Известное преимущество эффективности операционной семантики Рефала

И

его связь с распознаванием симметрии слов (+ об актуальной альфа-версии системы программирования Рефал-5)

Андрей П. Немытых Институт программных систем РАН г. Переславль-Залесский

Совместное рабочее совещание ИПС РАН и МГТУ имени Н.Э. Баумана по функциональному языку программирования Рефал

1-2 июля 2025 г., Переславль-Залесский

Данными языка программирования Рефал являются конечные последовательности термов.

```
Определение
```

```
data ::= term_data || \varepsilon \varepsilon ::= -- пустая последовательность; term ::= symbol || (data) symbol ::= CHARACTER || IDENTIFIER || MACRO-DIGIT
```

Элементарные инструменты для конструирования данных:

Данными языка программирования Рефал являются конечные последовательности термов.

Определение

```
data ::= term_data || \varepsilon \varepsilon ::= -- пустая последовательность; term ::= symbol || (data) symbol ::= CHARACTER || IDENTIFIER || MACRO-DIGIT
```

Элементарные инструменты для конструирования данных:

• Приписывание одной данной последовательности термов к другой последовательности термов.

Данными языка программирования Рефал являются конечные последовательности термов.

Определение

```
data ::= term_data || \varepsilon

\varepsilon ::= -- пустая последовательность;

term ::= symbol || (data)

symbol ::= CHARACTER || IDENTIFIER || MACRO-DIGIT
```

Элементарные инструменты для конструирования данных:

- Приписывание одной данной последовательности термов к другой последовательности термов.
- Построение узла (вершины) дерева.

Данными языка программирования Рефал являются конечные последовательности термов.

```
Определение
```

```
data ::= term_data || \varepsilon

\varepsilon ::= -- пустая последовательность;

term ::= symbol || (data)

symbol ::= CHARACTER || IDENTIFIER || MACRO-DIGIT
```

Элементарные инструменты для конструирования данных:

- Приписывание одной данной последовательности термов к другой последовательности термов.
 - Бинарный конструктор приписывания _ является ассоциативным.
 - Унарный конструктор (data) не путать с
 композиционными скобками, которые здесь не нужны.
 Рефал 2025/ 01-07-2025

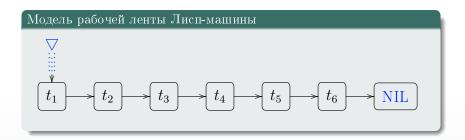
Рефал-машина (интерпретатор)

Данными языка программирования Рефал являются конечные последовательности термов.



Лисп-машина (потомки языка Лисп)

Данными языков программирования — потомков Лиспа являются конечные списки термов.



- Программа есть система переписывания термов.
 - На каждом шаге Рефал-интерпретатор выбирает единственное правило переписывания функции F, вызов которой

<F конкретные аргументы>
в данный момент находится на вершине стека вызовов.

 Это правило выбирается посредством отождествления аргументов

активного вызова с левыми частями правил переписывания функции F.

- Программа есть система переписывания термов.
 - На каждом шаге Рефал-интерпретатор выбирает единственное правило переписывания функции F, вызов которой

<F конкретные аргументы>
в данный момент находится на вершине стека вызовов.

• Это правило выбирается посредством отождествления аргументов

активного вызова

с левыми частями правил переписывания функции F.

Пример:

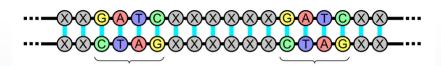
 Предикат распознавания симметрии последовательности символов.

Примеры палиндромов

- нежен
- Волгу див несет, тесен вид углов
- число-палиндром 121 является результатом:
 - произведения-палиндрома 11 * 11;
 - суммы-палиндрома 56 + 65;
- madam
- deified

Примеры палиндромов

В молекулах ДНК присутствует от 100 тысяч до 1 млн. коротких палиндромных последовательностей. Молекула ДНК состоит из двух комплементарных цепочек нуклеотидов, которые всегда соединяются одним и тем же образом (аденин (A) с тимином (T), цитозин (C) с гуанином (G)). Палиндромные участки распределены по ДНК неравномерно. Они играют важную роль в формировании некоторых типов нуклеиновых кислот, например, в случае транспортных РНК.



Пример:

 Предикат распознавания симметрии последовательности символов.

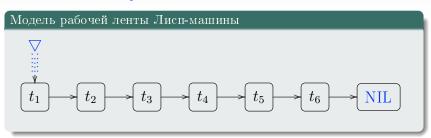
 Время распознавания (не)симметрии любой последовательности символов линейно от длины этой последовательности. Рефал-машина (интерпретатор)

Данными языка программирования Рефал являются конечные последовательности термов.



Лисп-машина (потомки языка Лисп)

Данными языков программирования — потомков Лиспа являются конечные списки термов.



- Найдётся последовательность символов, (не)симметрия которой не может быть распознана в этой модели за линейное время от длины этой последовательности.
- Таких последовательностей очень много «почти все».

Лисп-машина (потомки языка Лисп)

Данными языков программирования — потомков Лиспа являются конечные списки термов.

- Найдётся последовательность символов, (не)симметрия которой не может быть распознана в модели вычислений Лисп-машины за линейное время от длины этой последовательности.
- Таких последовательностей очень много «почти все».

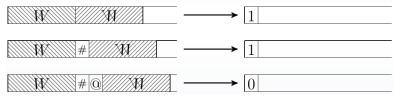
Определение

Пусть S(n) — число слов длины n, обладающих свойством S, $E_S(n)$ — число слов длины n, обладающих как свойством S так и свойством E. Говорят, что почти все слова, обладающие свойством S, обладают и свойством E, если $\lim_{n\to\infty}\frac{E_S(n)}{S(n)}=1$.

Краткая история вопроса

Модель вычислений: одноленточная машина Тьюринга.

• Для данного слова W длины n на лентах даны слова $U_1 = W\overline{W}, \ U_2 = W\prime\# \imath \overline{W}, \ U_3 = W\prime \# @ \imath \overline{W}.$



Задача распознавания симметрии слова.

С середины 1960-х годов известно, что решение этой задачи требует времени пропорционального n^2 для почти всех слов W.

• Т.е. для любой машины Тьюринга \mathcal{M} , решающей эту задачу для всех слов, $\exists \epsilon, 0 < \epsilon \in \mathbb{R}, \exists N_0 \in \mathbb{N}$ т.ч. $\forall n > N_0$, для почти всех слов W длины n распознавание симметрии слов U_1, U_2 (посредством \mathcal{M}) длится дольше $\epsilon \times n^2$.

Краткая история вопроса

Модель вычислений: одноленточная машина Тьюринга.

С середины 1960-х годов известно, что решение задачи требует времени пропорционального n^2 , где n — длина входного слова.

- Точная оценка:
 - $\Omega(n^2) \ni \mathrm{Time}(\mathcal{M}, W) \in \mathcal{O}(n^2)$, где |W| = n;
 - ullet для любой машины Тьюринга \mathcal{M} , решающей эту задачу для всех слов, $\exists \epsilon > 0$, т.ч. для почти всех слов W, где $|W| = n, \, \epsilon n^2 < \text{Time}(\mathcal{M}, W).$
 - существует машина Тьюринга \mathcal{M}_0 , решающая эту задачу для всех слов, $\exists C > 0$, т.ч. для всех слов W, где |W| = n, Time $(\mathcal{M}_0, W) < Cn^2$.

Теорема Я.М. Барздиня

Модель вычислений: одноленточная машина Тьюринга.

С середины 1960-х годов известно, что решение задачи требует времени пропорционального n^2 , где n — длина входного слова.

- Это один из первых результатов теории сложности вычислений.
 - Я.М. Барздинь. Сложность распознавания симметрии на машинах Тьюринга. Проблемы кибернетики. 1965. Т. 15. С. 245-248.

Теорема Я.М. Барздиня

Модель вычислений: одноленточная машина Тьюринга.

Пусть $T(\mathcal{M},x)$ — число шагов, которое требуется машине \mathcal{M} для вычисления предиката S(x), распознающего является слово $x\in\{0,1\}^*$ симметричным или нет.

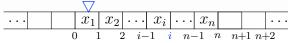
Теорема (Я. М. Барздинь, 1965 г.)

Для любой машины \mathcal{M} , которая реализует предикат симметрии, существует такая константа $\epsilon_{\mathcal{M}}$, что для почти всех симметричных слов x выполнено неравенство $T(\mathcal{M},x) \geq \epsilon_{\mathcal{M}}|x|^2$.

Определения

Модель вычислений: одноленточная машина Тьюринга.

• На ленте дано слово x длины n = |x|.



Определение

Пусть машина \mathcal{M} реализует предикат симметрии S. Пусть q_j есть состояние \mathcal{M} при вычислении S(x), когда она проходит через точку i в j-й раз.

- ullet Следом слова x в точке i назовём $\mathrm{tr}[i\mathcal{M}(x)] = q_1q_2\dots$
- Пусть $\operatorname{pref}_i(x)$ есть i-й собственный префикс слова x, т.е. $\operatorname{pref}_i(x) = x_1 x_2 \dots x_i$. Пусть $\operatorname{suff}_i(x) = x_{i+1} x_{i+2} \dots x_{|x|}$, т.е. $x = \operatorname{pref}_i(x) \operatorname{suff}_i(x)$. В начальный момент времени вычисления любая точка i на ленте находится между $\operatorname{pref}_i(x)$ и $\operatorname{suff}_i(x)$.

Лемма 1

Если х и у — симметричные слова и \exists общая для x,y точка i, т.ч. $\operatorname{pref}_i(x)\operatorname{suff}_i(y)$ — несимметричное слово, то $\operatorname{tr}[i\mathcal{M}(x)] \neq \operatorname{tr}[i\mathcal{M}(y)]$.

Доказательство (от обратного): Предположим, что для всех точек i выполнено равенство $\mathrm{tr}[i\mathcal{M}(x)] = \mathrm{tr}[i\mathcal{M}(y)]$. Рассмотрим слово $z = \mathrm{pref}_i(x)\mathrm{suff}_i(y)$. При вычислении S(z) в начальный момент на ленте дано слово z:

 $\mathrm{tr}[i\mathcal{M}(x)] = \mathrm{tr}[i\mathcal{M}(y)] \Longrightarrow \mathrm{tr}[i\mathcal{M}(z)] = \mathrm{tr}[i\mathcal{M}(x)] = \mathrm{tr}[i\mathcal{M}(y)]$ Следовательно, процесс вычисления S(z) левее точки i совпадает с проц. выч. S(x) левее точки i, а проц. выч. S(z) правее точки i— с проц. выч. S(y) правее точки i.

Это означает, что $S(z) \neq 0$, так как $S(x) \neq 0$ и $S(y) \neq 0$.

 $\Rightarrow S(z) = 1$, что противоречит несимметричности слова z. \square

Пусть ζ — некоторое слово в алфавите Q состояний машины \mathcal{M} . Обозначим через $A^n(i,\zeta)$ максимальное множество симметричных слов x длины n, т.ч. $tr[i\mathcal{M}(x)] = \zeta$.

Лемма 2

Для $i = 1, 2, \dots, \left\lceil \frac{n+1}{2} \right\rceil$ верно неравенство $|A^n(i, \zeta)| \leq 2^{\left\lceil \frac{n+1}{2} \right\rceil - i}$.

Доказательство (от обратного):

Предположим, что $|A^n(i,\zeta)| > 2^{\left[\frac{n+1}{2}\right]-i}$. Из этого неравенства и из того, что число симметричных слов z длины n, для которых одинаковы $\operatorname{pref}_i(z)$ $(i \leq \lceil \frac{n+1}{2} \rceil)$, равно $2^{\lceil \frac{n+1}{2} \rceil - i}$, вытекает, что \exists два симметричных слова $x, y \in A^n(i, \zeta)$, такие что $\operatorname{pref}_i(x) \neq \operatorname{pref}_i(y)$. Следовательно, слово $\operatorname{pref}_{i}(x)\operatorname{suff}_{i}(y)$ несимметричное. Поэтому, из предположения следует, что \exists симметричные слова

x, y, для которых $\operatorname{tr}[i\mathcal{M}(x)] = \operatorname{tr}[i\mathcal{M}(y)] = \zeta$, а слово $\operatorname{pref}_i(x)\operatorname{suff}_i(y)$ несимметрично. Что противеречит лемме 1. \square

Лемма 3

Пусть $C \in \mathbb{N}$ и $A_C^n(i)$ обозначает максимальное множество симметричных слов x длины n, для которых $|\mathrm{tr}[i\mathcal{M}(x)]| < Cn$.

Лемма 3 (Следствие леммы 2)

Для $i=1,2,\ldots,\left[\frac{n+1}{2}\right]$ верно неравенство $|A_C^n(i)|<2^{\left\lfloor\frac{n+1}{2}\right\rfloor-i+Cn\log r},$ где r — число внутренних состояний машины \mathcal{M} .

Доказательство:

Число различных следов длины, меньшей Cn, меньше r^{Cn} . Применение леммы 2 доказывает утверждение леммы 3. \square

• Т.е. $A_C^n(i)$ есть множество всех симметричных слов x длины n, т.ч. при вычислении S(x) головка машины М проходит через точку i небольшое число раз (его порядок не более, чем линейный от n).

Пусть $C \in \mathbb{R}$ и B_C^n обозначает максимальное множество симметричных слов x длины n, для которых выполнено свойство:

если $x \in B_C^n$, тогда \exists точка $i \in \left[\left[\frac{n}{4}\right], \left[\frac{n+1}{2}\right]\right]$ такая, что $|\mathrm{tr}[i\mathcal{M}(x)]| < Cn$.

Лемма 4

$$|B^n_C|<\left(\left[\frac{n+1}{2}\right]-\left[\frac{n}{4}\right]
ight) imes 2^{\left[\frac{n+1}{2}\right]-\left[\frac{n}{4}\right]+Cn\log r},$$
 где r — число внутренних состояний машины \mathcal{M} .

Доказательство: По определению B_C^n имеем:

$$B_C^n = A_C^n \left(\left[\frac{n}{4} \right] \right) \bigcup A_C^n \left(\left[\frac{n}{4} + 1 \right] \right) \bigcup \ldots \bigcup A_C^n \left(\left[\frac{n+1}{2} \right] \right) \Longrightarrow |B_C^n| \le \sum_{i=\left[\frac{n}{4} \right]}^{i=\left[\frac{n+1}{2} \right]} |A_C^n(i)|$$

По лемме 3 для $i \in \left\lceil \left\lceil \frac{n}{4} \right\rceil, \left\lceil \frac{n+1}{2} \right\rceil \right\rceil$ выполняется неравенство

$$|A_C^n(i)| < 2^{\left[\frac{n+1}{2}\right] - i + Cn\log r},$$

из которого следует утверждение леммы. \square

Лемма 4

Пусть $C \in \mathbb{R}$ и B_C^n обозначает максимальное множество симметричных слов x длины n, для которых выполнено свойство: если $x \in B_C^n$, тогда \exists точка $i \in \left[\left[\frac{n}{4}\right], \left[\frac{n+1}{2}\right]\right]$ такая, что $|\operatorname{tr}[i\mathcal{M}(x)]| < Cn$.

Лемма 4

 $|B_C^n| < (\lceil \frac{n+1}{2} \rceil - \lceil \frac{n}{4} \rceil) \times 2^{\lceil \frac{n+1}{2} \rceil - \lceil \frac{n}{4} \rceil + Cn \log r}$, где r — число внутренних состояний машины \mathcal{M} .

• Т.е. B_C^n есть множество всех симметричных слов x длины n, т.ч. при вычислении S(x) головка машины $\mathcal M$ проходит через некоторую точку (хотя бы одну) на отрезке $\left[\left[\frac{n}{4}\right],\left[\frac{n+1}{2}\right]\right]$ небольшое число раз (его порядок не более, чем линейный от n).

Лемма 5 (Следствие леммы 4)

Пусть S^n есть множество всех симметричных слов x длины n, а D^n_C — множество $S^n \backslash B^n_C$, где $C \in \mathbb{R}; \ |S^n| = 2^{\left[\frac{n+1}{2}\right]}$.

Лемма 5

Существует такая константа $C \in \mathbb{N}$, что $\lim_{n \to \infty} \frac{|D_C^n|}{|S^n|} = 1$

Доказательство: По лемме 4 получаем:

$$\frac{|D_C^n|}{|S^n|} = 1 - \frac{|B_C^n|}{|S^n|} > 1 - \frac{\left(\left[\frac{n+1}{2}\right] - \left[\frac{n}{4}\right]\right) \times 2^{\left[\frac{n+1}{2}\right] - \left[\frac{n}{4}\right] + Cn\log r}}{2^{\left[\frac{n+1}{2}\right]}}$$

где r — число внутренних состояний машины \mathcal{M} . Машина \mathcal{M} произвольная, но фиксированная, следовательно r фиксировано. \square

• Т.е. D_C^n есть множество всех симметричных слов x длины n, т.ч. при вычислении S(x) головка машины \mathcal{M} проходит через каждую точку на отрезке $\left[\left[\frac{n}{4} \right], \left[\frac{n+1}{2} \right] \right]$ большое число раз (его порядок Рефал 2025/ 01-07-2025 23 / 28 больше, чем линейный от n).

Пусть $T(\mathcal{M},x)$ — число шагов, которое требуется машине \mathcal{M} для вычисления предиката S(x), распознающего является слово $x\in\{0,1\}^*$ симметричным или нет.

Теорема (Я. М. Барздинь, 1965 г.)

Для любой машины \mathcal{M} , которая реализует предикат симметрии, существует такая константа $\epsilon_{\mathcal{M}}$, что для почти всех симметричных слов x выполнено неравенство $T(\mathcal{M}, x) \geq \epsilon_{\mathcal{M}} |x|^2$.

Доказательство: Пусть $C_0 \in \mathbb{R}$ есть некоторая константа, для которой $\lim_{n\to\infty} \frac{|D^n_{C_0}|}{|S^n|} = 1$.

- Из определения множества $D_{C_0}^n \Longrightarrow$ что, если $x \in D_{C_0}^n$, то $\forall i \in \left[\left[\frac{n}{4}\right], \left[\frac{n+1}{2}\right]\right]$. $|\text{tr}[i\mathcal{M}(x)]| \geq C_0 n$.
- Каждая буква каждого следа слова x соответствует определенному шагу машины \mathcal{M} при вычислении S(x). $\Longrightarrow \forall x \in D^n_{C_0}$. $T(\mathcal{M}, x) \geq \left(\left[\frac{n+1}{2}\right] \left[\frac{n}{4}\right]\right) \times C_0 n$.
- $\bullet \Longrightarrow \exists \epsilon_{\mathcal{M}} \in \mathbb{R}. \ T(\mathcal{M}, x) \geq \epsilon_{\mathcal{M}} |x|^2.$
- Следовательно, учитывая лемму 5, теорема доказана. 🗆

Рефал-5 / Version-ПЗ / Windows-10 / $x86_64$

Unico de-версия Рефала-5 перенесена под ОС Windows-10 / x86_64.

- cmd-консольное приложение.
- Использует кодировку UTF-8.
 - Активная кодовая страница должна быть установлена как 65001 (UTF-8).
 - Исходные тексты программ на Рефале-5 должны быть в кодировке UTF-8.
- Пока недостаточно протестирована.
- Для желающих потестировать:

 $http://refal.botik.ru/refal5/ref5_windows-10_x86_64_untested_250617.zip$

- Далее два снимка демонстрационных экранов.
 - И живая демонстрация.

Поддержка Кириллицы и китайских иероглифов

Исходный текст программы и его результат работы.

Церковно-славянский шрифт

Фрагмент исходного текста программы.

Церковно-славянский шрифт

Результат работы программы.

```
c:\nemytykh\Bureaucrat\Project2025\refal-64-unicode\ref5_win10_x64_untested>refc examples\Sirin.ref
c:\nemytykh\Bureaucrat\Project2025\refal-64-unicode\ref5_win10_x64_untested>refgo examples\Sirin
Любй доброе собесбдованїє, ѿ sлы́хь же бесбдь о́жальйсь:
Занє́ ни волше́бникь, ни разбойникь, ни гроборазори́тель тако роди́шась,
но тако най чишась ѿ человькивь, растле́нныхь о́жо́м ѿ сатаны.

В этой цитате 196 символов, включая пробелы и символы перевода строки.

И місоже тѣло, а́ше хльба не прійметь, жи́во бы́ти не мо́жеть:
тако и душа, а́ше не прійметь пищи своєм, мертва е́сть. Йбо человькь сугубь е́сть,
йз души и тьла. И сегм ради гластолаше Стьс:
не w хльбь единомь жи́вь будеть человькь.

В этой цитате 246 символов, включая пробелы и символы перевода строки.

c:\nemytykh\Bureaucrat\Project2025\refal-64-unicode\ref5_win10_x64_untested>
```