

УДК 004.043, 631.3.004.58

*Молородов Ю.И., Ходорченко Д.А.,  
Институт вычислительных технологий Сибирского отделения Российской академии наук,  
Новосибирск*

## РАЗВЕРТЫВАНИЕ ИНФРАСТРУКТУРЫ ИНТЕЛЛЕКТУАЛЬНОГО НАУЧНОГО ИНТЕРНЕТ-РЕСУРСА

В статье рассматривается один из способов развертывания инфраструктуры интеллектуального научного Интернет-ресурса (ИНИР) на основе онтологического подхода. При этом онтология служит для систематизации и формализации знаний и информационных ресурсов. На основе онтологии организована удобная навигация и содержательный поиск. Пользователь может осуществлять навигацию по контенту ИНИР, используя дерево понятий онтологии. Важной особенностью онтологического подхода является возможность предоставления доступа не только к описаниям методов обработки информации и решения задач, но и к их конкретным реализациям.

**Ключевые слова:** портал знаний, информационная система, система информационного обеспечения, информационные ресурсы, интеграция данных.

### Введение

Важной особенностью развития современного общества является экспоненциальный рост объема и быстрого устаревания научной информации. Поэтому актуальна проблема понятного и удобного отображения информации, и взаимодействия с потенциальным потребителем этой информацией. Любое научное исследование начинается с поиска научной информации об исследованиях в данной предметной области, но этот поиск затруднен постоянно возрастающим объемом статей, книг, монографий и пр. Ученый вынужден тратить много времени на поиск и обработку информации, позволяющей быстро ознакомиться с результатами других исследований, с тем, чтобы исключить дублирование результатов [1]. Не прекращается разработка новых методов хранения и отображения научной информации. Одним из достаточно новых методов работы с информационными потоками являются онтологии. Они используются, как средство объединения информации от нескольких источников в одной области знаний.

Интеграция информационных ресурсов в единую информационную среду и организация доступа к информационным ресурсам, позволяет унифицировать процесс обмена научными результатами, что повышает эффективность общения групп исследователей. В настоящее время в Интернете активно используются семантические сети и онтологии. Однако, несмотря на это, многие области знаний до сих пор не охвачены. Пользователи не имеют удобного доступа к упорядоченной информации о предметной области, и им приходится собирать данные из множества источников для составления полной картины знаний. Поэтому возникает необходимость создания интеллектуальных научных интернет-ресурсов по определенным тематикам с развитой онтологической структурой для эффективного доступа к информации.

Создание онтологий особо популярно, в частности, в биоинформатике. Они позволяют структурировать данные, обмениваться результатами исследований в конкретной предметной области между представителями различными научных организаций, а также вносить новые данные. В свою очередь использование онтологии в качестве навигационной системы позволяет не только отображать информацию в виде четкой структуры взаимосвязей, но и дополнять имеющиеся данные распределенным методом в соответствии с существующими аксиомами.

Онтология, как основа ресурса, позволяет, помимо структуризации данных, производить первоначальную верификацию данных, исходя из заданных в ней правил. Это помогает улучшить качество вносимой информации, и, в ряде случаев, исключить противоречивые факты.

### Определение онтологического подхода

Онтология – это спецификация концептуализации [2]. Концептуализацией называется некоторая форма представления знаний, содержащая описание объектов предметной области, а также всю информацию, имеющую отношение к объектам (атрибуты, отношения, ограничения, аксиомы, утверждения).

Формально, онтология портала представляет собой семерку вида  $O = \{C, R, T, D, A, F, Ax\}$ , где  $C$  – конечное непустое множество классов, описывающих понятия некоторой предметной области;  $R$  – конечное множество бинарных отношений, заданных на классах (понятиях);  $T$  – множество стандартных типов;  $D$  – множество доменов (множеств значений стандартного типа string);  $A$  – конечное множество атрибутов, описывающих свойства понятий  $C$  и отношений  $R$ ;  $F$  – множество ограничений на значения атрибутов понятий и отношений;  $Ax$  – множество аксиом, определяющих семантику классов и отношений онтологии.

В нашем случае онтология является ядром, базовым компонентом информационной модели портала. Она не только описывает систему знаний портала, но и задает формальные структуры для представления его контента. Онтология содержит понятия моделируемой области, связывающие их отношения, атрибуты понятий и отношений, ограничения на значения атрибутов, а также аксиомы, определяющие семантику понятий и отношений. Формализм, используемый в технологии построения порталов научных знаний, обеспечивает компактное и непротиворечивое описание понятий проблемной и предметной областей портала и разнообразных семантических связей между ними, а также выстраивание понятий в иерархию «общее-частное» и поддержку наследования свойств по этой иерархии. Так на базе информационной модели реализуется навигация, поиск и хранение знаний и данных. Что позволяет многократно использовать знания, пополнять онтологию, редактировать ее при изменении знаний.

### Систематизация информации предметной области

Социально значимой информацией являются данные об инфекционных заболеваниях, переносимых иксодовыми клещами, вызванных их разнообразием и сложной структурой природных очагов этих заболеваний [3]. Для интеграции сведений, полученных разными группами исследователей полученных в различных географических регионах, необходим современный подход, основанный на онтологии. Этот подход мы используем для разработки информационного ресурса по клещевой опасности в Новосибирской области. Этот ресурс позволит не только осветить проблему заражения трудноизлечимыми заболеваниями передаваемых путем укуса клещами, но и предоставить площадку для сбора и обмена информацией по данной проблеме в научных кругах. Различное территориальное расположение и отсутствие единой базы в знаниях являются серьезной помехой в продвижении научной и практической деятельности. Становится актуальной организация эффективного доступа не только к публикациям, описывающим методы и подходы к исследованию генетического разнообразия инфекционных агентов, переносимых клещами разного типа.

Построение интеллектуального ресурса научных знаний – длительный процесс. Важным этапом является изучение предметной области и систематизация полученной информации. Важным звеном является создание информационной модели. Она позволяет унифицировать процесс обмена результатами исследований, а также оперировать с данными и документами, интегрированными в открытое семантическое пространство. Модель позволяет предоставить сервисы преобразования разнородных ресурсов, реализующих средства описания, представления, автоматического связывания ресурсов и их взаимодействие с поисковыми и классификационными механизмами, в соответствии с потребностями пользователя [1]. Известным подходом при создании модели является составление таксономии. Таксономия – это предметная классификация, которая группирует термины и упорядочивает их в виде иерархических структур. Основу классификации составляет выделение ключевых терминов и установление парадигматических отношений между ними (например, типа родитель-потомок) [4, 5]. Таким образом, построенная таксономия будет являться иерархией классов онтологии по отношению вложения.

Объектом нашего внимания является проблема клещевой опасности. На основе таксономии были получены следующие ключевые термины: клещи, территории и местности их распространения, переносимые ими патогены и заболевания. Таксономия также должна включать персон (первооткрывателей, исследователей и других людей); а также методы исследования и полученный ими научный результат.

### Онтология предметной области

Онтология является базовой компонентой ресурса. Для каждого из объектов онтологии задаются отношения, атрибуты и ограничения (рисунок 1).



Рисунок 1 – Онтология информационной модели ресурса

Объектом нашего внимания являются те виды клещей, которые переносят возбудителей заболеваний, наиболее опасных для человека. Это – *Ixodespersulcatus*, *Ixodesricinus*, *Dermacentorreticulatus*, *Dermacentormarginatus* и др. Особое внимание обратим на виды, что обитают в Новосибирской области. Информация о них будет пополняться текущей статистикой укусов и заражений, а также анализом и возможным прогнозом ситуации.

При разработке онтологии информационной модели ресурса особое внимание уделено связям между экземплярами классов онтологии, поскольку данные, указанные в атрибутах и описании, являются основным источником знаний для пользователей. С этой точки зрения важно поддерживать не только навигацию через иерархию наследования, определяемую отношениями *is-A*, *subClassOf* и т.п., но и другими видами отношений [6].

Для каждого вида клещей необходимо определить связанные с ним понятия из иерархий других ключевых терминов. Схема связей между понятиями выглядит следующим образом (рисунок 2):

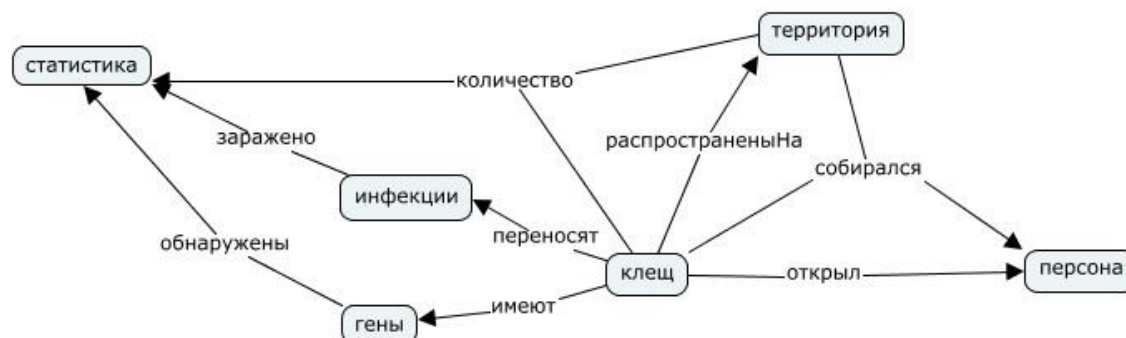


Рисунок 2 – Схема связей между ключевыми терминами

### Архитектура создаваемого ресурса

В основе архитектуры ресурса положена концепция MVC [7], позволяющая произвести разделение данных приложения, пользовательского интерфейса и управляющей логики на три отдельных компонента: модель, представление и контроллер (рисунок 3):

–модель (Model) предоставляет данные и реагирует на команды контроллера, изменяя свое состояние;

–представление (View) отвечает за отображение данных модели пользователю, реагируя на изменения модели;

– контроллер (Controller) интерпретирует действия пользователя, оповещая модель о необходимости изменений.

Выбранный подход позволяет производить независимые манипуляции как с разработкой, модификацией так и с заменой каждого из компонент независимо друг от друга. Функционально ресурс можно разделить на:навигационную поисковую системы, и также редакторы данных и онтологии.

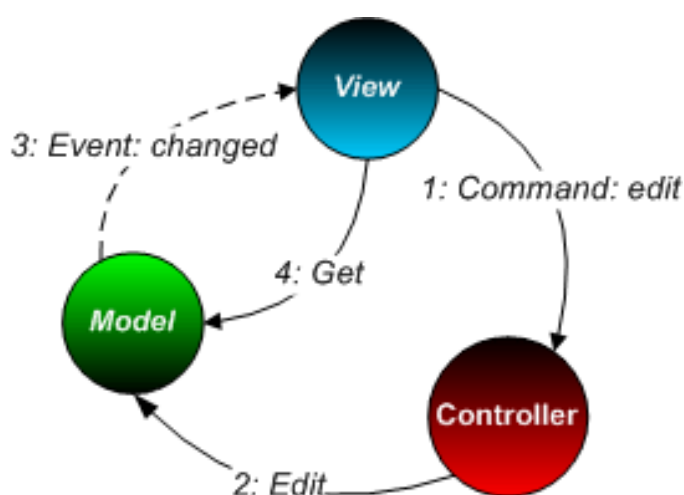


Рисунок 3 – Шаблон проектирования MVC

### Хранилище данных

Для того чтобы контент ресурса и онтология хранились в одном месте необходимо использовать специализированное хранилище данных. Так как должна быть возможность редактирования данных и онтологии после окончания проектирования ресурса, то необходимо использовать RDF-хранилище.

Модель RDF представляет собой утверждение о ресурсе в виде тройки (триплета) – субъект-предикат-объект, описывающего направленную связь от субъекта к объекту, пригодного для машинной обработки. Для обозначения субъектов, отношений объектов используются Uniform Resource Identifier (URI) [8]. В качестве RDF-хранилища был выбран Jena Fuseki Web Server, так как он является бесплатным, мультиплатформенным средством, поддерживающим запрос данных через SPARQL Query Language, изменение данных через SPARQL Update, а также логический вывод для проверки согласованности базы знаний. Взаимодействие с ним проходит через SPARQL HTTP соединение, что позволяет при необходимости перейти на другое RDF-хранилище, поддерживающие такие же функции.

Так как модель RDF служит только для описания данных, но не их обработки, то необходимо использовать сторонние средства. SPARQL – язык запросов, рекомендованный консорциумом W3C. SPARQL-запрос условно можно разделить на две части – префикс и шаблон подграфа, который необходимо найти или изменить в графе.

Таким образом, функциональные компоненты ресурса можно разделить на две группы по типу SPARQL-запросов. Навигационная и поисковая системы только обращаются к базе данных, тогда как редактор данных и онтологии еще и модифицируют ее.

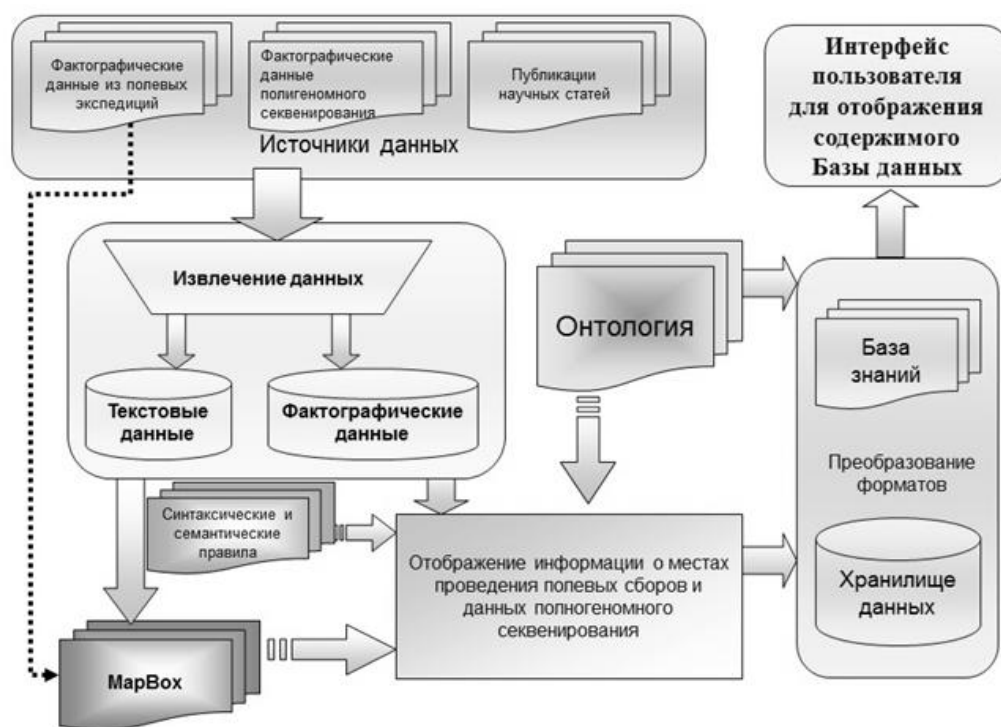


Рисунок 4 – Структура ресурса

### Структура информационной системы

Концептуальным базисом информационной системы создаваемого интернет-ресурса (портала знаний) является описанная выше онтология. Онтология портала вводит формальные описания понятий предметной области в виде классов объектов и отношений между ними, тем самым задавая структуры для представления реальных объектов и их связей. В соответствии с этим данные на портале представлены в виде семантической сети, т.е. как множество разнотипных взаимосвязанных информационных объектов. Содержательный доступ к систематизированным знаниям и информационным ресурсам обеспечивается с помощью информационной системы (ИС), предоставляющей развитые средства навигации и поиска [9]. Архитектуру ИС определяют ее компоненты, их функции и взаимодействие.

Система построена на основе технологии «клиент – сервер» и состоит из клиентской части, серверной, и СУБД MySQL. В описываемой архитектуре клиентом выступает браузер (пользователь), а сервером – веб-сервер. Логика системы распределена между сервером и клиентом, хранение данных осуществляется в СУБД MySQL, обмен информацией происходит по сети. Важным преимуществом подобного подхода является то обстоятельство, что клиенты не зависят от конкретной операционной системы пользователя, следовательно, система представляет собой межплатформенный сервис.

В основе клиентской части приложения построенного на основе структуры (рисунок 4) лежит использование картографического сервиса на основе библиотеки Leaflet. Это инновационная, с открытым исходным кодом, JavaScript библиотека. С ее помощью, отображается векторная карта, которая загружается с сервиса MapBox (<https://www.mapbox.com/>). Сервис MapBox предоставляет большой выбор различных карт, которые могут отличаться по дизайну, языку заголовков на карте и другим параметрам. В качестве источников, для формирования точных карт, MapBox использует сервис OpenStreetMap (<https://www.openstreetmap.org/>). OpenStreetMap — некоммерческий, открытый проект, который предоставляет точные координаты обо всех объектах, которые находятся на карте. Сервис использует OpenStreetMap для определения полигонов областей, которые формируют наглядный интерфейс для пользователя.

Отображение информации на географической карте дополняется элементами статистического анализа данных, накопленных в течение несколько лет [4]. Для этого мы используем библиотеку GoogleCharts (<https://developers.google.com/chart/>). Она написана на JavaScript,

построения комбинированных гистограмм, столбчатых диаграмм, календарных графиков, секторных диаграмм.

### Заключение

На основе концептуальной модели информационной системы определено абстрактное представление сущностей и отношений (связей между сущностями) и обеспечивается поддержка архитектуры универсальной информационной системы, связанной с конкретной областью научных знаний. Концептуальная модель включает в себя основные сущности: акторы (персоны, действующие лица, организации и др. субъекты деятельности, включая компьютерные приложения). Существенной составляющей концептуальной модели являются документы, публикации, словарные статьи, ключевые термины, данные и другие объекты деятельности, включая факты – особый вид документа. В свою очередь факты понимаются как характеристика сущностей, описываемых в онтологии информационной системы, представляемой как единичное значение данных.

По результатам изучения предметной области была выполнена структуризация информации. Вследствие чего была составлена онтология предметной области, являющаяся базовым компонентом при построении информационной модели ресурса.

Исследования выполнены при частичной поддержке гранта РФФИ № 18-07-01457, Интеграционного Проекта СО РАН № АААА-А18-118022190008-8 (№ 0316-2018-0002) и темы г. з. № АААА-А17-117120670141-7 (№ 0316-2018-0009).

### Библиографический список

1. Леонова Ю.В., Федотов А.М. Создание прототипа системы управления информационными ресурсами // Вестник Восточно-Казахстанского гос. Техн. Университета и журнала Вычислительные технологии ИВТ СО РАН. – CITech-2018, Усть-Каменогорск, Казахстан, 25-28 сентября 2018. – С. 47-56.
2. Gruber T. R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing // International Workshop on Formal Ontology, March, Padova, Italy, 1993.
3. Колонин Г.В. Мировое распространение иксодовых клещей. – М.: Наука, 1978. – 70 с.
4. Молородов Ю.И., Ходорченко Д.А. Онтологический подход для построения информационных систем // Материалы Всероссийской конференции с международным участием «Знания – Онтологии – Теории» (ЗОНТ-2017), 2 – 6 октября 2017г. – Новосибирск: ИМ СО РАН, 2017. – Т. 2. – С. 94-99.
5. Загоруйко Ю.А., Загоруйко Г.Б. Онтологии и их практическое применение в системах, основанных на знаниях // Всероссийская конференция с международным участием «Знания-Онтологии-Теории» (ЗОНТ-2011). – Новосибирск: Институт математики им. С.Л. Соболева СО РАН. – 2011. – Т. 1.
6. Грегер С.Э. Проектирование и реализация онтологии навигационной системы сайта // Объектные системы. – 2012. – №1 (6).
7. Model – view – controller. Wikipedia, the free encyclopedia. [Электронный ресурс]. URL: <https://en.wikipedia.org/wiki/Model-view-controller> (дата обращения: 15.05.2018).
8. Головков В., Портнов А., Чернов В. RDF – инструмент для неструктурированных данных // Открытые системы. – 2012. – №. 5. [Электронный ресурс]. URL: <https://www.osp.ru/os/2012/09/13032513> (дата обращения 20.05.2018).
9. Молородов Ю.И., Черненко В.В. Использование библиотеки Leaflet для визуализации и анализа зараженности территорий клещевыми инфекциями // Вычислительные технологии. – 2016. – Т. 21. – № Спец. 1. – С. 75-81.

*Molodov Yu.I., Khodorchenko D.A.,  
Institute of Computational Technologies of Siberian Branch of the Russian Academy of Sciences,  
Novosibirsk*

## DEPLOYING INFRASTRUCTURE INTELLIGENT SCIENTIFIC INTERNET RESOURCE

The article considers one of the ways to deploy the infrastructure of an intellectual scientific Internet resource (INIR) based on the ontological approach. At the same time, ontology serves both for systematization and for the formalization of knowledge and information resources. On the basis of ontology, it is possible to organize convenient navigation and a meaningful search. The user can navigate through the content of the INIR using the ontology tree of con-

cepts. An important feature of the ontological approach is the ability to provide access not only to descriptions of information processing methods and problem solving, but also to their specific implementations.

**Keywords:** knowledge portal, information system, information support system, information resources, data integration.

УДК 528.8.04

*Пестунов И.А., Мельников П.В., Рылов С.А., Синявский Ю.Н.,  
Институт вычислительных технологий Сибирского отделения Российской академии наук,  
Новосибирск*

## ОЦЕНКА КАЧЕСТВА ВСХОДОВ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР ПО RGB-ИЗОБРАЖЕНИЯМ С БПЛА

Предлагается метод, позволяющий по RGB-изображениям, полученным с беспилотных летательных аппаратов, обнаруживать сорняки и получать количественную и качественную оценку всходов сельскохозяйственных культур. Приводятся результаты тестирования, подтверждающие высокую эффективность предлагаемого метода.

**Ключевые слова:** RGB-изображения, БПЛА, индекс TGI, сорняки, автоматизированная оценка качества всходов.

**Введение.** Современные концепции точного земледелия требуют наличия точной информации о состоянии сельскохозяйственных культур. Получение такой информации традиционным способом (с помощью полевых наблюдений) сопряжено со значительными материальными и временными затратами.

В последние годы наблюдается стремительный прогресс в развитии беспилотных летательных аппаратов (БПЛА) и их масштабное проникновение во многие области человеческой деятельности, в том числе и в сельское хозяйство. Применение БПЛА позволяет существенно удешевить и ускорить процесс получения актуальных данных о состоянии посевов с высоким временным и пространственным разрешением. Однако существенным недостатком распространенных БПЛА является их малая грузоподъемность, поэтому на БПЛА зачастую устанавливаются легкие и недорогие фото- и видеокамеры, позволяющие получать изображения только в видимом диапазоне спектра. В связи с этим, в настоящее время RGB-изображения с БПЛА активно используются для оценки качества посевов [1, 2].

В данной работе предлагается автоматизированный метод, позволяющий по RGB-изображениям с БПЛА обнаруживать сорняки и получать количественную и качественную оценку всходов сельскохозяйственных культур.

**Исходные данные.** Для исследований были использованы RGB-изображения всходов льна и подсолнечника, полученные с помощью беззеркальной камеры SONY ILCE-6000 в июне 2017 г. в Алтайском крае специалистами ООО «Беспилотные технологии». Изображения сняты с высоты 150 м, пространственное разрешение – 1,5 см/пикс. Размер каждого изображения составил 6000×4000 пикселей. На рисунке 1 представлены примеры фрагментов таких изображений.

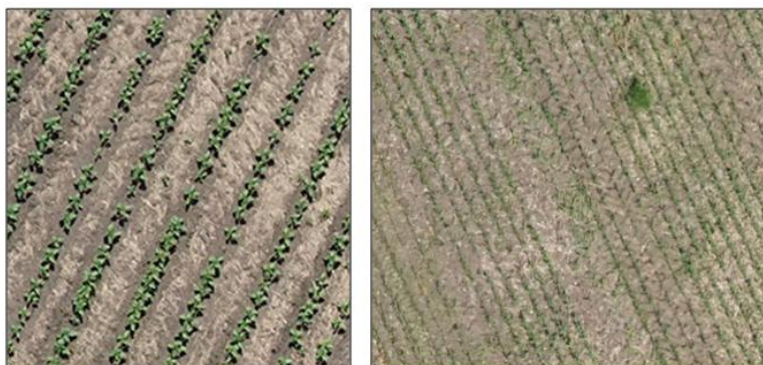


Рисунок 1 – Примеры фрагментов RGB-изображений всходов льна и подсолнечника