

УДК 004.934

DOI 10.25205/1818-7900-2018-16-4-107-114

Ю. И. Молородов^{1,2}, Д. А. Ходорченко²

¹ *Институт вычислительных технологий СО РАН
пр. Академика Лаврентьева, 6, Новосибирск, 630090, Россия*

² *Новосибирский государственный университет
ул. Пирогова, 1, Новосибирск, 630090, Россия*

yumo@ict.sbras.ru, daria.khodorchenko@gmail.com

РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ ПО КЛЕЩЕВОЙ ОПАСНОСТИ НА ОСНОВЕ ОНТОЛОГИИ ПРЕДМЕТНОЙ ОБЛАСТИ *

Предложен подход к разработке состава и структуры Интернет-ресурса на основе онтологии предметной области. Описана предметная область, связанная с распространением клещей и переносимыми ими заболеваниями, представлена ее семантическая структура, выделены основные разделы информационного наполнения и показаны способы навигации по ресурсу.

Ключевые слова: иксодовый клещ, инфекционные агенты, онтологический подход, информационная система.

Введение

Развитие научной деятельности в современном обществе приводит к росту роли компьютерных технологий. Возрос поток информации, что обусловило необходимость использования новых способов ее хранения, представления, формализации, систематизации и автоматической обработки. Появились способы создания баз знаний, позволяющих использовать их для различных практических целей.

Появления технологий Semantic Web¹ привело к появлению систем, способных без участия человека извлекать нужную информацию из текста. Гипертекстовые страницы Semantic Web имеют дополнительную разметку, которая содержит сведения о семантике элементов страницы. Важным компонентом Semantic Web является понятие онтологии, описывающее смысл семантической разметки.

Как правило, под онтологией понимают систему понятий некоторой предметной области, которая представляется как набор сущностей, соединенных различными отношениями [1].

Создание основанного на развитой онтологической структуре интеллектуального научно-го Интернет-ресурса (ИНИР) или портала по определенным тематикам позволит обеспечить эффективный доступ к информации и ее последующую обработку. Онтология как основа ресурса позволяет, помимо структуризации данных, производить первоначальную верификацию данных, исходя из заданных в ней правил [2; 3]. Это помогает улучшить качество вносимой информации и в ряде случаев исключить противоречивые факты.

Социально значимой информацией являются данные об инфекционных заболеваниях, переносимых иксодовыми клещами, вызванных их разнообразием и сложной структурой при-

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

¹ Semantic Web Wikipedia, the free encyclopedia. URL: https://en.wikipedia.org/wiki/Semantic_Web

родных очагов этих заболеваний. Интеграцию сведений в виде информационного ресурса, или информационной системы (ИС), ориентированной на анализ клещевой опасности на основе данных полевых и лабораторных работ, полученных разными группами исследователей, обеспечит использование подхода, основанного на онтологии. Этот ресурс позволит не только осветить проблему заражения трудноизлечимыми заболеваниями, передаваемыми через укус клеща, но и предоставить площадку для сбора и обмена информацией по данной проблеме в научных кругах. Различное территориальное расположение специалистов и отсутствие единой информационной базы являются серьезной помехой в 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, но и к фактическому материалу, связанному с клещами, территорией и средой обитания насекомых. Целью работы является описание подходов к разработке информационного ресурса, основанного на онтологическом подходе, ориентированном на эпидемиологическую опасность, на примере Новосибирской области.

Проблема клещевой опасности

Ареал обитания клещей очень широк. Встречаются виды, обитающие даже в Арктике и Антарктике. Они являются основными переносчиками множества вирусных, бактериальных и протозойных (паразитических простейших) возбудителей заболеваний человека и животных: клещевого энцефалита, боррелиоза, эрлихиоза и др. Вирус клещевого энцефалита (ВКЭ) и боррелии – наиболее социально опасный среди перечисленных возбудителей природно-очаговых инфекций человека [4]².

На территории России встречается около 60 видов иксодовых клещей, основными переносчиками *инфекционных заболеваний* (клещевой энцефалит и клещевой боррелиоз) являются таежный клещ *Ixodes persulcatus*, обитающий в азиатской части страны, и европейский лесной клещ *Ixodes ricinus*, обитающий в европейской части страны. Недавно обнаружена новая разновидность вируса Кемерово [5].

В зависимости от вида клещи могут содержать различные *инфекционные агенты* – вирусы, бактерии и пр. Часто их называют патогенами. Многие переносимые клещами заболевания имеют сходные ранние симптомы, что затрудняет точную диагностику и лечение³. Иногда клещи могут переносить несколько возбудителей заболеваний, а это еще сильнее осложняет постановку диагноза и оказание медицинской помощи. Только полная информация о клещах и переносимых ими инфекциях может помочь врачу правильно диагностировать заболевание и назначить лечение.

Для обеспечения эффективных мер профилактики вирусных заболеваний необходим пространственный и временной анализ распространения клещей, в том числе инфицированных теми или иными возбудителями. Отсутствие эффективной технологии раннего выявления известных и новых патогенов и прогнозирования их распространения является одной из важных и острых проблем. И в этой связи одним из наиболее перспективных способов контроля над возбудителями инфекций может стать система непрерывного наблюдения. Важным шагом к созданию подобной системы является пространственный и временной анализ на основе геоинформационных технологий [6]. Геоинформационная система (ГИС) – система сбора, хранения, анализа и графической визуализации пространственных (географических) данных и связанной с ними информации о необходимых объектах. Понятие ГИС также используется в более узком смысле – как инструмент (программный продукт), позволяющий пользователям искать, анализировать и редактировать как цифровую карту местности, так и дополнительную информацию об объектах.

² Иксодовые клещи // МЧС России. URL: http://www.mchs.gov.ru/dop/info/individual/Bitovie_CHS/item/7816051/ (дата обращения 05.05.2018).

³ Природно-очаговые инфекции / ФБУН Центральный НИИ Эпидемиологии Роспотребнадзора. URL: <https://www.cmd-online.ru/vracham/spravochnik-vracha/prirodno-ochagovye-infektsii/> (дата обращения 05.05.2018).

Онтология клещевой опасности

Основной задачей создаваемой информационной системы является получение, интеграция и предоставление данных и знаний в интересах научных исследований в предметной области. Концептуальной основой систематизации знаний и информации предметной области являются онтологии. Как правило, они исполняют роль модели предметной области [2; 3]. Онтология является ядром, базовым компонентом информационной модели. С ее помощью можно не только описать систему знаний ИНИР, но и создать 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 [7].

Важным моментом при разработке онтологии предметной области является построение таксономий. Таксономия – это предметная (тематическая) классификация, которая группирует термины в виде управляемых словарей (тезаурусов) и упорядочивает их в виде иерархических структур. Разработанная онтология предметной области по клещевой опасности включает четыре базовые иерархические структуры: иерархию объектов, иерархию предметов исследования, иерархию методов исследования и иерархию научных результатов [1]. Свойства каждого понятия описываются с помощью атрибутов и ограничений, налагаемых на область их значений. Понятия базовых онтологий связаны между собой ассоциативными отношениями, выбор которых осуществлялся исходя не только из полноты представления проблемной и предметной областей портала, но и из удобства навигации по его информационному пространству и поиска информации (рис. 1).

Одним из этапов создания ИНИР является изучение предметной области, связанной с клещевой опасностью, и формулирование некоторого представления структуры знаний об этой области. При работе со всем многообразием информации, имеющей отношение к клещам и связанной с ними эпидемиологической опасности, были выделены следующие основные объекты исследования: собственно клещи, местность их распространения (биотоп), инфекционные агенты, переносимые клещами, и гены, которые обнаруживают у клещей и у некоторых инфекционных агентов. Для каждого объекта исследования была составлена таксономия, содержащая элементы, актуальные для предметной области, связанной с клещевой опасностью, на примере Новосибирской области [5; 6].

В результате изучения предметной области были выделены основные понятия: клещ, территория и местность распространения, переносимые клещами патогены и вирусные заболевания. В онтологию были также включены персоны, профессионально связанные с этой областью: первооткрыватели, исследователи, специалисты санитарно-эпидемиологической службы и лица, принимающие решения.

Для каждого вида клещей были определены связанные с ним ключевые понятия. Схема связей между понятиями представлена на рис. 1. Для того чтобы формализовать структуру полученной информации о клещах, была построена иерархическая структура онтологии (рис. 2).

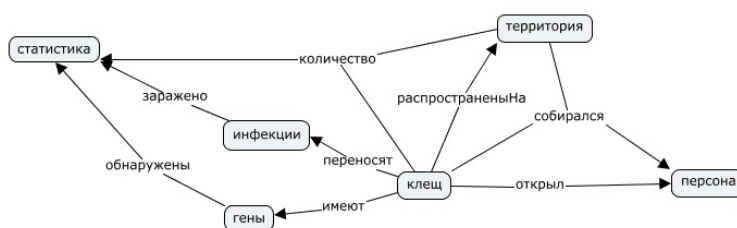


Рис. 1. Схема связей между ключевыми понятиями

Для конкретизации информации об исследуемых объектах, были определены методы исследования выбранных объектов. Наиболее широко используемыми методами являются секвенирование биологического материала и анализ генетической информации. Анализ подразделяется на несколько групп: анализ количества клещей по различным территориям, анализ клещей на одной территории по годам и филогенетический анализ генома клещей [6].

Онтология также содержит информацию о персонах, которые собирали исходный фактический материал, а потом использовали указанные методы. Этими персонами могут быть сборщики информации, которые собирали статистические данные о клещах, либо исследователи, которые в результате обработки и анализа данных получали научный результат, который, как правило, формулировался в виде публикаций.

Архитектура информационной системы

Концептуальным базисом 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 [7]. В соответствии с этим данные на портале представлены в виде семантической сети, т. е. как множество разнотипных взаимосвязанных информационных объектов [3]. Содержательный доступ к систематизированным знаниям и информационным ресурсам обеспечивается с помощью информационной системы, предоставляющей развитые средства навигации и поиска. Архитектуру ИС определяют ее компоненты, их функции и взаимодействие.

Система разработана на основе шаблона проектирования MVC⁴. Моделью является хранилище данных, представлением – пользовательский интерфейс, контроллером – интерпретатор действий пользователя (рис. 3). Использование данного подхода позволяет проводить разработку, модификацию или замену каждого компонента независимо друг от друга.

При работе с ИС можно выделить две важные части – получение данных и изменение данных. В каждой из них можно выделить следующие компоненты: *получение* – навигационная и поисковая системы, *изменение* – редакторы данных и онтологии.

Пользовательский интерфейс реализован с помощью технологии JavaServer Pages⁵. Функциями представления являются навигация по ресурсу, отображение контента и результатов поиска. Информация об объекте онтологии представляется в виде HTML-страницы, где показываются свойства и их значения, а также связи с другими объектами в виде гиперссылки. Отображение данных зависит от прав доступа пользователя. Так, незарегистрированному пользователю доступны только навигация и поиск, а эксперту знаний – весь функционал ресурса. На рис. 4 представлен интерфейс информационного ресурса.

Особое внимание при разработке информационной модели ресурса было уделено связям между экземплярами классов онтологии, поскольку данные, указанные в атрибутах и описаниях, 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 [8].

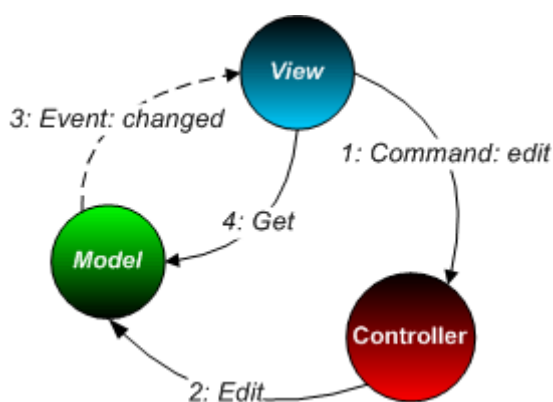


Рис. 3. Шаблон проектирования MVC

⁴ Model – view – controller. [Электронный ресурс] Wikipedia, the free encyclopedia. URL: <https://en.wikipedia.org/wiki/Model-view-controller/>.

⁵ JavaServer Pages Wikipedia, the free encyclopedia. URL: https://ru.wikipedia.org/wiki/JavaServer_Pages/.

[Вход](#) | [Редактор онтологии](#) | [Редактор данных](#)

The interface is divided into two main sections. On the left is a navigation tree with categories like 'Методы исследования', 'Объекты исследования', and 'Персона'. On the right, the 'Свойства объекта' (Object Properties) section shows details for 'Ixodes persulcatus', including its name 'Таежный клещ'. Below this, the 'Связи объекта' (Object Relationships) section lists 'Переносит' (Carries) with links to 'Инфекционные агенты' (Infectious agents) such as 'Borrelia burgdorferi' and 'Tick-borne encephalitis virus', and 'Распространена на' (Distributed in) with links to various Russian regions like 'Алтайский край' and 'Амурская область'.

Рис. 4. Интерфейс информационного ресурса

Онтология как базовый компонент информационной системы должна быть приведена к виду, пригодному для машинной обработки. Базовый компонент был разработан на языке веб-онтологий OWL⁶ в редакторе онтологий Protégé⁷. Для единообразного хранения онтологии и данных было принято решение использовать RDF-хранилище. Был выбран Jena Fuseki Server как бесплатное мультиплатформенное средство, поддерживающее запрос данных через SPARQL Query Language, изменение данных через SPARQL Update, а также логический вывод для проверки согласованности базы знаний⁸.

Контроллер связывает между собой модель и представление, интерпретируя действия пользователя для отображения или изменения данных. Контроллер реализован через Java Servlet интерфейс. Так как модель RDF служит только для описания данных, но не их обработки, то необходимо использовать сторонние средства. Был использован язык запросов SPARQL⁹. Функциональные компоненты ресурса были разделены по типу SPARQL-запросов на две группы. Навигационная и поисковая системы только обращаются к базе данных, а редактор данных и онтологии еще и модифицирует её.

Заключение

Построена концептуальная модель информационной системы. На ее основе определено представление сущностей и отношений (связей между сущностями), и обеспечивается поддержка архитектуры универсальной информационной системы, связанной с конкретной областью научных знаний, ориентированных на эпидемиологическую обстановку конкретного ареала. Концептуальная модель включает в себя основные сущности: методы и объекты исследований, научный результат, инфекционные агенты, иксодовые клещи, персоны и публикации. Важной составляющей концептуальной модели являются публикации и данные,

⁶ Semantic Web Wikipedia, the free encyclopedia. URL: https://en.wikipedia.org/wiki/Semantic_Web

⁷ URL: <http://protege.stanford.edu/>.

⁸ URL: http://jena.apache.org/documentation/serving_data/.

⁹ URL: <https://en.wikipedia.org/wiki/SPARQL/>.

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, ориентированной на эпидемиологическую проблему, вызванную инфицированными иксодовыми клещами, была выполнена структуризация информации. На основе структуризации составлена онтология предметной этой области, которая стала базовым компонентом при построении информационной модели ресурса.

Список литературы

1. Gruber T. R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing // International Workshop on Formal Ontology, March. Padova, Italy, 1993.
2. Загорулько Ю. А., Загорулько Г. Б., Шестаков В. К., Кононенко И. С. Концепция и архитектура тематического интеллектуального научного интернет-ресурса // Тр. XV Всерос. науч. конф. RCDL'2013. Ярославль: ЯрГУ, 2013. С. 57–62.
3. Загорулько Ю. А. Технология разработки интеллектуальных научных Интернет-ресурсов, ориентированная на экспертов предметной области // Инфраструктура научных информационных ресурсов и систем: Сб. избр. науч. ст. Труды Четвертого Всероссийского симпозиума / Под ред. Е. В. Кудашева, В. А. Серебрякова. М.: ВЦ РАН, 2014. Т. 1. С. 69–86.
4. Колонин Г. В. Мировое распространение иксодовых клещей. М.: Наука, 1978. 70 с.
5. Tkachev S., Panov V., Dobler G., Tikunova N. First detection of Kemerovo virus in *Ixodes pavlovskyi* and *Ixodes persulcatus* ticks collected in Novosibirsk region, Russia // Ticks tick-borne dis. 2014. No. 5. P. 494–496.
6. Molorodov Yu. I., Chernenko V. V., Fedotov A. M. Web-based repository for spatial monitoring viral tick-borne pathogens // Eurasian Journal of Mathematical and Computer Applications. 2016. Vol. 4, № 4. P. 37–43.
7. Загорулько Ю. А., Загорулько Г. Б. Онтологии и их практическое применение в системах, основанных на знаниях // Всероссийская конференция с международным участием «Знания – Онтологии – Теории» (ЗОНТ-2011) / Ин-т математики им. С. Л. Соболева СО РАН. Новосибирск, 2011. Т. 1.
8. Грегер С. Э. Проектирование и реализация онтологии навигационной системы сайта // Объектные системы. 2012. № 1 (6).

Материал поступил в редколлегию 17.06.2018

Yu. I. Molorodov^{1,2}, D. A. Khodorchenko²

¹ Institute of Computational Technologies SB RAS
6 Academician Lavrentiev Ave., Novosibirsk, 630090, Russian Federation

² Novosibirsk State University
1 Pirogov St., Novosibirsk, 630090, Russian Federation

yumo@ict.sbras.ru, daria.khodorchenko@gmail.com

DEVELOPMENT OF THE INFORMATION SYSTEM FOR TICK-HAZARD ON THE BASIS OF THE OBJECTIVE ONTOLOGY

The paper proposes an approach to the development of the composition and structure of an Internet resource, based on the domain ontology. The subject area associated with the spread of ticks and the diseases they transmit is described, its semantic structure is presented, the main sections of the content are highlighted, and ways to navigate the resource are shown.

Keywords: ixodes tick, pathogens, ontology, information system, portal of knowledge.

References

1. Gruber T. R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. *International Workshop on Formal Ontology*. March, Padova, Italy, 1993.
2. Zagorulko Yu. A., Zagorulko G. B., Shestakov V. K., Kononenko I. S. Concept and architecture of thematic intelligent scientific Internet resource. *Proc. of the XV All-Russian Scientific Conference RCDL'2013*. Yaroslavl, YarSU Publ., 2013, p. 57–62. (in Russ.)
3. Zagorulko Yu. A. Technology of development of intellectual scientific Internet resources, focused on experts of the subject domain. *Infrastructure of scientific information resources and systems. Collection of selected scientific articles. Proceedings of the Fourth All-Russian Symposium* Eds. E. V. Kudasheva, V. A. Serebryakov. Moscow, Computing Center of the Russian Academy of Sciences, 2014, vol. 1, p. 69–86. (in Russ.)
4. Kolonin G. V. World distribution of ixodic ticks. Moscow, Science, 1978, 70 p. (in Russ.)
5. Tkachev S., Panov V., Dobler G., Tikunova N. First detection of Kemerovo virus in *Ixodes pavlovskiyi* and *Ixodes persulcatus* ticks collected in Novosibirsk region, Russia. *Ticks tick-borne dis*, 2014, no. 5, p. 494–496.
6. Molorodov Yu. I., Chernenko V. V., Fedotov A. M. Web-based repository for spatial monitoring viral tick-borne pathogens. *Eurasian Journal of Mathematical and Computer Applications*, 2016, vol. 4, № 4, p. 37–43.
7. Zagorulko Yu. A., Zagorulko G. B. Ontologies and their practical application in systems based on knowledge. *All-Russian Conference with international participation "Knowledge-Ontologies-Theories" (ZONT-2011)*. Novosibirsk, S. Sobolev Institute of Mathematics SB RAS, 2011, vol. 1. (in Russ.)
8. Greger S. E. Design and implementation of the ontology of the navigation system of the site. *Object systems*, 2012, № 1 (6). (in Russ.)

Received 17.06.2018

Acknowledgements:

The work is supported by RFBR (grant 18-07-01457), projects AAAA-A18-118022190008-8 (0316-2018-0002) and AAAA-A17-117120670141-7 (0316-2018-0009).

For citation:

Molorodov Yu. I., Khodorchenko D. A. Development of the Information System for Tick-Hazard on the Basis of the Objective Ontology. *Vestnik NSU. Series: Information Technologies*, 2018, vol. 16, no. 4, p. 107–114. (in Russ.) DOI 10.25205/1818-7900-2018-16-4-107-114