Formal language guided data analysis group results

Semyon Grigorev

JetBrains Research, Programming Languages and Tools Lab
St. Petersburg University

18.12.2021

Research Interests

- Formal language theory, parsing algorithms
  - Linear algebra based algorithms for formal language constrained path querying development and evaluation
  - Time complexity of context-free path querying
- Formal semantics of graph query languages (in collaboration with Anton Podkopaev)
- Fast sparse linear algebra
  - Parallel and GPGPU programming, and techniques to do it fast and safe
  - Optimization techniques and specialized hardware: distillation, lambda-precessors (in collaboration with Daniil Berezun)
- Application of all above for
  - Static code analysis
  - Graph databases
  - Biological data analysis
  - ...

Team

- PhD students
  - Rustam Azimov
  - Ekaterina Shemetova
- Master students
  - Alexandra Istomina
  - Egor Orachev
  - Ilya Epelbaum
  - Vladimir Kutuev
  - Arseniy Terekhov → CLion team
- Bachelor students
  - Vlada Pogozhelskaya
  - Vadim Abzalov
  - Timur Zinnatulin
  - Dmitriy Panfilyonok
  - Artem Chernikov
- About six 2nd, 3rd, 4th year students: graduate projects, semester practices, etc

Conferences

- **EDBT-2021** (CORE A)
  - Arseniy Terekhov, Vlada Pogozhelskaya, Vadim Abzalov, Timur Zinnatulin, *Semyon Grigorev*. Multiple-Source Context-Free Path Querying in Terms of Linear Algebra
  - Scopus
- **LDBC TUC 2021**
  - *Semyon Grigorev*. Context-Free Path Querying: Obstacles on the Way to Adoption
  - Invited by Gabor Szarnyas
- **GrAPL-2021**
  - *Egor Orachev*, Maria Karpenko, Artem Khoroshev, Semyon Grigorev. SPbLA: The Library of GPGPU-Powered Sparse Boolean Linear Algebra Operations
  - Scopus

Conferences

- ✔ **GRADES-NDA 2021**
  - *Rustam Azimov*, Ilya Epelbaum, Semyon Grigorev. Context-Free Path Querying with All-Path Semantics by Matrix Multiplication
  - Scopus
- ✔ **VLDB PhD Workshop 2021**
  - *Rustam Azimov*. Context-Free Path Querying In Terms of Linear Algebra
  - Scopus
- ⌛ **EDBT-2022** (CORE A)
  - Vlada Pogozhelskaya, Anna Vlasova, Semyon Grigorev. GLL-based Context-Free Path Querying for Neo4j
  - Submitted

Collaboration

- Internal
  - Daniil Berezun: distillation of linear algebra based algorithms
  - Anton Podkopaev: Graph Query Language semantics formalization and mechanization in Coq
- External
  - Alexander Okhotin, RSF grant
    - ★ Semyon Grigorev,Ekaterina Shemetova
  - LDBC community
    - ★ Formal languages constrained path querying algorithms
    - ★ Competition of FL constrained path querying algorithms
    - ★ Rustam Azimov, Timur Zinnatulin, Vladimir Kutuev
  - Neo4j team
    - ★ CFPQ for Neo4j
    - ★ Graduation project of Vlada Pogozhelskaya
  - Selectel
    - ★ FPGA-equipped server for our algorithms evaluation
    - ★ Master thesis of Maria Karpenko

Teaching

- Formal language theory (lectures, seminars): SPbU
  - Lecture notes (in collaboration with Ekaterina Verbitskaia): `https://github.com/JetBrains-Research/FormalLanguageConstrainedReachability-LectureNotes`
  - Exercises and supplementary materials (in collaboration with Egor Orachev and Vadim Abzalov): `https://github.com/JetBrains-Research/formal-lang-course`
- Graph theory: SPbU
- Formal language theory seminar
- Graduation projects, practices, semester projects for students from CSC, HSE, SPbU, ITMO, etc

Grants

- RSF
  - “Sparse linear algebra: from specialized hardware to applied solutions”
  - Semyon Grigorev, Daniil Berezun, Anton Podkopaev, Timofey Briksin, Rustam Azimov, Egor Orachev, Alexey Turin, Arceniy Terekhov

Work in progerss: publications

- Ekaterina Shemetova, Alexander Okhotin, Semyon Grigirev. Rational index of bounded-oscillation languages. Arxiv: `https://arxiv.org/abs/2012.03567`
- Ekaterina Shemetova, Rustam Azimov, Egor Orachev, Ilya Epelbaum, Semyon Grigorev. One Algorithm to Evaluate Them All: Unified Linear Algebra Based Approach to Evaluate Both Regular and Context-Free Path Queries. Arxiv: `https://arxiv.org/abs/2103.14688`
- Polina Lunina, Vadim Abzalov, Semyon Grigorev. Genegram: RNA Secondary Structure Prediction Using Formal Grammars and Residual Neural Networks

Work in progerss: publications

- Ekaterina Shemetova, Alexander Okhotin, Semyon Grigirev. Rational index of bounded-oscillation languages. Arxiv: `https://arxiv.org/abs/2012.03567`
- Ekaterina Shemetova, Rustam Azimov, Egor Orachev, Ilya Epelbaum, Semyon Grigorev. One Algorithm to Evaluate Them All: Unified Linear Algebra Based Approach to Evaluate Both Regular and Context-Free Path Queries. Arxiv: `https://arxiv.org/abs/2103.14688`
- Polina Lunina, Vadim Abzalov, Semyon Grigorev. Genegram: RNA Secondary Structure Prediction Using Formal Grammars and Residual Neural Networks
- Egor Orachev and Gleb Mar'in. On multi-GPU sparse linear algebra
- Dmitriy Panfilyonok and Artem Chernikov. On On functional languages based design of generic sparse linear algebra routines
- Alexey Turin, Ekaterina Vinnik, Daniil Berezun. On distillation of sparse lianear algebra routines
- ...