



Структурное представление результата поиска путей с контекстно-свободными ограничениями

Семён Григорьев

JetBrains Research, лаборатория языковых инструментов
Санкт-Петербургский университет

21.10.2017

- Анализ графов
 - ▶ Запросы к графовым базам данных
 - ▶ Анализ сетей (социальных, интернет и т.д.)
- Статический анализ программ
 - ▶ Анализ алиасов
 - ▶ Taint analysis
 - ▶ Статический анализ динамически формируемого кода
- ...

Поиск путей с контекстно-свободными ограничениями

- $\mathbb{G} = (\Sigma, N, P)$ — контекстно-свободная грамматика, $L \subseteq \Sigma^*$
- $p = (v_0, l_0, v_1), \dots, (v_{n-1}, l_{n-1}, v_n)$ — путь в графе G
- $w(p) = w((v_0, l_0, v_1), \dots, (v_{n-1}, l_{n-1}, v_n)) = l_0 l_1 \dots l_{n-1}$
- $G = (V, E, L)$ — ориентированный граф, $E \subseteq V \times L \times V$
- $R = \{p \mid \exists N_i \in N (w(p) \in L(\mathbb{G}, N_i))\}$
 - ▶ Стартовый нетерминал можно зафиксировать заранее
 - ▶ **Проблема:** множество R может быть бесконечным

- В качестве основы используют СΥК
 - ▶ Необходимо преобразовывать грамматику в нормальную форму Хомского
 - ▶ Только проверка наличия пути (Zhang X. et al. Context-free path queries on RDF graphs. 2016)
 - ▶ Грамматика в качестве представления результата (Hellings J. Conjunctive context-free path queries. 2014)

Алгоритм выполнения КС запросов к графам

- Основан на обобщённом LL (Generalized GLL, GLL)
 - ▶ Scott E., Johnstone A. GLL parsing
- Поддерживает произвольные контекстно-свободные грамматики (неоднозначные, леворекурсивные)
- Не требует преобразования грамматики в нормальную форму Хомского
- Строит сжатое представление леса разбора (Sharep Packed Parse Forest, SPPF)

Пример

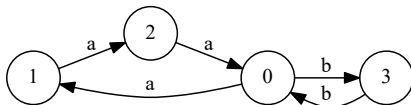


Рис.: Входной граф

0 : $S \rightarrow a S b$

1 : $S \rightarrow Middle$

2 : $Middle \rightarrow a b$

Рис.: Запрос: грамматика G для языка $L = \{a^n b^n; n \geq 1\}$ с явным выделением середины пути

Пример: SPPF

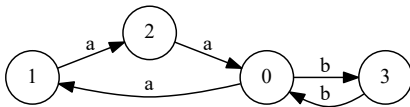
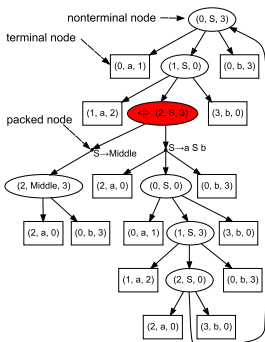
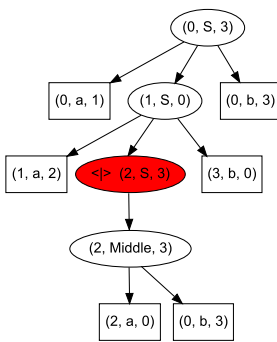


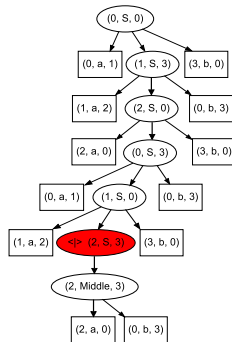
Рис.: Входной граф



Результат

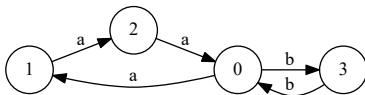


Структура пути1



Структура пути2

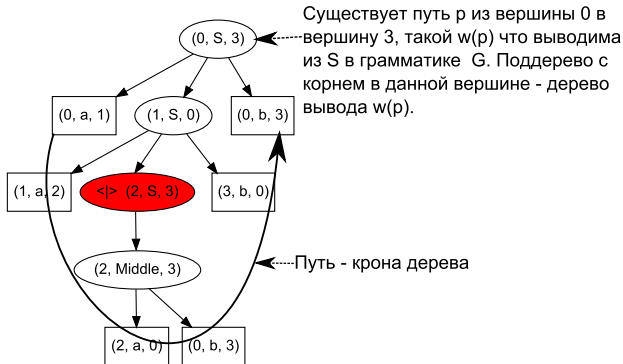
Пример: извлечение путей



0 : $S \rightarrow a S b$

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

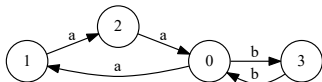
2 : $\textit{Middle} \rightarrow a b$



Путь: $(0, a, 1); (1, a, 2); (2, a, 0); (0, b, 3); (3, b, 0); (0, b, 3)$

Пример: почему это работает

Замкнутость КС языков относительно пересечения с регулярными



Регулярный язык

0 : $S \rightarrow a S b$

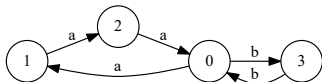
1 : $S \rightarrow \textit{Middle}$

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

Контекстно-сводный язык

Пример: почему это работает

Замкнутость КС языков относительно пересечения с регулярными



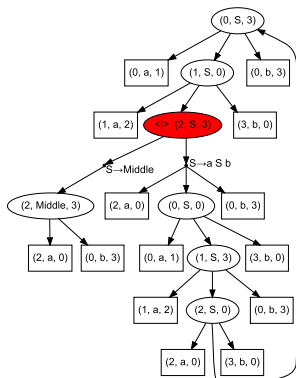
Регулярный язык

0 : $S \rightarrow a S b$

1 : $S \rightarrow \text{Middle}$

2 : $\text{Middle} \rightarrow a b$

Контекстно-сводный язык



$(0, S, 3) \rightarrow (0, a, 1) (1, S, 0) (0, b, 3)$

$(1, S, 0) \rightarrow (1, a, 2) (2, S, 3) (3, b, 0)$

$(2, S, 3) \rightarrow (2, a, 0) (0, S, 0) (0, b, 3)$

$(2, S, 3) \rightarrow (2, \text{Middle}, 3)$

$(0, S, 0) \rightarrow (0, a, 1) (1, S, 3) (3, b, 0)$

$(1, S, 3) \rightarrow (1, a, 2) (2, S, 0) (0, b, 3)$

$(2, S, 0) \rightarrow (2, a, 0) (0, S, 3) (3, b, 0)$

$(0, \text{Middle}, 3) \rightarrow (2, a, 0) (0, b, 3)$

Характеристики алгоритма

Пусть на входе граф $M = (V, E, L)$, тогда

- Пространственная сложность предложенного алгоритма равна $O(|V|^3 + |E|)$
- Временная сложность предложенного алгоритма равна $O\left(|V|^3 * \max_{v \in V} (deg^+(v))\right)$
- Результирующий SPPF имеет размер $O(|V'|^3 + |E'|)$ где $M' = (V', E', L')$ — это подграф M содержащий только искомые пути

Экспериментальное исследование: запросы

- 0 : $S \rightarrow subClassOf^{-1} S subClassOf$
- 1 : $S \rightarrow type^{-1} S type$
- 2 : $S \rightarrow subClassOf^{-1} subClassOf$
- 3 : $S \rightarrow type^{-1} type$

Рис.: Грамматика для запроса Query 1

- 0 : $S \rightarrow B subClassOf$
- 1 : $S \rightarrow subClassOf$
- 2 : $B \rightarrow subClassOf^{-1} B subClassOf$
- 3 : $B \rightarrow subClassOf^{-1} subClassOf$

Рис.: Грамматика для запроса Query 2

Экспериментальное исследование: результаты

Ontology	#edg	Query 1			Query 2	
		time CYK (ms)	time (ms)	#result	time (ms)	#result
skos	252	1044	10	810	1	1
generations	273	6091	19	2164	1	0
travel	277	13971	24	2499	1	63
univ-bench	293	20981	25	2540	11	81
foaf	631	–	39	4118	2	10
people-pets	640	82081	89	9472	3	37
funding	1086	–	212	17634	23	1158
atom-primitive	425	515285	255	15454	66	122
biomedical- measure-primitive	459	420604	261	15156	45	2871
pizza	1980	3233587	697	56195	29	1262
wine	1839	4075319	819	66572	8	133

- Почта: `semen.grigorev@jetbrains.com`
- GitHub-сообщество YaccConstructor:
`https://github.com/YaccConstructor`