

ТЕХНОЛОГИИ СЕМАНТИЧЕСКОГО ВЕБА ДЛЯ ПОДДЕРЖКИ ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ В ГЕОЛОГИИ

И. В. Бычков¹ [0000-0002-1765-0769], Е. А. Черкашин² [0000-0003-2428-2471],
Цзинь Чжан³ [0000-0003-0009-824X], Т. Ю. Черкашина⁴ [0000-0002-1443-6839],
В. А. Попова⁵ [0000-0002-7764-1995], О. А. Мазаева⁶ [0000-0002-3053-5511],
О. В. Лунина⁷ [0000-0001-7743-8877]

^{1, 2, 5} *Институт динамики систем и теории управления им. В. М. Матросова
Сибирского отделения РАН, г. Иркутск, Россия*

^{1, 2, 4} *Институт математики и информационных наук, Яньтайский
университет, г. Яньтай, Китай*

³ *Институт гражданского строительства, Яньтайский университет,
г. Яньтай, Китай*

^{4, 6, 7} *Институт земной коры Сибирского отделения РАН, г. Иркутск, Россия*

^{1, 5} *Институт математики и информатики, Иркутский государственный
университет, г. Иркутск, Россия*

¹bychkov@icc.ru, ²eugeneai@icc.ru, ³zhangjin513@outlook.com,

⁴tcherk@crust.irk.ru, ⁵victorypopova1@gmail.com, ⁶moks@crust.irk.ru,

⁷lounina@crust.irk.ru

Аннотация

Представлена инновационная методология применения технологий семантического веба для поддержки фундаментальных геологических исследований. Рассмотрена проблема семантической интеграции разнородных геологических данных, характеризующихся масштабом разного уровня и междисциплинарностью. Разработана пятиэтапная методология, включающая анализ предметной области, онтологическое концептуальное моделирование, трансформацию данных в граф знаний, развертывание инфраструктуры распределенного доступа к данным на основе концептуальной модели, а также интеграцию с процедурами обработки и анализа. Практическая апробация проведена на трех кейсах: анализе геохимических данных для оценки уровня загрязнения территории, создании информационной системы о разломах и исследовании динамики береговой зоны

водохранилищ. Предложенный онтологический подход обеспечивает соответствие FAIR-принципам и преодоление «семантического барьера» в геологических исследованиях. Показано, что технологии семантического веба позволяют перейти от фрагментированных информационных массивов к целостному семантическому пространству геологических знаний, что открывает новые возможности для генерации комплексных научных гипотез и кросс-дисциплинарных исследований.

Ключевые слова: *семантический веб, графы знаний, онтологическое моделирование, семантическое пространство научных знаний, фундаментальные геологические исследования.*

ВВЕДЕНИЕ

Современная геология как фундаментальная наука переживает этап цифровой трансформации, характеризующийся большим ростом объемов и разнообразия данных. Этот процесс сопровождается фундаментальным парадоксом: несмотря на увеличение объема доступной информации, ее разрозненность и семантическая несовместимость существенно ограничивают возможности генерации новых комплексных гипотез для исследования. Существует проблема поиска и обработки полученных количественных данных в различных областях знаний. Ученый-исследователь вынужден тратить до 70–80% рабочего времени на поиск информации, приведение ее к единому формату и согласование данных, а не на их анализ и научную интерпретацию.

Проблема интеграции разнородных, разного масштаба и междисциплинарных данных является одной из ключевых в области наук о Земле. Ее актуальность и сложность подчеркиваются в фундаментальных работах, охватывающих последнее десятилетие. Так, еще в 2013 г. Крис Мэттманн в своей программной статье в Nature [1] обозначил необходимость создания автоматизированных методов для работы с тысячами форматов файлов и бесшовной интеграции алгоритмов. Развитие семантических технологий, подробно рассмотренное Паскалем Хитцлером [2, 3], предлагает мощный аппарат для решения этих задач через онтологии и графы знаний. Наконец, практическая реализация этих подходов на

национальном уровне продемонстрирована в таких проектах, как Геохронологический атлас-справочник (ГАС) ВСЕГЕИ [4], в котором решается задача в рамках централизованной системы с жесткой экспертной оценкой. В настоящей работе предложен подход, основанный на технологиях семантического веба (СВ), для решения обозначенных задач.

Технологии СВ предлагают принципиально новый подход к решению этой проблемы, позволяя преодолеть «семантический барьер» в геологических исследованиях. Особую актуальность данное направление приобретает в контексте глобальной научной инициативы FAIR-принципов [5] (Findable, Accessible, Interoperable, Reusable), направленной на обеспечение обнаружения, доступности, совместимости и многократного использования научных данных.

Фундаментальные геологические исследования характеризуются разнообразием используемых данных, таких как литостратиграфические колонки и палеонтологические определения; результаты петрографических и минералогических исследований; геохимические и изотопные анализы горных пород; геофизические данные различных методов (сейсмологии, гравиметрии, магнитотеллурического зондирования); структурные наблюдения и тектонические построения; данные дистанционного зондирования и цифровые модели рельефа; результаты экспериментальных исследований и математического моделирования.

Проблематика больших данных (Big Data) в геологии имеет выраженную предметную специфику. Наиболее ярко характеристики больших данных проявляются в области геохимии, где можно выделить следующие аспекты концепции Five-V [6, 7].

Variety (разнообразие) – геохимические данные существуют в принципиально различных форматах: табличные результаты физико-химических методов исследования, структурные данные, изотопные соотношения, пространственные распределения химических элементов, что создает значительную сложность в их интеграции и совместного анализа.

Value (ценность) - скрытые закономерности в геохимических данных имеют фундаментальное значение для понимания процессов петрогенезиса и рудообразования, геодинамических процессов, однако их изучение требует комплексного анализа разнородных данных.

Veracity (достоверность) – большое разнообразие аналитических методов, имеющих отличные друг от друга стандарты методических измерений, создают проблемы при интеграции полученных из разных лабораторий количественных данных, что требует учета их метрологических характеристик.

Velocity (скорость) – современные аналитические комплексы (ICP-MS, LA-ICP-MS, электронные сканирующие микроскопы) генерируют непрерывные потоки данных, требующих оперативной обработки и интерпретации прежде всего для публикации результатов в виде научных статей.

Volume (объем) – накопленные массивы аналитической информации достигают значительных объемов, создавая вызовы для хранения, индексации и эффективного поиска релевантных данных.

Актуальным становится требование сохранения научной автономии исследовательских коллективов в условиях необходимости интеграции данных. Каждое учреждение (геологический институт, университет, производственная организация) должно сохранять возможность управления собственными данными и знаниями, при этом обеспечивая их семантическую интероперабельность для совместных исследовательских проектов. Это требование приобретает особую значимость в контексте распределенного характера геологических исследований, когда данные собираются различными организациями в разных регионах, в разное время и для разных предметных областей.

Отметим, что существующие системы управления геологическими данными имеют фундаментальные ограничения, не позволяющие в полной мере решать перечисленные проблемы:

- жесткость схем реляционных баз данных не соответствует эволюционирующему характеру геологических знаний и постоянно уточняющимся классификациям;
- отсутствие семантической интеграции между различными геологическими дисциплинами приводит к созданию независимых, трудно интегрируемых «информационных силосов»;
- неспособность поддерживать сложные виды обработки информации для выявления скрытых взаимосвязей и генерации нового знания;

- ограниченность средств формализации сложных геологических концепций и теорий, имеющих множество исключений и частных случаев.

Для решения этих задач в настоящей работе предложено ориентироваться на технологический стек СВ, который используется при реализации трех ключевых направлений:

- 1) представление разнородных геологических данных в виде единого графа знаний [8] на основе разработанных онтологий, что обеспечивает их семантическую согласованность;
- 2) обеспечение федеративного доступа к распределенным графам знаний через стандартизированные SPARQL-эндпоинты, позволяющие интегрировать данные без их централизации;
- 3) реализация целевых приложений, использующих семантически интегрированные данные для решения конкретных научных задач, относящихся к многомерному статистическому анализу геохимических данных (пример 1), визуализации структурно-тектонической информации (пример 2) и идентификации моделей динамики береговой зоны (пример 3).

Технологии СВ обеспечивают уникальные возможности, принципиально отличающиеся от традиционных подходов, а именно:

- онтологическое моделирование позволяет формализовать сложные геологические концепции (стратиграфические подразделения, тектонические элементы), создавая единую семантическую основу для интеграции знаний из различных дисциплин;
- гибкость RDF-формата обеспечивает адаптацию к развивающимся научным парадигмам и методикам исследований без необходимости перманентного перепроектирования всей системы хранения данных;
- гибкое управление данными в соответствии с принципами графов знаний [8] – возможность откладывать строгие формализации на более поздние этапы исследования. Эмерджентность знаний – это проявление новых закономерностей по мере интеграции данных из различных источников. Преобразование форматов данных как преобразование графов через SPARQL-запросы.

Технологии СВ совместимы не только с процедурами многомерного статистического анализа данных и большими языковыми моделями, но и с механизмами логического вывода, позволяющими не просто находить различного рода корреляции, а генерировать новое знание – автоматически идентифицировать геодинамическую обстановку формирования горной породы на основе комплекса геохимических и минералогических признаков, формализованных в онтологии или базе знаний. Архитектура открытых связанных данных (Linked Data), специальный вариант СВ, поддерживает создание распределенной сети геологических знаний с сохранением автономии участников, что соответствует организационной структуре современной науки.

Целью настоящего исследования была разработка методологии применения технологий СВ для поддержки фундаментальных геологических исследований, обеспечивающей семантическую интеграцию распределенных геологических знаний и переход от фрагментированных информационных массивов к целостному семантическому пространству научных знаний.

В задачи исследования входило:

- 1) разработка онтологических моделей для представления фундаментальных геологических знаний, охватывающих как основные дисциплины, так и конкретные научные задачи, и их взаимосвязи;
- 2) создание механизмов семантической интеграции распределенных данных исследовательских организаций с сохранением их автономии;
- 3) разработка методов интеграции с механизмами логического вывода и системами, основанными на формализованных знаниях, для решения конкретных научных задач в различных областях фундаментальной геологии;
- 4) апробация разработанного подхода на практических примерах (интерпретация геохимических данных, естественно-языковая система доступа к данным о геологических разломах, идентификация модели изменения береговой зоны водохранилищ);
- 5) оценка эффективности предложенных решений и их соответствия FAIR-принципам.

Ниже рассмотрены три примера решения задач с помощью предложенного онтологического подхода.

Пример 1: Анализ геохимических данных для оценки уровня загрязнения территории (геохимия, геоэкология)

Исследование эколого-геохимического состояния различных природных территорий требует интеграции данных об измеренных концентрациях химических элементов, результатах мониторинга загрязнений, различных нормативных показателей, а также информации о наличии природных или техногенных источников поступления загрязняющих компонентов. Отсутствие единой семантической модели затрудняет комплексную оценку экологического состояния территории и выявление возможных источников загрязнения, особенно когда данные поступают из различных ведомств и исследовательских групп.

Пример 2: Информационная система о геологических разломах (сейсмология, структурная геология)

Необходима интеграция разрозненных данных о разрывных нарушениях: кинематические характеристики, возраст образования, современная активность, связь с сейсмичностью, геофизические аномалии. Отсутствие единой семантической модели препятствует созданию целостной картины тектонического строения регионов и адекватной оценке сейсмической опасности.

Пример 3: Исследование береговой зоны (инженерная геология)

Требуется интеграция данных инженерно-геологических изысканий, гидрологических наблюдений, мониторинга береговых процессов. Разрозненность этих данных, хранящихся в различных форматах и у разных организаций, затрудняет создание комплексных моделей и прогнозирование развития береговой зоны, что критически важно для обеспечения безопасности прибрежной инфраструктуры.

Таким образом, предлагаемый онтологический подход и архитектура семантической интеграции позволяют преодолеть традиционные ограничения в работе с геологическими данными. Научная новизна исследования заключается в разработке комплексной онтологической модели, охватывающей междисциплинарные связи между геохимией, геоэкологией, инженерной геологией и тектоникой, что отсутствует в существующих отраслевых решениях. В отличие от подходов, ориентированных на отдельные типы данных, предлагаемая методология

обеспечивает сквозную семантическую интеграцию, позволяя выявлять кросс-дисциплинарные закономерности.

Онтологический подход и графы знаний не просто решают проблему интеграции данных, но и становятся новым инструментом научного познания, позволяя формулировать и проверять гипотезы, которые были недоступны при работе с разрозненными информационными массивами. Это создает новые перспективы для фундаментальных открытий через выявление ранее скрытых взаимосвязей и закономерностей.

Структура статьи отражает логику исследования: последующие разделы представляют разработанную онтологическую модель, архитектуру системы, примеры реализации для различных геологических задач, оценку эффективности и анализ перспектив развития подхода.

МЕТОДОЛОГИЯ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ СЕМАНТИЧЕСКОГО ВЕБА

Разработанная методология представляет собой структурированный подход к семантической интеграции геологических данных, состоящий из пяти последовательных этапов, обеспечивающих переход от разрозненных информационных массивов к целостному семантическому пространству научных знаний.

Этап 1. Доменный анализ и идентификация семантических барьеров

Начальный этап предполагает комплексный анализ предметной области с выявлением ключевых семантических разрывов. Проводится инвентаризация источников данных, анализируются используемые классификации и таксономии, идентифицируются типичные сценарии исследовательских задач. Особое внимание уделяется выявлению неявных знаний и контекстных зависимостей, характерных для геологических дисциплин.

Этап 2. Онтологическое моделирование

На основании результатов доменного анализа разрабатывается модульная онтологическая модель, сочетающая общие геологические концепции со специализированными доменными расширениями. Модель строится по принципу се-

мантической совместимости с существующими стандартами и онтологиями. Формализуются классы, свойства, отношения и ограничения, обеспечивающие адекватное представление предметной области.

Этап 3. Трансформация данных в граф знаний

На данном этапе происходит преобразование разнородных геологических данных в RDF-формат с сохранением семантической целостности. Разрабатываются правила картирования для различных форматов исходных данных, обеспечивающие корректное отображение в термины онтологии. Реализуются механизмы контроля качества и верификации преобразованных данных.

Этап 4. Развертывание семантической инфраструктуры

В рамках этого этапа создается архитектура доступа к семантически интегрированным данным, которая включает SPARQL-эндпоинты, специализированные API и пользовательские интерфейсы. Обеспечивается поддержка федеративных запросов к распределенным источникам данных. Разрабатываются механизмы кэширования и оптимизации производительности для работы с крупномасштабными графами знаний.

Этап 5. Интеграция с исследовательскими процедурами

Завершающий этап направлен на внедрение семантических технологий в существующие исследовательские процедуры. Разрабатываются интерфейсы для интеграции с аналитическими инструментами, системами визуализации и средствами статистического анализа. Создаются механизмы логического вывода и семантического обогащения данных.

Сквозные принципы методологии:

- итеративный характер разработки с непрерывной валидацией. Ориентация на соответствие FAIR-принципам;
- поддержка эволюции онтологических моделей;
- обеспечение семантической интероперабельности;
- Сохранение автономии источников данных.

Данная методология обеспечивает системный подход к преодолению семантических барьеров в геологических исследованиях и созданию инфраструктуры для генерации нового знания на основе интегрированных разнородных данных.

Рассмотрим три примера решения геологических задач с использованием предложенного онтологического подхода.

ПРИМЕР 1: АНАЛИЗ ГЕОХИМИЧЕСКИХ ДАННЫХ ДЛЯ ОЦЕНКИ УРОВНЯ ЗАГРЯЗНЕНИЯ ПРИРОДНЫХ ТЕРРИТОРИЙ

Рассмотрим три примера решения геологических задач при помощи предложенного онтологического подхода.

Исследование эколого-геохимического состояния территорий требует комплексной интеграции данных об измеренных концентрациях химических элементов, методах анализа вещества, точках отбора образцов и их геологическом контексте [9–13]. Технологии СВ предоставляют принципиально новый подход к решению этой задачи через создание единого графа знаний на основе разработанной онтологии.

Разработанная онтологическая модель охватывает все аспекты геохимических исследований: от базовых классов образцов и измерений до сложных систем контроля качества и пространственного контекста. Ключевые классы включают `pt:Sample`, `pt:GeochemicalSample`, `pt:Measurement` со специализированными подклассами для элементных и суммарных измерений, что обеспечивает семантически богатое представление данных.

Приведенный ниже пример конкретного образца почвы в формате RDF/Turtle демонстрирует полноту представления информации:

```
pt:AGS-0085 a pt:GeochemicalSample ;  
  rdfs:label "AGS-0085" ;  
  wgs:lat 53.21610 ;  
  wgs:long 107.4481 ;  
  pt:sampleDepth 0.5 ;  
  pt:sampleDate "2023-07-15"^^xsd:date ;  
  pt:location pt:Ольхон_Хужир ;  
  pt:formation pt:Палеозойские_граниты ;
```

```
pt:hasMeasurement [  
  a pt:ElementMeasurement ;  
  pt:measuredElement mt:As ;  
  pt:measurementValue 3.4 ;  
  pt:measurementUncertainty 0.2 ;  
  pt:detectionLimit 0.1 ;  
  pt:analyticalMethod "ICP-MS" ;  
  pt:qualityFlag pt:Good  
] .
```

Для извлечения и анализа данных разработан специализированный модуль `kgquery.ext`, который предоставляет исследователям удобный интерфейс доступа к графу знаний. Модуль абстрагирует сложность SPARQL-запросов, позволяя формулировать запросы на концептуальном уровне, близком к предметной области геологии.

Типичный SPARQL-запрос для анализа распределения мышьяка в почвах на территории острова Ольхон демонстрирует высокую эффективность семантического подхода:

```
PREFIX pt: <http://example.org/ontology/project#>  
PREFIX mt: <http://example.org/ontology/mineral#>  
PREFIX wgs: <http://www.w3.org/2003/01/geo/wgs84_pos#>  
  
SELECT ?sample ?lat ?long ?value ?method ?quality ?formation  
WHERE {  
  ?sample a pt:GeochemicalSample ;  
    wgs:lat ?lat ;  
    wgs:long ?long ;  
    pt:location/pt:formation ?formation ;  
    pt:hasMeasurement ?measurement .  
  
  ?measurement pt:measuredElement mt:As ;  
    pt:measurementValue ?value ;  
    pt:analyticalMethod ?method ;  
    pt:qualityFlag ?quality .
```

```
FILTER(?value > 2.0 && ?quality IN (pt:Excellent, pt:Good))  
}  
ORDER BY DESC(?value)
```

Особенностью такого подхода является возможность семантического обогащения данных через механизмы логического вывода. Система автоматически классифицирует пробы как «аномальные» на основе формализованных правил, учитывающих абсолютные концентрации элементов и их соотношения, а также геологический контекст.

Полученные количественные данные обрабатываются с применением методов статистического анализа, включая метод главных компонент (РСА), для выявления скрытых закономерностей в распределении химических элементов и межэлементных взаимосвязей. Пространственная интерполяция на основе регрессионных моделей позволяет построить карты поверхностного распределения токсичных элементов, отражающие различные уровни загрязнений окружающей среды с учетом физических параметров территории.

Система контроля качества, реализованная в онтологии через классы `pt:DataQuality` и `pt:QualityLevel`, обеспечивает достоверность аналитических результатов. Это особенно важно при интеграции данных из различных лабораторий, использующих для измерения разные аналитические методы с отличными друг от друга метрологическими характеристиками.

Интеграция с внешними источниками данных, такими как цифровые модели рельефа SRTM, позволяет учитывать влияние геоморфологических факторов на распределение загрязняющих веществ. Автоматическое вычисление производных параметров (уклонов, экспозиций) расширяет аналитические возможности системы.

Предложенный подход успешно апробирован на примере количественной оценки концентраций мышьяка в почвах острова Ольхон, показав возможность выявления пространственных закономерностей и аномалий в распределении этого элемента. Семантическая интеграция данных позволила учесть не только химические концентрации, но и геоморфологические характеристики территории (см. рис. 1).

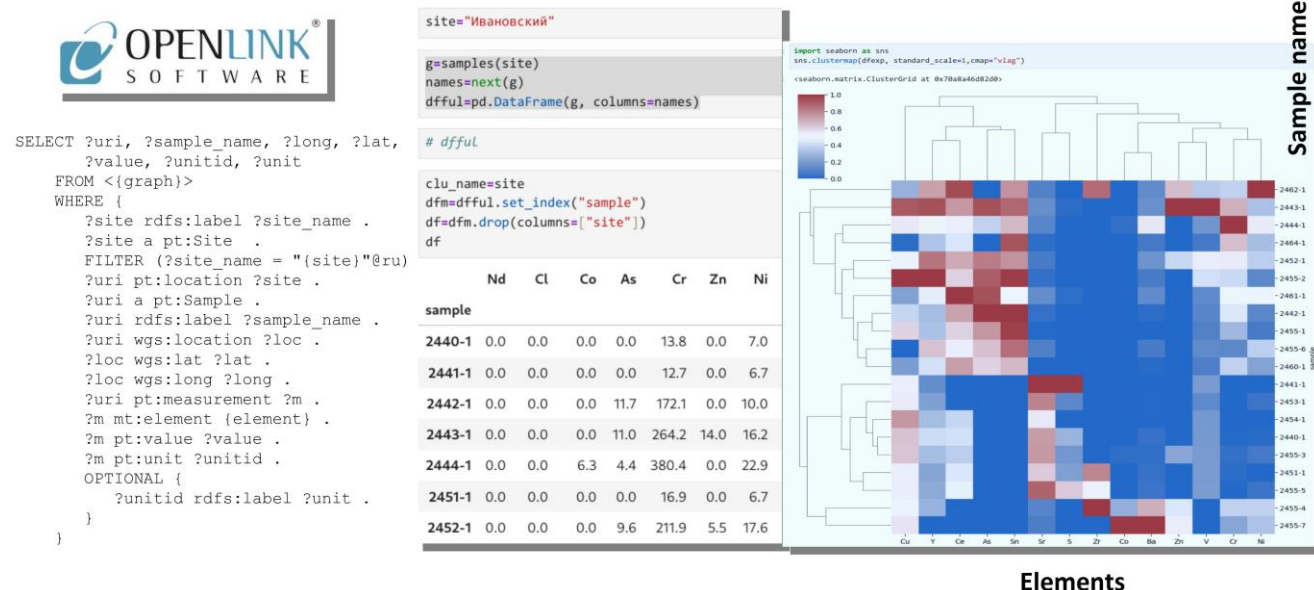


Рис. 1. Корреляционный анализ геохимических данных для почв острова Ольхон

Разработанный алгоритм демонстрирует переход от фрагментированных массивов геохимических данных к целостному семантическому пространству, где каждый новый факт автоматически обогащает общее знание и становится доступным для последующего анализа и открытия новых закономерностей в распределении загрязняющих веществ.

ПРИМЕР 2: ИНФОРМАЦИОННАЯ СИСТЕМА СЕМАНТИЧЕСКИХ ДАННЫХ О ГЕОЛОГИЧЕСКИХ РАЗЛОМАХ

Исходные данные о геологических разломах хранились в реляционной базе DBF-формата, содержащей 72 поля [14]. Эта структура включала идентификатор разлома, географическое название, различные характеристики с соответствующими уточнениями, уровень сейсмической активности, имя исследователя-компилятора и дату уточнения. Номера записей связывали записи базы данных с пространственными объектами слоя разломов. Важным ограничением формата DBF являлось то, что имена полей не могли быть длиннее десяти символов, должны были быть заглавными, а количество полей не могло превышать 255.

Идентификаторы атрибутов с общими префиксами определяли одно значение с уточнением. Например, атрибуты "Depth" и "Length" разлома, измеряемые в километрах, уточнялись атрибутом "quality" с суффиксом "_Q" в конце. Содер-

жание таблицы базы данных разломов было заполнено разреженно: многие атрибуты содержали нулевые значения. Поля “Geomorphol...” и некоторые другие заполнялись значениями из predetermined наборов. Согласованность данных контролировалась алгоритмически с помощью модулей расширения QGIS. Эта плоская структура и формат представления были предназначены для легкого доступа и реализации с помощью стандартных инструментов реляционных баз данных, но общие нарушения стандартных нормальных форм вынуждали разработчика реализовывать подпрограммы контроля содержания записи в дополнение к команде UPDATE DML.

Процесс преобразования табличных данных в граф знаний начинался с определения T-Box, включающего базовый класс `geo:Fault` и специализированные классы измерений: `geo:Strike`, `geo:Slip`, `geo:Rate`, `geo:Feature`, `geo:Activity`, `geo:Event`, которые являются подклассами `geo:Measurement`. Все записи о разломах последовательно конвертировали в RDF-триплеты с присвоением соответствующих классов, формируя таким образом A-Box (ящик экземпляров) – семантически структурированный граф знаний экземпляров разломов [15].

Универсальные свойства онтологии организованы вокруг объектного свойства `geo:hasMeasurement`, связывающего разломы с их измерениями, и набора дата-свойств для описания характеристик: `geo:value` (значение), `geo:type` (тип), `geo:quality` (качество), `geo:unit` (единица измерения) и `geo:grade` (балльная оценка). Дополнительные свойства `geo:id`, `geo:name` и `geo:location` описывают базовые атрибуты разломов, а свойства `geo:compiledBy`, `geo:compilerName` и `geo:date` фиксируют метаданные о компиляторах информации, обеспечивая полноту и прослеживаемость данных [15].

Для организации пространства имен использовали префикс `geo:` <<http://crust.irk.ru/ontologies/fault/terms/1.0/>>. Техническая реализация процесса осуществлена с помощью специализированной Python-программы, которая загружала исходные DBF-файлы и генерировала OWL2 XML-документы, после чего полученные онтологии T-Box и A-Box проходили визуальную верификацию в среде Protégé с последующей конвертацией в формат Turtle и загрузкой в семантические серверы GraphDB и Jena для последующего использования.

Ниже представлен пример преобразованного объекта – разлом «Главный Саянский» (RUAF_1), который демонстрирует насыщенность семантической модели:

```
<http://.../RUAF_1> a geo:Fault ;
  geo:id "RUAF_1" ;
  geo:name "Главный Саянский" ;
  geo:location "Вдоль юго-западной границы Сибирской платформы" ;

# Strike measurements
geo:hasMeasurement [
  a geo:Strike ;
  geo:value "301" ;
  geo:quality "ЛС"
];

# .....
# Activity
geo:hasMeasurement [
  a geo:Activity ;
  geo:value "аномально высокая" ;
  geo:grade 32
], [
  a geo:Activity ;
  geo:type "LastActivationAge" ;
  geo:value "Голоценовое" ;
  geo:grade 2
];

# Events
geo:hasMeasurement [
  a geo:Event ;
  geo:type "PotentialMagnitude" ;
  geo:value "8" ;
  geo:quality "ЛС"
];
```

```
# Compiler
geo:compiledBy [
  a geo:Compiler ;
  geo:compilerName "Лунина О.В." ;
  geo:date "2011-02-23"^^xsd:date
] .
```

АРХИТЕКТУРА ПРИЛОЖЕНИЯ WEB-GIS

Программа приложения Web-GIS реализована как одностраничное Web 2.0 приложение, построенное на основе React-компонентов с управлением состоянием через Redux. Данная архитектура обеспечивает эффективное управление интерфейсом и состоянием приложения, позволяя динамически обновлять содержимое виджетов при изменении набора визуализируемых геологических разломов.

Для визуализации картографических данных (рис. 2.) используется React-обертка библиотеки leaflet.js, которая предоставляет возможности отображения интерактивных графических объектов на топологической основе. Система поддерживает различные источники картографических подложек, включая OpenStreetMap в качестве базового варианта. Контурные разломов были преобразованы из формата KML в JSON-структуры, совместимые с leaflet, и могут быть загружены приложением по мере необходимости с веб-сервера. Важной особенностью этой системы является сопоставление контуров через International Resource Identifiers (IRI) разломов, что обеспечивает семантическую связь между графическими представлениями и данными в графе знаний.

Доступ к семантическим данным организован через интерпретацию результатов выполнения SPARQL-запросов к конечной точке СБ-сервера. Пользовательский интерфейс включает форму для задания условий фильтрации, которая автоматически модифицирует стандартные SPARQL-запросы, добавляя соответствующие ограничения. Для целей отладки и разработки предусмотрена возможность выполнения свободных SPARQL-запросов с визуализацией результатов на карте. Отображение атрибутов разломов реализовано через формы с RDFa-разметкой, что обеспечивает семантически обогащенное представление данных.

```
PREFIX geo: <http://crust.irk.ru/ontologies/fault/terms/1.0/>
SELECT ?fault ?name ?magnitude
WHERE {
```

```
?fault a geo:Fault ;
    geo:name ?name ;
    geo:hasMeasurement [
        a geo:Event ;
        geo:type "PotentialMagnitude" ;
```

```
    geo:value ?magnitude  
  ].  
  FILTER(xsd:decimal(?magnitude) > 7.0)  
}
```

ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ ИНФОРМАЦИОННОЙ СИСТЕМЫ

Система предоставляет естественно-языковой интерфейс доступа к данным графа знаний, автоматически преобразуя запросы на естественном языке в формализованные SPARQL-запросы. Это позволяет исследователям, не владеющим семантическими технологиями, формулировать сложные запросы, аналогичные приведенным в примере 1, а также запросы о разломах с голоценовой активизацией и идентификации левосдвиговых разломов длиной более 500 км с аномально высокой активностью.

По умолчанию система отображает все разломы из графа знаний, что соответствует результату SPARQL-запроса, выбирающего все объекты класса `geo:Fault`. Пользовательский интерфейс организован таким образом, чтобы обеспечить интуитивно понятное взаимодействие с данными. Он включает топооснову с наложенными контурами разломов, панель фильтрации для задания критериев отбора, всплывающие подсказки с названиями разломов при наведении курсора, а также детализированную форму отображения атрибутов выбранного разлома.

Система успешно преодолевает «семантический барьер» в работе с данными о разломах, обеспечивая гибкое представление неполных и эволюционирующих данных, соответствующих принципам Linked Open Data.

Ключевым достижением является создание естественно-языкового интерфейса, доступного для пользователей без специализированной подготовки в области семантических технологий. Система обеспечивает глубокую интеграцию пространственных и атрибутивных данных через семантические связи, что открывает новые возможности для комплексного анализа геологической информации.

Полученный опыт подтвердил практическую возможность построения Web-GIS-систем для научных сообществ, поддерживающих постоянное накопление, агрегацию и анализ данных благодаря фундаментальным свойствам графов зна-

ний. Такие системы способны адаптироваться к изменяющимся требованиям исследований и обеспечивать долгосрочную поддержку научной деятельности в области геологии и сейсмологии.

ПРИМЕР 3: СЕМАНТИЧЕСКАЯ ИНТЕГРАЦИЯ ДАННЫХ ДЛЯ ИССЛЕДОВАНИЯ БЕРЕГОВОЙ ЗОНЫ ВОДОХРАНИЛИЩ

Исследование динамики береговой зоны крупных равнинных водохранилищ характеризуется необходимостью интеграции разнородных данных инженерно-геологических изысканий, гидрологических наблюдений и мониторинга береговых процессов [16, 17]. Как отмечено выше, неупорядоченность этих данных затрудняет создание комплексных прогнозных моделей и оценку рисков для прибрежной инфраструктуры. Традиционные подходы, основанные на реляционных базах данных [18, 19] с жесткой схемой (в основном таблицы Excel), неспособны эффективно аккумулировать эволюционирующие классификации типов берегов и многообразие экзогенных геологических процессов, что ограничивает возможности комплексного анализа, мониторинга и прогнозирования.

Для решения обозначенных проблем была разработана онтологическая модель, представляющая береговую зону как единое семантическое пространство, в котором взаимосвязаны природные объекты, геологические процессы и условия их развития (см. рис. 3). Ключевой архитектурной идеей является процессоцентричная модель, где экзогенный процесс выступает активным агентом, меняющим состояние берега.

береговой зоны водохранилищ

```
# Ключевые классы (Class Hierarchy)
:Природный_объект rdf:type owl:Class .
:Берер rdfs:subClassOf :Природный_ОБЪЕКТ .
:Водоем rdfs:subClassOf :Природный_объект .
```

:Водохранилище rdfs:subClassOf :Водоем .

:Участок rdfs:subClassOf :Природный_объект .

:Геологический_процесс rdf:type owl:Class .

:Абразия rdfs:subClassOf :Геологический_процесс .

:Оползень rdfs:subClassOf :Геологический_процесс .

:Карстовый rdfs:subClassOf :Геологический_процесс .

:Тип_берега rdf:type owl:Class .

:Абразионный rdfs:subClassOf :Тип_берега .

:Аккумулятивный rdfs:subClassOf :Тип_берега .

Свойства отношений (Object Properties)

:изменяет rdf:type owl:ObjectProperty ;

 rdfs:domain :Геологический_процесс ;

 rdfs:range :Берег .

:привел_к rdf:type owl:ObjectProperty ;

 rdfs:domain :Геологический_процесс ;

 rdfs:range :Природный_объект .

:является_частью rdf:type owl:ObjectProperty, owl:TransitiveProperty ;

 rdfs:domain :Участок ;

 rdfs:range :Берег .

Предложенный подход приобретает важную практическую значимость в контексте решения задачи идентификации доминирующей модели динамики береговой зоны, эта задача требует комплексного анализа следующих факторов:

- литолого-петрографического состава пород берегового уступа;
- гидродинамического режима водоема;
- морфометрических характеристик берегового склона;
- наличия и степени активности экзогенных геологических процессов.

Для формализации этой задачи в онтологии были введены дополнительные аксиомы и правила логического вывода. В частности, свойство 'доминирую-

щая_модель' связывает участок берега с конкретным типом динамики (абразионная, аккумулятивная, абразионно-оползневая и др.) на основе комплекса диагностических признаков.

Пример SPARQL-запроса для идентификации модели динамики:

```
PREFIX : <http://www.semanticweb.org/moks/ontologies/2023/1/geolprocesses#>
SELECT ?Участок ?МодельДинамики ?Обоснование
WHERE {
    ?Участок :доминирующая_модель ?МодельДинамики .
    ?Участок :имеет_обоснование ?Обоснование .
    FILTER(?МодельДинамики IN (:Абразионная, :Абразионно-Оползневая, :Аккумулятивная))
}
```

А-Вох-онтологии. Наполнение онтологии конкретными данными и их интеграция в граф знаний позволили реализовать сценарий семантического запроса для решения конкретной научно-прикладной задачи. В качестве примера рассмотрим задачу выявления всех участков Братского водохранилища, на которых развиваются оползневые процессы, с указанием типа берега, SPARQL-запрос:

```
PREFIX : <http://www.semanticweb.org/moks/ontologies/2023/1/geolprocesses#>
SELECT ?Участок ?ТипБерега ?Процесс
WHERE {
    ?Участок :является_частью/:является_частью? :Береговая_зона_Братского_водохр .
    ?Участок rdf:type ?ТипБерега .
    FILTER(?ТипБерега IN (:Абразионный, :Аккумулятивный))
    ?Процесс rdf:type :Оползень ;
        :изменяет ?Участок .
}
```

Результат выполнения запроса: Система возвращает конкретные участки (например, Хадахан), для которых установлена связь между типом берега (Абразионный) и активным экзогенным процессом (Оползень_Хадахан). Это позволяет исследователю не просто видеть разрозненные данные, а идентифицировать зоны повышенного риска.

Таким образом, применение онтологического подхода к задаче интеграции данных о береговой зоне позволило преодолеть традиционную разрозненность информации. Построенный граф знаний обеспечил семантическую связность данных, что предоставило возможность формулировать комплексные запросы, расширяющие традиционные дисциплинарные границы между инженерной геологией и геоморфологией. Важным результатом является возможность идентификации доминирующей модели динамики береговой зоны, что критически важно для прогнозирования развития береговых процессов и обеспечения безопасности прибрежной инфраструктуры. Данный пример демонстрирует, каким образом технологии СВ превращают разрозненные массивы данных в инструмент для поддержки принятия решений при оценке рисков и планировании хозяйственной деятельности в береговой зоне.

ТЕХНИЧЕСКАЯ РЕАЛИЗАЦИЯ МЕТОДОЛОГИИ

Реализация предложенной методологии основана на современном технологическом стеке, обеспечивающем полный цикл работы с семантическими данными. Архитектура системы построена по многоуровневому принципу, где каждый уровень решает определенный класс задач. На уровне хранения и обработки данных используются графовые базы данных GraphDB, Apache Jena Fuseki, Openlink Virtuoso, которые обеспечивают надежное хранение RDF-данных и предоставление SPARQL-эндпоинтов для доступа к семантически структурированной информации.

Для представления данных применяются стандартизированные форматы: RDF/Turtle для онтологий, OWL2 XML для обмена метаданными между системами, JSON-LD для веб-интеграции. Серверная часть реализована на Python 3.x с использованием специализированных библиотек RDFLib и SPARQLWrapper, которые обеспечивают выполнение ETL-процессов по преобразованию геологических данных в семантические форматы.

Разработка онтологических моделей осуществляется с помощью инструментария Protégé, который служит основной средой для визуального проектирования и валидации T-Box-онтологий. Для выполнения запросов к данным используется язык SPARQL 1.1, стандартно поддерживающий федеративные запросы к распределенным источникам информации.

Пользовательские интерфейсы веб-приложений построены на основе React.js и React-оберток библиотек, обеспечивающих отображение табличных, а также пространственных данных на топологических основах. Семантическое обогащение веб-интерфейсов достигается за счет применения RDFS-разметки, а прямой доступ к графам знаний предоставляется через специализированные SPARQL-интерфейсы.

Обеспечение соответствия FAIR-принципам [5] является ключевым аспектом технической реализации. Для реализации принципа находимости (Findable) всем объектам присваиваются устойчивые URI через зарегистрированные пространства имен, а онтологии регистрируются в специализированных репозиториях. Доступность (Accessible) обеспечивается через открытые SPARQL-эндпоинты с современными механизмами аутентификации и поддержкой стандартных форматов данных.

Совместимость (Interoperable) достигается за счет использования стандартных онтологий в качестве основы и поддержки кросс-доменных запросов. Многообразие (Reusable) обеспечивается модульной архитектурой онтологий с четким разделением доменов и детальными метаданными о происхождении и качестве данных. Все компоненты системы лицензируются в соответствии с открытыми стандартами, что способствует их широкому распространению в научном сообществе.

СРАВНЕНИЕ С АНАЛОГАМИ

Проведенный анализ современных подходов к интеграции геологических данных выявил несколько ключевых направлений, каждое из которых обладает определенными преимуществами и ограничениями в контексте решения задач семантической интеграции разнородных геологических данных.

В области стандартизации геологических данных сформировались два принципиально различных подхода. GeoSciML [20] представляет собой практико-ориентированный стандарт обмена геологической информацией, основанный на языке GML (Geography Markup Language). Этот стандарт, разработанный Комиссией по управлению и применению геонаучной информации (CGI), обеспечивает

синтаксическую интероперабельность данных через интеграцию с OGC-сервисами, что делает его удобным инструментом для организаций, работающих в рамках существующей геоинформационной инфраструктуры. Однако GeoSciML не решает задач глубокой семантической интеграции и ограничивается стандартизацией форматов обмена.

Альтернативный подход представлен онтологией SWEET (Semantic Web for Earth and Environmental Terminology) [21], разработанной NASA для семантической разметки данных о Земле и окружающей среде. SWEET обеспечивает формальное описание понятий и их взаимосвязей через ортогональные концепции (пространство, время, физические свойства) и синтетические знания (явления, события). Это позволяет автоматически расширять поисковые запросы синонимами и связанными понятиями, значительно повышая релевантность результатов. Однако SWEET остается преимущественно теоретическим конструктом, слабо адаптированным для работы с реальными полевыми данными и решением конкретных прикладных задач.

Международные инициативы в области интеграции геологических данных демонстрируют различные стратегические подходы. OneGeology [22] представляет собой глобальную распределенную систему создания цифровой геологической карты планеты с участием 117 стран. Проект успешно реализует модель распределенной архитектуры, где данные сохраняются у организаций-владельцев, которые обеспечивают их актуальность и контроль. Однако фокус OneGeology ограничен преимущественно картографическими данными, что не позволяет решать задачи глубокой семантической интеграции разнородных геологических данных из различных дисциплин.

Более комплексные подходы демонстрируют проекты EarthCube (США) [23] и EPOS (Европейский союз). EarthCube ориентирован на создание киберинфраструктуры в области «Науки о Земле» через построение «системы систем» с распределенным управлением по аналогии с интернетом. Проект сталкивается с комплексными вызовами, включая концептуальные (время, масштаб), социально-культурные (мотивация, сотрудничество) и технические (интероперабельность) аспекты. European Plate Observing System (EPOS) [24] представляет собой панъевропейскую инициативу по интеграции исследовательских инфраструктур в

области наук о Земле, объединяя более 250 инфраструктур из 18 стран. Особое внимание в EPOS уделяется мониторингу тектонических процессов и связанных с ними природных опасностей, включая специализированные решения для работы в арктическом регионе.

В российской практике преобладает парадигма централизованных систем, характерным примером которой является ГАС ВСЕГЕИ [4]. Данный проект демонстрирует высокую эффективность в решении узкоспециализированной задачи унификации геохронологической информации для территории России. Ключевыми преимуществами системы являются строгий контроль качества через экспертную валидацию (отбраковка ~20% исходных определений) и интеграция с серийными легендами Государственных геологических карт. Централизованная архитектура через аккредитованный Центр изотопных исследований ВСЕГЕИ обеспечивает единые стандарты качества и методологии. Однако такая модель создает существенные ограничения для интеграции разнородных данных из смежных дисциплин и обладает высоким «порогом входа» для независимых исследователей.

Проведенный анализ выявил устойчивый разрыв между теоретическими возможностями семантических технологий (SWEET) и их практической реализацией в существующих системах (GeoSciML, OneGeology, EarthCube, EPOS, ГАС). Собственно, этот разрыв и был преодолен в данной работе через разработку методологии, которая синтезирует семантическую строгость онтологических подходов с практической ориентированностью существующих стандартов. Предложенное решение обеспечивает гибкость распределенных архитектур и позволяет интегрировать разнородные алгоритмы и данные для решения конкретных междисциплинарных научных задач в геологии.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Представленный подход доказал свою практическую ценность в различных геологических дисциплинах, что подтверждается успешным применением в геохимии, тектонике и инженерной геологии. Возможность кросс-дисциплинарного использования свидетельствует о продуманной архитектуре, способной работать с разнородными геологическими данными.

Использование стандартизированных технологий СВ (RML, SPARQL) обеспечивает совместимость систем и снижает порог внедрения за счет повторно используемых компонентов.

Однако следует учитывать существенные ограничения. Высокие требования к экспертизе в области СВ создают серьезный барьер для внедрения, особенно в геологических коллективах без соответствующей подготовки. Разработанная методология проверена на данных умеренного объема, тогда как ее масштабируемость для работы с петабайтными массивами геологической информации требует дополнительного изучения.

Эффективность предложенного подхода сильно зависит от исходного качества данных. Как показала практика, неоднородность единиц измерений, методов анализа и таксономий требует значительных усилий по их нормализации. Особенно критичен этап онтологического моделирования: ошибки проектирования на этой стадии негативно влияют на все последующие этапы.

Сложности интеграции с унаследованным геологическим ПО и специализированными инструментами также заслуживают внимания. Отсутствие готовых решений для подключения требует разработки собственных адаптеров, что увеличивает сложность и стоимость сопровождения.

Несмотря на отмеченные ограничения, методология успешно решает ключевые задачи семантической интеграции и открывает новые возможности для исследований благодаря федеративным запросам и логическому выводу. Итеративный процесс разработки с непрерывной валидацией обеспечивает необходимую гибкость для адаптации к конкретным предметным областям.

Перспективные направления развития включают создание масштабируемых архитектур для работы с большими данными, разработку учебных материалов для геологов и стандартизацию интеграции с распространенным геологическим ПО.

Предложенная методология применима в исследовательских проектах средней сложности, требующих интеграции разнородных данных. Для промышленного использования понадобятся дополнительные проверки и оптимизация производительности.

ЗАКЛЮЧЕНИЕ

Разработана и апробирована методология применения технологий СВ для поддержки фундаментальных геологических исследований. Основным результатом работы стало создание части комплексной онтологической модели, обеспечивающей семантическую интеграцию разнородных геологических данных из различных дисциплин: геохимии, тектоники и инженерной геологии. Практическая реализация подхода продемонстрирована на трех кейсах: анализе геохимических данных для оценки уровня загрязнения территории, создании информационной системы о разломах и изменении береговой зоны. Представленная архитектура на основе графов знаний и SPARQL-эндпоинтов подтвердила свою эффективность в преодолении «семантического барьера» и обеспечении соответствия FAIR-принципам.

Полученный опыт эксплуатации приложений, разработанных на основе предложенного подхода, выявил как значительные преимущества, так и определенные ограничения. К достоинствам относятся гибкость онтологической модели, способность адаптироваться к эволюционирующим научным парадигмам, поддержка распределенной архитектуры с сохранением автономии исследовательских коллективов, а также возможность создания подсистем естественно-языковых запросов на основе интеграции онтологических моделей и больших языковых моделей (LLM). Однако отмечена высокая сложность начального этапа разработки онтологий, необходимость обучения исследователей работе с семантическими технологиями, а также вычислительная нагрузка при обработке крупных графов знаний. Эти факторы могут ограничивать широкое внедрение подхода в геологическом сообществе.

Перспективы развития исследования связаны с несколькими направлениями. Во-первых, это расширение онтологической модели для охвата дополнительных геологических дисциплин и интеграции с международными стандартами типа GeoSciML. Во-вторых, разработка специализированных инструментов автоматического преобразования традиционных геологических данных в RDF-формат. Особый интерес представляет создание когнитивных систем поддержки принятия решений на базе разработанной платформы, которые будут объединять се-

мантически интегрированные данные, механизмы логического вывода и предиктивную аналитику для решения таких сложных задач, как прогнозирование развития деградационных процессов на определенных участках береговой зоны и экспертная оценка хозяйственного использования этой зоны, оценка сейсмической опасности и оптимизация работ по экологическому мониторингу. Таким образом, дальнейшая работа будет направлена на создание открытой платформы для семантической интеграции геологических знаний с вовлечением широкого круга заинтересованных лиц.

Благодарности

Исследование выполнено при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках гранта № 075-15-2024-533 на реализацию крупных научных проектов по приоритетным направлениям научно-технологического развития, проект «Фундаментальные исследования Байкальской природной территории на основе системы взаимосвязанных базовых методов, моделей, нейронных сетей и цифровой платформы для экологического мониторинга окружающей среды».

СПИСОК ЛИТЕРАТУРЫ

- 1) *Mattmann C.* A vision for data science // *Nature*. 2013. Vol. 493. P. 473–475. <https://doi.org/10.1038/493473a>
- 2) *Hitzler P., Janowicz K.* Semantic Web for Earth and Environmental Science: Current state and future directions // *Semantic Web*. 2010. Vol. 1, No. 1–2. P. 85–98. <https://doi.org/10.3233/SW-2010-0012>
- 3) *Hitzler P.* A review of the semantic web field // *Communications of the ACM*. 2021. Vol. 64, No. 2. P. 76–83. <https://doi.org/10.1145/3397512>
- 4) *Синькова Е.А., Петров О.В., Ханчук А.И.* Геохронологический атлас-справочник основных структурно-вещественных комплексов России – базовый информационный ресурс для геологической отрасли страны // *Разведка и охрана недр*. 2022. № 90. С. 5–14. https://doi.org/10.52349/0869-7892_2022_90_5-14
- 5) *Wilkinson M.D., Dumontier M., Aalbersberg I.J. et al.* The FAIR Guiding Principles for scientific data management and stewardship // *Scientific Data*. 2016. Vol. 3. P. 160018. <https://doi.org/10.1038/sdata.2016.18>

- 6) *Laney D.* 3D Data Management: Controlling Data Volume, Velocity, and Variety // META Group Research Note. 2001. Vol. 6, No. 70.
URL: <https://diegonogare.net/wp-content/uploads/2020/08/3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- 7) *Gandomi A., Haider M.* Beyond the hype: Big data concepts, methods, and analytics // International Journal of Information Management. 2015. Vol. 35, No. 2. P. 137–144. <https://doi.org/110.1016/j.ijinfomgt.2014.10.007>
- 8) *Hogan A., Blomqvist E., Cochez M., D'Amato C. et al.* Knowledge Graphs // ACM Computing Surveys. 2021. Vol. 54, No. 4. <https://doi.org/110.1145/3447772>
- 9) *Pellinen V.A., Cherkashina T.Y., Gustaitis M.A.* Assessment of metal pollution and subsequent ecological risk in the coastal zone of Olkhon Island, Lake Baikal, Russia // Science of the Total Environment. 2021. Vol. 786. P. 147441. <https://doi.org/110.1016/j.scitotenv.2021.147441>
- 10) *Zerizghi T., Yang Y., Wang W., Zhou Y., Zhang J., Yi Y.* Ecological risk assessment of heavy metal concentrations in sediment and fish of a shallow lake: a case study of Baiyangdian Lake, North China // Environmental Monitoring and Assessment. 2020. Vol. 192. P. 154. <https://doi.org/110.1007/s10661-020-8078-8>
- 11) *Zhang J., Wang K., Yi Q., Zhang T., Shi W., Zhou X.* Transport and partitioning of metals in river networks of a plain area with sedimentary resuspension and implications for downstream lakes // Environmental Pollution. 2022. Vol. 294. P. 118668. <https://doi.org/110.1016/j.envpol.2021.118668>
- 12) *Gopal V., Krishnamurthy R.R., Vignesh R., Nathan C.S. et al.* Assessment of heavy metal contamination in the surface sediments of the Vedaranyam coast, Southern India // Regional Studies in Marine Science. 2023. Vol. 65. P. 103081. <https://doi.org/110.1016/j.rsma.2023.103081>
- 13) *Chubarov V., Cherkashina T., Maltsev A., Chuparina E., Amosova A., Prosekin S.* Investigation of Soils and Pine Needles Using WDXRF and TXRF Techniques for Assessment of the Environmental Pollution of Shelekhov District, Eastern Siberia, by the Aluminum Industry and Heat Power Engineering // Agronomy. 2022. Vol. 12. P. 454. <https://doi.org/110.3390/agronomy12020454>

14) *Lunina O.V.* The digital map of the Pliocene–Quaternary crustal faults in the Southern East Siberia and the adjacent Northern Mongolia // *Geodynamics and Tectonophysics*. 2016. Vol. 7, No. 3. P. 407–434.

<https://doi.org/110.5800/GT-2016-7-3-0215>

15) *Cherkashin E.A., Lunina O.V., Demyanov L.O., Tsygankov A.V.* Web-GIS viewer for active faults data represented as a knowledge graph // *The 4th Scientific-practical Workshop Information Technologies: Algorithms Models, Systems*. September 14, 2021, Irkutsk, Russia / *CEUR Workshop Proceedings*. 2021. Vol. 2984. P. 55–65. URL: <https://ceur-ws.org/Vol-2984/paper8.pdf>

16) *Mazaeva O., Babicheva V., Kozyreva E.* Geomorphological process development under the impact of man-made reservoir operation, a case study: Bratsk reservoir, Baikal-Angara hydroengineering system, Russia // *Bulletin of Engineering Geology and the Environment*. 2019. Vol. 78. P. 4659–4672.

<https://doi.org/110.1007/s10064-018-1428-x>

17) *Овчинников Г.И., Павлов С.Х., Тржцинский Ю.Б.* Изменения геологической среды в зонах влияния ангаро-енисейских водохранилищ. Новосибирск: Наука, 1999. 254 с.

18) *Ranatunga S., Ødegård R.S., Jetlund K., Onstein E.* Use of Semantic Web Technologies to Enhance the Integration and Interoperability of Environmental Geospatial Data: A Framework Based on Ontology-Based Data Access // *ISPRS International Journal of Geo-Information*. 2025. Vol. 14, No. 2. P. 52.

<https://doi.org/110.3390/ijgi14020052>

19) *Moura A.-M., Porto F., Vidal V., Magalhães R.P. et al.* A semantic integration approach to publish and retrieve ecological data // *International Journal of Web Information Systems*. 2015. Vol. 11, No. 1. P. 87–119.

<https://doi.org/110.1108/IJWIS-08-2014-0028>

20) *Marcus S., Tim D.* GeoSciML: Development of a generic GeoScience Markup Language // *Computers & Geosciences*. 2005. Vol. 31, No. 9. P. 1095–1103.

<https://doi.org/10.1016/j.cageo.2004.12.003>

21) *Raskin R., Pan M.* Knowledge representation in the semantic web for Earth and environmental terminology (SWEET) // *Computers & Geosciences*. 2005. Vol. 31, No. 9. P. 1119–1125. <https://doi.org/110.1016/j.cageo.2004.12.004>

22) *Simons B.A., Raymond O., Jackson I., Lee K.* OneGeology – Improving global access to geoscience // The 5th Global Workshop on Digital Soil Mapping. April 10–13, 2012, Sydney, Australia / Digital Soil Assessments and Beyond: Proceedings of CRC Press, 2012. P. 265–275.

23) *Allison M.L., Ahern T., Arctur D. et al.* EarthCube Governance Framework: A Proposal to the Community (Version 1.0) // EarthCube Governance Working Group Technical Report. 2012. 237 p.

24) *Atakan K., Bjerrum L.W., Bungum H. et al.* The European Plate Observing System and the Arctic // Arctic. 2015. Vol. 68, Suppl. 1.
<https://doi.org/10.14430/arctic4446>

SEMANTIC WEB TECHNOLOGIES FOR SUPPORTING FUNDAMENTAL RESEARCH IN GEOLOGY

E. A. Cherkashin¹ [0000-0003-2428-2471], I. V. Bychkov² [0000-0002-1765-0769],
Jin Zhang³ [0000-0003-0009-824X], T. Yu. Cherkashina⁴ [0000-0002-1443-6839],
V. A. Popova⁵ [0000-0002-7764-1995], O. A. Mazaeva⁶ [0000-0002-3053-5511],
O. V. Lunina⁷ [0000-0001-7743-8877]

^{1, 2, 5} V. M. Matrosov Institute of System Dynamics and Control Theory of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, 664033, Russia

^{1, 2, 4} School of Mathematics and Information Science, Yantai University, Yantai, 264005, China

³School of Civil Engineering, Yantai University, Yantai, 264005, China

^{4, 6, 7} Institute of the Earth's Crust of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, 664033, Russia

^{1, 5}Institute of Mathematics and Informatics, Irkutsk State University, Irkutsk, 664003, Russia

¹eugeneai@icc.ru, ²bychkov@icc.ru, ³zhangjin513@outlook.com, tcherk@crust.irk.ru,

⁵victorypopova1@gmail.com, ⁶moks@crust.irk.ru, ⁷lounina@crust.irk.ru

Abstract

The article presents an innovative methodology for applying Semantic Web technologies to support fundamental geological research. The problem of semantic integration of heterogeneous geological data, characterized by different scales and interdisciplinarity, is considered. A five-stage methodology is developed, including domain analysis, ontological conceptual modeling, data transformation into a knowledge graph, deployment of a distributed data access infrastructure based on the conceptual model, and integration with processing and analysis procedures. Practical testing was conducted on three case studies: analysis of geochemical data for assessing territory pollution levels, creation of an information system about faults, and research on reservoir shoreline dynamics. The proposed ontological approach ensures compliance with FAIR principles and overcoming the "semantic barrier" in geological research. It is shown that Semantic Web technologies enable a transition from fragmented infor-

mation arrays to a holistic semantic space of geological knowledge, opening new opportunities for generating comprehensive scientific hypotheses and cross-disciplinary research.

Keywords: *Semantic Web, knowledge graphs, ontological modeling, semantic environment of scientific knowledge, fundamental geological research.*

REFERENCES

1. *Mattmann C.* A vision for data science // *Nature*. 2013. Vol. 493. P. 473–475. <https://doi.org/10.1038/493473a>
2. *Hitzler P., Janowicz K.* Semantic Web for Earth and Environmental Science: Current state and future directions // *Semantic Web*. 2010. Vol. 1, No. 1–2. P. 85–98. <https://doi.org/10.3233/SW-2010-0012>
3. *Hitzler P.* A review of the semantic web field // *Communications of the ACM*. 2021. Vol. 64, No. 2. P. 76–83. <https://doi.org/10.1145/3397512>
4. *Sinkova E.A., Petrov O.V., Khancuk A.I.* Geochronological Atlas and Reference Book of the Main Structural-Material Complexes of Russia – a Basic Information Resource for the Country's Geological Industry // *Exploration and Protection of Mineral Resources*. 2022. No. 90. P. 5–14. https://doi.org/10.52349/0869-7892_2022_90_5-14
5. *Wilkinson M.D., Dumontier M., Aalbersberg I.J. et al.* The FAIR Guiding Principles for scientific data management and stewardship // *Scientific Data*. 2016. Vol. 3. P. 160018. <https://doi.org/10.1038/sdata.2016.18>
6. *Laney D.* 3D Data Management: Controlling Data Volume, Velocity, and Variety // *META Group Research Note*. 2001. Vol. 6, No. 70. URL: <https://diegonogare.net/wp-content/uploads/2020/08/3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
7. *Gandomi A., Haider M.* Beyond the hype: Big data concepts, methods, and analytics // *International Journal of Information Management*. 2015. Vol. 35, No. 2. P. 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
8. *Hogan A., Blomqvist E., Cochez M., D'Amato C. et al.* Knowledge Graphs // *ACM Computing Surveys*. 2021. Vol. 54, No. 4. <https://doi.org/10.1145/3447772>

9. *Pellinen V.A., Cherkashina T.Y., Gustaitis M.A.* Assessment of metal pollution and subsequent ecological risk in the coastal zone of Olkhon Island, Lake Baikal, Russia // *Science of the Total Environment*. 2021. Vol. 786. P. 147441.
<https://doi.org/110.1016/j.scitotenv.2021.147441>
10. *Zerizghi T., Yang Y., Wang W., Zhou Y., Zhang J., Yi Y.* Ecological risk assessment of heavy metal concentrations in sediment and fish of a shallow lake: a case study of Baiyangdian Lake, North China // *Environmental Monitoring and Assessment*. 2020. Vol. 192. P. 154. <https://doi.org/110.1007/s10661-020-8078-8>
11. *Zhang J., Wang K., Yi Q., Zhang T., Shi W., Zhou X.* Transport and partitioning of metals in river networks of a plain area with sedimentary resuspension and implications for downstream lakes // *Environmental Pollution*. 2022. Vol. 294. P. 118668. <https://doi.org/110.1016/j.envpol.2021.118668>
12. *Gopal V., Krishnamurthy R.R., Vignesh R., Nathan C.S. et al.* Assessment of heavy metal contamination in the surface sediments of the Vedaranyam coast, Southern India // *Regional Studies in Marine Science*. 2023. Vol. 65. P. 103081. <https://doi.org/110.1016/j.rsma.2023.103081>
13. *Chubarov V., Cherkashina T., Maltsev A., Chuparina E., Amosova A., Prosekin S.* Investigation of Soils and Pine Needles Using WDXRF and TXRF Techniques for Assessment of the Environmental Pollution of Shelekhov District, Eastern Siberia, by the Aluminum Industry and Heat Power Engineering // *Agronomy*. 2022. Vol. 12. P. 454. <https://doi.org/110.3390/agronomy12020454>
14. *Lunina O.V.* The digital map of the Pliocene–Quaternary crustal faults in the Southern East Siberia and the adjacent Northern Mongolia // *Geodynamics and Tectonophysics*. 2016. Vol. 7, No. 3. P. 407–434.
<https://doi.org/110.5800/GT-2016-7-3-0215>
15. *Cherkashin E.A., Lunina O.V., Demyanov L.O., Tsygankov A.V.* Web-GIS viewer for active faults data represented as a knowledge graph // *The 4th Scientific-practical Workshop Information Technologies: Algorithms Models, Systems*. September 14, 2021, Irkutsk, Russia / *CEUR Workshop Proceedings*. 2021. Vol. 2984. P. 55–65. URL: <https://ceur-ws.org/Vol-2984/paper8.pdf>

16. *Mazaeva O., Babicheva V., Kozyreva E.* Geomorphological process development under the impact of man-made reservoir operation, a case study: Bratsk reservoir, Baikal-Angara hydroengineering system, Russia // *Bulletin of Engineering Geology and the Environment*. 2019. Vol. 78. P. 4659–4672.

<https://doi.org/110.1007/s10064-018-1428-x>

17. *Ovchinnikov G.I., Pavlov S.Kh., Trzhtsinsky Yu B.* Changes in the Geological Environment in the Zones of Influence of the Angara–Yenisei Reservoirs. Novosibirsk: Nauka, 1999. 254 p.

18. *Ranatunga S., Ødegård R.S., Jetlund K., Onstein E.* Use of Semantic Web Technologies to Enhance the Integration and Interoperability of Environmental Geospatial Data: A Framework Based on Ontology-Based Data Access // *ISPRS International Journal of Geo-Information*. 2025. Vol. 14, No. 2. P. 52.

<https://doi.org/110.3390/ijgi14020052>

19. *Moura A.-M., Porto F., Vidal V., Magalhães R.P. et al.* A semantic integration approach to publish and retrieve ecological data // *International Journal of Web Information Systems*. 2015. Vol. 11, No. 1. P. 87–119.

<https://doi.org/110.1108/IJWIS-08-2014-0028>

20. *Marcus S., Tim D.* GeoSciML: Development of a generic GeoScience Markup Language // *Computers & Geosciences*. 2005. Vol. 31, No. 9. P. 1095–1103.

<https://doi.org/10.1016/j.cageo.2004.12.003>.

21. *Raskin R., Pan M.* Knowledge representation in the semantic web for Earth and environmental terminology (SWEET) // *Computers & Geosciences*. 2005. Vol. 31, No. 9. P. 1119–1125. <https://doi.org/110.1016/j.cageo.2004.12.004>

22. *Simons B.A., Raymond O., Jackson I., Lee K.* OneGeology – Improving global access to geoscience // *The 5th Global Workshop on Digital Soil Mapping*. April 10–13, 2012, Sydney, Australia / *Digital Soil Assessments and Beyond: Proceedings of CRC Press*, 2012. P. 265–275.

23. *Allison M.L., Ahern T., Arctur D. et al.* EarthCube Governance Framework: A Proposal to the Community (Version 1.0) // *EarthCube Governance Working Group Technical Report*. 2012. 237 p.

24. Atakan K., Bjerrum L.W., Bungum H. et al. The European Plate Observing System and the Arctic // Arctic. 2015. Vol. 68, Suppl. 1. <https://doi.org/110.14430/arctic4446>

СВЕДЕНИЯ ОБ АВТОРАХ



БЫЧКОВ Игорь Вячеславович – доктор технических наук, профессор, академик РАН, директор Института динамики систем и теории управления им. В.М. Матросова Сибирского отделения РАН, член Президиума РАН, член бюро Отделения нанотехнологий и информационных технологий РАН, почетный профессор Института математики и компьютерных наук Яньтайского университета (Китай).

Научные интересы: искусственный интеллект, геоинформационные системы, web-технологии, системы интеллектуального анализа данных, математическое моделирование, облачные вычисления, большие данные, многоагентные системы, информационно-вычислительные и телекоммуникационные технологии.

Igor Vyacheslavovich BYCHKOV – Doctor of Technical Sciences, Professor, Member of the Russian Academy of Sciences, Director of the V.M. Matrosov Institute of System Dynamics and Control Theory of Siberian Branch of the Russian Academy of Sciences, Member of the Presidium of RAS, Member of the Bureau of the Department of Nanotechnology and Information Technologies of RAS, honorary professor of Institute of Mathematics and Computer Science at Yantai University (China).

Research interests: artificial intelligence, geoinformation systems, web technologies, data mining systems, mathematical modeling, cloud computing, big data, multi-agent systems, information-computing and telecommunication technologies.

email: bychkov@icc.ru

ORCID: 0000-0002-1765-0769



ЧЕРКАШИН Евгений Александрович – кандидат технических наук, доцент, заведующий кафедрой информационных технологий Института математики и информатики Иркутского государственного университета, старший научный сотрудник Института динамики систем и теории управления им. В.М. Матросова Сибирского отделения РАН, профессор Института математики и компьютерных наук Яньтайского университета (Китай).

Научные интересы: синтез программных объектов на основе абстрактных моделей (Model driven architecture); логическое программирование; системы, основанные на формализованных знаниях; обработка естественного языка; системы поддержки принятия решений на основе математического моделирования; психология педагогического процесса.

Evgeny Alexandrovich CHERKASHIN – Candidate of Technical Sciences, Associate Professor, Head of the Department of Information Technologies of Institute of Mathematics and Computer Science at Irkutsk State University, Senior Researcher at the V.M. Matrosov Institute of System Dynamics and Control Theory of Siberian Branch of the Russian Academy of Sciences.

Research interests: software objects synthesis based on abstract models (Model driven architecture), logical programming, systems based on formalized knowledge, natural language processing, decision support systems based on mathematical modeling, psychology of pedagogical process, honorary professor of Institute of Mathematics and Computer Science at Yantai University (China).

email: eugeneai@icc.ru

ORCID: 0000-0003-2428-2471



ЧЖАН Цзинь – профессор, заместитель декана Института гражданского строительства Яньтайского университета.

Научные интересы: экологическая инженерия, фосфорный цикл в водных системах, мембранное загрязнение, очистка сточных вод с использованием микроводорослей.

Jin ZHANG – Professor, Vice Dean of the School of Civil Engineering, Yantai University.

Research interests: environmental engineering, phosphorus cycling in aquatic systems, membrane fouling, microalgae-based wastewater treatment.

email: zhangjin513@outlook.com

ORCID: 0000-0003-0009-824X



ЧЕРКАШИНА Татьяна Юрьевна – кандидат геолого-минералогических наук, старший научный сотрудник Центра коллективного пользования «Геодинамика и геохронология» Института земной коры Сибирского отделения РАН, профессор Института математики и компьютерных наук Яньтайского университета (Китай).

Научные интересы: рентгенофлуоресцентный анализ горных пород, геохимия осадочных бассейнов юга Сибирской платформы, оценка эколого-геохимического состояния почвенного покрова и береговой зоны Байкальской природной территории.

Tatiana Yurievna CHERKASHINA – Candidate of Geological and Mineralogical Sciences, Senior Researcher at the Center for Geodynamics and Geochronology of the Institute of the Earth's Crust, Siberian Branch of the Russian Academy of Sciences, professor of Institute of Mathematics and Computer Science at Yantai University (China).

Research interests: X-ray fluorescence analysis of rocks, geochemistry of sedimentary basins of the Siberian Platform, assessment of the ecological and geochemical state of the soil cover and the coastal zone of the Baikal natural territory.

email: therk@crust.irk.ru

ORCID: 0000-0002-1443-6839



ПОПОВА Виктория Алексеевна – ассистент кафедры алгебраических и информационных систем Института математики и информационных технологий Иркутского государственного университета, инженер-программист.

Научные интересы: разработка программного обеспечения, криптографические методы защиты информации, статические анализаторы программного кода, разработка web-приложений.

Victoria Alexeevna POPOVA – Assistant of the Department of Algebraic and Information Systems, Institute of Mathematics and Information Technology of Irkutsk State University, Software Engineer.

Research interests: software development, cryptographic methods of information protection, static code analyzers, web-application development.

email: victorypopova1@gmail.com

ORCID: 0000-0002-7764-1995



МАЗАЕВА Оксана Анатольевна – кандидат геолого-минералогических наук, научный сотрудник лаборатории инженерной геологии и геоэкологии Института земной коры Сибирского отделения РАН.

Научные интересы: исследование эрозионных и других экзогенных геологических процессов в природных и природно-технических геосистемах юга Сибирской платформы, изучение состава и свойств рыхлых отложений.

Oksana Anatolievna MAZAEVA – Candidate of Geological and Mineralogical Sciences, Researcher at the Laboratory of Engineering Geology and Geoecology of the Institute of the Earth's Crust, Siberian Branch of the Russian Academy of Sciences.

Research interests: study of erosion and other exogenous geological processes in natural and natural-technical geosystems in the south of the Siberian Platform, investigation of the composition and properties of loose sediments.

email: moks@crust.irk.ru

ORCID: 0000-0002-3053-5511



ЛУНИНА Оксана Викторовна – доктор геолого-минералогических наук, главный научный сотрудник лаборатории тектонофизики Института земной коры Сибирского отделения РАН.

Научные интересы: тектоника и геодинамика, структурная геология, тектонофизика, сейсмотектоника, изучение строения земной коры и опасных природных процессов.

Oksana Viktorovna LUNINA – Doctor of Geological and Mineralogical Sciences, Chief Researcher at the Laboratory of Tectonophysics of the Institute of the Earth's Crust, Siberian Branch of the Russian Academy of Sciences.

Research interests: tectonics and geodynamics, structural geology, tectonophysics, seismotectonics, study of the Earth's crust structure and hazardous natural processes.

email: lounina@crust.irk.ru

ORCID: 0000-0001-7743-8877

Материал поступил в редакцию 14 октября 2025 года