

На правах рукописи

Азимов Рустам Шухратуллович

Эффективное вычисление запросов к большим графам

Специальность 05.13.11 —
Математическое и программное обеспечение вычислительных
машин, комплексов и компьютерных сетей

Автореферат
диссертации на соискание учёной степени
кандидата физико-математических наук

Санкт-Петербург — 2022

Работа выполнена на кафедре системного программирования Санкт-Петербургского государственного университета

Научный руководитель: кандидат физико-математических наук, доцент
Кознов Дмитрий Владимирович

Официальные оппоненты: Оппонент1,
доктор физико-математических наук, профессор,
федеральное государственное бюджетное учреждение науки Институт систем информатики им.
А.П. Ершова Сибирского отделения Российской академии наук, директор

Оппонент2,
кандидат технических наук, доцент,
федеральное государственное автономное образовательное учреждение высшего образования
“Санкт-Петербургский политехнический университет Петра Великого”, исполняющий обязанности заведующего кафедрой

Ведущая организация: Федеральное государственное бюджетное учреждение науки Институт системного программирования Российской академии наук (ИСП РАН)

Защита состоится _____ г. в _____ часов на заседании диссертационного совета Д 212.232.51 на базе Санкт-Петербургского государственного университета по адресу: 198504, Санкт-Петербург, Петродворец, Университетский пр., 28, математико-механический факультет, ауд. 405.

С диссертацией можно ознакомиться в Научной библиотеке Санкт-Петербургского государственного университета по адресу: 199034, Санкт-Петербург, Университетская наб., д. 7/9, а также на сайте <http://spbu.ru/science/disser/>.

Автореферат разослан _____ 20____ года

Ученый секретарь

диссертационного совета

Д 212.232.51, д.ф.-м.н., профессор

Демьянович Юрий Казимирович

Общая характеристика работы

Актуальность темы исследования

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 вычислять запросы к большим графам с целью выявления сложных зависимостей между их вершинами. Результатом вычисления таких запросов является множество неявных отношений между вершинами графа, то есть путей в графе. Естественным пометить ребра графа символами из некоторого конечного алфавита и выделять пути с помощью формальных грамматик над тем же алфавитом (регулярные выражения, контекстно-свободные грамматики). В настоящее время активно исследуются запросы к графам в виде контекстно-свободных (КС) грамматик, так как они позволяют описывать более широкий класс запросов, чем регулярные выражения. Также интерес представляют вложенные регулярные запросы к графам, так как они расширяют выразительную мощность регулярных выражений добавлением фильтров с логическими операциями, которые в свою очередь содержат регулярные выражения.

Однако большинство существующих алгоритмов в данной области имеют низкую производительность на больших графах, что затрудняет их анализ. Хорошую производительность показал матричный подход к вычислению КС-запросов к графам. Данный подход позволяет нагрузить основную вычислительную сложность на вычисление матричных операций. Кроме того, в процессе анализа может быть применен широкий класс матричных оптимизаций, например, разреженное представление матриц, параллельное вычисление. Но существующий матричный алгоритм в данной области позволяет лишь установить факт наличия между двумя вершинами пути определенного вида, при этом сам путь не предоставляется, хотя во многих областях нахождение пути необходимо. Поэтому для вычисления КС-запросов к графам с некоторыми семантиками не существует эффективного алгоритма. Таким образом, для большинства типов запросов к графам необходима разработка алгоритмов, эффективно работающих на больших графах.

Степень разработанности темы исследования

TODO

Объект исследования

Объектом исследования являются алгоритмы вычисления запросов к графам.

Цель и задачи диссертационной работы

Целью данной работы является разработка эффективных алгоритмов вычисления запросов к графам.

Достижение поставленной цели обеспечивается решением следующих задач.

1. Разработать матричный алгоритм вычисления КС-запросов к графам, позволяющий предоставлять по одному искомому пути для каждой пары вершин, если они существуют.
2. Разработать матричный алгоритм вычисления КС-запросов к графам, использующий произведение Кронекера.
3. Разработать алгоритм вычисления вложенных регулярных запросов к графам, транслирующий запрос в программу на даталоге.

Методология и методы исследования

TODO

Положения, выносимые на защиту

1. Разработан матричный алгоритм вычисления КС-запросов к графам, позволяющий предоставлять по одному искомому пути для каждой пары вершин, если они существуют. Доказана завершаемость и корректность предложенного алгоритма.
2. Разработан матричный алгоритм вычисления КС-запросов к графам, использующий такую матричную операцию, как произведение Кронекера. Доказана завершаемость и корректность предложенного алгоритма.
3. Разработан алгоритм вычисления вложенных регулярных запросов к графам, транслирующий запрос в программу на даталоге. Доказана завершаемость и корректность предложенного алгоритма.

Научная новизна

Научная новизна полученных в ходе исследования результатов заключается в следующем.

1. Алгоритм, предложенный в диссертации, отличается от аналогов (работы Семёна Григорьева, Джелле Хеллингса, Сяованга Чжана) активным использованием матричных операций в процессе вычисления запросов и отличается от матричного алгоритма Азимова Рустама возможностью построения по одному искомому пути для каждой пары вершин, если они существуют. Это позволяет как применять широкий класс матричных оптимизаций, так и предъявлять пути в качестве доказательства отношения определенного вида между парами вершин, что является важным результатом анализа во многих областях.
2. Алгоритм, предложенный в диссертации, отличается от аналогов (работы Семёна Григорьева, Джелле Хеллингса, Сяованга Чжана) активным использованием матричных операций в процессе вычисления запросов и отличается от матричного алгоритма Азимова Рустама использованием в процессе вычисления запросов произведения Кронекера и представлением КС-грамматики запроса в виде рекурсивного автомата. Это позволяет более эффективно вычислять сложные запросы к большим графам.
3. Алгоритм, предложенный в диссертации, отличается от аналогов (работы Хорхе Переса, Георга Готтлоба, Марсело Аренааса) трансляцией запроса в программу на даталоге и её последующим вычислением с помощью существующих техник. Это позволяет контролировать процесс вычисления запросов и исследовать только необходимую часть графа, что критически важно при анализе больших графов.

Теоретическая и практическая значимость работы

TODO

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: ... Дополнительной апробацией является то, что разработка предложенных алгоритмов была поддержана ... (РНФ?, РФФИ?)

Публикации по теме диссертации

Все результаты диссертации изложены в 3 научных работах, из которых 3 [1,2,3] содержат основные результаты работы и индексируются Scopus. Работы [1,2,3] написаны в соавторстве. В [1] Р. Азимову принадлежит разработка алгоритма, доказательство его корректности и завершаемости, работа

над текстом. В [2] Р. Азимову принадлежит работа над доказательствами корректности и завершаемости алгоритма, работа над текстом. В [3] автору принадлежит схема трансляции запросов к графам в программу на даталоге, доказательство корректности и завершаемости алгоритма, работа над текстом.