УДК 510.65

О РАЗРЕШИМЫХ И НЕРАЗРЕШИМЫХ ТЕОРИЯХ ЯЗЫКОВ

Карлов Б.Н.
Тверской государственный университет, г. Тверь

Поступила в редакцию 01.12.2024, после переработки 16.12.2024.

В работе исследованы некоторые теории языков с различными операциями. Доказана разрешимость теорий языков с операциями дополнения, обращения, циклического сдвига, а также некоторых фрагментов теории конкатенации, когда один из языков фиксирован. Установлена неразрешимость теории ортогональной конкатенации.

Ключевые слова: формальный язык, теория, разрешимость, унар, ортогональная конкатенация.

Вестник ТвГУ. Серия: Прикладная математика. 2024. № 4. С. 30–39.
https://doi.org/10.26456/vtpmk725

Введение

Теория формальных языков изучает различные способы описания языков (автоматы, грамматики и пр.), а также свойства получающихся при этом классов. Одной из важных задач в этой области является исследование алгоритмических свойств различных семейств языков. В качестве классических результатов можно упомянуть неразрешимость проблемы эквивалентности для контекстно-свободных грамматик и PSPACE-полноту аналогичной проблемы для регулярных выражений. Однако можно рассматривать свойства, относящиеся не к отдельным языкам, а к некоторым их семействам в целом. Например, можно изучать теорию всех конечных или контекстно-свободных языков в некотором фиксированном алфавите с заданным набором операций. В статье [7] были получены некоторые результаты в этом направлении. В частности, было установлено, что теория всех регулярных языков в произвольном алфавите с операцией конкатенации неразрешима.

В настоящей статье мы продолжаем изучать алгоритмические свойства классов языков с некоторыми естественными операциями. В разделе 2 мы доказываем разрешимость теорий языков с операциями дополнения, обращения и циклического сдвига, а также разрешимость некоторых фрагментов теории конкатенации. В разделе 3 мы доказываем неразрешимость теории языков в алфавите мощности не менее 2 с операцией ортогональной конкатенации.

1. Предварительные сведения

Унар — это алгебра вида $\mathfrak{A} = (A, f)$, где f — одноместная функция. Если f разнозначна, то основное множество A распадается на несколько (возможно, бесконечно много) классов одного из трёх видов:

1. конечные классы мощности n:
$$C = \{ x_1, \dots, x_n \}, \quad f(x_i) = x_{i+1}, \quad f(x_n) = x_1;$$
2. бесконечные классы типа целых чисел $\mathbb{Z}$:
$$C = \{ \dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots \}, \quad f(x_i) = x_{i+1};$$
3. бесконечные классы типа натуральных чисел ω:
$$C = \{ x_0, x_1, x_2, \dots \}, \quad f(x_i) = x_{i+1}, \quad f(y) \neq x_0 \text{ ни для какого } y \in A.$$

С каждым унаром $\mathfrak{A}$ с разнозначной функцией f связывается функция
$$\chi_{\mathfrak{A}} \colon \omega \cup \{ \omega, \mathbb{Z} \} \to \omega \cup \{ \infty \},$$
определяемая следующим образом (через ω обозначено множество натуральных чисел, включая 0):

1. $\chi_{\mathfrak{A}}(n) = k$, если $\mathfrak{A}$ содержит конечное число $k \geqslant 0$ классов мощности n, в противном случае $\chi_{\mathfrak{A}}(n) = \infty$;
2. $\chi_{\mathfrak{A}}(\omega) = k$, если $\mathfrak{A}$ содержит конечное число $k \geqslant 0$ классов типа ω, в противном случае $\chi_{\mathfrak{A}}(\omega) = \infty$;
3. $\chi_{\mathfrak{A}}(\mathbb{Z}) = 1$, если для любого $n \geqslant 0$ в унаре $\mathfrak{A}$ имеется класс (конечный или бесконечный), содержащий более n элементов, в противном случае $\chi_{\mathfrak{A}}(\mathbb{Z}) = 0$.

Через $R_{\mathfrak{A}}$ обозначается следующее отношение на множестве натуральных чисел:
$$R_{\mathfrak{A}} = \{ (n, k) : \chi_{\mathfrak{A}}(n) \geqslant k \}.$$
Мы считаем, что $\infty > n$ для любого $n \in \omega$.

В работе [1] был доказан критерий элементарной эквивалентности двух унаров с разнозначными функциями, а в [2] был установлен критерий разрешимости теории такого унара.

Теорема 1. *Пусть $\mathfrak{A} = (A, f)$ и $\mathfrak{B} = (B, f)$ — два унара, в каждом из которых функция f разнозначна. $\mathfrak{A}$ и $\mathfrak{B}$ элементарно эквивалентны тогда и только тогда, когда $\chi_{\mathfrak{A}} = \chi_{\mathfrak{B}}$.*

Теорема 2. *Пусть $\mathfrak{A} = (A, f)$ — унар, в котором функция f разнозначна. Теория унара $\mathfrak{A}$ разрешима тогда и только тогда, когда отношение $R_{\mathfrak{A}}$ рекурсивно.*

Из теоремы 2 следует, что если множество мощностей классов унара $\mathfrak{A}$ с разнозначной функцией конечно, то теория такого унара разрешима.

2. Унары языков с разрешимыми теориями

В этом разделе мы рассматриваем унары $\mathfrak{A} = (A, f)$, в которых A — некоторое множество языков, замкнутое относительно операции f. Мы применяем теорему 2 к некоторым конкретным алгебраическим системам и устанавливаем разрешимость теорий некоторых «естественных» унаров языков. В следующих двух теоремах приведены простейшие примеры разрешимых унаров.

Теорема 3. *Если $f(x) = \overline{x}$ — дополнение языка x, то теория унара $\mathfrak{A} = (A, f)$ разрешима для любого основного множества A.*

Доказательство. Так как для любого множества L выполняется $\overline{L} \neq L$, $\overline{\overline{L}} = L$, то все классы унара имеют мощность 2. Следовательно, по теореме 2 его теория разрешима. □

Напомним, что *обращение* слова $w = a_1 a_2 \dots a_n$ — это слово $w^{-1} = a_n \dots a_2 a_1$, а обращение языка L — это язык $L^{-1} = \{\, w^{-1} : w \in L \,\}$.

Теорема 4. *Если $f(x) = x^{-1}$ — обращение языка x, то теория унара $\mathfrak{A} = (A, f)$ разрешима для любого основного множества A.*

Доказательство. Для любого языка L выполняется либо $L^{-1} = L$, либо $L^{-1} \neq L$, $(L^{-1})^{-1} = L$, поэтому все классы унара имеют мощность 1 или 2. Следовательно, по теореме 2 его теория разрешима. □

Теперь мы приведём пример разрешимого фрагмента теории языков с операцией конкатенации. Конкатенацию языков L_1 и L_2 мы будем обозначать $L_1 \cdot L_2$ или просто $L_1 L_2$. Мы зафиксируем один из языков L и разрешим выполнять конкатенацию произвольного языка только с L. При этом мы ограничим выбор L беспрефиксными и бессуффиксными языками.

Определение 1. *Язык L называется* беспрефиксным *(соответственно* бессуффиксным*), если ни одно слово из L не является префиксом (соответственно суффиксом) другого.*

Непосредственно из определения следует, что если беспрефиксный или бессуффиксный язык L содержит пустое слово, то $L = \{\, \varepsilon \,\}$. Докажем два простых свойства беспрефиксных языков.

Лемма 1. *Если $L \neq \varnothing$ — беспрефиксный язык, то $L \cdot L_1 = L \cdot L_2$ тогда и только тогда, когда $L_1 = L_2$ для любых языков L_1 и L_2.*

Доказательство. Пусть x — некоторое слово языка L. Рассмотрим произвольное слово $y \in L_1$ и докажем, что $y \in L_2$. Слово xy принадлежит языку $L \cdot L_1$, следовательно, $xy \in L \cdot L_2$. Это значит, что существуют слова $u \in L$ и $v \in L_2$ такие, что $xy = uv$. Если $x \neq u$, то одно из слов x, u является префиксом другого, что невозможно по определению беспрефиксного языка. Значит, $x = u$, $y = v$ и $y \in L_2$. Включение $L_2 \subseteq L_1$ проверяется симметрично. □

Лемма 2. *Если L — беспрефиксный язык, отличный от $\varnothing$ и $\{\varepsilon\}$, то для любого непустого языка L_0 все языки $L^i \cdot L_0$ попарно различны.*

Доказательство. Так как язык L беспрефиксный и $L \neq \{\varepsilon\}$, то $\varepsilon \notin L$. Следовательно, кратчайшие слова языков L^i и L^j имеют разную длину при $i \neq j$. Поэтому кратчайшие слова языков $L^i \cdot L_0$ и $L^j \cdot L_0$ также различны, а значит, языки не равны. □

Аналогичные свойства имеют место и для бессуффиксных языков.

Лемма 3. *Если $L \neq \varnothing$ — бессуффиксный язык, то $L_1 \cdot L = L_2 \cdot L$ тогда и только тогда, когда $L_1 = L_2$ для любых языков L_1 и L_2.*

Лемма 4. *Если L — бессуффиксный язык, отличный от $\varnothing$ и $\{\varepsilon\}$, то для любого непустого языка L_0 все языки $L_0 \cdot L^i$ попарно различны.*

В следующей теореме доказывается разрешимость некоторого фрагмента теории языков с операцией конкатенации.

Теорема 5. *Если $f(x) = L \cdot x$ — конкатенация с некоторым фиксированным беспрефиксным языком L, то теория унара $\mathfrak{A} = (A, f)$ разрешима для любого основного множества A.*

Доказательство. Если $L = \{\varepsilon\}$, то все термы вида $f^i(x)$ можно заменить на x. Следовательно, теория унара разрешима как теория чистого равенства (см. [8]).

Если $L = \varnothing$, то все термы, содержащие конкатенацию, можно заменить на $\varnothing$. Теория унара снова разрешима как теория чистого равенства с дополнительным символом константы.

Если $L \neq \varnothing$ и $L \neq \{\varepsilon\}$, то по лемме 1 функция f разнозначна. Из леммы 2 следует, что $\chi_{\mathfrak{A}}(n) = 0$ для всех чисел n. Следовательно, по теореме 2 теория унара $\mathfrak{A}$ разрешима. □

Аналогичный результат имеет место и для бессуффиксных языков.

Теорема 6. *Если $f(x) = x \cdot L$ — конкатенация с некоторым фиксированным бессуффиксным языком L, то теория унара $\mathfrak{A} = (A, f)$ разрешима для любого основного множества A.*

Следующая теорема является следствием теоремы 1.

Теорема 7. *Все унары с операциями $f(x) = L_1 \cdot x$ и $f(x) = x \cdot L_2$, где L_1 и L_2 — беспрефиксный и бессуффиксный языки соответственно, элементарно эквивалентны.*

Доказательство. Все эти унары содержат бесконечно много классов типа ω и не содержат других классов. Следовательно, их функции χ равны, а значит, унары элементарно эквивалентны. □

Приведём теперь пример унара языков с разрешимой теорией, содержащего бесконечно много конечных классов. Через $\mathrm{cycle}_k(w)$ обозначим операцию циклического сдвига слова w вправо на k позиций:

$$\mathrm{cycle}_k(a_1 \dots a_n) = a_{n-k+1} \dots a_n a_1 \dots a_{n-k}.$$

Циклический сдвиг влево на k позиций можно рассматривать как сдвиг вправо на $-k$ позиций. Нетрудно видеть, что операция cycle_k разнозначна.

Теорема 8. *Теория множества слов Σ^* с операцией cycle_k разрешима для любого k.*

Доказательство. Если алфавит Σ содержит единственный символ a, то $\mathrm{cycle}_k(w) = w$ для любого k и любого w, поэтому теория разрешима. Пусть Σ содержит два разных символа a и b. Рассмотрим слова $w_{n,i} = (a^{kn-1}b)^i$ для $n = 1, 2, 3, \dots$ и $i = 1, 2, 3 \dots$. Нетрудно видеть, что в результате n-кратного применения операции cycle_k к слову $w_{n,i}$ снова получится $w_{n,i}$, так как блок $a^{kn-1}b$ совместится с таким же блоком ровно после n сдвигов. Поэтому унар слов содержит бесконечно много классов мощности n для любого натурального n, а значит, его теория разрешима по теореме 2. □

Теорема 9. *Если множество языков $\mathfrak{L}$ содержит все одноэлементные языки, то теория множества $\mathfrak{L}$ с операцией cycle_k разрешима для любого k.*

Доказательство. Так как $\mathfrak{L}$ содержит все одноэлементные языки, то унар содержит бесконечно много классов всех конечных мощностей. Значит, по теореме 2 его теория разрешима. □

Из критерия элементарной эквивалентности унаров получается следующий результат.

Теорема 10. *Если множество языков $\mathfrak{L}$ содержит все одноэлементные языки, то теории слов с операцией cycle_k и теории множества $\mathfrak{L}$ с операцией cycle_l элементарно эквивалентны для всех k и l.*

Доказательство. Все унары не содержат классов типа ω и содержат бесконечно много классов всех конечных мощностей. Следовательно, они элементарно эквивалентны по теореме 1. □

3. Неразрешимость теории ортогональной конкатенации

Операция ортогональной конкатенации, введённая в статье [4], является аналогом прямой суммы линейных подпространств или прямого произведения подгрупп.

Определение 2. *Ортогональная конкатенация языков L_1 и L_2 определяется следующим образом: $L_1 \odot_\perp L_2 = L_1 \cdot L_2$, если для каждого слова $w \in L_1 \cdot L_2$ существуют единственные слова $u \in L_1$, $v \in L_2$ такие, что $uv = w$; в противном случае $L_1 \odot_\perp L_2 = \varnothing$.*

Отметим, что в работе [5] полагается, что $L_1 \odot_\perp L_2$ не определено при нарушении единственности разложения. Мы считаем, что в этом случае результат равен $\varnothing$, чтобы можно было использовать $\odot_\perp$ как функциональный символ. Далее мы покажем, что эти два варианта определимы друг через друга.

Лемма 5. *Следующая формула $E(x)$ определяет пустое множество:*

$$(\forall y) x \odot_\perp y = x.$$

Доказательство. Если $x = \varnothing$, то $xy = \varnothing$, а значит, и $x \odot_\perp y = \varnothing = x$. Наоборот, пусть $x \neq \varnothing$. Положим $y = \{\, a \,\}$, тогда $x \odot_\perp y \neq \varnothing$, поскольку каждое слово $w \in xy$ допускает только одно представление вида $w = va$, $v \in x$. Следовательно, кратчайшее слово языка $x \odot_\perp y$ длиннее кратчайшего слова языка x, поэтому $x \odot_\perp y \neq x$. □

Пусть теперь $\mathrm{Orth}(x, y, z)$ означает, что $x \odot_\perp y = z$ в смысле работы [5]. Пустое множество определимо с помощью Orth формулой $(\forall y)\mathrm{Orth}(x, y, x)$. Тогда следующие две эквивалентности показывают, как определить один из символов Orth и $\odot_\perp$ через другой:

$$x \odot_\perp y = z \equiv \mathrm{Orth}(x, y, z) \vee (\neg(\exists u)\mathrm{Orth}(x, y, u) \wedge z = \varnothing),$$
$$\mathrm{Orth}(x, y, z) \equiv x \odot_\perp y = z \wedge (z = \varnothing \to (x = \varnothing \vee y = \varnothing)).$$

Далее мы покажем, как определить беспрефиксные и бессуффиксные языки с помощью ортогональной конкатенации.

Лемма 6. *Следующая формула $P(x)$ определяет беспрефиксные языки:*

$$(\forall y)(y \neq \varnothing \to x \odot_\perp y \neq \varnothing).$$

Доказательство. Пусть x — беспрефиксный язык, и пусть $y \neq \varnothing$. Предположим, что некоторое слово $w \in xy$ представляется в виде конкатенации слов из x и y двумя способами: $w = u_1v_1 = u_2v_2$, где $u_1, u_2 \in x$, $v_1, v_2 \in y$. Если $u_1 \neq u_2$, то одно из этих слов является префиксом другого, что противоречит тому, что язык x беспрефиксный. Если же $u_1 = u_2$, то также $v_1 = v_2$, поэтому двух разных представлений не существует.

Предположим теперь, что язык x не является беспрефиксным. Тогда существуют слова u и v такие, что $u, uv \in x$, $v \neq \varepsilon$. Выберем в качестве y язык $\{\, \varepsilon, v \,\}$. Тогда язык xy содержит слово uv, которое представимо в виде конкатенации двумя способами: $u \cdot v = uv \cdot \varepsilon$. Следовательно, $x \odot_\perp y = \varnothing$. □

Аналогично определяются и бессуффиксные языки.

Лемма 7. *Следующая формула $S(x)$ определяет бессуфиксные языки:*

$$(\forall y)(y \neq \varnothing \to y \odot_\perp x \neq \varnothing).$$

Теперь мы можем доказать неразрешимость теории ортогональной конкатенации.

Теорема 11. *Если $|\Sigma| \geqslant 2$, то теория алгебры $(\mathsf{P}(\Sigma^*), \odot_\perp)$ неразрешима.*

Доказательство. Формула $x \neq \varnothing \wedge P(x) \wedge S(x)$ выделяет все непустые языки, являющиеся одновременно беспрефиксными и бессуффиксными. Поскольку теория конкатенации таких языков неразрешима (см. [3]), то и теория ортогональной конкатенации неразрешима. □

Заключение

В работе были приведены примеры унаров языков с разрешимыми теориями, в частности, доказана разрешимость теории конкатенации с фиксированным беспрефиксным или бессуффиксным языком. Кроме того было доказано, что алгебра языков в алфавите мощности не менее 2 с операцией ортогональной конкатенации неразрешима. Интерес представляет поиск других разрешимых фрагментов теории конкатенации, а также исследование разрешимости теорий языков с различными бинарными операциями.

Список литературы

[1] Карлов Б.Н. Об элементарной эквивалентности некоторых уноидов и уноидов их подмножеств // Вестник ТвГУ. Серия: Прикладная математика. 2021. № 3. С. 18–32. https://doi.org/10.26456/vtpmk620

[2] Карлов Б.Н. О неразрешимости теорий подмножеств некоторых унаров // Доклады Российской академии наук. Математика, информатика, процессы управления. 2024. Т. 516, № 1. С. 15–20. https://doi.org/10.31857/S2686954324020035

[3] Карлов Б.Н. О теории конкатенации одного класса языков // Материалы международной конференции «Алгебра и математическая логика: теория и приложения». Казань: КФУ, 2024. С. 206–207.

[4] Anselmo M., Restivo A. On languages factorizing the free monoid // International Journal of Algebra and Computation. 1996. Vol. 6, № 4. Pp. 413–427. https://doi.org/10.1142/S0218196796000246

[5] Daley M., Domaratzki M., Salomaa K. Orthogonal concatenation: language equations and state complexity // Journal of Universal Computer Science. 2010. Vol. 16, № 5. Pp. 653–675. https://doi.org/10.3217/jucs-016-05-0653

[6] Dudakov S.M. On undecidability of concatenation theory for one-symbol languages // Lobachevskii Journal of Mathematics. 2020. Vol. 41, № 2. Pp. 168–175. https://doi.org/10.1134/S1995080220020055

[7] Dudakov S., Karlov B. On decidability of theories of regular languages // Theory of Computing Systems. 2021. Vol. 65. Pp. 462–478. http://doi.org/10.1007/s00224-020-09995-4

[8] Monk J.D. Mathematical logic. New York, Heidelberg, Berlin: Springer-Verlag, 1976. 542 p.

Образец цитирования

Карлов Б.Н. О разрешимых и неразрешимых теориях языков // Вестник ТвГУ. Серия: Прикладная математика. 2024. № 4. С. 30–39. https://doi.org/10.26456/vtpmk725

Сведения об авторах

1. **Карлов Борис Николаевич**
доцент кафедры информатики Тверского государственного университета.

Россия, 170100, г. Тверь, ул. Желябова, 33, ТвГУ. E-mail: bnkarlov@gmail.com

ON DECIDABLE AND UNDECIDABLE THEORIES OF LANGUAGES

Karlov B.N.
Tver State University, Tver

Received 01.12.2024, revised 16.12.2024.

In this paper we study some theories of languages with different operations. We prove the decidability of such theories with the operations of complement, reversal, and cyclic shift, and also the decidability of some fragments of concatenation theory when one of the languages is fixed. We establish the undecidability of the theory of orthogonal concatenation.

Keywords: formal language, theory, decidability, unar, orthogonal concatenation.

Citation

Karlov B.N., "On decidable and undecidable theories of languages", *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2024, № 4, 30–39 (in Russian). https://doi.org/10.26456/vtpmk725

References

[1] Karlov B.N., "On elementary equivalence of some unoids and unoids of their subsets", *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2021, № 3, 18–32 (in Russian), https://doi.org/10.26456/vtpmk620.

[2] Karlov B.N., "On undecidability of subset theories of some unars", *Doklady Mathematics*, **109**:2 (2024), 112–116, https://doi.org/10.31857/S2686954324020035.

[3] Karlov B.N., "On the theory of concatenation of one class of languages", *Materialy mezhdunarodnoj konferentsii ≪Algebra i matematicheskaya logika: teoriya i prilozheniya≫ [Proceedings of the international conference "Algebra and Mathematical Logic: Theory and Applications"]*, KFU, Kazan, 2024, 206–207 (in Russian).

[4] Anselmo M., Restivo A., "On languages factorizing the free monoid", *International Journal of Algebra and Computation*, **6**:4 (1996), 413–427, https://doi.org/10.1142/S0218196796000246.

[5] Daley M., Domaratzki M., Salomaa K., "Orthogonal concatenation: language equations and state complexity", *Journal of Universal Computer Science*, **16**:5 (2010), 653–675, https://doi.org/10.3217/jucs-016-05-0653.

[6] Dudakov S.M., "On undecidability of concatenation theory for one-symbol languages", *Lobachevskii Journal of Mathematics*, **41**:2 (2020), 168–175, https://doi.org/10.1134/S1995080220020055.

[7] Dudakov S., Karlov B., "On decidability of theories of regular languages", *Theory of Computing Systems*, **65** (2021), 462–478, http://doi.org/10.1007/s00224-020-09995-4.

[8] Monk J.D., *Mathematical logic*, Springer-Verlag, New York, Heidelberg, Berlin, 1976, 542 pp.

Author Info

1. **Karlov Boris Nikolaevich**
Associate Professor at Computer Science Department, Tver State University.

Russia, 170100, Tver, 33 Zhelyabova str., TverSU. E-mail: bnkarlov@gmail.com