



Теория формальных языков на практике

Открытая лекция Computer Science Center

Семён Григорьев

Лаборатория языковых инструментов JetBrains
Санкт-Петербургский государственный университет
Математико-механический факультет

18 октября 2019г.

- Семён Григорьев
 - ▶ rsdpisuy@gmail.com
 - ▶ semyon.grigorev@jetbrains.com
 - ▶ <https://research.jetbrains.org/researchers/gsv>
- Исследовательская группа на Математико-Механическом факультете СПбГУ
- Исследовательская группа в лаборатории языковых инструментов JetBrains Research: :
https://research.jetbrains.org/groups/plt_lab
- Сферы интересов
 - ▶ Теория формальных языков
 - ▶ **Применение теории формальных языков для решения прикладных задач**

Поиск путей с ограничениями в терминах формальных языков

- Конечный ориентированный граф с метками на рёбрах
 $\mathcal{G} = (V, E, L)$
- Путь — это слово в алфавите L
$$\omega(p) = \omega(v_0 \xrightarrow{l_0} v_1 \xrightarrow{l_1} \dots \xrightarrow{l_{n-1}} v_n) = l_0 \cdot l_1 \cdot \dots \cdot l_{n-1}$$
- Язык \mathcal{L} (над алфавитом L)

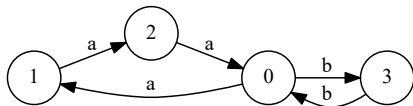
Поиск путей с ограничениями в терминах формальных языков

- Конечный ориентированный граф с метками на рёбрах
 $\mathcal{G} = (V, E, L)$
- Путь — это слово в алфавите L
$$\omega(p) = \omega(v_0 \xrightarrow{l_0} v_1 \xrightarrow{l_1} \dots \xrightarrow{l_{n-1}} v_n) = l_0 \cdot l_1 \cdot \dots \cdot l_{n-1}$$
- Язык \mathcal{L} (над алфавитом L)
- Задача достижимости: $Q = \{(v_i, v_j) \mid \exists p = v_i \dots v_j, \omega(p) \in \mathcal{L}\}$
- Задача поиска путей: $Q = \{p \mid \omega(p) \in \mathcal{L}\}$
 - ▶ Один путь, все пути, кратчайший путь...

Поиск путей с контекстно-свободными ограничениями

- \mathcal{L} — контекстно-свободный язык (КС язык)
- $G_{\mathcal{L}} = (N, \Sigma, R, S)$
- Задача достижимости: $Q = \{(v_i, v_j) \mid \exists p = v_i \dots v_j, S \xrightarrow[G_{\mathcal{L}}]{*} \omega(p)\}$
- Задача поиска путей: $Q = \{p \mid S \xrightarrow[G_{\mathcal{L}}]{*} \omega(p)\}$

Пример КС запроса



Входной граф

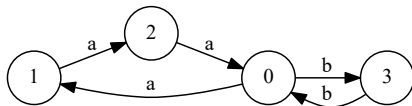
$S \rightarrow a S b$

$S \rightarrow \textit{Middle}$

$\textit{Middle} \rightarrow a b$

Запрос: язык $\{a^n b^n \mid n > 0\}$

Пример КС запроса



Входной граф

$$S \rightarrow a S b$$

$$S \rightarrow \textit{Middle}$$

$$\textit{Middle} \rightarrow a b$$

Запрос: язык $\{a^n b^n \mid n > 0\}$

Пример путей:

$$2 \xrightarrow{a} 0 \xrightarrow{b} 3$$

$$1 \xrightarrow{a} 2 \xrightarrow{a} 0 \xrightarrow{b} 3 \xrightarrow{b} 0$$

$$p_1 = 0 \xrightarrow{a} 1 \xrightarrow{a} 2 \xrightarrow{a} 0 \xrightarrow{b} 3 \xrightarrow{b} 0 \xrightarrow{b} 3$$

$$p_2 = 0 \xrightarrow{a} 1 \xrightarrow{a} 2 \xrightarrow{a} 0 \xrightarrow{a} 1 \xrightarrow{a} 2 \xrightarrow{a} 0 \xrightarrow{b} 3 \xrightarrow{b} 0 \xrightarrow{b} 3 \xrightarrow{b} 0 \xrightarrow{b} 3 \xrightarrow{b} 0$$

...

Контекстно-свободная достижимость для статического анализа кода

- *Thomas Reps et al.* “Precise interprocedural dataflow analysis via graph reachability.” 1995
- *Jakob Rehof and Manuel Fahndrich.* “Type-base flow analysis: from polymorphic subtyping to CFL-reachability.” 2001
- *Dacong Yan et al.* “Demand-driven context-sensitive alias analysis for Java.” 2011
- *Qirun Zhang et al.* “Efficient subcubic alias analysis for C.” 2014
- *Kai Wang et. al.* “Graspan: a single-machine disk-based graph system for interprocedural static analyses of large-scale systems code.” 2017
- *Qirun Zhang and Zhendong Su.* “Context-sensitive data-dependence analysis via linear conjunctive language reachability.” 2017

Контекстно-свободная достижимость для графовых баз данных

- *M. Yannakakis* “Graph-theoretic methods in database theory.” 1990
- *C. Barrett, R. Jacob, and M. Marathe* “Formal-language-constrained path problems.” 2000
- *Sevon P., Eronen L.* “Subgraph queries by context-free grammars.” 2008
- *Hellings J.* “Conjunctive context-free path queries.” 2014
- *Zhang X. et al.* “Context-free path queries on RDF graphs.” 2016
- *Jochem Kuijpers, George Fletcher, Nikolay Yakovets, and Tobias Lindaaker* “An Experimental Study of Context-Free Path Query Evaluation Methods.” 2019

Субкубический алгоритм для решения задачи КС-достижимости

- *L. G. Valiant* “General context-free recognition in less than cubic time.” 1975
- *Ph. G. Bradford* “Efficient exact paths for dyck and semi-dyck labeled path reachability.” 2017
- *Krishnendu Chatterjee* “Optimal Dyck reachability for data-dependence and alias analysis.” 2017