

На правах рукописи



Sign

Азимов Рустам Шухратуллович

**Алгоритмы выполнения навигационных запросов к
графам с использованием операций линейной алгебры**

Специальность **XX.XX.XX** —

**«Технология обработки, хранения и переработки злаковых,
бобовых культур, крупяных продуктов, плодоовощной продукции
и виноградарства»**

Автореферат

диссертации на соискание учёной степени
кандидата **физико-математических наук**

Город — 2020

Работа выполнена в учреждении с длинным длинным длинным длинным названием, в котором выполнялась данная диссертационная работа.

Научный руководитель: уч. степень, уч. звание
Фамилия Имя Отчество

Официальные оппоненты: **Фамилия Имя Отчество**,
доктор физико-математических наук, профессор,
Не очень длинное название для места работы,
старший научный сотрудник

Фамилия Имя Отчество,
кандидат физико-математических наук,
Основное место работы с длинным длинным
длинным длинным названием,
старший научный сотрудник

Ведущая организация: Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования с длинным длинным длинным длинным названием

Защита состоится DD mmmmmmmm YYYY г. в XX часов на заседании диссертационного совета Д123.456.78 при Название учреждения по адресу: Адрес.

С диссертацией можно ознакомиться в библиотеке Название библиотеки.

Отзывы на автореферат в двух экземплярах, заверенные печатью учреждения, просьба направлять по адресу: Адрес, ученому секретарю диссертационного совета Д123.456.78.

Автореферат разослан DD mmmmmmmm2020 года.
Телефон для справок: +7 (0000) 00-00-00.

Ученый секретарь
диссертационного совета
Д123.456.78,
д-р физ.-мат. наук

Sign
Фамилия Имя Отчество

Общая характеристика работы

Актуальность темы исследования. В современном мире становится всё больше данных, которые требуют анализа. При этом графы являются одной из самых распространённых структур данных. С их помощью большие объёмы информации компактно и удобно представляются для анализа и обработки. Графы используются в программной инженерии и биоинформатике, в телекоммуникациях, социальных сетях и сетевом анализе и т.д. Также, в настоящее время активно развиваются графовые базы данных, используемые для хранения и анализа больших объёмов данных в виде графов. Следует упомянуть самые популярные графовые базы данных: RedisGraph, Neo4j, HyperGraphDB и OrientDB. При этом важной является задача поиска различных путей в графах. Чтобы описать свойства искомых путей, необходимо задать определенные ограничения на них. Данные ограничения формулируются в виде запроса к графу, а ответом на запрос является информация о существовании соответствующих путей, удовлетворяющих данным ограничениям. Говорят, что такие запросы вычислены в relational семантике. Кроме того, в некоторых областях, в качестве доказательства существования таких путей необходимо предъявить все или хотя бы один из них, тогда говорят, что такие запросы вычислены в all-path и single-path семантиках соответственно.

Для описания ограничений на пути в помеченном графе естественно использовать формальные грамматики над некоторым алфавитом. Заданная грамматика ограничивает множество слов, полученных конкатенацией меток на рёбрах рассматриваемых путей. В настоящее время активно исследуются ограничения в виде контекстно-свободных (КС) грамматик, так как они позволяют описывать широкий класс запросов (КС-запросов), более широкий, чем, например, регулярные выражения. Однако большинство существующих алгоритмов вычисления КС-запросов имеют низкую производительность на больших графах, что затрудняет их применение на практике.

Одним из распространённых способов улучшения производительности алгоритмов на графах является их переформулирование в терминах линейной алгебры. Для тех алгоритмов, которые позволяют найти такую формулировку, имеется возможность применить разреженное представление матриц и параллельные вычисления, в частности, на GPU. Так, например, Лейуань Ван (Leyuan Wang) сформулировал алгоритм подсчёта количества треугольников в графе на языке линейной алгебры и реализовал полученный алгоритм на GPU. Полученная реализация показала лучшую производительность по сравнению с реализациями на CPU. Кроме того, такого рода алгоритмы зачастую просты в реализации, так как позволяют использовать существующие библиотеки линейной алгебры

(GraphBLAS::SuiteSparse, cuSPARSE, cuBLAS, m4ri, Scipy и др.). Таким образом, перспективным является направление формулирования алгоритмов вычисления КС-запросов к графам на языке линейной алгебры.

Однако возможность использования линейной алгебры в задачах поиска путей с КС-ограничениями в графах в настоящее время не исследована. Таким образом, актуальной задачей является разработка алгоритмов вычисления КС-запросов к графам, использующих различные операции линейной алгебры, и исследование их свойств.

Степень разработанности темы исследования. Множество работ посвящены переформулированию классических алгоритмов на графах в терминах операций линейной алгебры. Например, Айдын Булук (Aydin Buluç), Упасана Шридхар (Upasana Sridhar), Питер Чжан (Peter Zhang) и Арифул Азад (Ariful Azad) в своих работах показывают как можно свести к линейной алгебре такие алгоритмы, как поиск в ширину, алгоритм Дейкстры, алгоритм Беллмана-Форда и поиск наибольшего паросочетания в двудольном графе. Кроме того, существует стандарт GraphBLAS, который определяет базовые строительные блоки на языке линейной алгебры для алгоритмов на графах. GraphBLAS основан на том, что графы могут быть представлены в виде матрицы смежности или матрицы инцидентности. Также, ввиду того, что данные на практике разрежены, целесообразно использовать разреженный формат матриц. Однако, не каждый алгоритм на графах можно переформулировать на языке линейной алгебры. Так, например, до сих пор не смогли это сделать для алгоритма поиска в глубину. Также, в настоящее время, такая формулировка не найдена и для алгоритмов вычисления КС-запросов на графах.

Лесли Вэлиант (Leslie Valiant) провёл исследование, посвященное синтаксическому анализу КС-языков с использованием матричных операций. Однако предложенный им алгоритм позволяет проводить анализ только над строками, что эквивалентно анализу лишь специальных, линейно помеченных, графов. Кроме того, Филип Брэдфорд (Philip Bradford) исследовал задачу поиска кратчайших путей в графе с КС-ограничениями. Но предложенные им алгоритмы хоть и сформулированы на языке линейной алгебры, но предназначены только для частного случая КС-грамматик и/или специализированных графов. Таким образом, данные алгоритмы не применимы к поиску путей с произвольными КС-ограничениями в произвольных графах.

Задаче поиска путей с КС-ограничениями в произвольных графах посвящены работы таких учёных, как Семён Григорьев, Джелле Хеллингс (Jelle Hellings), Сюованг Чжан (Xiaowang Zhang) и Мартин Мюзиканте (Martin Musicante). В данных исследованиях используются подходы, основанные на различных алгоритмах синтаксического анализа (LR, LL, GLL, СΥΚ). Однако, данные алгоритмы не представлены в терминах линейной

алгебры. Впервые вопрос о возможности нахождения матричного алгоритма вычисления КС-запросов к графам исследовал Михалис Яннакакис (Mihalis Yannakakis). Он указывал, что алгоритм Вэлианта может быть расширен для обработки графов без циклов, однако сомневался в возможности расширения алгоритма Вэлианта до произвольных графов.

Таким образом, на текущий момент не существует алгоритма вычисления произвольных КС-запросов к произвольным графам, выраженного на языке линейной алгебры. Поэтому необходимо дальнейшее исследование возможности разработки таких алгоритмов.

Целью данной работы является исследование применимости линейной алгебры в задаче вычисления КС-запросов к графам.

Целью данной работы является исследование применимости линейной алгебры в задаче вычисления КС-запросов к графам и получение реализаций данных алгоритмов с улучшенной производительностью.

Целью данной работы является исследование применимости линейной алгебры в задаче вычисления КС-запросов к графам и получение более высокопроизводительных реализаций для данных алгоритмов с использованием параллельных вычислительных систем.

Достижение поставленной цели обеспечивается решением следующих **задач**:

1. Разработать подход к вычислению КС-запросов к графам в терминах линейной алгебры.
2. Разработать семейство алгоритмов, использующих предложенный подход и вычисляющих КС-запросы с relational, single-path и all-path семантиками.
3. Разработать семейство алгоритмов вычисления КС-запросов к графам, использующих предложенный подход, вычисляющих КС-запросы с relational, single-path и all-path семантиками, и не требующих преобразований входной КС-грамматики.
4. Реализовать предложенные алгоритмы на CPU и GPU, провести их экспериментальное исследование на синтетических и реальных данных, сравнить их между собой и с существующими реализациями.

Научная новизна:

1. Алгоритмы вычисления КС-запросов к графам, которые используют подход, предложенный в диссертации, отличаются от алгоритмов, предложенных в работах Семёна Григорьева, Сяованга Чжана, Джелле Хеллингса, Филиппа Брэдфорда и Мартина Мюзиканте тем, что первые сформулированы в терминах линейной алгебры и вычисляют произвольные КС-запросы к произвольным графам. С практической точки зрения, предложенный подход, позволяет применять широкий класс оптимизаций для вычисления

матричных операций и дает возможность автоматически распараллеливать вычисления за счёт существующих решений.

2. Впервые получены алгоритмы вычисления произвольных КС-запросов к произвольным графам, сформулированные в терминах линейной алгебры.
3. Алгоритмы, предложенный в диссертации, отличается от алгоритмов, предложенных в работах Семёна Григорьева, Сяованга Чжана, Джелле Хеллингса, Филиппа Брэдфорда и Мартина Мюзиканте возможностью работать с произвольными КС-грамматиками без необходимости их преобразования. Таким образом, удаётся избежать значительного увеличения размеров входной грамматики, от которого напрямую зависит временная сложность данных алгоритмов.
4. Экспериментальное исследование алгоритмов вычисления КС-запросов с различной семантикой к произвольным графам, использующих операции линейной алгебры проводится впервые и позволяет судить о применимости на практике разработанных алгоритмов.

Теоретическая и практическая значимость работы.

Теоретическая значимость диссертационного исследования заключается в разработке подхода к вычислению КС-запросов к графам, использующего операции линейной алгебры, в разработке формальных алгоритмов, использующих данный подход, а также в формальном доказательстве завершаемости, корректности и оценок временной сложности разработанных алгоритмов.

Полученные в ходе исследования реализации позволили оценить применимость на практике алгоритмов вычисления КС-запросов, сформулированных в терминах линейной алгебры. Кроме того, данные реализации могут быть использованы для интеграции с такими графовыми базами данных, как RedisGraph. Это добавит возможность вычислять КС-запросы к этим базам данных.

Методология и методы исследования.

Методология исследования основана на линейной алгебре и теории графов. Подход, предлагающий использовать операции линейной алгебры при анализе графов, начал активно развиваться со второй половины 20-го века. В 2013 году был создан стандарта GraphBLAS, определяющий для алгоритмов на графах базовые строительные блоки на языке линейной алгебры. На текущий момент множество классических алгоритмов из теории графов были сформулированы в терминах линейной алгебры.

Кроме того, в исследовании использовалась теория формальных языков. Одной из задач, для которых до сих пор не найдена формулировка в терминах линейной алгебры, является поиск путей в графе с ограничениями в виде КС-грамматик. Данная задача использует подход к анализу

строк, который начал активно развиваться в 50-х годах 20-го века в связи с изучением естественных языков (работы Н. Хомского). В последствии этот подход получил широкое распространение в различных областях, в том числе и связанных с анализом графов.

Доказательство завершаемости, корректности и оценок временной сложности предложенных алгоритмов проводится с применением линейной алгебры, теории формальных языков, теории графов и теории сложности алгоритмов.

Основные положения, выносимые на защиту:

1. Разработан подход к вычислению КС-запросов к графам в терминах линейной алгебры.
2. Разработано семейство алгоритмов, использующих предложенный подход и вычисляющих КС-запросы с relational, single-path и all-path семантиками. Доказана завершаемость и корректность предложенных алгоритмов. Получена теоритическая оценка сверху временной сложности алгоритмов.
3. Разработано семейство алгоритмов, использующих предложенный подход и вычисляющих КС-запросы с relational, single-path и all-path семантиками, а также не требующих преобразований входной КС-грамматики. Доказана завершаемость и корректность предложенных алгоритмов. Получена теоритическая оценка сверху временной сложности алгоритмов.
4. Проведено экспериментальное исследование разработанных алгоритмов. Предложенные алгоритмы реализованы на CPU и GPU. Дальше детали.

Степень достоверности и апробация результатов. Достоверность и обоснованность результатов исследования опирается на использование формальных методов исследуемой области, выполнение формальных доказательств и инженерные эксперименты.

Основные результаты работы были представлены на ряде международных научных конференций: GRADES'18, GRADES'20, ADBIS'20, SIGMOD'21(еще не приняли). А также разработка предложенных алгоритмов была поддержана грантом РНФ №18-11-00100 и грантом РФФИ №19-37-90101.

Публикации. Все результаты диссертации изложены в 7 [1; 2; 3; 4; 5; 6; 7] научных работах. Из них 2 работы [6; 7] индексируются ВАК и 6 работ [1; 2; 3; 4; 5; 6] индексируются Scopus. Работы [1; 2; 3; 4; 6; 7] написаны в соавторстве. В [1; 6; 7] Р. Азимову принадлежит разработка алгоритма, доказательство его корректности и завершаемости, реализация алгоритма, работа над текстом. В [2] Р. Азимову принадлежит разработка алгоритма, доказательство его корректности и завершаемости, работа над

текстом. В [3; 4] Р. Азимову принадлежит работа над доказательствами корректности и завершаемости алгоритма, работа над текстом.

Список литературы

- [1] Azimov R. Context-Free Path Querying by Matrix Multiplication / Azimov R., Grigorev S. // In Proceedings of the 1st Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA) (GRADES-NDA'18)
- [2] Azimov R. Context-Free Path Querying with Single-Path Semantics by Matrix Multiplication / Terekhov A., Khoroshev A., Azimov R., Grigorev S. // In Proceedings of the 3rd Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA) (GRADES-NDA'20)
- [3] Azimov R. Context-Free Path Querying by Kronecker Product / Orachev E., Epelbaum I., Azimov R., Grigorev S. // In Proceedings of the 24th European Conference on Advances in Databases and Information Systems (ADBIS'20)
- [4] Azimov R. Context-Free Path Querying by Kronecker Product большая версия / Orachev E., Epelbaum I., Azimov R., Grigorev S. // In Proceedings of the (SIGMOD'21)
- [5] Azimov R. Ненаписанная работа матричный алгоритм по всем путям
- [6] Azimov R. Path Querying with Conjunctive Grammars by Matrix Multiplication / Azimov R., Grigorev S. // Programming and Computer Software. – 2019. – Vol. 45. – №. 7. – pp. 357-364.
- [7] Азимов Р. Ш. Синтаксический анализ графов с использованием конъюнктивных грамматик / Азимов Р., Григорьев С. // Труды ИСП РАН, 2018, том 1 вып. 2, с. 3-4.

Содержание работы

Во введении обосновывается актуальность исследований, проводимых в рамках данной диссертационной работы, приводится обзор научной литературы по изучаемой проблеме, формулируется цель, ставятся задачи работы, излагается научная новизна и практическая значимость представляемой работы. В последующих главах сначала описывается общий

принцип, позволяющий ..., а потом идёт апробация на частных примерах: ... и

Первая глава посвящена ...

картинку можно добавить так:

L^AT_EX

а) L^AT_EX



б) Knuth

Рис. 1 — Подпись к картинке.

Формулы в строку без номера добавляются так:

$$\lambda_{T_s} = K_x \frac{dx}{dT_s}, \quad \lambda_{q_s} = K_x \frac{dx}{dq_s},$$

Вторая глава посвящена исследованию

Третья глава посвящена исследованию

Можно сослаться на свои работы в автореферате. Для этого в файле `Synopsis/setup.tex` необходимо присвоить положительное значение счётчику `\setcounter{usefootcite}{1}`. В таком случае ссылки на работы других авторов будут подстрочными. Изложенные в третьей главе результаты опубликованы в [vakbib1; vakbib2]. Использование подстрочных ссылок внутри таблиц может вызывать проблемы.

В четвертой главе приведено описание

В заклучении приведены основные результаты работы, которые заключаются в следующем:

1. Разработан подход к вычислению КС-запросов к графам, использующий операции линейной алгебры.
2. Разработано семейство алгоритмов вычисления КС-запросов к графам, использующих предложенный подход и позволяющих предоставлять искомые пути. Доказана завершаемость и корректность предложенных алгоритмов.
3. Разработано семейство алгоритмов вычисления КС-запросов к графам, использующих предложенный подход и не требующего преобразований входной КС-грамматики. Доказана завершаемость и корректность предложенных алгоритмов.

4. Проведено экспериментальное исследование разработанных алгоритмов.

При использовании пакета `biblatex` список публикаций автора по теме диссертации формируется в разделе «Публикации.» файла `common/characteristic.tex` при помощи команды `\nocite`

Азимов Рустам Шухратулович

Алгоритмы выполнения навигационных запросов к графам с использованием
операций линейной алгебры

Автореф. дис. на соискание ученой степени **канд. физ.-мат. наук**

Подписано в печать _____._____._____. Заказ № _____

Формат 60×90/16. Усл. печ. л. 1. Тираж 100 экз.

Типография _____

