

## Интеграция экологических знаний с помощью Data Science и онтологий

Верещагина Е. А.<sup>1</sup>, Фролов А. В.<sup>2</sup>, Титова А. А.<sup>1</sup>

<sup>1</sup>Дальневосточный федеральный университет  
Владивосток, Россия

vereschagina.ea@dvfu.ru, titova.aana@dvfu.ru

<sup>2</sup>Морской государственный университет имени адмирала Г. И. Невельского  
Владивосток, Россия

Frolov@msun.ru

В экологии актуальны проблемы точности, определенности и своевременности принимаемых решений, поэтому ошибки управления для экологии должны минимизироваться. В рамках данного исследования были проведены комплексные анализы связей на различных уровнях – логическом (онтологическом), предметно-ориентированном и аналитическом с использованием методов системного анализа-синтеза и онтологий. В результате исследований была разработана структурная (структурно-алгебраическая) модель и предложена схема онтологий для представления знаний в экологии. Ключевые результаты исследования включают следующие аспекты: формализация требований к экологическим моделям, включающих релевантность, адекватность, формализованность, модифицируемость, технологичность и верифицируемость; разработка общей схемы представления знаний, состоящей из интента, прецедента и оценки; описание алгоритмов онтологической модели, реализующих функциональные процедуры диагностики риск-ситуаций, идентификации параметров критических экологических ситуаций, классификации и многомерного шкалирования; анализ применения Data Analytics в экологии с примерами построения тепловых карт экосреды, прогнозирования развития экологических процессов и временного анализа данных со спутников; выявление проблем при работе с большими данными в экологии, таких как стандартизация метаданных, единиц измерения и протоколов, а также сложности визуализации и анализа. Полученные результаты позволяют более эффективно использовать семантические технологии и онтологии для представления и анализа экологических знаний. Созданная структурная модель и схема онтологий могут быть применены в практической деятельности для минимизации рисков и повышения качества принятия решений в экологической сфере.

**Ключевые слова:** экологические знания; методы Data Mining и Advanced Data Analytics; онтологическая модель; инструменты Big Data; проектирование и применение; экологические данные; синтез моделей.

### ВВЕДЕНИЕ

В экологии существует много проблем, относящихся как к самой экологии, так и междисциплинарных проблем, в которых встречаются нечеткость и неточность, неопределенность знаний, особенно, новых. Это особенно заметно при рассмотрении проблем зеленого и бережливого производства – направлений, в которых эколого-экономические связи понимаются в зависимости от степени устойчивости экологической среды и оптимальности (рациональности) производства (Фюкс, 2016; Кожевников, Лебедева, 2019). Ошибки менеджмента в данном случае должны быть минимизированы путем их формализации.

Знания, извлекаемые из контента предметной области, могут рассматриваться при различных условиях их представления – на логическом (теоретическом, часто и аксиоматическом), аналитическом и синтетическом (предметно-ориентированном) (Пальчунов, 2008), что согласовано с классификацией истинности суждений:

1) на логическом уровне истинные высказывания – истинны всегда, независимо от значений (смысла) входящих в них понятий. На этом уровне моделируются потенциальные возможности средств онтологического моделирования, несколько абстрагируясь от языков представления онтологий и удобств пользователей.

2) на аналитическом уровне истинность аналитических выражений зависит лишь от смысла понятий, составляющих выражение;

3) на синтетическом уровне истинность предложений (синтетических) определяется их связями и свойствами предметной области. Структура предметной онтологии отображается в виде семантической сети, вершинами которой являются базовые понятия, а стрелки выражают отношения (связи) между ними.

Онтологии используются для представления знаний в различных областях независимо от используемых технологий. Они предоставляют общий словарь и содержательную информацию об отношениях между вещами/данными/информацией в области экологии.

Цель работы – разработать методологические основы для минимизации ошибок управления экологическими процессами через создание структурной (структурно-алгебраической) модели и онтологической схемы представления информации в области экологии.

## МАТЕРИАЛ И МЕТОДИКА

В рамках данного исследования использовались методы системного анализа-синтеза и онтологического моделирования для изучения связей на различных уровнях представления знаний: логическом (онтологическом), предметно-ориентированном и аналитическом. Эти методы позволили формализовать подход к анализу экологических данных и разработать структурную модель для представления знаний в экологии.

Основными методами, примененными в работе, являются:

- 1) системный анализ-синтез: для выявления взаимосвязей между компонентами экосистем и их влияния на принятие решений;
- 2) онтологическое моделирование: для формализации знаний о сложных экологических процессах и создания общей структуры представления информации;
- 3) Data Analytics: для анализа больших массивов данных, выявления закономерностей и прогнозирования экологических ситуаций.

Материалом исследования послужили:

- 1) экологические данные, включая наблюдения за состоянием окружающей среды, результаты мониторинга и параметры экосистем;
- 2) структурированные и неструктурированные данные из различных источников (файлы, таблицы, графическая информация);
- 3) теоретические основы онтологий, такие как язык OWL для описания классов и отношений в предметной области.

Для описания модели были использованы следующие принципы:

- 1) структурно-алгебраическая модель, представленная в виде алгебры;
- 2) общая схема представления знаний.

При разработке моделей были учтены следующие ключевые требования:

- 1) релевантность и адекватность целям исследования;
- 2) формализованность (математическая или информационно-логическая);
- 3) модифицируемость для адаптации к меняющимся условиям;
- 4) технологичность и реализуемость в практической деятельности;
- 5) верифицируемость через тестирование и доказательства.

Для реализации исследовательских задач применялись:

- 1) язык OWL для описания онтологий;
- 2) методы Data Mining и Advanced Data Analytics для анализа больших данных;
- 3) инструменты безопасного хранения и обработки Big Data (например, Hadoop, Apache).

Таким образом, методология исследования включала комплексный подход, сочетающий теоретические основы онтологий, методы анализа данных и практические инструменты для решения актуальных экологических проблем.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ

**Data analytics, онтологии и модели в экологии.** С формальной позиции или позиций формально рассматриваемой предметной области, модель – это алгебра  $O = \langle A, B \rangle$ , где  $A$  – совокупность аналитических предложений для описания совокупности, смысла понятий в предметной области или носитель, а  $B$  – сигнатура или совокупность операций, определенных над элементами  $A$  алгебры  $O$ . Для описания классов и отношений между ними часто используется язык OWL (использования онтологий) и др.

Теорией Т предметной области можно считать все предложения, верные для каждого теста предметной области. На практике их невозможно установить, идентифицировать полностью и сразу, поэтому теории бесконечны и развиваются, а мы имеем дело лишь с конечной совокупностью онтологического знания  $B$ .

В экологии работать приходится с моделями различных сигнатур, хотя аксиоматическая теория такое не предполагает, она работает с моделями одной сигнатуры, часто даже неопределенной. Например, при извлечении интента (знаний, смысла) из контента, сигнатура алгебраической системы – динамична, расширяема. Рассматривают пару  $O_n = \langle A_n, B \rangle$ , где  $A_n \subseteq A$  – конечное подмножество.

При интеграции интента интегрируются и сигнатуры, возникает проблема различного смысла в текстах одного содержательного смысла, что при отслеживании непротиворечивости и полноты системы весьма сложно реализовать. Можно лишь с определенной вероятностью формулировать гипотезы, истинные для всей предметной области. Общая схема представления знаний состоит из интента, прецедента и оценки (рис. 1).

К предметным моделям предъявляются ключевые требования (Казиев, Казиев, 2004):

- 1) релевантность, адекватность, соответствие целям, гипотезам, ресурсам;
- 2) ценность (полезность, отражение необходимых связей);
- 3) формализованность (математическая, инфологическая);
- 4) модифицируемость (экстраполяция на следующий период рассмотрения);
- 5) технологичность (реализуемость);
- 6) функциональность, возможность использования в имитационном (сценарном, ситуационном) варианте моделирования, в том числе, автоматизированном;
- 7) верифицируемость (тестируемость, доказательность) и др.

Экологическая модель должна учитывать внешнюю среду посредством наблюдаемых признаков, «прямых» воздействий окружения, а также внутреннюю среду, определяемую воздействиями экологических факторов, образа жизни (физиологическими особенностями). Эти воздействия, причинно-следственные связи и определяют индивидуальное реагирование на экологические меры.

Структура предметной онтологии отображается в виде семантической сети, вершинами которой являются базовые понятия, а стрелки выражают отношения (связи) между ними. Приоритет, важность симптома, причины и причинно-следственной связи обычно идентифицируются экспертизно-эвристически, при определенных гипотезах моделирования, которые также зависят от системы экологического наблюдения, мониторинга, а также инструментария (методики, планы мониторинга, аппаратное и компьютерное обеспечение, базы знаний, ПО и др.).

Онтологическая модель в экологии позволяет структурировать и формализовать экологическую область, класс экологических задач и прецедентов (эмпирики), используя оценочные знания (вероятностные критерии).

Онтологическая модель в экологической области формализует ключевые термы, их отношения, связи явлений и процессов в экологии.



Рис. 1. Схема онтологической модели

Для того чтобы расширить возможность описания предметных областей за счет включения в онтологию не только объектов, но и их качественных и количественных характеристик, в конструкторе онтологий реализованы средства создания свойств объектов и добавления атрибутов для всех концептов. В данном случае свойства отвечают за качественные характеристики, а атрибуты – за количественные. Наличие такой возможности увеличивает объем знаний, которые пользователь может описать с помощью онтологии, так как задание свойств объектов влечет за собой появление следующего уровня детализации онтологии, а именно создание и описание процессов, в которые могут вступать объекты. Прецеденты рассматриваемой области формализуемы булевозначными функциями (моделями), а оценочные критерии, знания – нечеткими моделями.

Алгоритмы (методы) онтологической модели реализуют функциональные процедуры. Основными из них являются:

- 1) алгоритмы диагностики риск-ситуаций, кризисов, например, из-за экологических противопоказаний;
- 2) алгоритмы идентификации (прогноза) параметров возникновения и развития критических экологических ситуаций;
- 3) алгоритмы интерфейса класса «база знаний – эколог», «эколог – предприятие» и др., обеспечения связей с экологическими паспортами, прецедентами;
- 4) алгоритмы классификации, многомерного шкалирования, предварительной статистической обработки и оценки статистических гипотез (например, нормальности по Гауссу);
- 5) алгоритмы оценки противопоказаний и негативных влияний в системе.

С увеличением количества концептов и связей семантические сети становятся нечитаемыми и трудными для восприятия, поэтому возникла необходимость выделения подмножества объектов из общей сети онтологии. В экологических системах актуально использование Data или больших массивов данных, которые плохо структурируемы, представимы и обрабатываемы с помощью классических (например, реляционных) базах данных (БД). Они также плохо обрабатываемы в единых приложениях, с помощью единого интерфейса, в основном, из-за своей мультиформатности, объемности, целевой необходимости извлечения новых связей, априори неизвестных (Маурер, 2020).

Большинство экологических данных представлено файлами, таблицами, графикой. Без связок на уровне описаний, онтологий, экспертных суждений. Это делает их малоценным «вагоном данных», не приспособленным для извлечения знаний (причинно-следственных связей) с помощью Data Mining, Advanced Data Analytics.

Проблема возникает не только с потоком данных, но и с инфологическими моделями их извлечения, анализа и принятия решений. Из-за роста асимптотической сложности алгоритмов и необходимости обеспечения безопасности персональных данных в условиях резкого роста БД и актуализации знаний в доказательной медицине. Влияет и развитие облачно-туманной инфраструктуры систем медицины и здравоохранения.

Главные стимулы использования Data Science – не только выгоды, финансовые или материальные, но и выгоды организационные, технологические, качества, оперативности обслуживания населения.

Примеры использования Advanced Data Analytics в экологии:

- 1) построение тепловых карт экосреды, экосостояния;
- 2) анализ данных Data Mining) с прогнозированием развития экологического процесса;
- 3) экологическое образование (анализ среды, распознавание объектов, оценка заражений и др.);
- 4) временной анализ (например, вейвлет-анализ данных со спутников в реальном режиме;
- 5) дистанционный мониторинг.

Интеграция цифровых методов повышает экологический, эволюционный потенциал и экономит ресурсы, особенно, временные и организационные (всей экосреды).

Экологи и ученые, изучающие экосистемы, в частности, имеют долгую историю работы с концепциями больших данных. Основными проблемами при этом являются стандартизация метаданных, единиц измерения и протоколов, «обнаруживаемость» данных и сложность визуализации и анализа внутри или между подключенными киберинфраструктурными и аналитическими платформами (например, DataONE и статистическое программное обеспечение). В то время как ученые-экосистемологи разрабатывают и используют подходы, основанные на больших данных, наука об экосистемах по-прежнему в значительной степени является дисциплиной, которая объясняет закономерности и процессы, а не занимается их прогнозированием. Отчасти это объясняется тем, что в структуре и процессе все еще очень многое требует объяснения.

Ученые-экосистемологи стремятся понять биофизические процессы, а также сложные области применения и последствия этих процессов (например, функции экосистем, стабильность и услуги). Большие данные могут существенно расширить возможности понимания и прогнозирования экосистемной науки в течение следующего десятилетия. Хотя качество данных и проницательность исследователей по-прежнему будут иметь решающее значение при любом анализе, использование технологий и философии больших данных может обеспечить надежные подходы и инструменты для продвижения к ответам на некоторые сложные вопросы экосистемной науки, такие как, например, как изменение климата на местном и региональном уровнях влияет на функционирование экосистем, как стабильность экосистемных процессов влияет на экосистемную безопасность, поддерживается биологическим разнообразием и другими факторами. Кроме того, четкая связь между подходами к работе с большими данными и сетевыми науками указывает на потенциальный сдвиг в сторону будущего, в котором преподаватели, начинающие карьеру,

должны как развивать сотрудничество, так и получать вознаграждение за его успешное выполнение.

С ростом потоков данных, стандартные методы обработки (анализа) становятся слабыми и «жесткими» (мало адаптивными). Data Analytics имеет огромный потенциал в экологии и других областях, но требует новых подходов и моделей обработки данных.

Для решения проблем с Big Data применяются специализированные технологии и инструменты (Hadoop, Apache и др.). Они позволяют оперативно обрабатывать и эффективно анализировать большие массивы данных. Хотя многие продукты, основанные на больших данных, уже используются, например, в экосистемных моделях глобальных циклов углерода и питательных веществ, они, возможно, менее ценятся за их ценность для развития теории и стимулирования открытия новых гипотез.

Существует два общих аналитических подхода к получению выводов на основе больших данных. Если имеется достаточное понимание для определения модели, для оценки параметров и проверки гипотез используются методы, основанные на статистической регрессии. Иерархическая (часто байесовская) статистика позволяет интегрировать различные источники данных для понимания более сложных экосистемных процессов.

Однако, когда априорного понимания недостаточно для описания модели процесса, методы интеллектуального анализа данных могут выявить надежные многомерные закономерности, определить важные движущие силы и сгенерировать прогнозы на основе самих данных. В экологии исследовательские подходы исторически означали тщательное изучение окружающей среды для выявления экологических взаимосвязей и выработки проверяемых гипотез. Визуализация данных и интеллектуальный анализ также использовались в качестве шага к подтверждающим (статистическим выводам) подходам. Более формально интеллектуальный анализ данных относится к изучению больших многомерных наборов данных для выявления надежных закономерностей данных (например, кластеров точек данных, аномальных сигнатур в данных или зависимостей между переменными), которые заслуживают дальнейшего изучения с помощью целенаправленной проверки гипотез (подтверждающий анализ).

Использование вычислительных алгоритмов, включая ансамблевую регрессию и деревья классификации или анализ ассоциативных правил позволяет обойти ограниченную способность априори назначать взаимодействия и статистические распределения (возможно, очень большому) количеству переменных. Такие подходы становятся все более востребованными для получения информации о системах, которые являются очень сложными или в которых трудно поддается репликации, но этот подход требует некоторой переориентации наших представлений об анализе данных.

Например, Data Mining позволяет компаниям выявлять тренды и потребности клиентов, улучшать принятие решения и др. Кроме Big Data, развиваются и машинное обучение, искусственный интеллект. Они позволяют глубоко анализировать данные, делать прогнозы заданного уровня приемлемости. Развиваются также методы безопасного хранения-обработки Big Data.

## ОБСУЖДЕНИЕ

Полученные результаты исследования позволяют сделать вывод о том, что применение онтологий и методов Data Analytics в экологии открывает новые возможности для решения актуальных задач управления природными ресурсами. Обсуждение этих результатов можно разделить на несколько ключевых аспектов.

**Эффективность онтологической модели.** Разработанная структурная (структурно-алгебраическая) модель и схема онтологий показали свою применимость для формализации знаний в экологии. Использование онтологий позволяет создавать общий словарь терминов и четко определять отношения между различными компонентами экосистем. Это особенно важно при работе с большими данными, где стандартизация метаданных и единиц измерения становится критически важной задачей.

**Применение Data Analytics в экологии.** Анализ данных с использованием методов Data Mining и Advanced Data Analytics доказал свою эффективность в выявлении закономерностей и прогнозировании экологических процессов. Например, построение тепловых карт экосреды и анализ данных со спутников помогают лучше понять состояние окружающей среды и принимать более обоснованные решения. Однако остается проблема интеграции различных источников данных, что требует дальнейшего развития технологий обработки Big Data.

**Человеческий фактор и минимизация ошибок.** Внедрение онтологической модели способствует снижению влияния человеческого фактора при принятии решений в экологии. Это достигается за счет формализации процессов анализа данных и использования вероятностных подходов для оценки рисков. Такой подход особенно важен при управлении критическими экологическими ситуациями, где точность и своевременность решений играют решающую роль.

**Сложности при работе с большими данными.** Несмотря на успехи в применении Big Data в экологии, остаются значительные вызовы. Стандартизация метаданных, единиц измерения и протоколов все еще является актуальной проблемой. Кроме того, сложность визуализации и анализа больших массивов данных требует разработки новых методов и инструментов. Развитие облачных технологий и туманной инфраструктуры может стать одним из путей решения этих проблем.

**Перспективы развития.** Будущее применения онтологий и Data Analytics в экологии видится в направлении создания более интеллектуальных систем анализа данных. Интеграция машинного обучения и искусственного интеллекта позволит глубже исследовать сложные экосистемные процессы и делать более точные прогнозы. Также важно продолжать работу по стандартизации данных и развитию интероперабельных платформ для их обмена между исследователями.

Таким образом, полученные результаты подтверждают важность применения современных семантических технологий и методов анализа данных в экологии. Они предоставляют надежную основу для дальнейшего развития подходов к управлению экологическими системами и минимизации ошибок в принятии решений.

## ЗАКЛЮЧЕНИЕ

На основе проведенного исследования можно сделать вывод о том, что использование онтологической модели и методов Data Analytics позволяет существенно повысить эффективность управления экологическими процессами. Созданная структурная (структурно-алгебраическая) модель и схема онтологий обеспечивают формализацию знаний в экологии, что помогает минимизировать ошибки принятия решений. Ключевые преимущества подхода включают возможность интеграции различных типов данных, учет сложных причинно-следственных связей и повышение точности прогнозирования критических экологических ситуаций. При этом сохраняется актуальность решения проблем обработки больших данных, таких как стандартизация метаданных и развитие адаптивных методов анализа. Онтологическая модель позволяет априори снижать вероятность рисковых ситуаций и учитывать человеческий фактор при управлении экосистемами.

## Список литературы

- Фюкс Р. Зеленая революция: экономический рост без ущерба для экологии. – М: Альпина нон-фикшн. – 2016. – 330 с.
- Кожевников С. А., Лебедева М. А. Проблемы перехода к зеленой экономике в регионе (на материалах Европейского Севера России) // Проблемы развития территории. – 2019. – № 4 (102). – С. 72–88.
- Пальчунов Д. Е. Решение задачи поиска информации на основе онтологии // Бизнес-информатика. – 2008. – № 1. – С. 3–13.
- Казиев В. М., Казиев К. В. Основы математического и инфологического моделирования в примерах // Информатика и образование. – 2004. – № 1. – С. 18–23.
- Маурер Р. Метод кайдзен: Шаг за шагом к достижению цели. – М: Альпина Паблишер. – 2020. – 179 с.

Фролов А. В., Титова А. А., Верещагина Е. А. Мониторинг рисков и безопасности перевозок нефти-газа железнодорожным транспортом // Экологические системы и приборы. – 2022. – № 9. – С. 51–53. DOI: 10.25791/esip.9.2022.1319. – EDN IHZQEE

Фролов А. В., Титова А. А. Системная инженерия в комплексе улучшения судовождения // Эксплуатация морского транспорта. – 2022. – № 2. – С. 69–71. DOI: 10.34046/aumsuomt103/15. – EDN HODVBT

Фролов А. В., Титова А. А. Цифровые технологии судовождения // Эксплуатация морского транспорта. – 2022. – № 3 (104). – С. 152–153. DOI: 10.34046/aumsuomt104/24. – EDN SNMPPS

Фролов А. В. Моделирование выделения полезного линейночастотного сигнала в системе распределенных приемников в воднотранспортном потоке // Прикладная физика и математика. – 2022. – № 1. – С. 6–9. DOI: 10.25791/pfim.01.2022.1219. – EDN KQYXTJ

Фролов А. В. Моделирование воднотранспортных потоков с учетом инфраструктуры порта // Промышленные АСУ и контроллеры. – 2021. – № 11. – С. 9–11. DOI: 10.25791/asu.11.2021.1323. – EDN AKNNHM

**Vereshchagina E. A., Frolov A. V., Titova A. A. Integration of Environmental Knowledge through Data Science and Ontologies //** Ekosistemy. 2025. Iss. 43. P. 53–60.

In contemporary environment ecology, the problems of accuracy, certainty and timeliness of decision-making are relevant, therefore management errors should be minimized. This study presents a comprehensive analysis of the relationships at various levels – logical (ontological), subject-oriented, and analytical – employing methods of systems analysis, synthesis, and ontologies. As a result of the research, a structural (structural-algebraic) model was developed and an ontology scheme was proposed for representing environmental knowledge. The key findings of the study include the following aspects: formalization of requirements for ecological models, including relevance, adequacy, formalization, modifiability, manufacturability and verifiability; development of a general scheme for representing knowledge, consisting of intent, precedent and assessment; description of algorithms of the ontological model that implement functional procedures for diagnosing risk situations, identifying parameters of critical environmental situations, classification and multidimensional scaling; analysis of the application of Data Analytics in environment with examples of constructing heat maps of the eco-environment, forecasting the development of environmental processes and time analysis of satellite data; identification of problems when working with big data in environment, such as standardization of metadata, units of measurement and protocols, as well as the complexity of visualization and analysis. The obtained results facilitate more efficient use of semantic technologies and ontologies for the presentation and analysis of ecological environmental knowledge. The created structural model and ontology scheme can be applied in practical activities to minimize risks and improve the quality of decision-making in the environmental sphere.

*Key words:* environmental knowledge, Data mining; Advanced data analysis, ontological model, Big Data tools, design engineering and application; environmental data; model synthesis.

*Поступила в редакцию 08.04.25  
Принята к печати 26.06.25*