



Parsing techniques for graph analysis

Ekaterina Verbitskaia

JetBrains Research, Programming Languages and Tools Lab
Saint Petersburg University

Oktober 22, 2017

Language-constrained paths filtering

- $\mathbb{G} = (\Sigma, N, P)$ — context-free grammar
- $G = (V, E, L)$ — directed graph, $E \subseteq V \times L \times V$, $L \subseteq \Sigma$
- $p = (v_0, l_0, v_1), \dots, (v_{n-1}, l_{n-1}, v_n)$ — path in G
- $\omega(p) = \omega((v_0, l_0, v_1), \dots, (v_{n-1}, l_{n-1}, v_n)) = l_0 l_1 \dots l_{n-1}$
- $R = \{p \mid \exists N_i \in N (\omega(p) \in L(\mathbb{G}, N_i))\}$
- Other possible variants

Language-constrained paths filtering: more formal

- $\mathbb{G} = (\Sigma, N, P)$ — context-free grammar
- $G = (V, E, L)$ — directed graph, $E \subseteq V \times L \times V$, $L \subseteq \Sigma$
- $p = (v_0, l_0, v_1), \dots, (v_{n-1}, l_{n-1}, v_n)$ — path in G
- $\omega(p) = \omega((v_0, l_0, v_1), \dots, (v_{n-1}, l_{n-1}, v_n)) = l_0 l_1 \dots l_{n-1}$
- $R = \{p \mid \exists N_i \in N(\omega(p) \in L(\mathbb{G}, N_i))\}$
- Other possible variants

- Graph analysis
 - ▶ Graph database querying
 - ▶ Network graph analysis
 - ▶
- Code analysis
 - ▶ Static analysis CFL(linear conjunctive) reachability: alias analysis, points-to analysis, etc
 - ▶ Dynamically generated strings analysis
 - ▶ Multiple input parsing
- ...

- Effective algorithm creation
- Result representation for debugging, futer processing
- GPGPU utilization

Bar-Hillel theorem!

- Bar-Hillel theorem!
- Parsing algorithms are constructive proof of BH theorem for one simple case...

Example

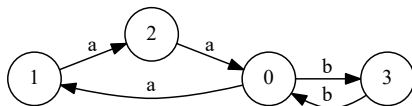


Figure: An example: the map of School (input graph M)

0 : $S \rightarrow a S b$

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

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

Figure: An example: grammar G_1 for language $L = \{a^n b^n; n \geq 1\}$ with additional marker for the middle of a path

Example

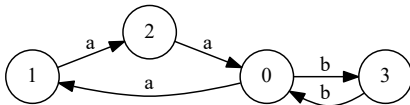
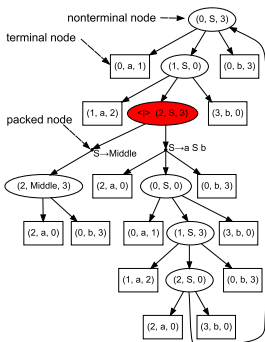
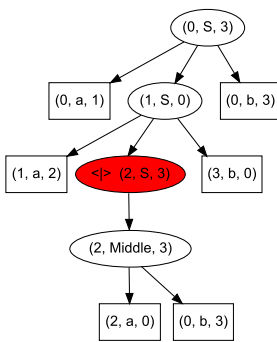


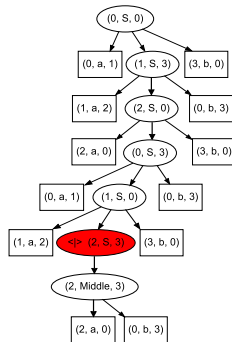
Figure: An example: the map of School (input graph M)



sppf



дерево



дерево

- Relaxed parsing of dynamically generated SQL-queries. RNGLR-based
- Context-free path querying with structural representation of result. GLL-based
- Combinators for context-free path querying
 - ▶ Based on Meerkat (Izmailova, Afroozeh)
- Context-free path querying by Matrix multiplication

- Other grammars and language classes intersection
 - ▶ Context-free grammars intersection
 - ★ Mark-Jan Nederhof, *The language intersection problem for non-recursive context-free grammars*
 - ▶ Approximated intersection of regular and conjunctive/boolean languages
 - ▶ ...
- Mechanization in Coq
 - ▶ Bar-Hillel theorem
 - ▶ GLL-based algorithms
 - ▶ ...
- Other areas for application

- Ekaterina Verbitskaia: kajigor@gmail.com
- Semyon Grigorev: semen.grigorev@jetbrains.com
- YaccConstructor: <https://github.com/YaccConstructor>