

Parsing Techniques for Context-Free Path Querying

Semyon Grigorev

JetBrains Research, Programming Languages and Tools Lab
Saint Petersburg University

April 05, 2019

- https://research.jetbrains.org/groups/plt_lab

Formal languages for data analysis

- Semyon Grigorev

Topics of interest

- Formal language theory
- Parsing algorithms

Formal language constrained path querying

- Finite directed edge-labelled graph $\mathcal{G} = (V, E, L)$
- The path is a world over L :
$$\omega(p) = \omega(v_0 \xrightarrow{l_0} v_1 \xrightarrow{l_1} \dots \xrightarrow{l_{n-1}} v_n) = l_0 \cdot l_1 \cdot \dots \cdot l_{n-1}$$
- The language \mathcal{L} (over L)

Formal language constrained path querying

- Finite directed edge-labelled graph $\mathcal{G} = (V, E, L)$
- The path is a world over L :
$$\omega(p) = \omega(v_0 \xrightarrow{l_0} v_1 \xrightarrow{l_1} \dots \xrightarrow{l_{n-1}} v_n) = l_0 \cdot l_1 \cdot \dots \cdot l_{n-1}$$
- The language \mathcal{L} (over L)
- Reachability problem: $Q = \{(v_i, v_j) \mid \exists p = v_i \dots v_j, \omega(p) \in \mathcal{L}\}$
- Path querying problem: $Q = \{p \mid \omega(p) \in \mathcal{L}\}$
 - ▶ Single path, all paths, shortest path ...

Context-Free path querying

- \mathcal{L} is a context-free language
- $G_{\mathcal{L}} = (N, \Sigma, R, S)$
- Reachability problem: $Q = \{(v_i, v_j) \mid \exists p = v_i \dots v_j, S \xrightarrow[G_L]{*} \omega(p)\}$
- Path querying problem: $Q = \{p \mid \omega(p) \in \mathcal{L}\}$

Context-Free path querying

- \mathcal{L} is a context-free language
- $G_{\mathcal{L}} = (N, \Sigma, R, S)$
- Reachability problem: $Q = \{(v_i, v_j) \mid \exists p = v_i \dots v_j, S \xrightarrow[G_L]{*} \omega(p)\}$
- Path querying problem: $Q = \{p \mid \omega(p) \in \mathcal{L}\}$
- Reachability problem: $Q = \{(v_i, v_j) \mid \exists p = v_i \dots v_j, \omega(p) \in \mathcal{L}\}$
- Path querying problem: $Q = \{p \mid \omega(p) \in \mathcal{L}\}$
 - ▶ Single path, all paths, shortest path ...

Example of CFPQ



Applicability

- Static code analysis
- Graph data bases
- !!!!

Open questions

- Subcubic (or to BMM)
- Engineering
- New field for parsing algorithms



- Subcubic (or to BMM)
- Engineering
- New field for parsing algorithms

- High performance
- New classes of grammars
- !!!!!

Contact Information

- Semyon Grigorev:
 - ▶ s.v.grigoriev@spbu.ru
 - ▶ Semen.Grigorev@jetbrains.com
- Polina Lunina:
 - ▶ lunina_polina@mail.ru
- Trained models: <https://github.com/YaccConstructor/YC.Bio>

Thanks!