

# Синтаксический анализ графов с использованием конъюнктивных грамматик

Азимов Р. Ш.,  
rustam.azimov19021995@gmail.com,  
Санкт-Петербургский государственный университет,  
Лаборатория языковых инструментов JetBrains

13 ноября 2017 г.

## Аннотация

Графы используются в качестве структуры данных во многих областях, например, биоинформатика, графовые базы данных. В этих областях часто необходимо вычислять некоторые запросы к большим графам. Один из наиболее распространенных запросов к графам являются навигационные запросы. Результатом вычисления таких запросов является множество неявных отношений между вершинами графа, то есть путей в графе. Естественно выделять такие отношения — пометив ребра графа символами из некоторого конечного алфавита и выделив необходимые пути в графе с помощью формальных грамматик над тем же алфавитом. Наиболее популярны запросы, использующие контекстно-свободные грамматики. Ответом на такие запросы обычно является множество всех троек  $(A, m, n)$ , для которых существует путь в графе от вершины  $m$  до вершины  $n$  такой, что метки на ребрах этого пути образуют строку, выводимую из нетерминала данной контекстно-свободной грамматики  $A$ . Говорят, что такой тип запросов вычислен с использованием *реляционной семантики запросов*. Кроме того, существуют *конъюнктивные грамматики*, образующие более широкий класс грамматик, чем контекстно-свободные. Использование конъюнктивных грамматик в задаче синтаксического анализа графов позволит формулировать более сложные запросы к графу и решать более широкий круг задач. Известно, что задача вычисления запросов к графу с использованием реляционной семантики и конъюнктивных грамматик — неразрешима. В данной работе будет предложен алгоритм, вычисляющий приближенное решение данной задачи, а именно аппроксимацию сверху множества троек  $(A, m, n)$ . Предложенный алгоритм основан на матричных операциях, что позволяет повысить производительность, используя вычисления на графическом процессоре.

**Ключевые слова:** синтаксический анализ графов, конъюнктивные грамматики, транзитивное замыкание, матричные операции, вычисления на GPU

# 1 Введение

Графы используются в качестве структуры данных во многих областях, например, биоинформатика [6], графовые базы данных [4]. В этих областях часто необходимо вычислять некоторые запросы к большим графам. Один из наиболее распространенных запросов к графам являются навигационные запросы. Результатом вычисления таких запросов является множество неявных отношений между вершинами графа, то есть путей в графе. Естественно выделять такие отношения — пометив ребра графа символами из некоторого конечного алфавита и выделив необходимые пути в графе с помощью формальных грамматик (регулярные выражения, контекстно-свободные грамматики) над тем же алфавитом. Наиболее популярны запросы, использующие контекстно-свободные грамматики, так как КС-языки обладают большей выразительной мощностью, чем регулярные.

Ответом на запросы с использованием КС-грамматик обычно является множество всех троек  $(A, t, n)$ , для которых существует путь в графе от вершины  $t$  до вершины  $n$  такой, что метки на ребрах этого пути образуют строку, выводимую из нетерминала  $A$  данной КС-грамматики. Говорят, что такой тип запросов вычислен с использованием *реляционной семантики запросов* [3].

Существует ряд алгоритмов синтаксического анализа графов с использованием реляционной семантики запросов и КС-грамматик [2; 3; 7]. Данные алгоритмы демонстрируют низкую производительность на больших графах. Одной из самых популярных техник, используемых для увеличения производительности при работе с большими объемами данных, является использование графического процессора для вычислений, но перечисленные алгоритмы не позволяют эффективно применить данную технику.

Кроме того, существует алгоритм синтаксического анализа графов с использованием реляционной семантики запросов и КС-грамматик, вычисляющий матричное транзитивное замыкание. Активное использование матричных операций в данном алгоритме позволяет эффективно использовать вычисления на графическом процессоре [1].

Также существуют *конъюнктивные грамматики* [5], образующие более широкий класс грамматик, чем контекстно-свободные. Использование конъюнктивных грамматик в задаче синтаксического анализа графов позволит формулировать более сложные запросы к графу и решать более широкий круг задач. Известно, что задача вычисления запросов к графу с использованием реляционной семантики и конъюнктивных грамматик — неразрешима [3]. Один из распространенных способов найти приближенное решение неразрешимой задачи — найти аппроксимацию решения (сверху или снизу).

В данной работе будет предложен алгоритм, вычисляющий приближенное решение задачи синтаксического анализа графов с использованием реляционной семантики запросов и конъюнктивных грамматик, а именно аппроксимацию сверху множества троек  $(A, t, n)$ . Предложенный алгоритм основан на матричных операциях, что позволяет повысить производительность, используя вычисления на графическом процессоре.

- 2 Обзор
- 3 Существующие работы
- 4 Определения
- 5 Сведение синтаксического анализа графов к поиску транзитивного замыкания
- 6 Алгоритм
- 7 Апробация
- 8 Заключение

## Список литературы

1. *Che S., Beckmann B. M., Reinhardt S. K.* Programming GPGPU Graph Applications with Linear Algebra Building Blocks // International Journal of Parallel Programming. — 2016. — С. 1–23.
2. Context-free path queries on RDF graphs / X. Zhang [и др.] // International Semantic Web Conference. — Springer. 2016. — С. 632–648.
3. *Hellings J.* Conjunctive context-free path queries. — 2014.
4. *Mendelzon A., Wood P.* Finding Regular Simple Paths in Graph Databases // SIAM J. Computing. — 1995. — Т. 24, № 6. — С. 1235–1258.
5. *Okhotin A.* Conjunctive grammars // Journal of Automata, Languages and Combinatorics. — 2001. — Т. 6, № 4. — С. 519–535.
6. Quantifying variances in comparative RNA secondary structure prediction / J. W. Anderson [и др.] // BMC bioinformatics. — 2013. — Т. 14, № 1. — С. 149.
7. *Sevon P., Eronen L.* Subgraph queries by context-free grammars // Journal of Integrative Bioinformatics. — 2008. — Т. 5, № 2. — С. 100.