

Фёдоров Роман Константинович

**Сервис-ориентированная информационно-аналитическая среда
композиции сервисов обработки пространственных данных**

Специальность 2.3.5 – Математическое и программное обеспечение
вычислительных систем, комплексов и компьютерных сетей

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
доктора технических наук

Иркутск – 2024

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук (ИДСТУ СО РАН).

- Научный консультант:** **Бычков Игорь Вячеславович** – академик РАН, доктор технических наук, профессор, Институт динамики систем и теории управления имени В.М. Матросова СО РАН, директор
- Официальные оппоненты:** **Марченко Михаил Александрович** – доктор физико-математических наук, профессор, Институт вычислительной математики и математической геофизики СО РАН, директор
Шалфеева Елена Арефьевна – доктор технических наук, Институт автоматизации и процессов управления ДВО РАН, ведущий научный сотрудник
Ничепорчук Валерий Васильевич – доктор технических наук, Институт вычислительного моделирования СО РАН, старший научный сотрудник
- Ведущая организация:** **Институт автоматизации и электрометрии СО РАН (г. Новосибирск)**

Защита состоится «26» декабря 2024 г. в 14 часов на заседании диссертационного совета 24.1.060.01 при Федеральном государственном бюджетном учреждении науки Институте динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук (ИДСТУ СО РАН) по адресу: 664033, г. Иркутск, ул. Лермонтова, 134.

С диссертацией можно ознакомиться в библиотеке и на официальном сайте www.idstu.irk.ru ИДСТУ СО РАН.

Автореферат разослан «__» _____ 202 г.

Ученый секретарь
диссертационного совета, к.ф.-м.н.

Т.В. Груздева

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Для современного этапа решения крупномасштабных фундаментальных и прикладных задач актуально повышение эффективности и надежности процессов обработки и передачи данных в вычислительных системах, комплексах и компьютерных сетях путём внедрения сервис-ориентированной архитектуры (СОА, англ. service-oriented architecture, SOA), на основе которой реализуют интеллектуальные технологии и машинное обучение, а также системы обработки больших объемов пространственных данных (ПД).

Сервис-ориентированная архитектура является развитием подходов, применявшихся при создании пакетов прикладных программ (ППП), а затем параллельных и распределенных вычислительных систем, и базируется на теоретических и практических результатах исследований ведущих российских и зарубежных ученых, в том числе С.М. Абрамова, А.И. Аветисяна, А.П. Афанасьева, В.Б. Бетелина, И.В. Бычкова, А.В. Бухановского, Вл.В. Воеводина, В.П. Гергеля, Б.М. Глинского, В.П. Иванникова, И.А. Каляева, В.Н. Коваленко, Д.А. Корягина, В.В. Коренькова, В.Д. Корнеева, И.И. Левина, А.И. Легалова, Л.В. Массель, Ю.Е. Малашенко, В.Э. Малышкина, А.В. Манциводы, Г.А. Опарина, Г.И. Радченко, Г.М. Ружникова, С.И. Смагина, И.А. Соколова, Л.Б. Соколинского, О.В. Сухорослова, В.В. Топоркова, А.Г. Феоктистова, В.Г. Хорошевского, Б.Н. Четверушкина, М.В. Якобовского, A. Ballatore, G. Baryannis, J. Bih, D. Churches, F. Curbera, E. Deelman, L. Di, D. Edmond, M. Farnaghi, D. Fellows, I. Foster, A. Friis-Christensen, G. Gombas, R. Haines, A. Harrison, M. Hinz, A. Hofstede, W. Huang, J. Pereira, G. Juve, Y.-K. Kwok, H. Li, Yu. Liyang, R. Lucchi, M. Lutz, D. Martin, L. Miao, T. Nixon, N. Ostländer, S. Scheider, P. Schut, D. Silva, Z. Sun, H. Topcuoglu, D. Ulutaş Karakol, K. Vahi, K. Wolstencroft, W. Yang, P. Yue, P. Zhao, H. Zhi-Wei, Q. Zhu, C. Zhuang и других известных специалистов.

В 70-х годах прошлого столетия ППП включали различные реализации численного решения вычислительных задач и средств системного обеспечения (программных и языковых). ППП позволяли исследователю получить целый набор различных взаимосвязанных методов, ориентированных на определенную предметную область. Взаимодействие программ происходило только в рамках пакета через файловую систему с использованием обменных форматов, либо через заданные программные интерфейсы. За рубежом аналогом ППП является система управления научными рабочими процессами (англ., Workflow Management System – WMS). При этом схема или план решения задачи в ППП коррелирует с вышеупомянутым понятием научного рабочего процесса.

В связи с активным использованием Интернета развивались методы интеграции и взаимодействия пакетов программ через сети передачи данных (СПД), что привело к созданию нового направления «распределенных пакетов программ». В том числе были реализованы методы планирования и выполнения распределенных пакетов программ на гетерогенных вычислительных ресурсах и на суперкомпьютерах. Но взаимодействие программ через СПД производилось, в основном, в рамках пакета.

СОА появилась в конце 1980-х и берёт своё начало в идеях, изложенных в CORBA, DCOM, DCE и т. д. Реализация методов в виде сервисов упрощает использование программного обеспечения (ПО), которое сводится к вызову сервиса через Интернет, используя его программный интерфейс. Программные системы,

реализованные с помощью данного подхода, обладают следующими преимуществами: сервисы кроссплатформенны, тестируемы, доступны по сети. В рамках СОА нет необходимости устанавливать, конфигурировать, обновлять программное обеспечение. Сервисы имеют низкий порог вхождения для их использования (пользователю нет необходимости изучать подробности реализации метода, достаточно изучить интерфейс).

В настоящее время тенденция развития WMS базируется на разработке сервис-ориентированных научных процессов. Кроме того, в связи с высокой актуальностью задач, решаемых, например, в рамках экологического мониторинга, активно развиваются специализированные системы управления сервис-ориентированными научными процессами в области геоинформатики. Известными примерами таких систем являются BPEL Designer Project (<https://projects.eclipse.org/projects/soa.bpel>) и GeoJModelBuilder (<https://github.com/geoprocessing/GeoJModelBuilder>).

В современном мире активно растет количество сервисов, реализующих предоставление, обработку и публикацию данных. Например, это сервисы предоставления данных дистанционного зондирования земли, обработки пространственных данных, расшифровки генома и т. д. Созданные сервисы значительно упрощают решение многих задач. В области обработки ПД определены и активно используются стандарты Open Geospatial Consortium (OGC). При этом возникает ряд сложных научно-технических задач нахождения сервисов, построения их связей между собой, проверки корректности совместного выполнения этих сервисов и др.

Объединение сервисов, т. е. создание их композиции, позволяет решать большое количество задач. СОА значительно упрощает и ускоряет интеграцию программного обеспечения, созданного разработчиками из разных предметных областей, за счет упрощения и стандартизации интерфейсов. Композиции сервисов обеспечивают повышение уровня автоматизации решения задач, начиная от ввода данных и заканчивая публикацией результатов.

В то же время создание композиций сервисов является нетривиальной задачей. Наличие большого количества сервисов, с одной стороны, увеличивает возможности исследователей, а с другой стороны, значительно усложняет поиск нужных сервисов для решения конкретной задачи. Составление найденных сервисов в композиции в некоторых случаях может оказаться комбинаторно сложной задачей, при этом возможна генерация достаточно большого числа альтернативных композиций сервисов. Все эти альтернативы необходимо будет оценить, выбрать наиболее релевантные для решаемой задачи, а также их проверить на возможность взаимодействия. Часто композиция из двух потенциально возможных сервисов не реализуема из-за того, что структуры данных и форматы входных и выходных данных отличаются.

Композиции сервисов позволяют объединить результаты работы исследователей, но создание таких композиций в силу сложного процесса требует высокой квалификации. Поэтому актуальной научной проблемой является разработка новых моделей, алгоритмов, методов и технологии для создания сервис-ориентированной информационно-аналитической среды (СОИАС) с целью повышения эффективности процессов подготовки и проведения научных экспериментов по решению задач в области геоинформатики за счет автоматизации построения и применения композиций сервисов. Здесь под эффективностью понимается сокращение накладных расходов (стоимостных, временных и др.

трудозатрат) на выполнение экспериментов, а также обеспечение новых функциональных возможностей, необходимость которых обуславливается наличием вышеупомянутых сложных научно-технических задач.

Цель исследования заключается в повышении эффективности процессов подготовки и проведения научных экспериментов на основе сервис-ориентированной парадигмы за счет автоматизации построения и применения композиций сервисов, реализующих методы анализа и обработки ПД.

Основные задачи диссертационного исследования

1. Провести анализ инструментальных средств, технологий и существующих информационно-аналитических сред с использованием сервисов обработки ПД и композиций сервисов.
2. Исследовать и разработать модель СОИАС, обеспечивающую создание и обмен композициями сервисов между пользователями.
3. Разработать метод создания композиций сервисов обработки ПД.
4. Разработать методы выполнения композиции сервисов обработки ПД.
5. Разработать основные компоненты среды, реализующие модель СОИАС.
6. Провести апробацию СОИАС на задаче поддержки междисциплинарных научных исследований Байкальской территории.

Объектом исследования является распределенная гетерогенная информационно-вычислительная среда, функционирующая на основе СОА.

Предметом исследования являются модели, методы, алгоритмы и программное обеспечение автоматизации построения и применения композиций сервисов в распределенной гетерогенной информационно-вычислительной среде.

Методы исследования. В работе использовались методы информационного моделирования, теории графов, системного и объектно-ориентированного программирования, проектирования баз данных, построения распределенных комплексов проблемно-ориентированных программ, веб-технологий, планирования выполнения композиций сервисов в статических и динамических средах.

Научную новизну диссертации представляют следующие результаты исследования, выносимые на защиту и расширяющие существующий базис теории и практики сервис-ориентированных вычислений:

- 1) создана модель сервис-ориентированной информационно-аналитической среды обработки пространственных данных междисциплинарных исследований, которая в сравнении с подобными моделями обеспечивает оценку композиций сервисов на основе многопользовательской статистики их применения;
- 2) разработан метод создания композиций сервисов, базирующийся на применении предложенной модели, который в отличие от существующих методов проводит комплексный анализ метаданных, онтологий, экспертных знаний и статистики применения сервисов, что позволяет находить композиции сервисов на основе комбинации данных;
- 3) разработан оригинальный программный инструмент создания сервисов ввода и публикации реляционных данных, обеспечивающий предоставление метаданных, пользовательский и программный интерфейс редактирования данных, поддержку передачи данных WPS сервисам. Создание сервисов данных впервые производится на основе иерархической модели данных с возможностью задания асинхронного вычисления значений атрибутов с помощью сервисов. Создаваемые сервисы можно сразу включать во множество композиций;

- 4) разработан оригинальный программный компонент выполнения композиций сервисов, заданных на процедурном языке, с обработкой промежуточных данных с помощью средств языка и его библиотек, для которого в отличие от других подходов обеспечивается формирование DAG с помощью процедурного языка и одновременно его планирование и выполнение с учетом добавляемых в процессе выполнения новых заданий в гетерогенной динамической вычислительной среде;
- 5) разработан комплекс программных компонентов, реализующий модель сервис-ориентированной информационно-аналитической среды, который обеспечил создание композиций сервисов и их обмен между пользователями.

Соответствие диссертации паспорту научной специальности. Тема и основные результаты диссертации соответствуют следующим областям исследований паспорта специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»:

- модели, методы, архитектуры, алгоритмы, языки и программные инструменты организации взаимодействия программ и программных систем;
- модели, методы, алгоритмы, облачные технологии и программная инфраструктура организации глобально распределенной обработки данных.

Теоретическая значимость работы заключается в развитии методов и средств автоматизации построения и применения композиций сервисов на основе комплексного анализа метаданных, онтологий, экспертных знаний и статистики применения сервисов. Основные результаты диссертационного исследования использованы при выполнении государственных заданий и научных исследований:

- грантов Министерства науки и высшего образования РФ на выполнение крупного научного проекта по приоритетным направлениям научно-технического развития «Фундаментальные основы, методы и технологии цифрового мониторинга и прогнозирования экологической обстановки Байкальской природной территории» (№ 075-15-2020-787), «Фундаментальные исследования Байкальской природной территории на основе системы взаимосвязанных базовых методов, моделей, нейронных сетей и цифровой платформы экологического мониторинга окружающей среды» (№ 075-15-2024-533);
- проектов Программы фундаментальных исследований Президиума РАН: программа II.1, тема «Разработка и экспериментальное исследование эффективности методов оценки антропогенного воздействия на окружающую среду» ФНМ-51, 49, 27 (2012 г., 2013-2015 гг., 2018-2019 гг.);
- проектов РФФИ 18-07-00758_a (2018-2020 гг.), 17-57-44006 Монг_a (2018-2020 гг.), 17-47-380007_p-a (2017-2019 гг.), 17-29-05089_a (2017-2019 гг.), 16-57-44034 Монг_a (2016-2017 гг.), 16-07-00554-a (2016-2018 гг.), 16-07-00411_a (2016-2018 гг.), 15-47-04348_a (2015-2017 гг.), 14-07-00166_a (2014-2016 гг.), 14-47-04125 p_сибирь_a (2014-2016 гг.), 13-07-12080 офи_m (2014-2016 гг.), 13-05-41105 РГО_a (2013-2014 гг.);
- проекта Программы фундаментальных исследований Отделения нанотехнологий и информационных технологий РАН № 4.1 (2012-2014 гг.);
- проектов междисциплинарных интеграционных программ СО РАН № 17, 131 (2012-2014 гг.);
- проектов междисциплинарных интеграционных программ СО РАН и ДВО РАН № 73, 74 (2012-2014 гг.);
- базовых проектов Программы фундаментальных исследований СО РАН:

1. № IV.38.1.2. «Методы и технологии облачной сервис-ориентированной цифровой платформы сбора, хранения и обработки больших объёмов разноформатных междисциплинарных данных и знаний, основанные на применении искусственного интеллекта, модельно-управляемого подхода и машинного обучения» (2020-2021 гг.);

2. № IV.38.1.2 «Методы и технологии создания распределенной сервис-ориентированной среды сбора, хранения, обработки больших объёмов разноформатных междисциплинарных научных данных и знаний, основанные на конструктивных средствах спецификации, порождающем программировании и интеллектуализации» (2017-2019 гг.);

3. № IV.38.2.3 «Новые методы, технологии и сервисы обработки пространственных и тематических данных, основанные на декларативных спецификациях и знаниях» (2013-2015 гг.);

4. № IV.31.2.4 «Методы и технологии разработки программного обеспечения для анализа, обработки и хранения разноформатных междисциплинарных данных и знаний, основанные на применении декларативных спецификаций форматов представления информации и программных систем» (2010-2012 гг.).

Практическая значимость. Предложенные в рамках диссертационной работы методы, модели, алгоритмы и программное обеспечение позволяют снизить трудозатраты и сократить сроки разработки программного обеспечения за счет автоматизации построения композиций сервисов обработки междисциплинарных пространственных данных. Комплекс программных компонентов, реализующий модель сервис-ориентированной информационно-аналитической среды, активно используется на практике. В ходе выполнения различных проектов созданы более 200 сервисов предоставления данных, более 40 сервисов обработки данных и 250 сервисов публикации данных. Развернуты 6 различных геопорталов, ориентированных на различные предметные области и коллективы: ИДСТУ СО РАН; «Информационно-аналитическая система по фиторазнообразию Байкальской Сибири»; «Атлас ИГ СО РАН»; ИЗК СО РАН; «Очаги распространения иксодовых клещей»; «Информационная система (ИС) «L.». В рамках перечисленных геопорталов сформированы композиции сервисов, объединяющие сервисы данных, сервисы обработки и публикации, созданные разными коллективами. Практическая значимость результатов подтверждена полученными актами внедрения комплекса программных компонент ИГ СО РАН, СИФИБР СО РАН, ПАБСИ КНЦ РАН, ИППЭС КНЦ РАН, Самарский университет им. Королева, НЦ ПЗСРЧ. Автором в составе коллектива получено 8 свидетельств о регистрации программ для электронных вычислительных машин (ЭВМ).

Достоверность и обоснованность полученных в диссертации результатов подтверждается корректным применением классических методов исследования, анализом адекватности разработанных моделей и алгоритмов, решением прикладных и тестовых задач, индексацией полученных результатов в РИНЦ, Web of Science, Scopus, активной эксплуатацией СОАИС большим числом пользователей.

Апробация результатов исследования. Основные результаты диссертационного исследования докладывались на следующих научных мероприятиях: Международная конференция «3th Russia and Pacific Conference on Computer Technology and Applications» (Владивосток, 2017 г.); Международная научно-практическая конференция «Использование современных информационных технологий в ботанических исследованиях» (Апатиты, 2017 г.); Всероссийская

конференция «Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов» (Новосибирск, 2017 г.; Бердск, 2019 г.); XVIII Всероссийская конференция молодых ученых по математическому моделированию и информационным технологиям (Барнаул, 2013 г.; Новосибирск, 2017 г.); Национальный суперкомпьютерный форум (НСКФ, Переславль-Залесский, 2016 г.); Российско-монгольская конференция молодых ученых по математическому моделированию, вычислительно-информационным технологиям и управлению (Иркутск-Ханх (Монголия), 2013, 2015, 2016 гг.); III Всероссийская конференция «Математическое моделирование и вычислительно-информационные технологии в междисциплинарных научных исследованиях» (Иркутск, 2013 г.); Всероссийская конференция «Ляпуновские чтения» (Иркутск, 2014-2017 гг.); Международная конференция «Системный анализ и информационные технологии» (САИТ, Светлогорск, 2015 г.; Иркутск, 2019 г.); Scientific-practical Workshop Information Technologies: Algorithms, Models, Systems (ITAMS, Иркутск, 2018-2021 гг.); Всероссийская конференция с международным участием «Информационные и математические технологии в науке и управлении» (ИМТ, Иркутск, 2009, 2011, 2013, 2014, 2016, 2019-2021 гг.); Международная конференция «Современные проблемы дистанционного зондирования Земли из космоса» (Москва, 2019 г.); Всероссийская конференция «Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов» (Бердск, 2019 г.); Международная конференция «Математические и информационные технологии» (MIT-2011, Врнячка Баня, Сербия, 2011 г.); «Международная географическая конференция» (Иркутск, п. Листвянка, 2018 г.); VIII Всероссийская конференция «Безопасность и мониторинг природных и техногенных систем» (Красноярск, 2023 г.); Всероссийская конференция «Информационные технологии в управлении» (Санкт-Петербург, 2020 г.); 16-я Международная конференция «Системный анализ, управление и навигация» (Евпатория, 2011 г.); Всероссийская конференция «Распределенные информационно-вычислительные ресурсы. Наука – цифровой экономике» (DICR, Новосибирск, 2017 г.); Международный научно-технический конгресс «Интеллектуальные системы и информационные технологии» (Дивноморское, 2022 г.), а также семинарах ИДСТУ СО РАН.

Личный вклад автора. Все выносимые на защиту научные положения получены соискателем лично. Из совместных исследований в диссертацию включены только те результаты, которые принадлежат непосредственно автору. Модель сервис-ориентированной информационно-аналитической среды, метод создания композиций сервисов [4, 8, 18, 28, 41, 42], программный инструмент создания сервисов ввода и публикации реляционных данных, комплекс программных компонентов, реализующий модель сервис-ориентированной информационно-аналитической среды [1, 2, 3, 4, 6, 10, 11, 12, 15], программный компонент выполнения композиций сервисов, заданных на процедурном языке [5, 7], разработаны соискателем лично на всех этапах от постановки задач до разработки правил, моделей, программных компонентов. Программная реализация модуля интерпретации сценариев выполнена совместно с Шумиловым А.С. Часть компонентов и сервисов СОИАС, которые непосредственно не касаются создания композиций сервисов, но необходимы для полноценной работы среды, разработаны совместно с коллегами: Авраменко Ю.В., Хмельновым А.Е., Поповой А.К., Ветровым А.А., Парамоновым В.В. Совместно с научным консультантом Бычковым И.В. и Ружниковым Г.М. выделены проблемы, сформулированы направления, и

обобщены результаты исследований, сформулированы научные выводы и практические рекомендации.

Публикации. Результаты диссертационного исследования отражены в 45 научных работах. Основные публикации представлены в 2 монографиях, 9 статьях в российских журналах, рекомендованных ВАК для опубликования научных результатов диссертации, а также в 32 статьях, проиндексированных в международных базах цитирования Web of Science и Scopus. Автором в составе коллектива получено 8 свидетельств о регистрации программ для электронных вычислительных машин (ЭВМ).

Структура диссертации. Диссертация состоит из введения, пяти глав, заключения, библиографии из 201 наименования, списка принятых сокращений и 12 приложений. Объем основного текста работы – 195 страниц, включая 5 таблиц и 81 рисунок. Общий объем диссертации 271 страница.

Автор выражает глубокую благодарность научному консультанту академику И.В. Бычкову за чуткое неустанное сопровождение в работе, помощь в преодолении трудностей и постановку нестандартных, увлекательных и интересных задач, Ружникову Г.М. за полную поддержку всех идей и начинаний, своим коллегам – соавторам за помощь в разработке компонентов, полезные советы и конструктивную критику, а также сотрудникам ИДСТУ СО РАН за обсуждение и полезные замечания при выполнении диссертационной работы.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертационной работы, сформулированы цели и задачи исследования, отражена научная новизна и практическая значимость полученных результатов, приведены сведения об апробации диссертации, использовании результатов и опубликованных работах, дана общая характеристика работы.

В первой главе выполнен анализ существующих направлений развития технологий распределенных вычислений, методов и подходов к созданию композиций сервисов и их выполнения, который показал, что в настоящее время активно развивается СОА, появляется большое количество сервисов. В связи с этим, возникает необходимость решения следующих задач: поиск сервисов и реализация их взаимодействия, т. е. создание композиций сервисов. Композиции сервисов позволяют объединить методы и данные, созданные разными специалистами, и решать сложные задачи. Поиск и создание новых композиций сервисов является комбинаторно сложной задачей, в которой необходимо проанализировать большое количество сервисов и предоставить наиболее релевантные их комбинации. Существующие методы и подходы на основе анализа метаданных и онтологий не позволяют значительно упростить поиск из-за частого отсутствия метаданных и трудоемкости разработки онтологий. Создание композиций сервисов ограничено сложностью взаимодействия сервисов по данным из-за различий в программном интерфейсе доступа к данным, в структуре данных, в используемых справочниках и т. д. Поэтому является актуальной разработка сервис-ориентированной информационно-аналитической среды (СОИАС) обработки междисциплинарных пространственных данных, повышающей эффективность научных исследований за счет создания композиций сервисов и организации обмена ими.

Во второй главе представлена новая вычислительная модель композиции сервисов. На основе требований (**раздел 2.1**) к организации междисциплинарных

научных исследований в сервис-ориентированной среде разработана вычислительная модель композиции сервисов (раздел 2.2), определяемая структурой

$CM = \langle Z, S, O, A, T, CR, F \rangle$, где

Z – множество параметров. В стандарте WPS OGC обработки пространственно-временных данных параметры связаны с определенными типами данных.

S – множество сервисов. В работе рассматриваются только сервисы без хранения состояния (stateless), т. е. результат выполнения сервисов не должен зависеть от предыдущих вызовов. Этого достаточно для решения широкого круга задач. Взаимодействие сервисов с хранением состояния (stateful) в данной работе не рассматривается. Сервис $s \in S$ – это упорядоченная тройка $s = \langle name, Z^{Inp}, Z^{Out} \rangle$, где $name$ – имя сервиса, $Z^{Inp} \subset Z$ – множество входных параметров, $Z^{Out} \subset Z$ – множество выходных параметров сервиса. Среди выходных параметров сервиса имеется параметр успешности выполнения сервиса. Далее будем обозначать $s.Z^{Inp}$ и $s.Z^{Out}$ параметры, принадлежащие определенному сервису. Сервисы разделяются на сервисы получения данных $S^d \subset S$, вычислительные сервисы $S^c \subset S$ и сервисы публикации данных $S^p \subset S$.

$O \subset S \times S$ – множество отношений между сервисами, определяет возможность передачи данных между сервисами.

A – множество пользователей, которые взаимодействуют с СОИАС через применение сервисов S . Набор применяемых ими сервисов может варьироваться.

T – множество заданий выполнения сервисов из S . Каждое задание $t \in T$ определяется упорядоченной четверкой $t = \langle a, s, V_I, V_O \rangle$, т. е. вызов сервиса $s \in S$ пользователем $a \in A$ с множеством значений V_I входных параметров $s.I$ и результаты V_O выходных параметров $s.O$.

CR – множество вычислительных узлов, на которых выполняются сервисы. Каждый сервис может быть активирован на заранее определенных узлах. Автоматическое выделение новых узлов для сервисов в работе не рассматривается.

F – множество функций ранжирования сервисов.

Отметим, что композиция сервисов является также сервисом с заданным именем, входными и выходными параметрами. Общеизвестным стандартом определения композиций является использование направленного ациклического графа (Directed Acyclic Graph (DAG)) $s_i = DAG_i = \langle T_i, E_i \rangle$, $s_i \in S$, в котором вершинами являются задания $t \in T_i$, а ребра $e = \langle v_k, v_l \rangle$, $e \in E_i$ показывают передачу данных от выходного параметра одного задания к входному параметру другого задания, где $v_k \in V_I^i$, $v_l \in V_O^j$, $t_i = \langle a_i, s_i, V_I^i, V_O^i \rangle$, $t_j = \langle a_j, s_j, V_I^j, V_O^j \rangle$. Поскольку возможна передача данных одновременно по нескольким параметрам, то ребер между двумя вершинами (заданиями) также может быть несколько. В литературе чаще всего передача данных между заданиями обычно обозначается одним ребром. Для создания готовой к выполнению композиции необходимо определить все передаваемые данные между параметрами.

Определим типы данных, используемые для обработки пространственных данных. Частично примем обозначения, принятые в реляционной алгебре:

$D = \{D_1, \dots, D_l\}$ – множество скалярных типов, определенных в стандарте WPS;

$R = \{R_1, \dots, R_m\}$ – множество схем отношений;

$d = \{d_1, \dots, d_k\}$ – множество таблиц (отношений). В качестве параметров могут использоваться реляционные таблицы $d_i \in Z$.

Введем множество типов параметров $TZ = D \cup R$. Для каждого параметра определен тип $tz_j = type(z_i)$, $tz_j \in TZ$.

Для формирования композиций сервисов требуется информация о возможности передачи данных выходного параметра сервиса для использования во входном параметре другого сервиса. Однако соответствие типов параметров не является достаточным, например, тип «число» может иметь смысл – высота над уровнем моря, а в другом случае – температура. Поэтому соответствие типов является необходимым условием, но не достаточным. Получение информации о возможности передачи данных между параметрами может производиться на основе анализа метаданных, онтологий, экспертных знаний, статистики применения сервисов и т. д. Поэтому вводим понятие метки, которое позволяет абстрагироваться от способа получения информации о возможной передаче данных между параметрами. Метка используется для идентификации возможности передачи данных, т. е. если выходной параметр и входной параметр имеют одну и ту же метку, то между ними возможна передача данных. Каждому параметру может соответствовать ноль и более меток. Метками могут являться онтологические концепты, заданные с помощью URI, или некоторые сгенерированные уникальные значения.

$m \in M$ – множество меток. Для каждого параметра z_j определено подмножество меток M_j . Проверка возможности передачи данных между сервисами в рамках модели выполняется на основе соответствия типов параметров и меток. Для двух сервисов s^i и s^j возможна передача данных от параметра $z^i \in s^i.Z^{Out}$ к параметру $z^j \in s^j.Z^{Inp}$, если типы параметров совпадают $type(z_i) = type(z_j)$ и существует метка $m \in M$, которая имеется у обоих параметров ($m \in M_i$ и $m \in M_j$). Выполнение этого условия не гарантирует корректное выполнение сервиса s^j . Это является необходимым условием. Оно позволяет значительно сузить количество возможных композиций сервисов.

В разделе 2.3 предложен метод создания композиций. Процесс создания композиции сервисов состоит из трех основных этапов. На первом этапе производится построение модели предметной области, которое состоит из 5 шагов, выполняемых независимо друг от друга.

1.1. Создание и регистрация сервисов – предполагается разработка вычислительных сервисов, сервисов данных и сервисов публикации данных. Вычислительные сервисы разрабатываются в соответствии со стандартом WPS на вычислительных ресурсах среды или на любых других серверах, к которым имеется сетевой доступ. Для создания сервисов данных используется компонент «Фабрика сервисов ввода и редактирования реляционных данных», который поддерживает все основные этапы работы с табличными данными. «Фабрика сервисов отображения графиков» и «Фабрика сервисов отображения пространственных данных» предназначены для создания сервисов публикации данных. Все создаваемые сервисы регистрируются в каталоге, обладают программным интерфейсом и готовы стать частью композиции сервисов.

1.2. Регистрация сторонних сервисов, т.е. осуществляется подключение WPS сервисов, созданных на сторонних вычислительных ресурсах. В каталоге сервисов указывается необходимая для запуска сервиса информация и метаданные.

1.3. Описание онтологической модели сервисов – предназначено для определения возможности передачи данных между параметрами сервисов на основе

онтологий. В рамках описания онтологической модели сервисов предполагается определение концептов для параметров сервисов. Каждому параметру может соответствовать ноль и более меток. Метками могут являться онтологические концепты, заданные с помощью URI. Использование одного и того же концепта для выходного параметра сервиса и для входного параметра другого сервиса предполагает возможность передачи данных между ними и присвоение одной метки для них. Одним из основных отношений, которые применяются для формирования меток, является отношение частное – общее. В онтологии для задания этого отношения между концептами используется элемент `subClassOf`. Например, если сервис имеет входной параметр z_j с заданным некоторым концептом k , то в качестве входных данных могут использоваться данные сервисов, которые имеют концепты, являющиеся частными по отношению к k .

1.4. Формирование экспертных знаний – используется для определения возможности передачи данных пользователем. Разработка онтологий – достаточно сложный процесс. Часто с помощью онтологий не все сущности и отношения предметной области формализованы. Поэтому в качестве меток параметров сервисов могут использоваться некоторые сгенерированные уникальные значения, назначаемые пользователем для конкретных сервисов. Например, сервис построения карты плотности может применяться для всех точечных данных. Соответственно, выходным параметрам, которые предоставляют векторный точечный слой, можно назначить ту же самую метку, что и у входного параметра сервиса построения карты плотности.

1.5. Сбор статистики выполнения сервисов, т.е. производится сбор информации о произведенных вызовах сервисов, которые включают название сервиса и его адрес, значения входных и выходных параметров, время выполнения сервисов, успешность выполнения, ошибки выполнения и т. д. Данные сохраняются в таблице, где значения параметров хранятся в формате JSON. Далее данные этой таблицы будем обозначать как *Log* – множество заданий (успешных вызовов сервисов), совершенных пользователями. Причиной неудачного вызова сервиса могут быть некорректные значения параметров, поэтому данные фильтруются и оставляются только успешные вызовы. Производится анализ статистических данных, определение передачи данных между параметрами сервисов и присвоение меток. Определение связи между двумя вызовами сервисов может производиться на основе анализа значений параметров. Множество параметров можно разделить на входные и выходные. Предполагается, если значение выходного параметра совпадает со значением входного параметра, то между вызовами сервисов может быть передача данных. Значения параметров могут быть строковыми или числовыми, а также идентификаторами ресурсов URI, обычно это URL ссылки на файлы, сервисы предоставления данных и т. д. Каждый URI, используемый для получения данных WPS сервисом, однозначно идентифицирует передаваемые данные. Соответственно, по URI параметрам можно однозначно определить, что данные передаются от одного вызова сервиса другому. На текущий момент строковые и числовые параметры не рассматриваются. Разработан алгоритм проверки передачи данных между двумя вызовами сервисов. Данный алгоритм кроме наличия связи между вызовами сервисов определяет, по каким параметрам они связаны.

После построения модели предметной области на этапе 2 применяются два альтернативных способа создания композиций:

- автоматическое формирование композиций сервисов позволяет создать композицию сервисов на основе анализа графа связности заданий;
- разработка композиций сервисов пользователем с помощью языка JavaScript. Этот способ требуется при необходимости обработки промежуточных данных с помощью средств языка, или при заранее неизвестном количестве заданий в композиции $DA G_i$, которые формируются в зависимости от данных.

На этапе 3 производится регистрация полученных композиций сервисов в каталоге для последующего применения пользователями.

В разделе 2.4. описан разработанный алгоритм формирования сети связанных сервисов, который может быть использован для упрощения разработки композиции сервисов. Сеть связанных сервисов определяет возможные связи сервисов по передаваемым данным, т. е. дуги этой сети задают возможность элементарной композиции – цепочки выполнения из двух сервисов, когда результаты работы одного сервиса могут быть применены в качестве входных данных другого сервиса. Будем обозначать сеть сервисов с помощью графа $SG = \langle S, L \rangle$, в котором вершинами являются сервисы $s \in S$, а ребра $l = \langle s_k, s_l \rangle$, $l \in L$ показывают возможность передачи данных от одного сервиса к другому.

Для формирования сети связанных сервисов используют метаинформацию и онтологии. В частности, осуществляется построение семантической сети на основе WSDL. Следует отметить, что автоматическое построение сети связанных сервисов является сложной задачей по следующим причинам:

- сервисы имеют недостаточно метаинформации, либо она вообще отсутствует или рассчитана на человека;
- параметры некоторых сервисов могут быть связаны со многими другими сервисами. Например, сервисы алгебры растров могут использоваться во многих композициях сервисов для анализа пространственных данных. Связывание трех последовательных сервисов, где промежуточным сервисом являются сервисы алгебры изображений, требует проработанной онтологии предметной области. Ее построение – достаточно сложный процесс, в котором необходимо рассмотреть все возможные сущности и связи;
- требуется согласованность метаинформации сервисов и онтологии предметной области.

Для получения недостающей информации при построении сети связанных сервисов разработан алгоритм, основанный на анализе статистических данных о применении сервисов пользователями. Построение сети связанных сервисов сводится к задаче определения передачи данных между параметрами сервисов (causal link) $s.Z^{Inp}$ и $s.Z^{Out}$.

Разработан следующий алгоритм построения сети связанных сервисов.

Input: Log – множество заданий

Output: SG – связи между сервисами

```
1  for each ti in Log:
2      for each tj in Log:
3          for each vm in ti.VO:
4              for each vn in tj.VI:
5                  If (typeof vm = file and typeof vn = file and vm = vn) then
6                      add (ti, tj) into SG
7                  end if
8              end for
9          end for
10     end for
11 end for
```

Данный алгоритм производит поиск сочетания вызовов сервисов, в которых файл, являющийся результатом работы одного сервиса, передается в качестве параметра другому сервису. Предложенный метод построения сети связанных сервисов апробирован на небольшом фрагменте данных о применении сервисов пользователями. Источники данных представлены в виде сервисов, у которых присутствуют только исходящие дуги. На рисунке 1 представлена часть полученной сети. Несвязанные сервисы (одиночные вершины) не отображены. Необходимо отметить, что при таком подходе семантика всех параметров сервисов не определяется. Среди положительных моментов укажем, что вместе с сетью связанных сервисов получаем связанность параметров по данным, возможные значения параметров, частотные характеристики использования пользователями сервисов и их композиций, время выполнения сервисов и т. д.

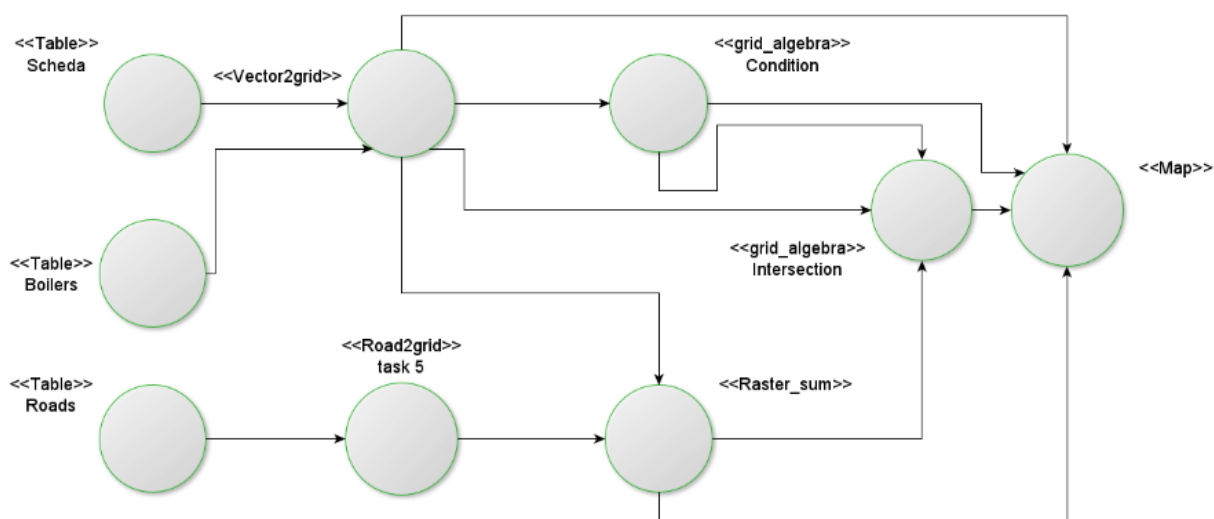


Рисунок 1 – Полученная сеть сервисов

Представленный метод позволяет сформировать сеть связанных сервисов на основе статистических данных применения сервисов. Сеть связанных сервисов значительно упрощает разработку композиции сервисов, предоставляя информацию о возможных цепочках из двух сервисов. Преимуществом данного алгоритма является то, что связь между двумя сервисами подтверждается работоспособным примером, демонстрирующим передачу данных от одного сервиса другому. Кроме того, вместе с сетью связанных сервисов получаем возможные значения параметров,

частотные характеристики использования пользователями сервисов и их композиций, время выполнения сервисов и т. д.

В разделе 2.5 приводится алгоритм автоматического формирования композиций сервисов. В СОА решение задачи может требовать многократного вызова сервисов, при этом пользователю приходится придерживаться определенной последовательности вызова сервисов, вводить одни и те же значения параметров для каждого сервиса и т. д. При многовариантных экспериментах это становится достаточно трудоемким процессом и требует много времени. Автоматизация процесса позволит значительно сократить время проведения эксперимента и сформирует готовый инструмент анализа и обработки данных для других пользователей. Автоматическое создание композиций производится на основе статистических данных, экспертных знаний и онтологий. Каждая композиция задается графом DAG , с заданными параметрами заданий и связями по данным. Создаваемые композиции сохраняются в каталоге сервисов, где пользователь может найти нужную композицию в соответствии со своим запросом. Автоматическое создание композиций состоит из следующих шагов:

Шаг 1. Создание графа заданий и их связей (граф DAG_{gen}).

Шаг 2. Декомпозиция графа заданий.

Шаг 3. Редукция множества композиций сервисов.

Создание графа заданий и их связей производится на основе анализа таблицы заданий. Разработан следующий алгоритм построения графа DAG_{gen} .

Input: Log – множество заданий

Output: DAG_{gen} – граф заданий и их связей

```
1  for each  $t_i$  in  $Log$ :
2      for each  $t_j$  in  $Log$ :
3          If islinked( $t_i, t_j$ ) then
4              add ( $t_i, t_j$ ) into DAGgen
5              newlabel() -> label
6              addlabel(label,  $t_i$ )
7              addlabel(label,  $t_j$ )
8          end for
9  end for
```

Функция **islinked** проверяет возможность передачи данных между двумя заданиями. После выполнения алгоритма из списка выполненных сервисов получим общий граф DAG_{gen} , где все вызовы пользователей сервисов t_i связаны по данным. Дополнительно получаем возможные значения параметров сервисов, являются ли они результатами работы других сервисов или их ввел пользователь. Функция **islinked** определяет возможность передачи данных на основе меток, которые формируются с помощью метаданных сервисов, экспертных знаний и онтологий.

Далее производится декомпозиция графа заданий. Граф DAG_{gen} содержит задания, применяемые для решения множества разных задач. Следовательно, его необходимо декомпонировать и выделить отдельные подграфы, соответствующие решаемым задачам. Пользователь может выполнить некоторую последовательность сервисов, задающую последовательность заданий. Если задания связаны по данным,

то они образуют связанный подграф D_i (компонента связности графа) в DAG_{gen} . Выполнение композиции сервисов всегда приводит к созданию связанного подграфа D_i . Если задания, сформированные композицией сервисов, не связаны, тогда такую композицию сервисов можно разделить. Для получения подграфа, соответствующего решению задачи, недостаточно найти компоненты связности в DAG_{gen} , так как в подграфе связности могут быть задания получения данных. Например, задания могут быть сервисами получения данных дистанционного зондирования Земли (ДЗЗ) или сервисами предоставления векторных и реляционных данных. Эти данные могут использоваться для решения большого количества различных задач. Например, данные ДЗЗ Landsat 8, которые можно получить с помощью сервисов, используются для задач оценки подтоплений из-за разлива рек, динамики изменения ледников, классификации породного состава лесов и т. д. Каждому решению задачи будет соответствовать свой подграф в DAG_{gen} , и они будут все связаны между собой через задание получения данных ДЗЗ, т. е. объединяться в компоненты связности. Поэтому при выделении подграфов, соответствующих только одной задаче, необходимо не рассматривать связи через задания получения данных, т. е. через сервисы данных $S^d \subset S$. Учет заданий, выполняемых сервисами предоставления данных, при выполнении анализа может быть полезен для определения параметров композиций сервисов, т. е. сервисов, которые могут быть заменены на другие при их повторном выполнении.

Для поиска связанных подграфов D_i применяется алгоритм поиска в ширину. При этом сложность поиска является линейной от суммы числа вершин и числа рёбер графа DAG_{gen} . Это обосновывается направленностью графа и отсутствием циклов.

Input: *nodes* – граф DAG_{gen}

Output: *subgraphs* – массив связанных подграфов

```

1  Stack <- [];
2  while(nodes is not empty):// перебираем все задания
3      new subgraph
4      add subgraph into subgraphs
5      first <- nodes.pop()
6      add first into subgraph
7      add first into stack
8      while(stack is not empty){//перебор всех связанных заданий
9          cur <- stack.pop()
10         for each n in cur.nexts:
11             if n exists in nodes then
12                 add n into subgraph.nodes
13                 add n into stack
14                 remove n in nodes
15             end if
16         end for
17         for each n in cur.prevs:
18             if n exists in nodes then
19                 add n into subgraph.nodes
20                 add n into stack
21                 remove n in nodes
22             end if
23         end for
24     end while
25 end while

```


В результате выполнения алгоритма получаем множество связанных подграфов D_i (рисунок 2). На рисунке задания получения данных выделены пунктирной линией, задания публикации данных имеют более темную заливку, каждый D_i является частным случаем реализации композиции сервисов.

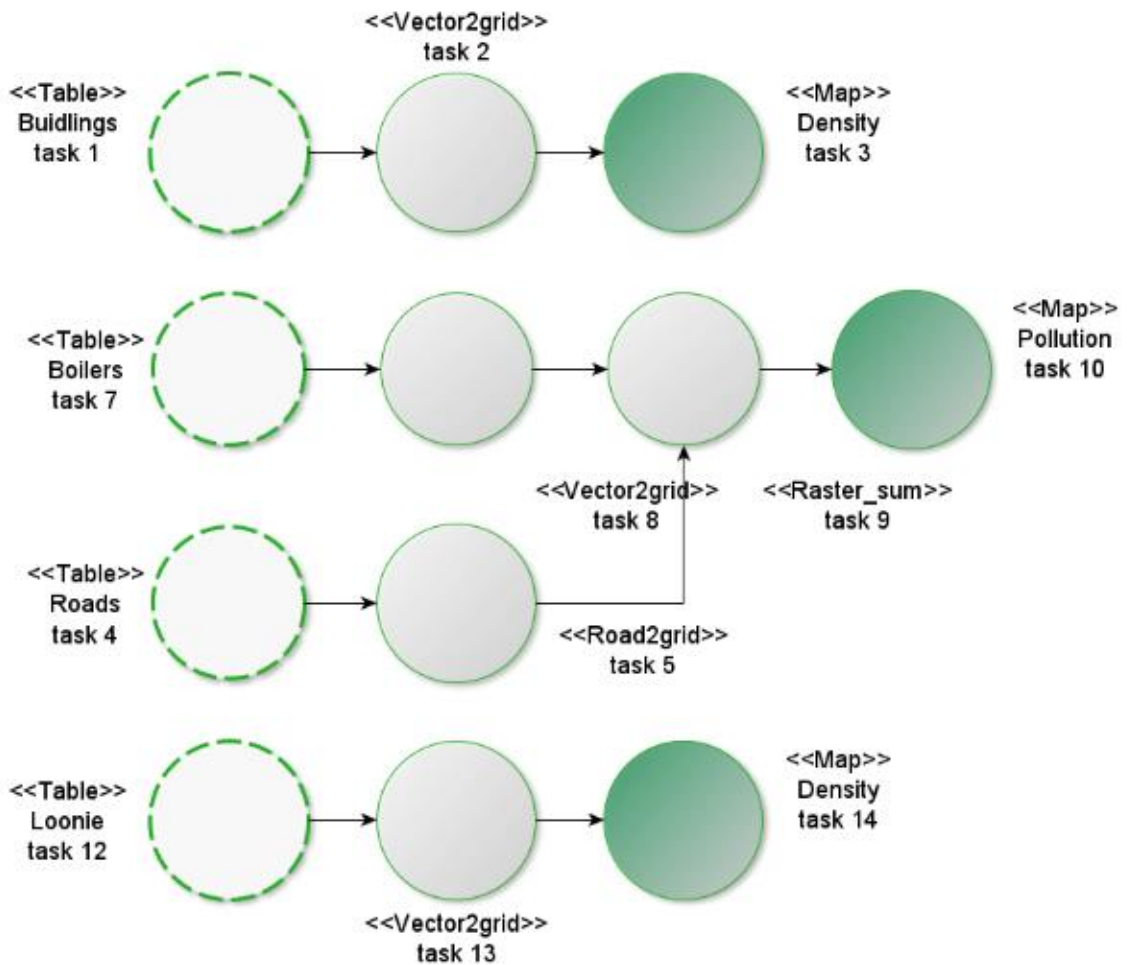


Рисунок 2 – Связанные подграфы D_i

Алгоритм поиска может получить достаточно много связанных подграфов D_i . Часть подграфов может соответствовать экспериментам, которые привели к недостаточно хорошим результатам. Также среди них может оказаться достаточно много дублирований, которые возникают из-за многократного решения задачи на разных данных и с разными параметрами. Поэтому возникает задача редуцирования множества D_i и разработки методов оценки, позволяющих ранжировать и выделить наиболее полезные композиции сервисов.

Композиции сервисов, состоящие из одной вершины (одного вызова сервиса), являются тривиальными, и для них нет необходимости формировать композиции. Поэтому подграфы D_i , состоящие из одной вершины, сразу отбрасываем. Одним из эвристических критериев необходимости связанного подграфа заданий может быть получение конечного для пользователя результата. Предполагается, что результат будет конечным, если пользователь его публикует. Например, на основе полученных данных создаются карты, графики, таблицы и т. д. В общем, среди всех связанных подграфов выделяются D_i , состоящие как минимум из двух вершин и имеющие хотя бы одно задание публикации данных, т. е. вызова сервиса $s \in S^p$.

Часто методики применения сервисов многократно повторяются на разных наборах данных. Соответственно, пользователь может повторять одну и ту же последовательность вызовов сервисов. При этом могут меняться входные данные, часть которых предоставляется сервисами. Это приводит к созданию нескольких подграфов D_i , в которых производятся вызовы одних и тех же вычислительных сервисов, отличающихся значениями параметров и сервисами предоставления данных. Заменяем в связанных подграфах D_i задания t на соответствующие сервисы s . Ребра оставим без изменений, и в новом графе SD_i эти ребра вместо заданий будут связывать сервисы. Повторение одной и той же последовательности вызовов сервисов приведет к созданию изоморфных подграфов SD_i относительно сервисов и связей по данным без учета сервисов предоставления данных. Проверка на изоморфность двух графов осуществляется нижеследующим алгоритмом.

Графы SD_i являются направленными, что значительно упрощает их сравнение. Предварительно для каждого графа добавляется вспомогательное начальное задание, не связанное с сервисом, с которого начинается выполнение композиции. Все задания, не зависящие по данным от других, добавляются к начальному заданию. Сравнение двух графов начинается с начальных заданий и выполняется рекурсивно. Два задания являются изоморфными, если название сервисов совпадают и являются изоморфными все зависящие от них задания.

Input: $n1, n2$ – графы SD_i .

Output: true или false, являются графы изоморфными или нет.

```

1  function isonodes(n1,n2)
2  if n1.nexts.length <> n2.nexts.length then
3    return false
4  end if
5  for each nx1 in n1.nexts:
6    found <- false
7    for each nx2 in n2.nexts:
8      if nx1.service = nx2.service and isonodes(nx1,nx2) then
9        found <- true
10     break
11     end if
12   end for
13   if not found then
14     return false
15   end if
16 end for
17 return true

```

Изоморфные подграфы SD_i не всегда соответствуют одной и той же композиции сервисов. Например, один и тот же набор сервисов может применяться для получения разных результатов при изменении входных параметров. При дешифрировании данных ДЗЗ могут использоваться параметры, задающие различные пороговые значения. Их изменения могут привести к получению принципиально разных результатов. С другой стороны, изменение параметров не всегда означает применение другой композиции вычислительных сервисов. Например, композиция сервисов может быть применена для другой территории, т. е. изменен параметр, задающий территорию, но решает ту же самую задачу. Возникает проблема определения, соответствуют ли два изоморфных графа решению одной задачи. Одним из способов решения этой проблемы является определение

соответствия семантики результатов изоморфных графов. Автоматическое определение семантики результатов является нетривиальной задачей, поэтому предлагается воспользоваться экспертными знаниями пользователей. В рамках среды пользователи могут указать, что два изоморфных графа друг другу соответствуют, с помощью присваивания меток.

В результате выполнения алгоритма множество компонентов связности разбивается на ряд подмножеств $IDAG_i$, где в каждом подмножестве все компоненты связности изоморфны. Количество подграфов D_j в $IDAG_i$ является оценкой частоты использования этой композиции сервисов. Далее проводится отбор изоморфных подграфов $IDAG_i$, у которых количество вершин более одной и имеется хотя бы один вызов сервиса для публикации данных.

На рисунке 3 приводится результат определения двух изоморфных подграфов SD_i , которые имеют одинаковую метку «Density» на сервисе публикации карт «Map».

Подграфы SD_i являются искомыми композициями сервисов. Для того чтобы выполнить композицию сервисов, представленную подграфом SD_i , пользователю необходимо задать все свободные входные параметры и переопределить сервисы предоставления данных. Все эти данные могут быть введены пользователем на генерируемой форме выполнения композиции сервисов. В качестве значений по умолчанию предлагаются наиболее часто используемые значения среди всех изоморфных подграфов SD_i .

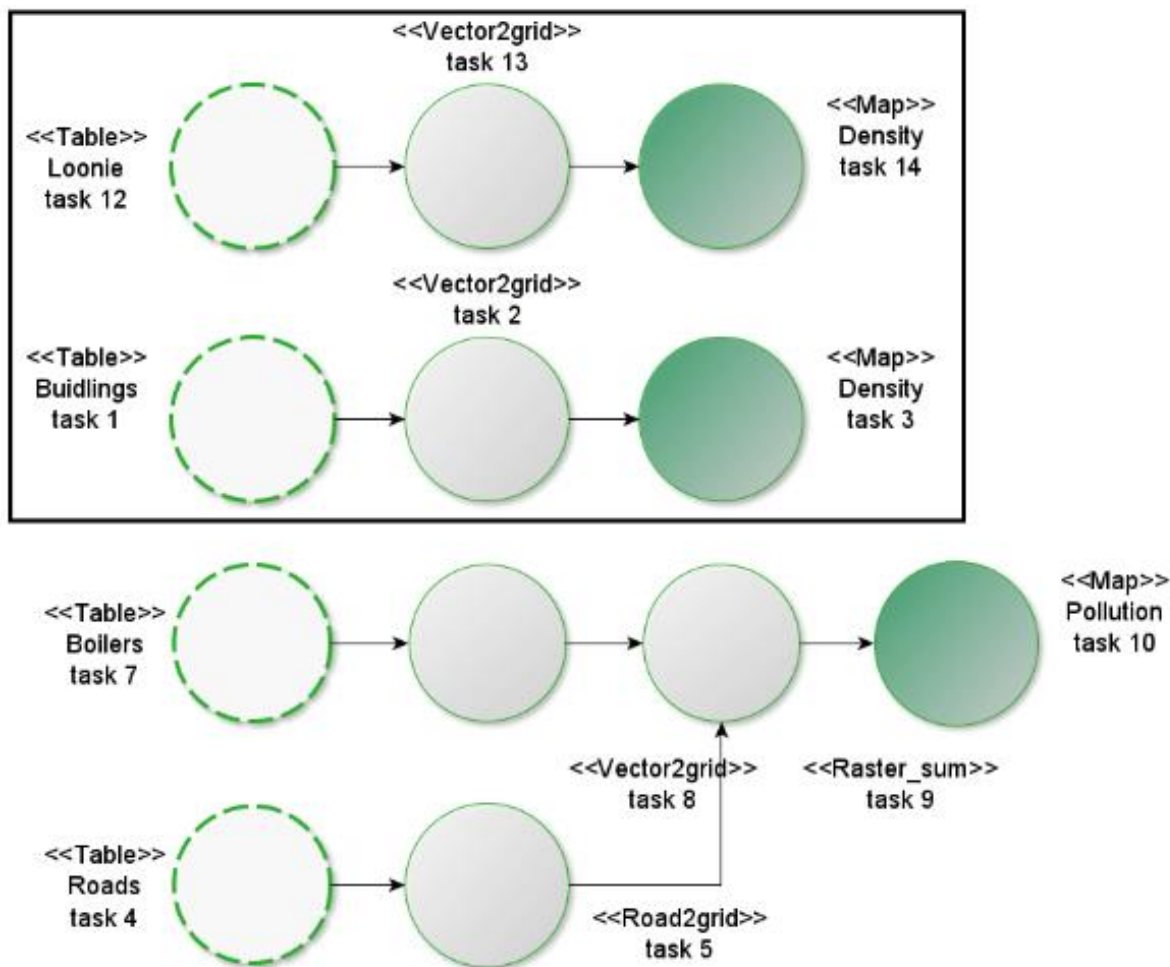


Рисунок 3 – Изоморфные компоненты связности выделены прямоугольником

Изоморфные подграфы SD_i с одинаковой семантикой результатов объединяются (рисунок 4). Количество $count(SD_i)$ изоморфных подграфов SD_i является оценкой частоты использования этой композиции сервисов. Частота использования может применяться, например, для отображения композиций сервисов в порядке, при котором отображаются сначала наиболее часто используемые.

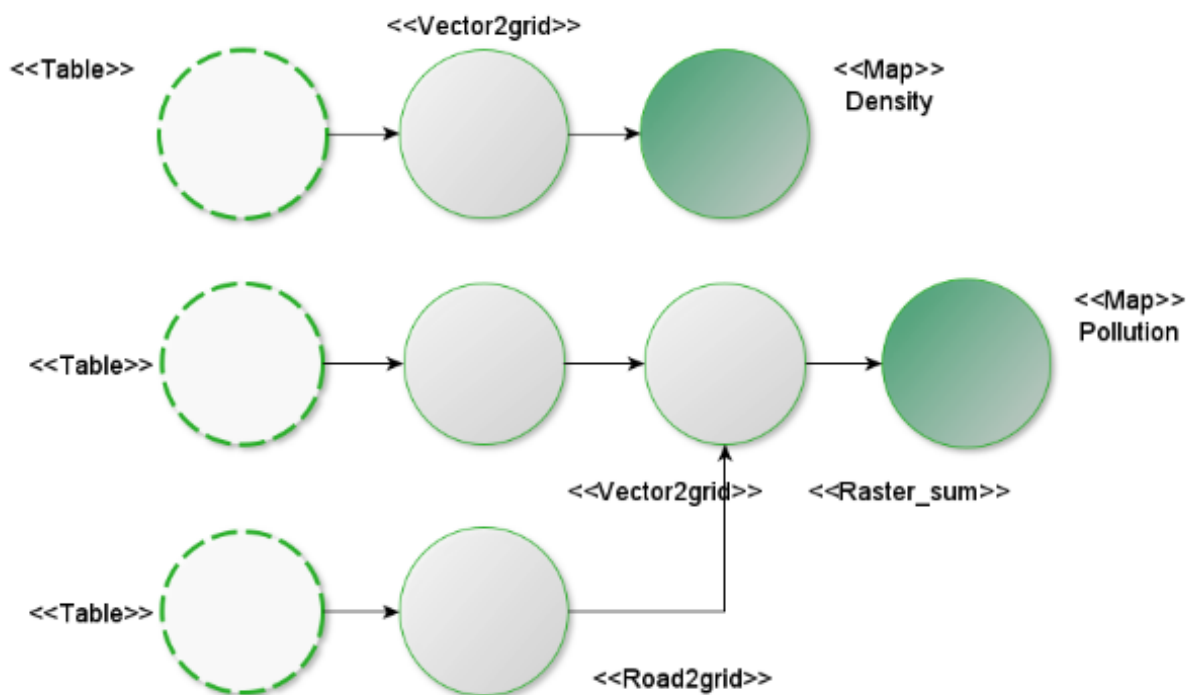


Рисунок 4 – Композиции сервисов

В разделе 2.6 приведена постановка задачи поиска композиций сервисов. Модель SM позволяет решать задачу создания композиций сервисов. Поиск композиций сервисов должен проводиться в соответствии с некоторым запросом пользователя. Пусть Q – множество запросов, запрос $q \in Q$ – это упорядоченная пара $q = \langle Input, Output \rangle$, где

- *Input* – описание начального сервиса;
- *Output* – описание конечного сервиса.

Нужно найти композицию сервисов s , представленную с помощью графа DAG, где вершины – это вызовы сервисов (задания), ребра – передача данных между заданиями. *Input* будет начальным сервисом, т. е. не будет получать данные от других сервисов, а *Output* – конечным сервисом, т. е. не будет передавать свои данные другим сервисам композиции. Все вызовы сервисов (задания) в композиции образуют связанный граф. Формирование композиций является более сложной задачей, чем поиск последовательности сервисов (*service chain*), так как часто сервисы имеют несколько параметров, данные которых могут быть получены от нескольких заданий (сервисов).

В запросе *Input* или *Output* может быть не задан. Например, когда есть определенные данные, и нужно узнать, что можно из них получить. Либо наоборот, известно, что нужно получить, но не знаем, как и из каких данных.

Запросу $q \in Q$ может соответствовать достаточно большое количество композиций сервисов. Поэтому их требуется ранжировать (упорядочить) по степени соответствия запросу. Способы ранжирования сервисов зависят от целей пользователя, ограничений и т. д. Функция $f \in F$, $f(q, s) \rightarrow [0, +\infty]$ сопоставляет каждому сервису и запросу число – вычисленную релевантность запросу пользователя (ranking function). Эта функция учитывает соответствие различных признаков сервиса запросу.

В разделе 2.7 приводится реализация поиска и эвристический метод ранжирования композиций сервисов.

Поиск композиции сервисов производится в каталоге сервисов. Представим реализацию метода поиска композиции сервисов в рамках заданной модели. Запрос пользователя $q = \langle Input, Output \rangle$ состоит из двух частей: описания начального и конечного сервисов. В рамках программной реализации модели описание может быть задано в виде текста или указанием конкретных сервисов из каталога. В случае описания в виде текста производится поиск по ключевым словам в каталоге сервисов. Первоначально производится поиск *Input* сервисов, затем поиск *Output* сервисов, если они заданы. В общем случае будем обозначать множество найденных сервисов S^{Input} для начальных, а S^{Output} для конечных сервисов. Если пользователь указал конкретные сервисы, то эти множества состоят из одного элемента.

Количество найденных композиций может быть большим. В этом случае требуется ранжировать найденные композиции для удобства пользователей. Функции ранжирования сервисов являются эвристическими. На текущий момент предложена функция, которая производит оценку по частоте использования композиции, т. е. по количеству изоморфных подграфов (см. раздел 2.5)

$$f(q, s) = count(DAG_i).$$

Данная оценка позволяет выделить композиции сервисов, которые активно применяются пользователями.

Существует множество предметных специалистов, которые производят сбор больших массивов данных с помощью сервисов. Поэтому в пользовательском интерфейсе предоставляется функция поиска, где в качестве *Input* будет сервис получения данных $s \in S^d$. В результате предметные специалисты могут оперативно получать все возможные композиции сервисов (методы), которые можно применить к их данным. Реализация такого поиска упрощает применение композиций. При этом список композиций сервисов будет пополняться по мере появления новых сервисов, добавления метаданных, онтологий или применения сервисов на аналогичных таблицах.

В разделе 2.8 приводится сравнение предлагаемой модели, алгоритмов и методов. Построение композиции сервисов включает автоматизацию следующих шагов:

- 1) поиск сервисов;
- 2) построение связей (определение параметров, по которым производится взаимодействие по данным);
- 3) проверка совместимости параметров;
- 4) построение композиций сервисов (создание выполняемого описания, например, DAG).

Применение статистики позволяет распространять среди пользователей успешный опыт проведения вычислительных экспериментов. Выполнено сравнение с точки зрения поддержки автоматизации формирования композиций сервисов

предлагаемой СОИАС с существующими системами управления научными рабочими процессами: BPEL Designer Project (<https://projects.eclipse.org/projects/soa.bpel>), GeoJModelBuilder (<https://github.com/geoprocessing/GeoJModelBuilder>), Everest (Сухорослов О.В.), CLAVIRE (Бухановский А.В.), DiVTB (Радченко Г.И.), Pegasus (<https://github.com/pegasus-isi/pegasus>). Результаты сравнения показывают, что предлагаемая реализация СОИАС при создании композиции сервисов на основе предложенной модели, алгоритмов и методов имеет ряд полезных ключевых функций: проверка совместимости параметров, использование статистики, автоматическое построение композиций сервисов.

Результаты исследований, представленные в данной главе, опубликованы в [4, 8, 18, 28, 41].

В **третьей главе** приводится алгоритмическая и программная реализация сервис-ориентированной информационно-аналитической среды. Разработана архитектура, реализующая вычислительную модель композиции сервисов *SM*, состоящая из ряда отдельных компонентов. Архитектура разработана таким образом, чтобы обеспечить функциональную расширяемость за счет сервисов, и дает возможность пользователю работать полностью удаленно. Рассмотрим основные компоненты СОАИС.

Типовой геопортал реализует функции ввода, редактирования, отображения и анализа данных с помощью сервисов. Геопортал является точкой входа для поиска и использования сервисов данных и геообработки. Для упрощения работы пользователей для каждой предметной области создается отдельный геопортал, который предоставляет быстрый доступ к наиболее часто используемым сервисам.

JupyterHub – многопользовательский сервер, предназначенный для анализа данных в среде Jupyter notebook. Позволяет проводить интерактивное программирование и создавать отчеты-ноутбуки. Предоставляет доступ к вычислительным узлам с графическими ускорителями и современным методам обработки данных. В JupyterHub автоматически подключается пользовательская директория системы хранения данных, предоставляются базовые пространственные данные, космоснимки и т. д.

Nextcloud – свободно распространяемая программная система (<https://nextcloud.com/>) для управления файловой системой хранения данных. При работе с сервисами геообработки часто приходится работать с данными в виде файлов. Выделяется каждому пользователю в системе хранения данных директория. Nextcloud предоставляет функцию регистрации пользователей и является сервером авторизации пользователей на основе протокола OAuth 2.0 для всех компонентов среды.

«Сервисы» — методы обработки данных, развернутые на вычислительных узлах облачной среды или в Интернет. Методы должны быть реализованы в соответствии со стандартом WPS. В рамках среды предоставляются виртуальные машины с предустановленным программным обеспечением для реализации WPS сервисов.

Разработан типовой геопортал на основе программной платформы Node.js, реализующий общие для информационных систем сбора и анализа данных функции на основе WPS сервисов. Исходный код геопортала выложен на Gitlab (<https://gitlab.com/fromul/geoservices>) под лицензией MIT. Геопортал обеспечивает единую точку доступа ко всем ресурсам и сервисам обработки. Его основными

функциями являются формирование готовых наборов карт, загрузка и выгрузка данных, регламентация доступа к ним, создание собственных данных, встроенный картографический веб-клиент для просмотра картографических веб-служб, применение сервисов обработки данных и т. д. (рисунок 5).



Рисунок 5 – Структура геопортала

Компоненты «Фабрика сервисов ввода и редактирования реляционных данных», «Фабрика сервисов отображения графиков», «Фабрика сервисов отображения пространственных данных» предназначены для создания новых сервисов, каждый из которых обладает программным интерфейсом и готов стать частью композиции сервисов. Созданные сервисы автоматически регистрируются в соответствующих каталогах.

«Каталог сервисов обработки данных» предназначен для хранения метаданных сервисов. Он обеспечивает поиск сервисов и их выполнение.

«Каталог данных и структурных спецификаций» предназначен для хранения метаданных сервисов для таблиц. Он обеспечивает поиск данных.

«Сервис конвертации данных» предназначен для преобразования реляционных данных с одной структуры к другой с функциями нормализации данных. Позволяет загрузить данные из разных форматов: CVS, SHAPE, MIF и т. д.

«PostgreSQL» – свободно распространяемая СУБД. Для обработки пространственных данных в СУБД установлено расширение PostGIS. Каждый пользователь геопортала может создавать таблицы. СУБД может функционировать на каждой виртуальной машине геопортала или на общей СУБД.

Фабрика сервисов ввода и редактирования реляционных данных (зарегистрирована под названием «Фарамант») предназначена для создания таблиц и сервисов работы с ними. Сервисы обеспечивают пользовательский интерфейс для ввода и редактирования разных по структуре пользовательских таблиц и реализуют добавление, чтение, изменение и удаление записей таблицы (CRUD) с поддержкой некоторых сложных типов данных, встречающихся в исследованиях. Они предоставляют программный интерфейс для обеспечения композиции с другими

сервисами. Это дает возможность интегрировать данные сервисов в информационную среду. Процесс создания сервисов тоже автоматизирован, так как таблицы могут использоваться для автоматической публикации результатов вычислений. Соответственно, геопортал предоставляет сервис для создания новых сервисов ввода и редактирования реляционных данных.

Работа сервисов основывается на модели таблицы. Модели таблиц упорядочиваются в виде иерархий и реализуются механизмы наследования и полиморфизма в терминах объектно-ориентированного подхода. Применение модели таблицы позволяет:

- 1) создавать таблицы на основе готовых и устоявшихся моделей предметной области;
- 2) унифицировать по структуре таблицы разных пользователей, содержащие общую модель или унаследованные от нее другие модели;
- 3) создавать WPS-сервисы не к конкретным таблицам, а к моделям, т. е. применять их к любым таблицам, созданным по данной модели;
- 4) проводить анализ и создавать отчеты по пользовательским таблицам, созданным на основе общей модели.

Большая часть данных является пространственно привязанной. Создаваемые сервисы позволяют работать с пространственными атрибутами: вводить, отображать на карте, выполнять пространственные операции и т. д. Большинство сервисов обработки пространственных данных поддерживают стандарт WPS. Табличные данные часто используются в качестве входных данных для сервисов. Реализована поддержка этого стандарта разработанными сервисами, что значительно повышает интероперабельность сервисов и возможность создания композиций сервисов.

Модель описывает используемую структуру данных и пользовательский интерфейс и хранится в формате JSON. Структура данных задается как множество реляционных кортежей, представленных в виде иерархии. В корне иерархии всегда находится один главный кортеж.

Каждый кортеж имеет имя, заголовок и множество атрибутов. Каждый атрибут в свою очередь специфицируется названием, именем в базе данных, типом данных, единицами измерения (для числовых данных), элементом управления и его свойствами. Атрибут специального типа details задает отношение один ко многим. С помощью этого атрибута строится иерархия документа. Элементы управления реализуют специфичные методы, необходимые для формирования пользовательского интерфейса добавления, редактирования и отображения данных атрибута. Разработано более двадцати различных элементов управления, позволяющих работать со стандартными типами данных: number, string, date, boolean и т. д.

Модели документов с метаданными (авторы, дата обновления и т. д.) хранятся в каталоге таблиц. Для создания таблицы пользователю необходимо создать модель в виде структурных спецификаций с помощью редактора модели, либо использовать готовую. На основе модели таблицы создаются таблицы в СУБД PostgreSQL, генерируется пользовательский интерфейс и определяется логика работы сервисов. Разработан REST интерфейс, который позволяет программно создать сервисы ввода и редактирования, готовые к работе, определить права доступа и т. д. Создаваемые сервисы реализуют CRUD операции, поддерживается иерархия таблиц.

В таблице значения атрибутов могут вводиться пользователями или автоматически вычисляться с помощью специальных выражений, которые могут

использовать данные сервисов. Вычисление значений атрибутов происходит при наступлении следующих событий:

- создание формы документа;
- сохранение документа;
- дублирование документа.

Вычисление атрибутов документа производится на основании выражения на специальном языке. Для создания парсера использован Jison генератор, который создает код разбора выражений и их вычислений на языке JavaScript. Выражения могут включать операторы получения значения с помощью REST сервисов. Выражения по умолчанию вычисляются на стороне сервера. Выражения, вычисляемые при создании или дублировании документа, вычисляются на стороне браузера. Из-за наличия в выражениях оператора получения значения из таблицы геопортала и оператора получения значения из REST сервиса вычисление выражение проходит асинхронно.

Разработаны компоненты для создания сервисов отображения данных:

- «Фабрика сервисов отображения пространственных данных» обеспечивает создание WMS сервиса;
- «Фабрика сервисов отображения диаграмм и графиков» предназначена для отображения графиков на основе реляционных данных, предоставляемых сервисами ввода и редактирования.

Способ отображения данных задается моделью, которая передается в параметре сервисов.

Разработан сервис конвертации реляционных данных из файлов различных ГИС форматов, поддерживаемых библиотекой GDAL/OGR. Кроме того, поддерживаются данные в формате CSV. Производится нормализация данных к существующим справочникам.

Каталог данных создан с помощью Фабрики сервисов ввода и редактирования пространственных данных, соответственно предоставляет программный и пользовательский интерфейс. Для каждого сервиса в каталоге хранятся метаданные в соответствие с ядром Dublin Core и структурные спецификации модели, необходимые для использования сервисов. Ключевые слова, которые входят в ядро метаданных, могут использоваться в качестве меток для сервисов. Кроме того, ключевые слова используются для поиска сервисов пользователем. При регистрации сервиса данных пользователь указывает набор метаданных. Поиск сервисов производится по ключевым словам. Для реализации поиска цифровых ресурсов и выполнения сложных запросов реализована поддержка спецификации GraphQL.

Каталог сервисов предназначен для сбора метаданных сервисов, их хранения, поиска, выполнения, сбора статистики и композиции сервисов. Поиск нужных сервисов и их использование являются сложной задачей из-за их распределенности, большого количества, часто отсутствия описания и т. д. Поэтому каталог сервисов является ключевым элементом среды, позволяющим производить поиск и расширять набор методов среды усилиями независимых разработчиков. Каталог создан с помощью Фабрики сервисов ввода и редактирования пространственных данных, соответственно, предоставляет программный и пользовательский интерфейс.

Разработанные программные компоненты нацелены на композицию сервисов, созданными разными разработчиками. Результаты исследований, представленные в данной главе, опубликованы в [1,2, 3, 4, 6, 10, 11, 12, 15].

В четвертой главе описывается разработанная подсистема выполнения композиций сервисов обработки пространственно-временных данных.

При увеличении количества пользователей или сложности решения задач выполнение сервисов сталкивается с ограничениями на вычислительные ресурсы, которые могут быть гетерогенными, т. е. обладать разными характеристиками по мощности, памяти и т. д. Композиция сервисов часто представлена сложным графом заданий. Задания могут быть зависимы по данным друг от друга и выполняться различное время на разных вычислительных узлах. Последовательное выполнение заданий композиций сервисов может привести к неэффективному использованию вычислительных ресурсов. Поэтому СОАИС должна обеспечить эффективное распределение вычислительных ресурсов при выполнении сервисов.

Ранее рассматривалось описание композиций в виде направленного ациклического графа (DAG). Для решения некоторых задач такой способ определения композиций не подходит, поскольку часто невозможно сформировать композицию сервисов в виде DAG из-за необходимости промежуточной обработки данных. Например, результатом работы сервиса является таблица, а на вход другого сервиса требуется передать отдельную запись таблицы. Для композиции двух сервисов требуется получить необходимую запись и преобразовать ее к нужному для второго сервиса виду. Создание отдельного сервиса для выполнения подобных действий приведет к лишним временным затратам. Такие сервисы будут использоваться только для этого сочетания сервисов из-за специфичности обработки. Кроме того, в DAG все задания (вызовы сервисов) заранее определены, но в некоторых задачах число заданий может зависеть от данных. Альтернативным способом определения композиций сервисов является использование процедурных языков программирования. Например, в Everest (Mathcloud) для задания композиций можно использовать язык Python. Применение процедурного языка позволяет обработать данные и вызвать сервисы в процессе выполнения программы (далее сценария). Количество заданий может быть большим, например, десятки тысяч. Вручную создать DAG с таким количеством заданий достаточно сложно.

Поэтому предлагается применение процедурного языка программирования для задания композиций, что позволяет проводить обработку промежуточных данных с помощью средств языка и его библиотек, использовать промежуточные данные в управляющих конструкциях языка. Применение процедурного языка значительно упрощает создание новых композиций сервисов. Существующие программные решения, использующие языки программирования для формирования композиций сервисов, как правило, не производят автоматическое планирование вызовов сервисов в условиях распределенной гетерогенной среды; процесс выполнения таких композиций сервисов неустойчив к изменениям, происходящим в среде. Методы планирования применяются к композициям сервисов, заданных только с помощью DAG.

Предполагается, что вычислительный узел, на котором развернут сервис, должен обрабатывать только одно задание в одну единицу времени. На практике запуск одновременного выполнения двух ресурсоемких заданий может привести к значительному увеличению общего времени выполнения. Поэтому на множестве вычислительных узлов CR необходимо отслеживать, чтобы два задания одновременно не выполнялись. Для реализации этого условия необходимо централизованное управление выполнением, которое реализует постановку заданий

на очередь и планирование вычислений. В среде централизованное управление осуществляет подсистема выполнения WPS сервисов.

При запуске сервиса в зависимости от его типа имеются следующие способы формирования заданий:

- 1) WPS сервис или REST: будет сформировано одно задание, которое каталогом сервисов будет отправлено подсистеме выполнения WPS сервисов;
- 2) DAG: будет сформировано множество заданий, которые также будут отправлены подсистеме выполнения;
- 3) JavaScript сценарий: каталог сервисов отправляет программу интерпретатору JavaScript сценариев, который формирует задания в процессе выполнения программы.

Подсистема выполнения WPS сервисов осуществляет планирование и выполнение WPS сервисов, обеспечивая эффективное по времени распределение заданий по вычислительным узлам. Подсистема предоставляет REST интерфейс для передачи заданий и получения результатов.

Основная идея, заложенная в задание композиций сервисов с помощью процедурных языков программирования, заключается в том, что в процессе выполнения программы (сценария) на процедурном языке программирования формируется DAG, который одновременно выполняется с учетом зависимостей по данным и существующих вычислительных ресурсов. Планирование выполнения текущего состояния DAG осуществляется существующими оптимизирующими методами. План выполнения DAG пересчитывается при добавлении новых заданий, завершении выполняющихся заданий, изменении списка вычислительных ресурсов и т.д. В качестве процедурного языка программирования для разработанного способа задания композиций выбран язык JavaScript.

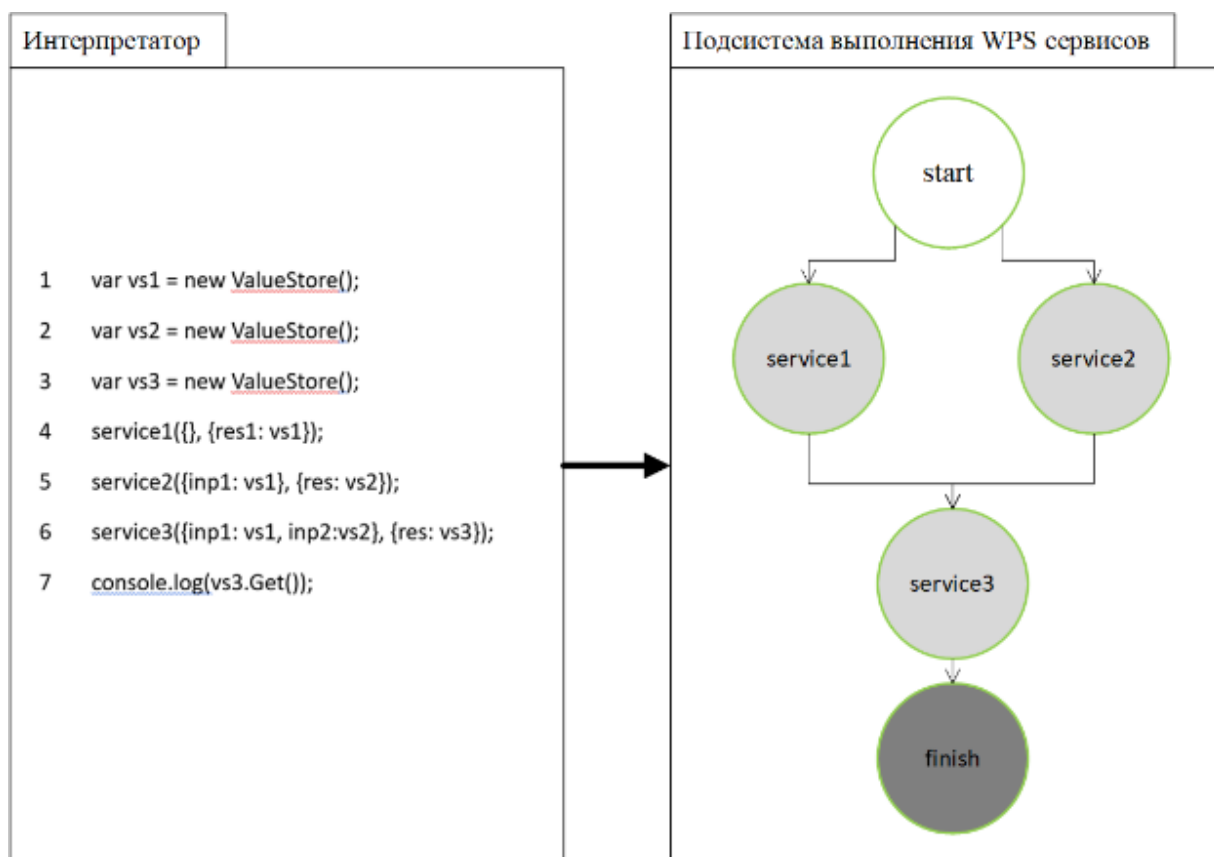


Рисунок 6 – Взаимодействие интерпретатора и модуля DAG

Выполнение программы (интерпретация JavaScript) производится интерпретатором, созданным на основе V8 – движок JavaScript с открытым исходным кодом. Интерпретация сценария и выполнение DAG осуществляется параллельно. Планирование и выполнение DAG осуществляет подсистема выполнения WPS сервисов. Интерпретатор в процессе выполнения сценария передает подсистеме выполнения новые задания с зависимостями и обработанные в сценарии данные, в том числе промежуточные. Подсистема выполнения предоставляет интерпретатору промежуточные данные при необходимости их обработки или использования в управляющих конструкциях (рисунок 6).

Передача заданий модулю DAG осуществляется с помощью вызова определенных функций, соответствующих зарегистрированным в каталоге сервисам (функции-обертки). При вызове функций-обертки происходит добавление задания в очередь с указанием сервиса, его характеристик, передаваемых и получаемых данных. Функция-обертка одного и того же сервиса может вызываться произвольное количество раз, в каждом случае это будет отдельное задание со своими зависимостями от других заданий. Формирование ребер DAG, т. е. зависимостей между заданиями по данным, производится при добавлении узла DAG. Получение данных внутри сценария для последующей обработки средствами выбранного языка программирования производится с помощью вызова метода подсистемы выполнения. Если данные еще не готовы, то блокируется выполнение сценария до получения данных.

В рамках сервис-ориентированной информационно-аналитической среды реализована модель MapReduce для параллельной обработки пространственных данных с помощью WPS сервисов. Рассмотрим особенности обработки растровых пространственных данных. Обработка данных в рамках модели MapReduce состоит из двух шагов: Map и Reduce. В шаге Map пространственные данные необходимо разделить по пространственному положению на N частей и произвести обработку каждой части с помощью WPS сервиса, затем на шаге Reduce собрать полученные данные. Операции над пространственными данными, которые используются в шагах Map и Reduce, повторяются для различных инструментов геообработки ввиду общности обрабатываемых данных. Предлагается метод, включающий в себя обработчики для операций Map и Reduce и спецификации, на основе которых будет происходить процесс распределения и сбора данных среди вычислительных узлов. Отличительной чертой данного метода является возможность использования сервисов обработки пространственных данных в распределенной вычислительной среде без их модификации.

Подсистема выполнения WPS сервисов производит планирование, непосредственный вызов и получение результатов выполнения сервисов. Структура подсистемы представлена на рисунке 7. Взаимодействие с каталогами сервисов и JavaScript интерпретаторами производится с помощью модуля HTTP сервер, который реализует REST интерфейс. Добавление задания через REST приводит к постановке задания в очередь DAG, содержащего задания разных пользователей. При постановке производится запрос к базе данных для уточнения списка вычислительных узлов, на которых данный сервис может выполняться, и времени работы задания.

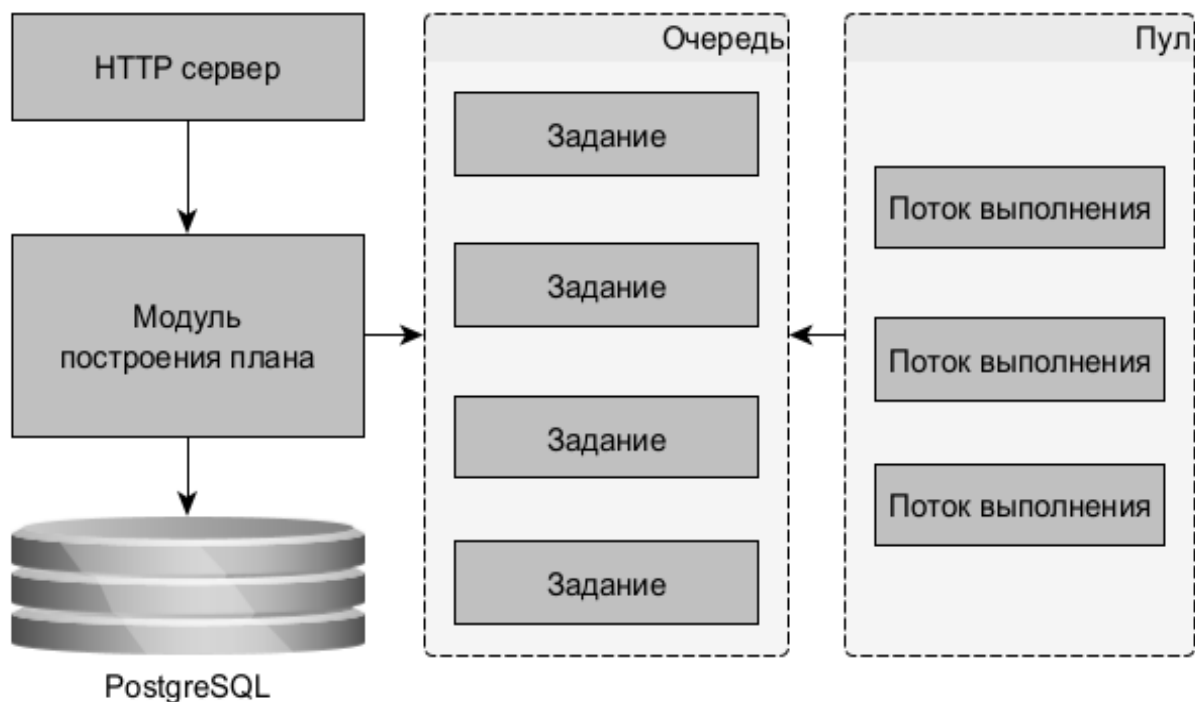


Рисунок 7 – Структура подсистемы выполнения WPS сервисов

Для нахождения плана, приближенного к оптимальному по времени выполнения, используется модификация спискового эвристического алгоритма составления расписания HEFT (Heterogeneous Earliest Finish Time), реализующего метод поиска в глубину. Поток выполнения запускает задания, проверяет их состояние, отправляя запрос на получение файла статуса выполнения. Поточков выполнения может быть несколько для того, чтобы параллельно запускать и отслеживать несколько сервисов.

При выполнении DAG необходимо учитывать изменение состояния вычислительной среды и добавление новых вершин в DAG. Все эти изменения требуют перестроения плана. Целесообразно проводить перестроение только при определенных событиях:

- 1) включение или отключение вычислительного узла, поддерживающего выполнение хотя бы одного сервиса из DAG;
- 2) добавление нового задания в DAG;
- 3) отличие фактического времени выполнения одного из заданий DAG от ожидаемого;
- 4) завершение выполнения задания с ошибкой.

При регистрации события происходит перестроение плана. Например, в случае добавления задания в DAG и назначения его на менее загруженный вычислительный узел общее время плана не меняется, перестроение не требуется.

Результаты, представленные в данной главе, опубликованы в [5, 7].

В **пятой главе** представлены реализации методик, созданных на основе метода создания композиций, и апробация сервис-ориентированной информационно-аналитической среды обработки пространственно-временных данных.

СОИАС создана на основе сетевой и вычислительной инфраструктуры ИДСТУ СО РАН. В рамках среды создано 7 геопорталов, ориентированных на разные предметные области. СОИАС обеспечивает множество проектов, среди которых

наиболее крупный «Фундаментальные основы, методы и технологии цифрового мониторинга и прогнозирования экологической обстановки Байкальской природной территории», работающий с 2020 года. Он объединяет 13 институтов Сибирского отделения РАН. В рамках проекта постоянно развивается геопортал «Цифровой мониторинг Байкальской природной территории» для проведения комплексного экологического мониторинга и прогнозирования на основе цифровых платформ, комплекса математических и информационных моделей, сервисов и методов машинного обучения, обеспечивающих сбор, хранение, обработку, анализ больших массивов разнородных пространственных данных. Геопортал находится по адресу <https://geos.icc.ru/>.

Разработаны методики создания композиций сервисов для:

- 1) ботанических исследований;
- 2) моделирования загрязнений воздуха;
- 3) анализа ПД на основе JavaScript;
- 4) обработки данных ДЗЗ.

В рамках каждой методики выполнено построение модели предметной области, созданы модели общеиспользуемых данных, с помощью Фабрики сервисов ввода и редактирования реляционных данных созданы сервисы данных, разработаны и зарегистрированы предметные сервисы обработки данных, определены онтологические модели предметных областей и сформированы экспертные знания в виде множества меток. Также собрана статистика применения сервисов.

На основе представленных методик автоматически сформирован ряд композиций сервисов. Например, для проведения ботанических исследований автоматически создана композиция сервисов прогнозирования изменения флористического состава Байкальского региона (рисунок 8).

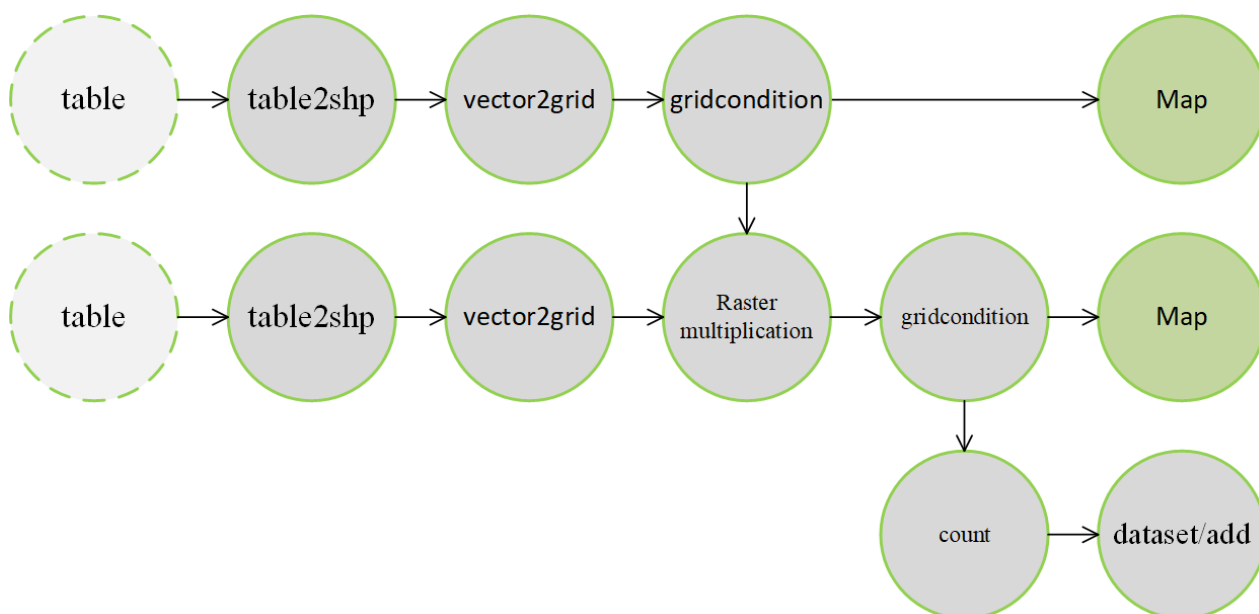


Рисунок 8 – Композиция сервисов анализа распространения вида

Разработка композиции сервисов с помощью языка JavaScript применялась для расчета временной доступности инфраструктурных объектов. Код композиции сервисов представлен ниже. Первоначально производится сохранение данных в формат GeoJSON. Затем формируется цикл по всем объектам, представленным в

GeoJSON. Каждый объект сохраняется в файл формата SHAPE. Далее строится буферная область вокруг объекта в виде растра и вызывается сервис road_analysis, который строит искомый растр.

```
1  function school_avialablity(input, output){
2      var dir = input.dir.substring(0, input.dir.lastIndexOf("\\")+1);
3      var schools_json = new ValueStore();
4      table2json({source: input.schools}, {ResultJSON: schools_json});
5      var schools = JSON.parse(schools_json.get());
6      for(var l = 0; l < schools.features.length; i++){
7          var sch = new ValueStore();
8          var sch_jsn={ "type": "FeatureCollection", "features": [school_json.features[i]]}
9          sch.set(sch_jsn);
10         var shape = new ValueStore();
11         geojson_to_shp_converter({json:sch }, {shape: shape });
12         var tiffbuf = new ValueStore();
13         bufferize_vector({input: shape, extent: input.extent, cellsize: input.cellsize },
14             {buf: tiffbuf });
15         var resltfile = new ValueStore();
16         resltfile.set(dir+"\\ "+i+".TIFF");
17         road_analysis({targets: tiffbuf, roads: input.roads, barrier: input.barrier,
18             extent: input.extent, cellsize: input.cellsize }, {res: resltfile })
19     }
20 }
```

Создание композиции и ее сохранение в каталоге сервисов позволяет любому пользователю повторить последовательность сервисов с теми же параметрами на других данных. Указанные при построении модели метки сервисам данных позволяют рекомендовать пользователям сервисы.

Комплекс программных компонентов, реализующий модель сервис-ориентированной информационно-аналитической среды, активно используется на практике. Созданы более 200 сервисов данных, более 40 сервисов обработки данных. Развернуты 7 геопорталов, ориентированных на различные предметные области и коллективы. На основе созданных сервисов сформированы композиции сервисов, объединяющие сервисы данных, сервисы обработки и публикации. Впервые разработана среда, которая обеспечивает создание композиций сервисов и их обмен между пользователями.

Результаты исследований, представленные в данной главе, опубликованы в [9, 12, 13, 14, 16, 17, 19–27, 29-40].

Заключение

В ходе выполнения диссертационного исследования проведен анализ инструментальных средств, технологий и существующих информационно-аналитических сред с использованием сервисов обработки пространственно-временных данных и средств создания композиций сервисов. Выявлено, что поиск новых композиций сервисов является комбинаторно сложной задачей, в которой необходимо проанализировать большое количество сервисов и предоставить наиболее релевантные их комбинации. Создание композиций сервисов ограничено сложностью взаимодействия сервисов по данным из-за различий в программном интерфейсе доступа к данным, в структуре, в используемых справочниках и т. д. Существующие методы и подходы на основе анализа метаданных и онтологий не позволяют значительно упростить поиск и создание композиций сервисов, поэтому разработка сервис-ориентированной информационно-аналитической среды обработки междисциплинарных пространственных данных, повышающей эффективность научных исследований за счет создания композиций сервисов и организации обмена ими, является актуальной задачей для развития междисциплинарных исследований. В процессе ее решения получены следующие новые результаты:

- вычислительная модель сервис-ориентированной информационно-аналитической среды обработки пространственно-временных данных междисциплинарных исследований, которая в сравнении с подобными моделями обеспечивает оценку композиций сервисов на основе многопользовательской статистики их применения;
- оригинальные алгоритмы и метод автоматического создания композиций сервисов на основе статистических данных использования сервисов, базирующиеся на применении предложенной модели. Использование метода может значительно упростить работу пользователя и автоматизирует часто повторяющиеся его действия. В результате применения модели и метода создания композиций опыт применения пользователем сервисов автоматически распространяется среди всех пользователей;
- оригинальный программный компонент выполнения композиций сервисов, заданных на процедурном языке, с обработкой промежуточных данных с помощью средств языка и его библиотек, в процессе выполнения которого формируется DAG, для которого одновременно обеспечивается динамическое планирование с применением алгоритма HEFT и выполнение в гетерогенной динамической вычислительной среде;
- основные компоненты среды, реализующие модель СОИАС и метод автоматического создания композиций сервисов, в частности, оригинальные программные инструменты «Фабрика сервисов ввода и редактирования реляционных данных», «Фабрика сервисов отображения пространственных данных», каталог данных и структурных спецификаций, каталог сервисов обработки данных, сервис конвертации реляционных данных;
- апробация СОИАС на задаче поддержки междисциплинарных научных исследований Байкальской территории. Комплекс программных компонентов, реализующий модель сервис-ориентированной информационно-аналитической среды, активно используется на практике. Созданы более 200 сервисов данных, более 40 сервисов обработки данных. Развернуты 6 различных геопорталов, ориентированных на различные предметные области и коллективы. На основе

созданных сервисов сформированы композиции сервисов, объединяющие сервисы данных, сервисы обработки и публикации. Впервые среда обеспечивает создание композиций сервисов и их обмен между пользователями.

Результаты диссертации апробированы, опубликованы, обсуждены на семинарах и конференциях, использованы для решения ряда важных практических задач. В настоящее время разработанные программные компоненты специалистами из различных научных и образовательных организаций.

Предложенная в диссертации сервис-ориентированная технология проведения научных экспериментов допускает свое естественное развитие и обобщение применительно к другим предметным областям научных исследований.

СПИСОК ОСНОВНЫХ ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

Монографии

1. Маджара Т.И., Бычков И.В., Ружников Г.М., Хмельнов А.Е., Федоров Р.К., Парамонов В.В., Шигаров А.О., Фереферов Е.С., Гаченко А.С., Михайлов А.А., Шумилов А.С., Авраменко Ю.В. Инфраструктура информационных ресурсов и технологии создания информационно-аналитических систем территориального управления: моногр. 2016.
2. Бычков И.В., Ружников Г.М. и др. Фундаментальные основы, методы и технологии цифрового мониторинга и прогнозирования экологической обстановки Байкальской природной территории: моногр. Новосибирск. 2022.

В изданиях, рекомендованных ВАК:

3. Бычков И.В., Ружников Г.М., Хмельнов А.Е., Федоров Р.К., Маджара Т.И., Шигаров А.О., Дорж Т., Нергуй Б. Технологические основы развития инфраструктуры пространственных данных Монгольской академии наук // Вычислительные технологии. 2013. Т. 18, № 5. С. 16–26.
4. Федоров Р.К., Шумилов А.С., Бычков И.В., Ружников Г.М. Компоненты среды WPS-сервисов обработки геоданных // Вестник Новосибирского гос. ун-та. Сер. Информационные технологии. 2014. Т. 12, № 3. С. 16-24.
5. Федоров Р.К., Бычков И.В., Шумилов А.С., Ружников Г.М. Система планирования и выполнения композиций веб-сервисов в гетерогенной динамической среде // Вычислительные технологии. 2016. Т. 21, № 6. С. 18-35.
6. Бычков И.В., Ружников Г.М., Парамонов В.В., Шумилов А.С., Фёдоров Р.К. Инфраструктурный подход к обработке пространственных данных в задачах управления территориальным развитием // Вычислительные технологии. 2018. Т. 23, № 4. С. 15-31. DOI: 10.25743/ICT.2018.23.16488.
7. Бычков И.В., Ружников Г.М., Фёдоров Р.К., Шумилов А.С. Выполнение JAVASCRIPT-композиций WPS-сервисов в распределенной гетерогенной среде // Вычислительные технологии. 2019. Т. 24, № 3. С. 44-58. DOI: 10.25743/ICT.2019.24.3.004.
8. Фёдоров Р.К., Бычков И.В., Ружников Г.М. Формирование композиций сервисов на основе статистических данных пользователей // Вестник Новосибирского гос. ун-та. Сер. Информ. технологии. 2021. Т. 19, № 2. С. 115-130. DOI: 10.25205/1818-7900-2021-19-2-115-130.
9. Бычков И.В., Ружников Г.М., Федоров Р.К., Попова А.К., Будээбазар А., Балт Б., Ууганбаатар Д. Цифровая трансформация экологического мониторинга оз.

- Хубсугул и Прихубсугулья // Вычислительные технологии. 2022. Т. 27. № 5. С. 14-29
10. Бычков И.В., Федоров Р.К., Фереферов Е.С. Инструментальные компоненты цифровой платформы экологического мониторинга Байкальской природной территории // Вычислительные технологии. 2023. Т. 28. № 6. С. 95-107
11. Бычков И.В., Маджара Т.И., Новопашин А.П., Фереферов Е.С., Феокистов А.Г., Федоров Р.К. Информационно-вычислительные ресурсы ИРНОК: инфраструктура, данные, приложения // Вычислительные технологии. 2023. Т. 28. № 3. С. 117-135

В изданиях, входящих в базы цитирования Web of Science и Scopus:

12. Bychkov I.V., Plyusnin V.M., Ruzhnikov G.M., Fedorov R.K., Khmel'nov A.E., Gachenko A.S. The creation of a spatial data infrastructure in management of regions (exemplified by Irkutsk oblast) // Geography and Natural Resources. 2013. Vol. 34, № 2. P. 191–195. DOI: 10.1134/S1875372813020133 (Scopus).
13. Paramonov V., Fedorov R., Ruzhnikov G., Shumilov A. Web-Based Analytical Information System for Spatial Data Processing // Communications in Computer and Information Science. 2013. Vol. 403. P. 93-101 (Web of Science Emerging Sources Citation Index).
14. Bychkov I.V., Ruzhnikov G.M., Paramonov V.V., Shumilov A.S., Fedorov R.K., Sanjaa B. Infrastructural approach to spatial data processing in applications to territorial development management // CEUR Workshop Proceedings. 2017. Vol. 2033. P. 7-9 (Scopus).
15. Fedorov R.K., Shumilov A.S., Ruzhnikov G.M. Geoportal cloud // CEUR Workshop Proceedings. 2017. Vol. 2033. P. 305-308 (Scopus).
16. Bychkov I.V., Fedorov R.K., Avramenko Y.V., Shumilov A.S., Shigarov A.O., Ruzhnikov G.M. et al. Information-analytical environment supporting interdisciplinary research of natural resources in the Baikal region // CEUR Workshop Proceedings, Proc. 1st Scientific-Practical Workshop on Information Technologies: Algorithms, Models, Systems (ITAMS-2018). 2018. Vol. 2221. P. 42-52. URL: <http://ceur-ws.org/Vol-2221/paper8.pdf> (Scopus).
17. Fedorov R.K., Shumilov A.S. Service compositions in problems of urban planning // Proc. 1st Scientific-Practical Workshop on Information Technologies: Algorithms, Models, Systems (ITAMS-2018). 2018. Vol. 2221. P. 1-6. URL: <http://ceur-ws.org/Vol-2221/paper1.pdf> (Scopus).
18. Fedorov R.K., Shumilov A.S., Voskoboynikov M.L. Analysis of service calls for construction of the semantic network of services // Proc. 1st Scientific-Practical Workshop on Information Technologies: Algorithms, Models, Systems (ITAMS-2018). 2018. Vol. 2221. P. 20-24. URL: <http://ceur-ws.org/Vol-2221/paper4.pdf> (Scopus).
19. Bychkov I.V., Ruzhnikov G.M., Paramonov V.V., Shumilov A.S., Fedorov R.K., Levi K.G., Demberel S. Infrastructural approach and geospatial data processing services in the tasks of territorial development management // Proceedings 1st Intern. Geographical Conference of North Asian Countries on China-Mongolia-Russia Economic Corridor: Geographical and Environmental Factors and Territorial Development Opportunities. 2018. Vol. 190, № 1. DOI: 10.1088/1755-1315/190/1/012048 (Web of Science Emerging Sources Citation Index).

20. Fereferov E.S., Gachenko A.S., Hmelnov A.E., Fedorov R.K. Information Technologies for Monitoring of Anthropogenic Impacts to Lake Baikal // Proceedings for First Scientific-practical Workshop Information Technologies: Algorithms, Models, Systems (ITAMS-2018). 2018. Vol. 2221. P. 61-69. URL: <http://ceur-ws.org/Vol-2221/paper10.pdf> (Scopus).
21. Avramenko Y.V., Fedorov R.K. The technology of classification geospatial data based on WPS standard // Proceedings 1st Scientific-Practical Workshop on Information Technologies: Algorithms, Models, Systems (ITAMS-2018). 2018. Vol. 2221. P. 37-41. URL: <http://ceur-ws.org/Vol-2221/paper7.pdf> (Scopus).
22. Bychkov I.V., Gachenko A.S., Hmelnov A.E., Fedorov R.K., Fereferov E.S. Geological Information System of environmental and human intervention impact assessment on bodies of water of the Irkutsk region // Proceedings 1st Intern. Geographical Conference of North Asian Countries on China-Mongolia-Russia Economic Corridor: Geographical and Environmental Factors and Territorial Development Opportunities. 2018. Vol. 190, № 1. DOI: 10.1088/1755-1315/190/1/012027 (Web of Science Emerging Sources Citation Index).
23. Bychkov I.V., Ruzhnikov G.M., Khmelnov A.E., Fedorov R.K., Madzhara T.I. Digital monitoring of the ecosystem of Lake Baikal // CEUR Workshop Proceedings. Volume: All-Russian Conference «Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes» (SDM 2019). 2019. Vol. 2534. P. 8-14 (Scopus).
24. Bychkov I.V., Ruzhnikov G.M., Khmelnov A.E., Fedorov R.K., Madzhara T.I., Popova A.K. Digital monitoring of Lake Baikal and its coastal area // CEUR Workshop Proceedings: Proceedings of 2nd Scientific-Practical Workshop Information Technologies: Algorithms, Models, Systems (ITAMS-2019). 2019. Vol. 2463. P. 13-23. DOI: 2-s2.0-85073518338 (Scopus).
25. Bychkov I., Ruzhnikov G., Paramonov V., Mikhailov A., Fedorov R., Klyuchevskii A., Dem'yanovich V., Demberel S. The Framework of the Digital Environment for Analysing of Seismic Hazards of Lithosphere Blocks in Baikal-Mongolian Region // CEUR Workshop Proceedings: Proceedings of 2nd Scientific-Practical Workshop Information Technologies: Algorithms, Models, Systems (ITAMS-2019). 2019. Vol. 2463. P. 84-92 (Scopus).
26. Batuev A.R., Batuev D.A., Beshentsev A.N., Bogdanov V.N., Dashpilov T.B., Korytniy L.M., Tikunov V.S., Fedorov R.K. Atlas information system for providing socio-economic development of the Baikal region // InterCarto, InterGIS. 2019. Vol. 25. P. 66-80. DOI: 10.35595/2414-9179-2019-1-25-66-80 (Scopus).
27. Fedorov R.K., Kitov A.D., Avramenko Y.V. Automation of forming a database of the glaciers based on remote sensing // CEUR Workshop Proceedings: All-Russian Conf. «Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes» (SDM 2019). 2019. Vol. 2534. P. 207-211 (Scopus).
28. Fedorov R. Building service composition based on statistics of the services use // CEUR Workshop Proceedings: Proc. of 2nd Scientific-Practical Workshop Information Technologies: Algorithms, Models, Systems (ITAMS-2019). 2019. Vol. 2463. P. 40-46. (Scopus).
29. Ruzhnikov G.M., Klyuchevskii A.V., Paramonov V.V., Mikhailov A.A., Fedorov R.K., Dem'yanovich V.M., Demberel S. Service-oriented Information and Analytical the System of Estimation of Influence of the Lithosphere Model on Dynamic Parameters of Rocky Soil Oscillations From Earthquakes of Southern Baikal Region // CEUR Workshop Proceedings: Proceedings Information Technologies in Earth Sciences and

- Applications for Geology, Mining and Economy (ITES&MP-2019). 2019. Vol. 2527. P. 42-46 (Scopus).
30. Bychkov I., Feoktistov A., Gorsky S., Edelev A., Sidorov I., Kostromin R., Fereferov E., Fedorov R. Supercomputer Engineering for Supporting Decision-making on Energy Systems Resilience // Proceedings of the 14th IEEE International Conference on Application of Information and Communication Technologies (IEEE, 2020). 2020. P. 1-6 (Scopus).
 31. Bychkov I.V., Ruzhnikov G.M., Fedorov R.K., Khmelnov A.E., Popova A.K. Digital environmental monitoring technology Baikal natural territory // CEUR Workshop Proceedings: 3rd Scientific-Practical Workshop Information Technologies: Algorithms, Models, Systems (ITAMS-2020). 2020. Vol. 2677 (Scopus).
 32. Avramenko Y.V., Fedorov R.K. Applied digital platform for remote sensing data processing // CEUR Workshop Proceedings: 3rd Scientific-Practical Workshop Information Technologies: Algorithms, Models, Systems (ITAMS-2020). 2020. Vol. 2677 (Scopus).
 33. Bychkov I.V., Ruzhnikov G.M., Fedorov R.K., Popova A.K. Digital platform for forest resources monitoring in the BAIKAL natural territory // Journal of Physics: Conference Series: 13th Multiconference on Control Problems (MCCP 2020). 2021. Vol. 1864, № 1. DOI: 10.1088/1742-6596/1864/1/012111 (Scopus).
 34. Bychkov I.V., Ruzhnikov G.M., Fedorov R.K., Popova A.K. A platform approach to the organization of digital forest monitoring of the Baikal natural territory // IOP Conference Series: Earth and Environmental Science: 11th Intern. Conf. and Early Career Scientists School on Environmental Observations, Modeling and Information Systems (ENVIROMIS 2020). 2021. Vol. 611. P. 012056. DOI: 10.1088/1755-1315/611/1/012056 (Scopus).
 35. Bychkov I.V., Ruzhnikov G.M., Fedorov R.K., Khmelnov A.E., Popova A.K. Organization of digital monitoring of the Baikal natural territory // IOP Conference Series: Earth and Environmental Science. 2021. DOI: 10.1088/1755-1315/629/1/012067 (Scopus).
 36. Bychkov I.V., Ruzhnikov G.M., Fedorov R.K., Popova A.K. Digital technologies for forest monitoring in the Baikal natural territory // CEUR Workshop Proceedings: 4th Scientific-Practical Workshop Information Technologies: Algorithms, Models, Systems (ITAMS-2021). 2021. Vol. 2984. P. 1-5 (Scopus).
 37. Avramenko Y.V., Popova A.K., Fedorov R.K. Cloud service of Geoportal ISDCT SB RAS for machine learning // CEUR Workshop Proceedings: 4th Scientific-Practical Workshop Information Technologies: Algorithms, Models, Systems (ITAMS-2021). 2021. Vol. 2984. P. 6-10 (Scopus).
 38. Bychkov I.V., Paramonov V.V., Ruzhnikov G.M., Mikhailov A.A., Fedorov R.K., Klyuchevskii A.V., Dem'yanovich V.M., Demberel S. Russian-Mongolian scientific initiative for assessing the seismic hazards of the Baikal region and Mongolia // CEUR Workshop Proceedings: 4th Scientific-Practical Workshop Information Technologies: Algorithms, Models, Systems (ITAMS-2021). 2021. Vol. 2984. P. 112-119 (Scopus).
 39. Zubova E., Kashulin N., Terentyev P., Melekhin A., Fedorov R.K., Shalygin S. Occurrence of fish species in the inland water of Murmansk Region (Russia): research in 1972-2021 // Biodiversity Data Journal. 2021. Vol. 9. DOI: 10.3897/BDJ.9.e68131 (Web of Science Q3).
 40. Bychkov I.V., Ruzhnikov G.M., Fedorov R.K., Popova A.K., Avramenko Y.V. Classification of Sentinel-2 satellite images of the Baikal Natural Territory //

Computer Optics. 2022. Vol. 46, № 1. P. 90-96. DOI: 10.18287/2412-6179-CO-1022. URL: <http://www.computeroptics.ru/KO/PDF/KO46-1/460111.pdf> (Web of Science Q2).

41. Feoktistov A., Gorsky S., Kostromin R., Fedorov R., Bychkov I. Integration of web processing services with workflow-based scientific applications for solving environmental monitoring problems // ISPRS International Journal of Geo-Information. 2022. Vol. 11, № 1 DOI: 10.3390/ijgi11010008 (Web of Science Q2).
42. Bychkov I.V., Feoktistov A.G., Gorsky S.A., Kostromin R.O., Fedorov R.K. Automating the Integration of Services for the Web Processing of Environmental Monitoring Data with Distributed Scientific Applications // Optoelectronics, Instrumentation and Data Processing. 2022. Vol. 58, №4. P. 373–379. DOI: 10.3103/S8756699022040045 (Scopus Q3)

Регистрации программ для ЭВМ

43. Шумилов А.С., Фёдоров Р.К., Ружников Г.М., Ветров А.А., Михайлов А.А. Интернет-система ввода и редактирования пространственных данных «Фарамант»: Свидетельство о государственной регистрации программ для ЭВМ № 2014610274 от 09.01.2014. М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2014.
44. Бычков И.В., Федоров Р.К., Шумилов А.С., Ружников Г.М. Система планирования и выполнения композиций веб-сервисов в гетерогенной динамической среде: Свидетельство о государственной регистрации программ для ЭВМ № 2016663724 от 18.10.2016. М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2016.
45. Федоров Р.К., Бычков И.В., Шумилов А.С., Ружников Г.М. Среда выполнения сервисов и их сценариев: Свидетельство о государственной регистрации программ для ЭВМ № 2017617913 от 17.07.2017. М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2017.
46. Федоров Р.К., Парамонов В.В., Ружников Г.М., Данчинова Г.А., Хасантинов М.А., Ляпунов А.В. WEB-сервис импорта данных из реляционных таблиц в CSV формате: Свидетельство о государственной регистрации программ для ЭВМ № 2017615230 от 05.05.2017. М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2017.
47. Данчинова Г.А., Ляпунов А.В., Хасантинов М.А., Болотова Н.А., Манзарова Э.Л., Соловаров И.С., Петрова И.В., Шулёпова С.Ю., Лазарева Е.Л., Федоров Р.К., Парамонов В.В. Информационно-справочная система «Обращаемость людей, пострадавших от присасывания иксодовых клещей на Территории Бурятии» (ИСС «Бурятия-клещи»): Свидетельство о государственной регистрации программ для ЭВМ № 2017620505 от 04.05.2017. М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2017.
48. Федоров Р.К., Авраменко Ю.В. Программа расчета спектральных индексов и построения композитного отображения для космоснимков Sentinel-2: Свидетельство о государственной регистрации программ для ЭВМ № 2022615805 от 04.04.2022. М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2022.
49. Фёдоров Р.К. Программная система выполнения WPS сервисов: Свидетельство о государственной регистрации программ для ЭВМ № 2024611171 от 20.12.2023. М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2023.

50. Фёдоров Р.К., Авраменко Ю.В. WPS сервис классификации космоснимков Sentinel-2: Свидетельство о государственной регистрации программ для ЭВМ № 2024610919 от 21.12.2023. М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2023.

Научно-организационный отдел
Федерального государственного бюджетного учреждения науки
Института динамики систем и теории управления имени В.М. Матросова СО
РАН
664033, Иркутск, ул. Лермонтова, д. 134
e-mail: rio@icc.ru
Подписано к печати 24.09.2024 г.
Формат бумаги 60×84 1/16, объем 1 п.л.
Заказ № 5. Тираж 100 экз.

Отпечатано в ИДСТУ СО РАН