

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 происходит значительный рост данных, которые требуют анализа. Графы используются в качестве структуры данных для представления больших объемов информации в компактной и удобной для анализа форме во многих областях, например, в биоинформатике, в графовых базах данных, при статическом анализе программ. При этом оказывается необходимым выявлять сложные зависимости между вершинами графа, которые выражаются путями между ними. Поэтому очень важен широкий класс задач, посвященный поиску путей в графах. Чтобы описать свойства искомым путей в графе, необходимо задать определенные ограничения на них. Данные ограничения формулируются в виде запроса к графу, а ответом на запрос является информация о существовании путей, удовлетворяющих данным ограничениям между каждой парой вершин. Кроме того, в некоторых областях, в качестве доказательства существования таких путей необходимо предъявить все или хотя бы один из них.

Для описания ограничений на пути в помеченном графе естественно выделять пути с помощью формальных грамматик (регулярные выражения, контекстно-свободные грамматики) над алфавитом, содержащим метки на ребрах этого графа. В настоящее время активно исследуются ограничения в виде контекстно-свободных (КС) грамматик, так как они позволяют описывать более широкий класс запросов (КС-запросов), чем 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 формулирование в терминах линейной алгебры. Для тех алгоритмов, которые позволяют найти такую формулировку, имеется возможность применить широкий класс оптимизаций, например, разреженное представление матриц и параллельные вычисления. Кроме того, данные алгоритмы зачастую просты в реализации, так как позволяют использовать существующие библиотеки линейной алгебры. Таким образом, формулирование различных алгоритмов на графах в терминах линейной алгебры является перспективным направлением для оптимизации вычисления КС-запросов к графам.

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 формулированию классических алгоритмов на графах в

терминах операций линейной алгебры. Например, такая формулировка была найдена для таких алгоритмов, как поиск в ширину, алгоритм Дейкстры, алгоритм Беллмана-Форда, поиск наибольшего паросочетания в двудольном графе и т.д. Кроме того, существует стандарт GraphBLAS, который определяет базовые строительные блоки для графовых алгоритмов на языке линейной алгебры. GraphBLAS основан на том, что разреженная матрица может использоваться для представления графов в виде матрицы смежности или матрицы инцидентности. Однако формулировка в терминах операций линейной алгебры для алгоритмов вычисления КС-запросов на графах в настоящее время не найдена.

Существует классическое исследование Лесли Вэлианта, посвященное синтаксическому анализу КС-грамматик с использованием матричных операций. Однако изложенный в нём алгоритм позволяет производить анализ только над строками, что эквивалентно анализу частного случая графов — линейных помеченных графов. Кроме того, существуют исследования Филиппа Брэдфорда посвященные поиску кратчайших путей в графе с КС-ограничениями. Алгоритмы, представленные в данных исследованиях предназначены для частного случая КС-грамматик и/или частного случая 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 Семёна Григорьева, Джелле Хеллингса, Сяованга Чжана, Сиро Медейроса. В данных исследованиях используются подходы, основанные на различных алгоритмах синтаксического анализа (LR, LL, GLL, СΥΚ). Однако, данные алгоритмы сформулированы не в терминах операций линейной алгебры над матрицами и векторами. Впервые вопросом о возможности нахождения матричного алгоритма вычисления КС-запросов к графам занялся Яннакакис в своем исследовании о применимости методов из теории графов в теории баз данных. Он указывал на то, что алгоритм Вэлианта может быть расширен для работы с графами без циклов, однако сомневался в возможности расширения до произвольных графов.

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

Целью данной работы является разработка алгоритмов вычисления КС-запросов к графам, использующих операции линейной алгебры.

Достижение поставленной цели обеспечивается решением следующих задач:

1. Разработать подход к вычислению КС-запросов к графам, использующий операции линейной алгебры.

2. Разработать семейство алгоритмов вычисления КС-запросов к графам, использующих предложенный подход и позволяющих предоставлять искомые пути.
3. Разработать семейство алгоритмов вычисления КС-запросов к графам, использующий предложенный подход и не требующего преобразований входной КС-грамматики.
4. Провести экспериментальное исследование разработанных алгоритмов.

Научная новизна:

1. Подход, предложенный в диссертации, отличается от подходов, использованных в работах Семёна Григорьева, Сяованга Чжана, Джелле Хеллингса, Филиппа Брэдфорда и Сиро Медейроса активным использованием матричных операций в процессе вычисления запросов. Это позволяет применять широкий класс оптимизаций для вычисления матричных операций и дает возможность автоматически распараллеливать вычисления за счёт существующих решений.
2. Задача вычисления КС-запросов к графам впервые сведена к вычислению операций линейной алгебры и получены соответствующие алгоритмы.
3. Алгоритмы, предложенный в диссертации, отличается от алгоритмов, предложенных в работах Семёна Григорьева, Сяованга Чжана, Джелле Хеллингса, Филиппа Брэдфорда и Сиро Медейроса возможностью работать с произвольными КС-грамматиками без необходимости приводить их в нормальную форму. Таким образом, удаётся избежать значительного увеличения размеров входной грамматики, от которой напрямую зависит временная сложность данных алгоритмов.
4. Экспериментальное исследование алгоритмов вычисления КС-запросов к 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 и корректности разработанных алгоритмов.

Полученные в ходе исследования 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 подходе использования операций линейной алгебре при

анализе графов, который начал активно развиваться с 2013 года в связи с созданием стандарта GraphBLAS. Данный стандарт определяет базовые строительные блоки для графовых алгоритмов на языке линейной алгебры. GraphBLAS основан на том, что разреженная матрица может использоваться для представления графов в виде матрицы смежности или матрицы инцидентности. На текущий момент множество классических алгоритмов из теории графов были сформулированы в терминах линейной алгебры.

Однако для большинства задач анализа графов такая формулировка ещё не найдена. Одной из таких задач является поиск путей в графе с ограничениями в виде КС-грамматик. Данная задача использует подход к анализу строк с помощью теории формальных языков и грамматик, 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. При этом основными элементами данного подхода являются алфавит и грамматика исследуемого языка, выступающего в качестве ограничения на искомые пути в графе. Решаемые в связи с этим задачи связаны с поиском эффективных алгоритмов нахождения путей, удовлетворяющих заданным ограничениям.

Доказательство завершаемости и корректности предложенного алгоритма 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, 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, использующий предложенный подход и не требующего преобразований входной КС-грамматики. Доказана завершаемость и корректность предложенных алгоритмов.
4. Проведено экспериментальное исследование разработанных алгоритмов.

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 доложены 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, которые содержат основные результаты работы. 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, использующий 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 _____