



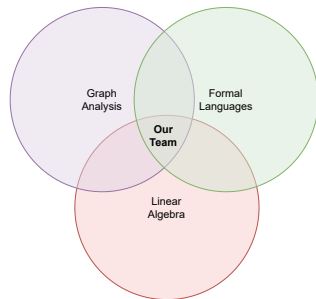
Анализ данных с применением теории формальных языков

Семён Григорьев

Санкт-Петербургский Государственный Университет

2 апреля 2025

- Поиск путей в графах с метками на рёбрах с использованием формальных языков в качестве ограничений¹
 - Регулярные языки (Regular path querying, RPQ): графовые базы данных, часть языка GQL, SMT-решатели
 - Контекстно-свободные языки (Context-free path querying, CFPQ, CFL-r): графовые БД, статический анализ кода
- Высокопроизводительный анализ графов с использованием линейной алгебры
 - Разработка алгоритмов на основе линейной алгебры
 - Разработка библиотек линейной алгебры, в том числе, с использованием GPGPU
 - Создание специализированных решений для разработки оптимизации алгоритмов на основе линейной алгебры



¹Для языка \mathcal{L} и графа с метками на рёбрах \mathcal{G} построить $R = \{\pi \mid \pi \text{ — путь в } \mathcal{G}, \omega(\pi) \in \mathcal{L}\}$,
 $\omega(v_0 \xrightarrow{l_0} v_1 \xrightarrow{l_1} v_2 \dots) = l_0 l_1 \dots$

Формальные языки как ограничения на пути⁴: алгоритмы

- ✓ Диссертация Рустама Азимова про CFPQ с использованием линейной алгебры
- ✓ Оптимизации CFPQ (Владимир Кутев, Илья Муравьев, Егор Орачев, Арсений Терехов и многие другие)
- ✓ Базовая версия CFPQ LAGraph² (Ильхом Комбаев)
- ✓ Базовая версия RPQ LAGraph³ (Георгий Белянин)
- ⚙ Перенести в LAGraph различные вариации алгоритмов CFPQ и RPQ и их оптимизации
- ⚙ Диссертация Ольги Бачище про CFPQ с использованием обобщённого LL алгоритма
- ⚙ Интеграция алгоритмов в инструменты статического анализа кода

²<https://github.com/GraphBLAS/LAGraph/pull/265>

³<https://github.com/GraphBLAS/LAGraph/pull/261>

⁴Наше сообщество на GitHub: <https://github.com/FormalLanguageConstrainedPathQuerying>

Формальные языки как ограничения на пути: сопутствующие материалы

- Расширение курса по теории формальных языков (Вадим Абзалов, Екатерина Шеметова, Екатерина Вербицкая, Егор Орачев, Николай Пономарев, Ефим Кубышкин, и многие другие)
 - ✓ Задачи, система проверки⁵
 - ⚙ Разрабатывается конспект лекций⁶
- Набор данных⁷
 - ⚙ Интеграция в SuiteSparse matrix collection

⁵<https://github.com/FormalLanguageConstrainedPathQuerying/formal-lang-course>

⁶<https://github.com/FormalLanguageConstrainedPathQuerying/FormalLanguageConstrainedReachability-LectureNotes>

⁷CFPQ_Data: https://github.com/FormalLanguageConstrainedPathQuerying/CFPQ_Data

Линейная алгебра и анализ графов¹¹

- Линейная алгебра на GPGPU (Егор Орачев, Владимир Кутуев, Дмитрий Козенко и многие другие)
 - ✓ cuBool⁸: разреженная булева линейная алгебра на Cuda
 - ✓ Spla⁹: обобщённая разреженная линейная алгебра на OpenCL
 - ⚙ Перенос Spla на RISC-V
- Вклад в экосистему GraphBLAS
 - ✓ Векторизация умножения матриц на RISC-V RVV1.0¹⁰ (Родион Суворов)
 - ⚙ Кросс-сборка
- ⚙ Эксперименты с RISC-V GPGPU Vortex (Даниил Власенко)

⁸<https://github.com/SparseLinearAlgebra/cuBool>

⁹<https://github.com/SparseLinearAlgebra/spla>

¹⁰Реквест с изменениями: <https://github.com/DrTimothyAldenDavis/GraphBLAS/pull/381>

¹¹Наше сообщество на GitHub: <https://github.com/SparseLinearAlgebra>

- ❄️ Brahma.FSharp¹²: программирование GPGPU с использованием F#/.Net
- ❄️ GraphBLAS-sharp¹³: обобщённая разреженная линейная алгебра в функциональном стиле (Игорь Ерин, Кирилл Гарбар, и многие другие)
 - ▶ Как решить проблему неявных нулей?
 - ▶ Можно ли упростить GraphBLAS API?
- ❄️ LamaGraph¹⁴: программно-аппаратный комплекс для ускорения обобщённой разреженной линейной алгебры (Николай Пономарев, Ефим Кубышкин)

¹²<https://github.com/YaccConstructor/Brahma.FSharp>

¹³<https://github.com/SparseLinearAlgebra/GraphBLAS-sharp>

¹⁴<https://github.com/Lamagraph/interaction-nets-in-fpga>

- Графовые нейронные сети и символьное исполнение
 - ▶ PySymGym¹⁵: инфраструктура для обучения нейронных сетей управлению символьной машиной (Анна Чистякова, Давид Ахмедов, Данил Парфёнов, и многие другие)
- Биоинформатика (Анна Чистякова, Полина Лунина, Юлия Сусанина)
 - ▶ Предсказание вторичной структуры РНК
 - ▶ Анализ метагеномных сборок

¹⁵<https://github.com/PySymGym/PySymGym>