура диссертации.

В разделе благодарностей автор благодарит людей, без которых создание диссертационной работы было бы невозможно.

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. При формулировке результатов используются такие классические понятия из теории автоматов как *абстрактный конечный автомат*, *инициальный абстрактный конечный автомат*, *недетерменированный конечный автомат*, *инициальный недетерменированный конечный автомат*, *представимые* конечными автоматами *языки, регулярные языки* и *регулярные выражения*. Кроме того, вводятся следующие обозначения.

Алфавитное кодирование. Пусть A, B - некоторые конечные непустые алфавиты. Произвольное отображение $f : A \to B^* \setminus \{\Lambda\}$:

$$\begin{gathered} f(a_1) = \beta_1 \\ f(a_2) = \beta_2 \\ \dots \\ f(a_r) = \beta_r \end{gathered}$$

называем *схемой кодирования из алфавита A в алфавит B*. Множество всех схем кодирования из алфавита A в алфавит B обозначаем через $F(A, B)$. Доопределим отображение f до отображения $\tilde{f} : A^* \to B^*$ следующим образом:

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

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.

Автоматные обозначения. Для произвольного $n \in \mathbb{N}$ через $K(A, B, n)$ обозначается множество всех инициальных абстрактных конечных автоматов с входным алфавитом A, выходным алфавитом B и алфавитом состояний мощности n. Через $K_{\leqslant}(A, B, n)$ обозначается множество всех инициальных абстрактных конечных автоматов с входным алфавитом A, выходным алфавитом B и алфавитом состояний мощности не выше n. Для произвольного автомата V_q через $[V_q]$ обозначаем множество инициальных абстрактных конечных автоматов, полученных из V_q изменением начального состояния. Сам автомат V_q тоже входит в $[V_q]$. Для произвольного автомата V через $1(V)$ обозначаем множество слов, представимое эти автоматом последним символом 1.

Множественные обозначения. Для произвольных $n \in \mathbb{N}$, $P \subseteq A^*$ через $P_{\leqslant}(n)$ обозначаем множество слов из P, длина которых не превосходит n. Через E_2 обозначаем множество $\{0, 1\}$.

Основные результаты главы 1.

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 $W_1 \in K_{\leqslant}(B, E_2, m)$, для которого

$1(W_1) = \tilde{f}(1(V_1))$.

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

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

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

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

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

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

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

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

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

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. При формулировке результатов главы используются следующие определения.

Тонкие языки. Для $s \in \mathbb{N}$ регулярное множество $P \subseteq A^*$ называем *s-тонким в алфавите A*, если в P есть s несовпадающих слов одинаковой длины, но нет $s+1$ несовпадающих слов одинаковой длины. 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.*

Функция роста. Для $P \subseteq A^*$ через $T_n(P)$ обозначаем мощность множества $P_{\leqslant}(n)$:

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

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).

Спектральная независимость. Для произвольного $P \subseteq A^*$ называем его *спектром* множество

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

Для множеств $P_1, P_2 \subseteq A^*$ говорим, что они *спектрально независимы*, если их спектры не пересекаются. Для $r \geqslant 1$ множеств $P_1, \ldots, P_r \subseteq A^*$ говорим, что они *спектрально независимы в совокупности*, если любые два из них спектрально независимы.

Прогрессивные и общепрогрессивные множества. Если для непустого слова β в алфавите A 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*. 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. Называем множество $P \subseteq A^*$ *общепрогрессивным*, если оно является конечным объединением прогрессивных множеств с одинаковым основанием.

Основные результаты главы 2.

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.

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.

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 II класса $\mathfrak{T}(A)$ с классом регулярных языков с не более чем линейной функцией роста. При формулировке результатов главы используются следующие определения.

Линейный вид и правильный линейный вид. 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 равенство

$$\mathfrak{P} = \alpha_1 \cdot (\beta_1)^* \cdot \alpha_2 \ldots \alpha_s \cdot (\beta_s)^* \cdot \alpha_{s+1}$$

выполнено для некоторых $s \in \mathbb{N}_0$, $\alpha_1, \alpha_{s+1} \in A^*$, $\alpha_2, \ldots, \alpha_s \in A^* \setminus \{\Lambda\}$, $\beta_1, \ldots, \beta_s \in A^* \setminus \{\Lambda\}$ таких, что первые буквы (если они есть, а для α_{s+1} это не всегда так) слов β_i, α_{i+1} при всех $1 \leqslant i \leqslant s$ различны. Называем множество $P \subseteq A^*$ *множеством 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.

Основные результаты главы 3.

Теорема 3.1 *Пусть A - конечный алфавит. Любое множество $P \in RP(A)$ может быть представлено в виде конечного объединения 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.

В четвертой главе приводится решение проблемы ОАД для класса $\mathfrak{T}(A)$ тонких языков в некотором произвольном алфавите A. При формулировке результатов главы используются следующие определения.

Функция сложности L и L_p. Пусть A - конечное множество. Если выражение $\mathfrak{P}$ в алфавите A является конечной дизъюнкцией выражений вида

$$\alpha \cdot (\beta)^* \cdot \gamma, \quad \text{где} \quad \alpha, \beta, \gamma \in A^*,$$

то его *сложностью* называем максимальную из сложностей $|\alpha|+|\beta|+|\gamma|$ 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 $L_p(P)$:

$$L_p(P) := \min_{\mathfrak{P} \in \mathfrak{X}(A), |\mathfrak{P}| = P} L(\mathfrak{P}).$$

Класс $U^n(A)$**.** Для произвольного натурального $n \in \mathbb{N}$ через $U^n(A)$ обозначаем класс

$$U^n(A) := \{P \in U(A) | L_p(P) \leqslant n\}.$$

Основные результаты главы 4.

Теорема 4.1 *Пусть A, 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 $P_{\leqslant}((n+2)^2 + 4n^2 l_f^2 + l_f) \in I(f)$.*

В пятой главе приводится решение проблемы ОАД для класса $RP(A)$ регулярных языков с полиномиальной функцией роста в некотором произвольном алфавите A. При формулировке результатов главы используются следующие определения.

Классы $(W)RP^1(A), (W)RP_n^1(A), (W)RP_n(A)$. 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 $RP_n^1(A)$ 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 $RP_n^1(A)$. Для произвольного $n \in \mathbb{N}$ обозначаем через $WRP_n^1(A)$ 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 $WRP_n^1(A)$.

Основные результаты главы 5.

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 $P_{\leqslant}((n+2)^2 + (4n^2 l_f^2 + 2)^2 + l_f) \in I(f)$.*

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 $P_{\leqslant}((n^2+2)^2+(4n^4l_f^2+2)^2+l_f) \in I(f)$.*

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. При формулировке результатов главы используются следующие определения.

Проблемы ВКД_1 и ВКД_2. Пусть A, B - конечные алфавиты и есть схема $f \in F(A,B)$. 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 $f_1 \leqslant f_2$, если

$$\mathbb{R}(f_1) \subseteq \mathbb{R}(f_2).$$

Проблемой ВКД_1 в алфавитах A, B называется проблема проверки свойства

$$f_1 \leqslant f_2$$

для произвольных $f_1, f_2 \in F(A,B)$.
Пусть A, B - конечные алфавиты и есть регулярный язык $P \in R(A)$. 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 $P_1 \leqslant P_2$, если

$$\mathbb{F}(P_1) \subseteq \mathbb{F}(P_2).$$

Проблемой ВКД_2 в алфавитах A, B называется проблема проверки свойства

$$P_1 \leqslant P_2$$

для произвольных $P_1, P_2 \in R(A)$.

Основные результаты главы 6.

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.

Заключение.

В диссертации получены следующие основные результаты.

- Получены верхние полиномиальные оценки на сложность решения проблемы однозначности алфавитного кодирования в классах тонких языков и регулярных языков с полиномиальной функцией роста.
- Построена полная неизбыточная классификация внутренней структуры элементов из класса тонких языков в терминах конечных объединений прогрессивных множеств.
- Построена полная неизбыточная классификация внутренней структуры элементов из класса регулярных языков с полиномиальной функцией роста 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 сравнения полезности произвольной пары функций алфавитного кодирования в общих алфавитах.
- Доказана алгоритмическая разрешимость проблемы сравнения полезности произвольной пары регулярных языков в общем алфавите, когда мощность этого алфавита равна двум.
- Предложена модель, демонстрирующая применение полученных результатов на практике.

Полученные результаты демонстрируют, что алфавитное кодирование элементов из классов тонких языков и регулярных языков с полиномиальной функцией роста может быть применено как в области сжатия информации, так и при передаче информации по каналу связи на длительные расстояния.

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.

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 сверхъязыков.* 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.