Лабораторная работа №4

Лиганкина Анна
Вариант 20

17 января 2026 г.

Содержание

1 Задание

Дан язык:

$$L = \hat{}((?: a|b)^*)c((?: a|b)^*)(?= \backslash 1ab\backslash 2)\backslash 2ba\backslash 1\$$$

Необходимо:

- проанализировать язык на КС-свойство, в случае его наличия - на регулярность;
- построить «наивный» парсер слов для языка, используя рекурсивный разбор с возвратами (парсер не должен зацикливаться);
- построить оптимизированный парсер слов для языка. Оценить сверху его вычислительную сложность;
- посредством фазз-тестирования проверить эквивалентность парсеров и построить сравнительные графики их времени работы на случайных словах, принадлежащих языку и не принадлежащих языку (два тестовых пула).

2 Исследование языка

Заметим, что исходный язык можно переписать в виде:

$$L = \{w \mid w = xcyybax \;\&\; x = (a|b)^* \;\&\; y = (a|b)^* \;\&\; xaby = ybax\}$$

так как длина предпросмотра совпадает с длиной конца слова ($|\backslash 1ab\backslash 2| = |\backslash 2ba\backslash 1|$).

Доказательство, что язык не является КС

Так как контекстно-свободные языки замкнуты относительно пересечения с регулярными языками, то пересечем L с $R = a^*ca^*ba^+$ и докажем, что получившийся язык $L_R = L \cap R$ не КС.

Слово x однозначно определяется разделителем c и состоит из букв a, либо является пустым словом. Пусть $x = a^n, n \geq 0$. Так как в словах из R содержится лишь одна буква b, то она соответствует букве b из обязательного блока ba, а значит, y тоже состоит из букв a. Пусть $y = a^m, m \geq 0$.

Рассмотрим условие $xaby = ybax$:

$$a^n aba^m = a^m baa^n$$

$$a^{n+1}ba^m = a^m ba^{n+1} \Rightarrow m = n+1$$

Таким образом, язык L_R выглядит так:

$$L = \{w \mid w = a^n ca^{2n+2} ba^{n+1} \; \& \; n \geq 0\}$$

По лемме о накачке (длина накачки N) рассмотрим слово $a^N ca^{2N+2} ba^{N+1}$ $\in L$:

- если $|vxy| < N$ попадает в один из блоков из букв a, то нарушается баланс с другими буквами a, и мы выходим из языка;
- если $|vxy| < N$ попадает на стык блоков букв a и накачка задевает букву c или букву b, то нарушается структура слова, где лишь одна буква b и одна буква c, и мы выходим из языка;
- если $|vxy| < N$ попадает на стык блоков букв a и мы качаем одновременно 2 блока букв a (3 не можем, так как $|vxy| < N$), то все равно рушится баланс с оставшимся блоком a, и мы выходим из языка.

Поэтому язык L_R не является контекстно-свободным, а, следовательно, и язык L.

3 Наивный парсер

Наивный парсер считывает и запоминает x до символа c. Далее, он пытается рекурсивно «угадать» длину слова y. Запоминает y, а затем бежит по хвосту слова и сравнивает его одновременно с $xaby$ и $ybax$. Если y достиг конца слова, то исходное слово не принадлежит языку. Таким образом, оценка сложности данного парсера - $o(n^2)$, где n - длина входного слова.

4 Оптимизированный парсер

Оптимизация парсера происходит за счет следующего:

- если длина слова четна, то слово автоматически не подходит языку;

- длина y однозначно вычисляется по формуле «(длина слова - 3)/2 - длина x», благодаря чему алгоритмическая сложность парсера - $o(n)$, где n - длина входного слова;
- слова x и y являются палиндромами, а также $xaby$ и $ybax$, так как $xaby = ybax$.

Доказательство того, что если $xaby = ybax$, то x, y и $xaby = ybax$ - палиндромы.

Докажем по индукции (по длине $|x| + |y|$).
База индукции. Пусть $x = \varepsilon$. Тогда $aby = yba \Rightarrow y \neq \varepsilon$. Т.к. левая часть начинается с a, то $y = ay_1$. Тогда

$$abay_1 = ay_1ba \Rightarrow bay_1 = y_1ba$$

Отсюда, либо $y_1 = \varepsilon$, либо y_1 начинается на b, т.е. $y_1 = by_2$

$$baby_2 = by_2ba \Rightarrow aby_2 = y_2ba$$

Далее, $y_2 \neq \varepsilon$ и начинается на a, т.е. $y_2 = ay_3$ и т.д. Получаем, что $y = a(ba)^k, k \geq 0$ - палиндром.
Аналогично, если $y = \varepsilon$, то $x = b(ab)^k, k \geq 0$ - палиндром.
Шаг индукции. Пусть $x \neq \varepsilon$ и $y \neq \varepsilon$. Пусть

$$w = xaby = ybax$$

Тогда более короткое слово из x, y является префиксом другого (при этом $|x| \neq |y|$, т.к. получится, что $ab = ba$). 2 случая:

1. $|x| < |y|$. Тогда x является префиксом y, т.е. $y = xy_1$, где $y_1 \neq \varepsilon$.

   $$xabxy_1 = xy_1bax \Rightarrow abxy_1 = y_1bax$$

   А значит, y_1 начинается с a, т.е. $y_1 = ay_2$

   $$abxay_2 = ay_2bax \Rightarrow bxay_2 = y_2bax$$

   Здесь также 2 случая:

   (a) $y_2 = \varepsilon$. Тогда $y = xa$ и

   $$bxa = bax \Rightarrow xa = ax \Rightarrow x = a^n, y = a^{n+1},$$

   и x, y - палиндромы; $w = a^naba^{n+1} = a^{n+1}baa^n = a^{n+1}ba^{n+1}$ - палиндром.

(b) $y_2 = by_3$. Тогда $y = xaby_3$ и

$$bxaby_3 = by_3bax \Rightarrow xaby_3 = y_3bax$$

Мы пришли к тому же уравнению вида $xaby = ybax$, но длина $|x| + |y_3|$ уменьшилась $\Rightarrow$ по предположению индукции x, y_3 и $xaby_3 = y_3bax$ - палиндромы. Отсюда $y = xaby_3$ - палиндром и

$$w^R = (xaby)^R = x^R(ab)^R y^R = (x = x^R, y = y^R) = xbay = w,$$

т.е. w - палиндром.

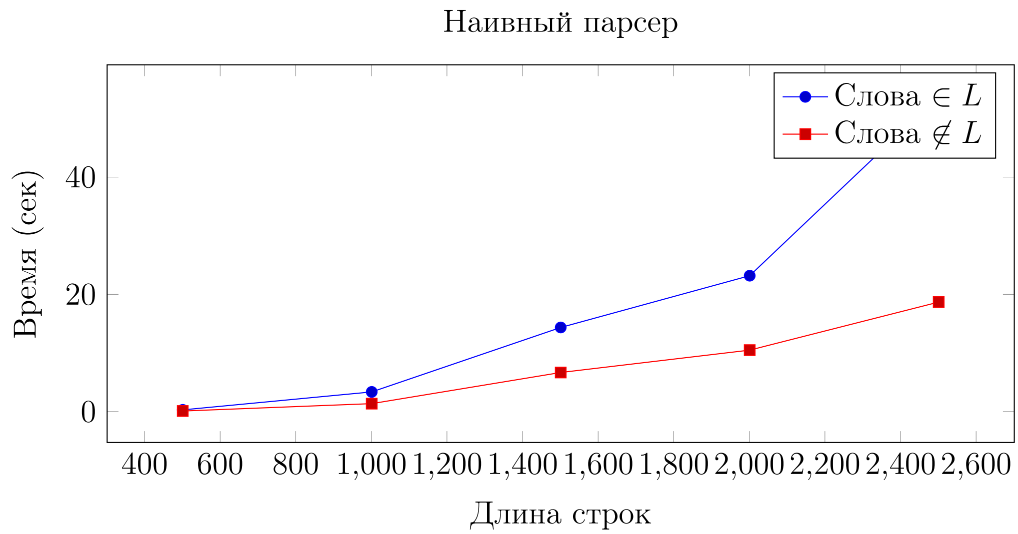
2. $|y| < |x|$ - симметричная ситуация. Тогда y является префиксом x, т.е. $x = yx_1$, где $x_1 \neq \varepsilon$. И 2 аналогичных исхода:

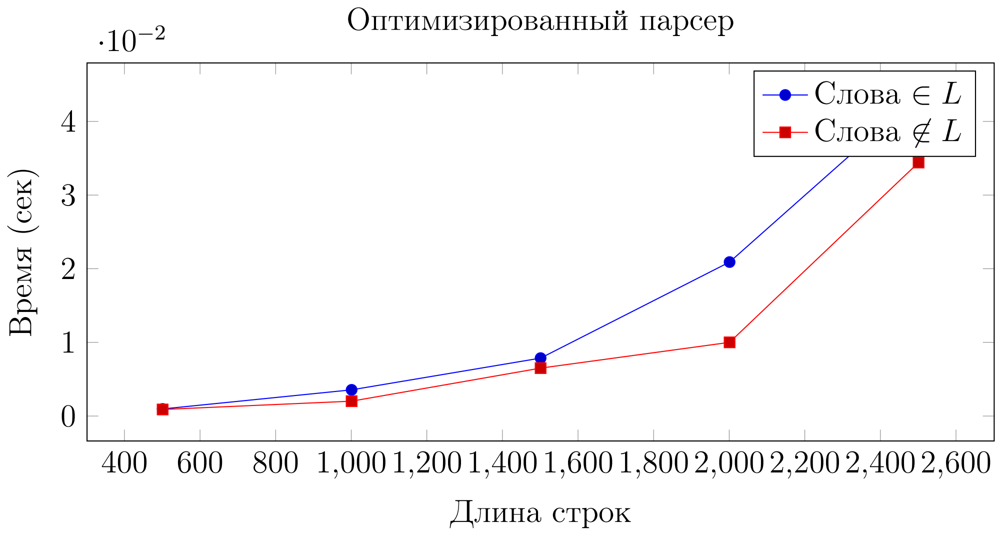
   (a) $x = b^{n+1}, y = b^n$ и $w = b^{n+1}abb^n = b^nbab^{n+1}$ - палиндромы.

   (b) $x = ybax_3, y$ и w - палиндромы.

Таким образом, x, y и w - палиндромы.

5 Оценка эффективности




Оптимизированный парсер
$\cdot 10^{-2}$
Время (сек)
0
1
2
3
4
400 600 800 1,000 1,200 1,400 1,600 1,800 2,000 2,200 2,400 2,600
Длина строк
Слова $\in L$
Слова $\notin L$