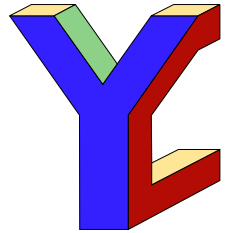


Parsing techniques for graph analysis

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, $E \subseteq V \times L \times V$, $L \subseteq \Sigma$
- $p = (v_0, l_0, v_1), \cdots, (v_{n-1}, l_{n-1}, v_n)$ — path in G
- $\omega(p) = \omega((v_0, l_0, v_1), \cdots, (v_{n-1}, l_{n-1}, v_n)) = l_0 l_1 \cdots l_{n-1}$
- $R = \{p | \exists N_i \in N(\omega(p) \in L(\mathbb{G}, N_i))\}$
- Other possible variants

Applications

- 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
- ...

Open Problems etc

- 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...

GLL-based

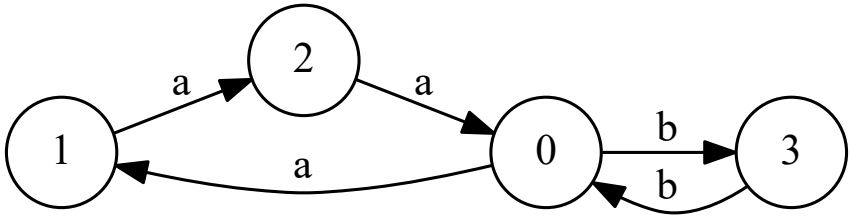


Рис.: An example: the map of School (input graph M)

$$
\begin{aligned}
0: &\quad S \to a\ S\ b \\
1: &\quad S \to Middle \\
2: &\quad Middle \to a\ b
\end{aligned}
$$

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

GLL-based

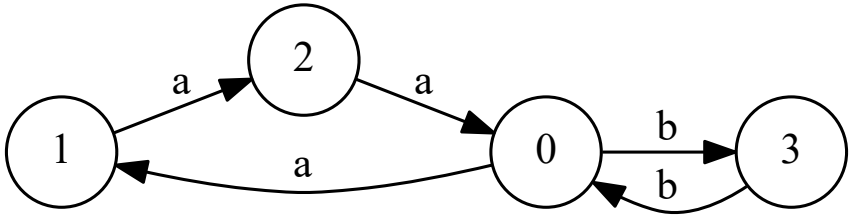
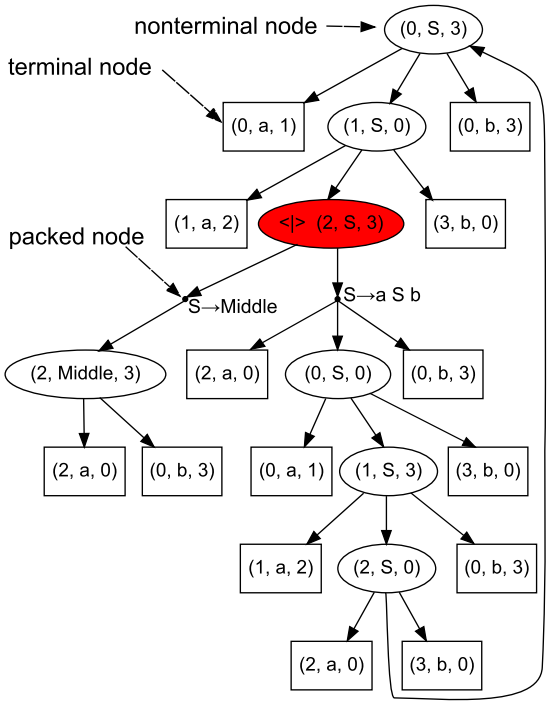
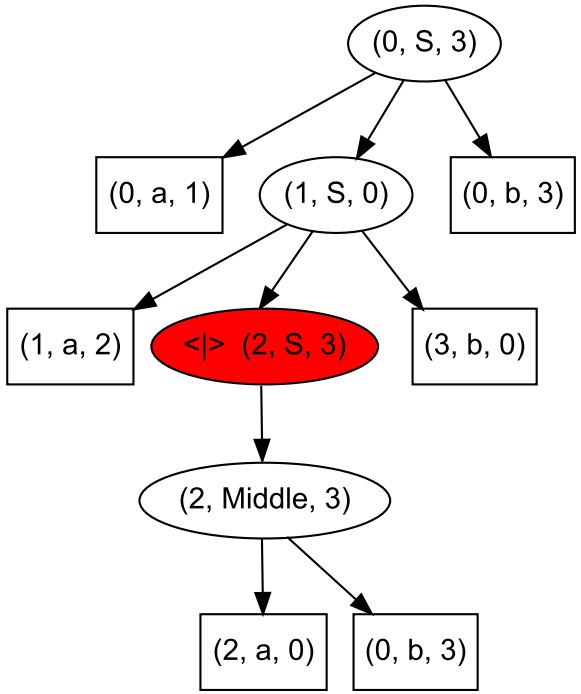


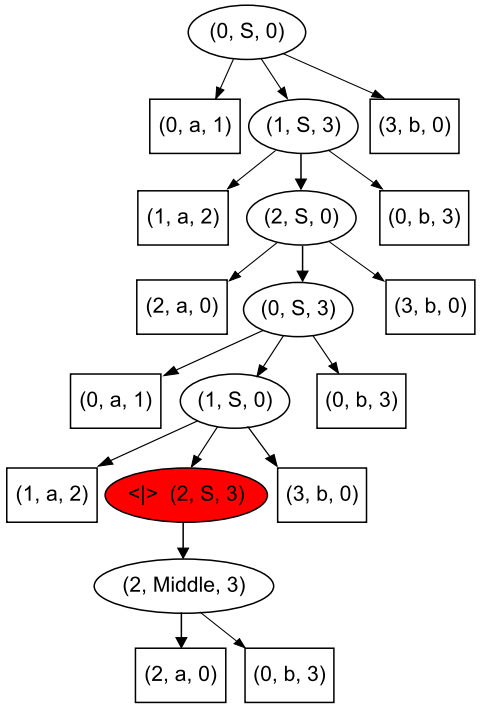
Рис.: An example: the map of School (input graph M)



sppf



дерево



дерево

Our solutions

- 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

Future work

- Other grammars and languages intersection
  - Nederhof
  - Reg + Conj (bool) on matrix
  - ...
- Mechaniation on Coq
  - Bar-Hillel theorem
  - GLL-based graph parsing
  - ...
- Applications

Information

- semen.grigorev@jetbrains.com
- kajigor@gmail.com
- YaccConstructor: https://github.com/YaccConstructor