«Утверждаю»
ВРИО директора ГЦ РАН
член-корреспондент РАН,
доктор физико-математических наук
Соловьев А.А.
28.11.18.

Отзыв ведущей организации
Федерального государственного бюджетного учреждение науки Геофизический центр Российской академии наук (ГЦ РАН)
на диссертацию Матерухина Андрея Викторовича
«Теоретические основы и методология обработки потоков пространственно-временных данных», представленную на соискание ученой степени доктора технических наук по специальности 25.00.35 — «Геоинформатика».

Отзыв составлен на основании заключения структурного подразделения ГЦ РАН — Лаборатории геоинформатики и геомагнитных исследований по результатам обсуждения диссертации, проведенного на семинаре Лаборатории геоинформатики и геомагнитных исследований.

Тематика рассматриваемой работы связана с развитием методологии и технологии создания геоинформационных систем (ГИС). В задачи геоинформатики как науки входит задача разработки методов, методик и технологических решений, которые позволяли бы создавать ГИС, функциональность которых соответствовала текущим потребностям наук о Земле и смежных с ними социально-экономических наук. В работе показано, что имеющий место в настоящее время лавинообразный рост объемов геопространственных данных, вызванных развитием сетей геофизических наблюдений и увеличением частоты регистрации широкого ряда физических параметров, требует адекватных подходов к их обработке, хранению, представлению и анализу. В настоящее время данная проблема становится одной из наиболее обсуждаемых в среде экспертов в области наук о данных и междисциплинарного взаимодействия. Последние достижения в области геоинформатики и системного анализа позволяют найти решение данной проблемы. Они демонстрируют достаточную универсальность для применения к пространственным данным, имеющим географическую привязку. Современные геоинформационные технологии являются одним из основных подходов к сбору, хранению, анализу и представлению больших объемов

пространственных данных. В этой связи актуальной является проблема обеспечения непрерывности обработки потоков пространственно-временных данных в ГИС.

В диссертационной работе показано, что современные требования к непрерывности обработки потоков пространственно-временных данных высокой интенсивности в ГИС не могут быть в полной мере удовлетворены в рамках традиционного подхода с использованием реляционных баз данных и модели «обработка после обязательного сохранения».

В диссертационной работе автор приводит две причины, которые обуславливают невозможность обеспечить непрерывность обработки:

- это несоответствие объекта обработки (потоков пространственно-временных данных) реляционной модели;
- это несоответствие требований к непрерывности обработки модели «обработка после сохранения».

В рассматриваемой диссертационной работе предложено решение проблемы обеспечения непрерывности обработки потоков пространственно-временных данных в ГИС в режиме реального времени.

Достигнутые результаты

В рамках проведенного диссертационного исследования автором получены следующие научные результаты:

1. Предложены новые понятия и термины геоинформатики, дополняющие имеющийся понятийный аппарат в области обработки пространственных и пространственно-временных данных.
2. Разработана математическая модель процессов непрерывной обработки потоков пространственно-временных данных в ГИС в режиме реального времени.
3. Разработана система типов пространственно-временных данных.
4. Разработана методология обработки потоков пространственно-временных данных в ГИС.

Тема диссертации соответствует следующим пунктам паспорта научной специальности 25.00.35 — «Геоинформатика»:

П.1 «Теоретические и экспериментальные исследования в области развития научных и методических основ геоинформатики»

П.6 «Математические методы, математическое, информационное, лингвистическое и программное обеспечение для ГИС»

П.7 «Геоинформационное картографирование и другие виды геомоделирования, системный анализ многоуровневой и разнородной геоинформации».

П.9 «Геоинформационные инфраструктуры, методы и технологии хранения и использования геоинформации на основе распределенных баз данных и знаний»

Значимость полученных автором диссертации результатов

Согласно паспорту специальности научной специальности 25.00.35— «Геоинформатика»: «Геоинформатика — область науки и техники, отражающая и изучающая природные и социально-экономические геосистемы, их взаимодействие и развитие посредством компьютерного моделирования на основе информационных систем и технологий, баз данных и баз знаний».

Моделирование и изучение поведения геосистем тем эффективнее, чем выше уровень пространственно-временной детализации данных, характеризующих эти геосистемы. Рост пространственно-временной детализации приводит к росту интенсивности потоков данных. Такое увеличение интенсивности потоков пространственно-временных данных затрудняет непрерывную обработку поступающих новых данных и даже, как показано в рассматриваемой диссертационной работе, приводит на определенном уровне интенсивности этих потоков к невозможности непрерывной обработки поступающих данных при использовании реляционной базы данных.

В диссертационной работе автором подчеркивается, что необходимость организации непрерывной геоинформационной диагностики состояния геосистем, которая возможна только при обеспечении непрерывности обработки данных, поступающих в геоинформационную систему, привела к тому, что в Программу фундаментальных научных исследований государственных академий наук на 2013—2020 годы (в редакции распоряжения Правительства Российской Федерации от 31 октября 2015 г. № 2217-р) в раздел «138. Научные основы разработки методов, технологий и средств исследования поверхности и недр Земли, атмосферы, включая ионосферу и магнитосферу Земли, гидросферы и криосферы; численное моделирование и геоинформатика (инфраструктура пространственных данных и ГИС-технологии)» в число основных ожидаемых результатов были включены ГИС, банки и базы данных для непрерывной геоинформационной диагностики состояния геосистем, процессов природопользования и возникновения опасных природных процессов в различных регионах России.

В Приложении 1 к диссертационной работе приводится Справка о внедрении, подписанной Научным руководителем проекта «Комплексное геоинформационное картографирование и оценка воздействия источников загрязнения на состояние окружающей среды России» членом-корреспондентом РАН, доктором технических наук, профессором Савиных В.П., в которой указывается, что научные результаты диссертации были использованы и продолжают использоваться при выполнении проекта № 17-05-41156

«Комплексное геоинформационное картографирование и оценка воздействия источников загрязнения на состояние окружающей среды России», поддерживаемого совместно федеральным государственным бюджетным учреждением «Российский фонд фундаментальных исследований» и Всероссийской общественной организацией «Русское географическое общество».

В Справке о внедрении также указано, что реализация разработанной в ходе выполнения диссертационной работы методологии обработки потоков пространственно-временных данных, решающей проблему обеспечения непрерывности такой обработки, сделало возможным создание пилотного проекта ГИС, предназначенной для непрерывной геоинформационной диагностики состояния геосистем, процессов природопользования и возникновения опасных природных процессов в различных регионах России. Таким образом, **практическая значимость** полученных автором диссертации результатов для развития геоинформатики состоит в том, что они в совокупности решают проблему обеспечения непрерывности обработки потоков пространственно-временных данных в геоинформационных системах и делают возможным создание ГИС, предназначенных для непрерывной геоинформационной диагностики состояния геосистем.

Теоретическая значимость полученных автором диссертации результатов для развития геоинформатики состоит в том, что они дополняют и развивают фундаментальные аспекты теории геоинформатики, относящиеся к обработке больших объемов пространственно-временных данных. В настоящее время разработчики программного обеспечения для ГИС традиционно используют реляционные системы управления базами данных (СУБД). И в настоящее время в области ГИС не произошло перехода от СУБД к системам управления потоками данных (СУПД), поскольку для непрерывной обработки потоков пространственно-временных данных в СУПД требуется разработка теоретической основы и методологии такой обработки.

Представленная в диссертационной работе математическая модель процессов обработки потоков пространственно-временных данных в ГИС в режиме реального времени в отличие от ранее предложенных моделей обработки данных в ГИС учитывает специфику потоков пространственно-временных данных и позволяет реализовать непрерывную обработку потоков пространственно-временных данных. Разработанная диссертантом и представленная в работе система типов пространственно-временных данных в виде расширенной сигнатуры многосортной алгебраической системы в отличие от других учитывает специфику потоков пространственно-временных данных, обеспечивает гибкость представления пространственно-временных данных в ГИС и дополнительные возможности построения запросов к этим данным.

Рекомендации по использованию результатов и выводов, приведенных в диссертации:

Рекомендуемые области применения полученных научных результатов — это те области исследования в науках о Земле и смежных с ними науках, в которых возникают задачи, требующие независимости установленного максимального времени задержки выполнения запросов пространственной аналитики к входящим потокам пространственно-временных данных от интенсивности этих потоков.

Результаты диссертационного исследования могут быть использованы:

- для разработки приложений операционной аналитики, обрабатывающих и анализирующих данные от спутниковых систем наблюдения Земли;
- при развитии сетей геофизического мониторинга, в т.ч. сейсмического, геомагнитного, океанического, атмосферного и др.
- при проектировании и разработке систем геодезического контроля крупных инженерных объектов;
- при проектировании и разработке систем экологического мониторинга и цифрового картографирования загрязнений;
- при проектировании и разработке новых систем контроля периметра на объектах большой площади.

Разработанный в диссертационном исследовании теоретический аппарат рекомендуется использовать исследователям в области геоинформатики, а также разработчиками программного обеспечения. Разработанная в диссертационном исследовании математическая модель процессов непрерывной обработки потоков пространственно-временных данных в ГИС в режиме реального времени рекомендуется для использования в качестве основы для разработки и внедрения приложений, ориентированных на обработку потоков пространственно-временных данных высокой интенсивности. Такой подход позволяет свести проблему обработки потоков пространственно-временных данных в геоинформационных системах в режиме реального времени к проблеме реализации пространственно-временных запросов монотонными операторами.

Разработанную в диссертационном исследовании систему типов пространственно-временных данных рекомендуется использовать при разработке ГИС-приложений, предназначенных для обработки больших потоков данных. Это обусловлено гибкостью предлагаемой системы типов, что существенно снижает сложность и трудоемкость реализации различных запросов пространственной аналитики к потокам пространственно-временных данных.

Разработанная в рамках диссертационного исследования методология обработки потоков пространственно-временных данных в ГИС рекомендуется для применения в условиях использования вычислительных кластераов. Для эффективной обработки интенсивных потоков пространственно-временных данных в соответствии с предложенной методологией в условиях вычислительного кластера необходима программная среда, которая поддерживает реализацию пространственно-временных типов данных, а также операций над ними, и может обеспечить легкое горизонтальное масштабирование.

При выполнении этих условий разработанная диссертантом методология обеспечивает независимость времени выполнения пространственных запросов к входящим потокам пространственно-временных данных от интенсивности этих потоков.

По диссертационной работе имеются следующие замечания

Несмотря на отмеченные достоинства работы, ее актуальность, она имеет ряд недостатков.

В диссертационной работе не приведены конкретные числовые значения показателей интенсивности потоков пространственно-временных данных, при которых становится невозможной непрерывная обработка данных, поступающих в геоинформационную систему, использующую системы обработки пространственно-временных данных на основе систем управления реляционными базами данных. По нашему мнению, указание этих значений помогло бы разработчикам геоинформационных систем в принятии решений об использовании предложенной методологии в конкретных проектах.

Главным, на взгляд рецензента, недостатком является то, что в работе уделено очень мало внимания примерам практической реализации предлагаемого подхода. Из 146 страниц текста (без учета списка литературы) 4 глава, в которой раскрываются вопросы применимости полученных результатов занимает всего 23 страницы. В тексте главы приведены примеры существующих ГИС и систем управления и анализа данными, разработанными сторонними организациями. Эту информацию следует отнести к 1 главе, к аналитическому обзору.

Из текста диссертации рецензенту не удалось получить представление о конкретных результатах практической реализации предложенного диссертантом теоретического аппарата. Возможно, это было сделано, но не приведено подробно в тексте диссертации. Приведены лишь ссылки на выполнение диссертантом определенных работ в рамках госзадания. Однако все ограничено исключительно декларированием наличия результатов. Конкретные результаты, технические параметры не приведены. Если они носят закрытый характер, то тогда следовало бы защищаться на специальном совете.

В диссертационной работе не приведено описание пилотного проекта ГИС, предназначенной для непрерывной диагностики состояния геосистем, процессов природопользования и возникновения опасных природных процессов в различных регионах России, о котором говорится о Справке о внедрении. По нашему мнению, описание деталей этого проекта позволило бы более детально рассмотреть вопросы практической применимости полученных научных результатов. Суммируя сказанное, рецензенту сложно однозначно ответить на простой вопрос, чем проведенное исследование может повысить качество проводимых исследований, проводимых в области геофизики, геодезии, экологии, природопользования и т.п. Текст диссертации, в целом, представляется аморфным, неконкретным и трудным для восприятия. Основная часть иллюстраций (например, 1.1, 2.1, 2.2, 2.3, 3.1, 3.9) оформлена неудачно. Надписи плохо читаемы и неудобны для восприятия. Автору следовало бы пожертвовать детальностью для достижения большей наглядности или привести иллюстрации в качестве отдельных приложений. Высказанные замечания носят рекомендательный характер и могут быть учтены диссертантом при подготовке доклада для публичной защиты диссертации.

Заключение о соответствии диссертационной работы критериям, установленным Положением о присуждении ученых степеней:

Диссертационная работа Матерухина А. В. «Теоретические основы и методология обработки потоков пространственно-временных данных» соответствует требованиям, предъявляемым к диссертациям на соискание ученой степени доктора наук, установленным Положением о присуждении ученых степеней, является научно-квалификационной работой, в которой решена научная проблема обеспечения непрерывности обработки потоков пространственно-временных данных в геоинформационных системах (ГИС) в режиме реального времени, имеющая важное практическое значение, а ее автор — Матерухин Андрей Викторович заслуживает присуждения ученой степени доктора технических наук по специальности 25.00.35 — «Геоинформатика».

Дата проведения заседания Лаборатории геоинформатики и геомагнитных исследований, на котором была обсуждена диссертационная работа – 3 октября 2018 (протокол б/н от 03.10.2018). Дата составления отзыва — 28 ноября 2018 года.

Агаян Сергей Мартикович, Главный научный сотрудник Лаборатории геоинформатики и геомагнитных исследований, доктор физико-математических наук. Шифр научной специальности, по которой защищена диссертация — 25.00.35 — «Геоинформатика».

Геофизический центр РАН г. Москва, 119296 ул. Молодежная, д.3, ГЦ РАН. Тел.: +7 (495) 930-05-46, e-mail: gcras@gcras.ru