

## Проблемы разработки онтологии для тематического интеллектуального научного интернет-ресурса<sup>1</sup>

<sup>1,2</sup> Загорулько Ю. А., <sup>1</sup> Загорулько Г. Б., <sup>1</sup> Боровикова О. И.  
<sup>1</sup> *Институт систем информатики имени А. П. Ершова, СО РАН  
просп. Лаврентьева, 6, г. Новосибирск, 630090, Российская Федерация*  
<sup>2</sup> *Новосибирский государственный университет  
ул. Пирогова, 2, г. Новосибирск, 630090, Российская Федерация  
zagor@iis.nsk.su*

Статья поступила 25 октября 2016 г.

**Аннотация:** *Рассматриваются проблемы и особенности разработки онтологии для тематического интеллектуального научного интернет-ресурса. В такого рода ресурсах онтология используется не только для формализации и систематизации разнородных знаний, данных и средств их обработки, но и для организации удобного доступа к ним и их использования. Построение онтологии, обеспечивающей указанные выше функции, является довольно сложным и трудоемким процессом. Для его облегчения предложена методология, включающая набор методик и паттернов разработки онтологий, а также метод построения онтологии на основе представительного множества базовых онтологий.*

**Ключевые слова:** *интеллектуальный научный интернет-ресурс, онтология, технология Semantic Web, паттерн разработки онтологии.*

**Для цитирования (ГОСТ 7.0.5—2008):** Загорулько Ю. А., Загорулько Г. Б., Боровикова О. И. Проблемы разработки онтологии для тематических интеллектуальных научных интернет-ресурсов // *Инфокоммуникационные и радиоэлектронные технологии*. 2018. Т. 1, № 2. С. 207—217.

**Для цитирования (ГОСТ 7.0.11—2011):** Загорулько, Ю. А. Проблемы разработки онтологии для тематических интеллектуальных научных интернет-ресурсов / Ю. А. Загорулько, Г. Б. Загорулько, О. И. Боровикова // *Инфокоммуникационные и радиоэлектронные технологии*. — 2018. — Т. 1, № 2. — С. 207—217.

---

<sup>1</sup> Статья является расширенной версией доклада, представленного на 26-й Международной Крымской конференции «СВЧ-техника и телекоммуникационные технологии» — КрыМиКо'2016 (Севастополь, РФ, 4—10 сентября 2016 г.).

# Problems of development of ontology for subject-based intelligent scientific Internet resource

Yu. A. Zagorulko<sup>1,2</sup>, G. B. Zagorulko<sup>1</sup>, and O. I. Borovikova<sup>1</sup>

<sup>1</sup>A. P. Ershov Institute of Informatics Systems, RAS  
6, Lavrientiev Ave., Novosibirsk, 630090, Russian Federation

<sup>2</sup>Novosibirsk State University  
2, Pirogova Str., Novosibirsk, 630090, Russian Federation  
zagor@iis.nsk.su

Received on October 25, 2016

**Abstract:** *The present paper concerns the problems and main features of the development of ontology for subject-based intelligent scientific Internet resources. In these resources the ontology is used not only for formalization and systematization of the heterogeneous knowledge, data and information processing facilities, but also for organization of a convenient access to them. Development of ontology providing the above-mentioned functions is a very difficult and laborious process. To facilitate it the methodology including a set of techniques and ontology design patterns, as well as ontology building method using a representative set of base ontologies, is suggested.*

**Keywords:** *intelligent scientific internet resources, ontology, Semantic Web technology, ontology design pattern.*

**For citation (IEEE):** Yu. A. Zagorulko, G. B. Zagorulko, and O. I. Borovikova “Problems of development of ontology for subject-based intelligent scientific Internet resource,” *Infocommunications and Radio Technologies*, vol. 1, no. 2, pp. 207–217, 2016. (In Russ.). doi: 10.15826/icrt.2018.01.2.16

## 1. Введение

В своей повседневной деятельности ученый должен иметь удобный содержательный доступ ко всем имеющимся на данный момент знаниям и данным, относящимся к области его исследований, а также к разработанным в ней средствам обработки и анализа информации. При этом желательно, чтобы вся эта информация представлялась в виде сети знаний и данных как наиболее естественной и удобной форме подачи информации для человека.

В такой сети первый слой (слой знаний) составляют понятия моделируемой области знаний, выстроенные в иерархию «общее — частное», а второй (слой данных) — связанные между собой различными отношения-

ми информационные объекты, т. е. экземпляры понятий, представляющие реальные или виртуальные объекты моделируемой области знаний. Например, научные организации, персоны, разработанные в рамках данной области знаний модели, методы анализа данных, включая их реализации в виде web-сервисов, и т. п. Если по такой сети организовать навигацию, то пользователь получит удобный доступ к знаниям и данным.

Для решения этой проблемы была предложена концепция и архитектура тематического интеллектуального научного интернет-ресурса (ИНИР) [1], обеспечивающего содержательный доступ к систематизированным научным знаниям и информационным ресурсам заданной области знаний, а также к средствам их интеллектуальной обработки и анализа.

Основу такого ИНИР составляет онтология [2], которая служит не только для формализации и систематизации различных видов знаний, данных и средств обработки и анализа информации, интегрируемых в ИНИР, но и для организации удобного содержательного доступа к ним.

Построение такой онтологии является довольно сложным и трудоемким процессом. Чтобы облегчить его, предложена методология, включающая набор методик и паттернов разработки онтологий, а также метод построения онтологии ИНИР на основе представительного множества базовых онтологий. В статье подробно описываются основные характеристики данной методологии.

## **2. Структура и методы построения онтологии ИНИР**

Как было сказано выше, онтология составляет ядро системы знаний ИНИР и наряду с описанием различных аспектов моделируемой области знаний содержит соотнесенное с ним описание структуры и типологии интегрируемых информационных ресурсов и методов обработки и анализа данных. В связи с этим онтология ИНИР состоит из системы взаимосвязанных онтологий, отвечающих за представление указанных выше трех компонентов знаний, а именно: онтологии области знаний ИНИР, онтологии задач и методов, онтологии научных интернет-ресурсов.

Онтология области знаний задает систему понятий и отношений, предназначенных для детального описания области знаний ИНИР и выполняемой в ее рамках научной и исследовательской деятельности.

Онтология задач и методов описывает задачи, для решения которых предназначен ИНИР, и методы их решения.

Онтология научных интернет-ресурсов служит для описания представленных в сети Интернет информационных ресурсов, релевантных области знаний ИНИР.

Построение таких сложных онтологий «с нуля» — непростая задача, поэтому был предложен метод построения онтологий ИНИР на основе небольшого, но представительного набора базовых онтологий, включающих только самые общие сущности, не зависящие от конкретных областей знаний ИНИР. В этот набор входят: (1) онтология научного знания, (2) онтология научной деятельности, (3) базовая онтология задач и методов, (4) базовая онтология научных информационных ресурсов.

Онтология научного знания (см. рис. 1) содержит классы, задающие структуры для описания понятий конкретных областей знаний, такие как *Раздел науки*, *Метод исследования*, *Объект исследования*, *Предмет исследования*, *Научный результат* и др. Эта онтология также включает отношения, связывающие между собой объекты указанных выше классов. Используя эти классы, можно выделить и описать значимые для моделируемой области знаний разделы и подразделы, задать типизацию методов и объектов исследования, описать результаты научной деятельности.

Онтология научной деятельности (см. рис. 1) базируется на онтологии, предложенной в [3] для описания научно-исследовательских проектов и расширенной для применения к более широкому классу задач. Эта онтология включает классы понятий, относящиеся к организации научной и исследовательской деятельности, такие как *Персона*, *Организация*, *Событие*, *Научная деятельность*, *Проект*, *Публикация* и др. Онтология научной деятельности включает также отношения, позволяющие связывать понятия данной онтологии не только между собой, но и с понятиями онтологии научного знания.

Базовая онтология научных информационных ресурсов включает класс *Информационный ресурс* в качестве основного класса. Набор атрибутов и связей этого класса основан на стандарте Dublin core [4]. Его атрибутами являются: название ресурса, язык ресурса, тематика ресурса, тип ресурса, дата создания ресурса и др. Для представления информации об источниках ресурса и его создателях, а также связанных с ним событиях, организациях, персонах, публикациях и других сущностях вводятся специальные отношения, связывающие класс *Информационный ресурс* с классами других онтологий ИНИР.

Базовая онтология задач и методов включает такие классы, как *Задача*, *Метод решения* и *Web-сервис*. С помощью понятий и отношений данной онтологии могут быть описаны задачи, для решения которых предназначен ИНИР, методы их решения и реализующие их web-сервисы.

На основе первых двух онтологий строится онтология области знаний ИНИР, на основе третьей — онтология задач и методов, на основе четвертой — онтология научных интернет-ресурсов (см. табл. 1).

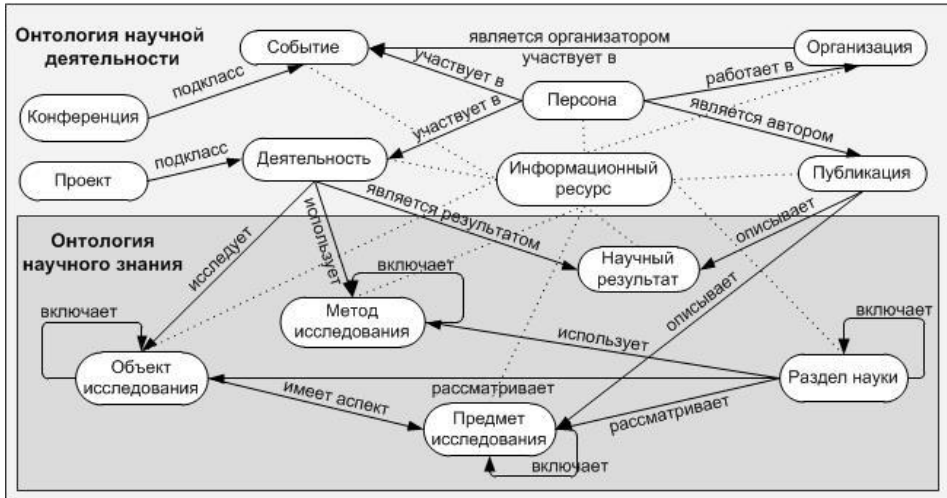


Рис. 1. Онтология научной деятельности и онтология научного знания.

Fig. 1. The ontology of research activity and the ontology of scientific knowledge

Таблица 1. Онтологии ИНИР и базовые онтологии.

Table 1. The ISIR ontologies and base ontologies

Онтологии ИНИР	Используемые базовые онтологии
Онтология области знаний	Онтология научной деятельности Онтология научного знания
Онтология задач и методов	Базовая онтология задач и методов
Онтология научных интернет-ресурсов	Базовая онтология научных интернет-ресурсов

Использование такой методологии значительно упрощает создание онтологии ИНИР и ее дальнейшее сопровождение.

Заметим, что система знаний ИНИР включает также тезаурус, который содержит описание терминов области знаний ИНИР, т. е. слова и словосочетания, с помощью которых понятия онтологии представляются в текстах и пользовательских запросах. Тезаурус также строится на основе специальной базовой онтологии — онтологии представления тезауруса.

### 3. Представление сложных сущностей

При реализации основных компонентов ИНИР, в том числе онтологии, были использованы средства технологии Semantic Web [5], хорошо

зареккомендовавшие себя при разработке не только интеллектуальных интернет-приложений, но и прикладных интеллектуальных систем.

Технология Semantic Web была выбрана прежде всего потому, что она предоставляет достаточно удобные и выразительные средства представления знаний и данных, в частности языки RDF, RDF(S) и OWL [2], которые стали уже стандартными средствами описания онтологий, что обеспечивает возможность понимания и переиспользования онтологий другими разработчиками.

Кроме того, применение языка OWL [2, 6] позволяет использовать существующие в свободном доступе машины вывода, с помощью которых можно не только выводить новые знания, но и контролировать корректность и целостность всей системы знаний ИНИР.

Несмотря на указанные выше неоспоримые преимущества средств технологии Semantic Web, при их использовании возникали проблемы, связанные с тем, что предложенная концепция представления знаний и данных в ИНИР и их визуализации пользователю не всегда удобно укладывалась в возможности этих средств.

Основная трудность спецификации онтологии средствами языка OWL состоит в том, что OWL обеспечивает представление только простых сущностей, в то время как в онтологии необходимо представлять сложные понятия и отношения. Для этих целей был разработан набор оригинальных методик и паттернов разработки онтологий, использование которых не только упрощает разработку онтологии, но и повышает уровень контроля вводимых в нее данных, а также обеспечивает удобное представление информации конечному пользователю.

Заметим, что описанный в настоящей статье набор методик и паттернов был использован при построении базовых онтологий ИНИР.

### **3.1. Описание областей допустимых значений**

При описании классов онтологии необходимо задавать область допустимых значений его атрибутов. Если такая область имеет конечное число значений, то ее называют доменом. Домен характеризуется названием и множеством элементарных значений. Введение доменов не только позволяет контролировать ввод информации, но и повышает удобство этой операции — инженер знаний или эксперт вместо того, чтобы каждый раз вспоминать (или придумывать) значение того или иного атрибута объекта, просто выбирает его из списка значений, заданного в домене.

Примерами таких доменов являются «Должность», «Тип организации», «Тип публикации», «Язык публикации», которые включают соот-

ответственно виды должностей в организации, типы организаций, типы и языки публикаций.

Так как в OWL для задания указанного вида доменов нет специальных средств, было предложено представлять каждый домен отдельным классом, заданным перечислением своих элементов (индивидов). Все такие классы являются наследниками служебного класса «Домен» (см. рис. 2), который, в свою очередь, является наследником класса «Служебный класс».

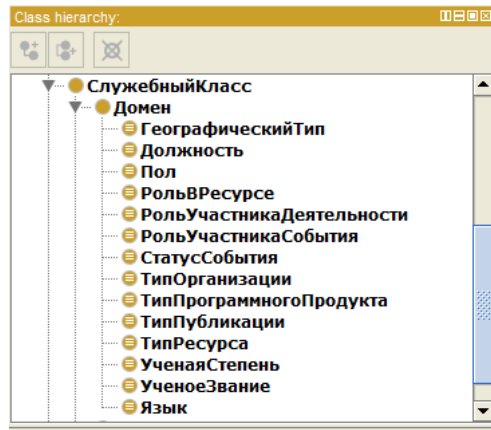


Рис. 2. Служебный класс *Домен*.

Fig. 2. *Domain* service class

Достоинством такого способа задания области допустимых значений является и то, что он позволяет поддерживать многоязычность онтологии. Действительно, с каждым значением домена могут быть связаны метки (*label*), представляющие его на разных языках.

### 3.2. Представление сложноструктурированных сущностей

При построении онтологии часто требуется представлять сущности, имеющие сложную структуру. Так, при описании научной деятельности часто возникает потребность в атрибутированных отношениях, т. е. в бинарных отношениях, снабженных атрибутами, уточняющими связь между аргументами отношения.

Примером атрибутированного отношения является отношение «участвует Персона в Деятельности» из онтологии научной деятельности:

*участвует Персона в Деятельности*  
*arg1: Персона: Нариньяни А. С.*  
*arg2: Деятельность: проект InDoc*

роль: научный руководитель,  
дата начала: 2000,  
дата окончания: 2003.

Аргументами такого отношения являются «персона» и «деятельность», а атрибутами — «роль», «дата начала» и «дата окончания», служащие для задания дополнительной информации об отношении.

Так как язык OWL не позволяет задавать атрибуты для отношений, нами был предложен специальный паттерн.

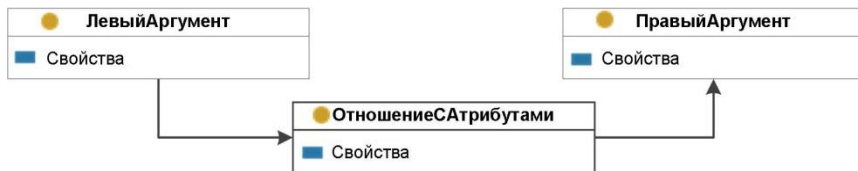


Рис. 3. Паттерн атрибутированного отношения.

Fig. 3. Attributed relation pattern

Для представления атрибутированных отношений данный паттерн предусматривает введение служебного класса «ОтношениеСАтрибутами», а для представления каждого конкретного отношения такого типа вводится класс, являющийся наследником этого класса.

Заметим, что этот паттерн отличается от паттерна “qualified relation” [7], предложенного в сообществе Linked Data, тем, что в нем явно указывается порядок аргументов отношения. Благодаря этому сохраняется информация об ориентированности отношения, что важно для представления пользователю полной информации о характере связи между объектами.

Аналогичным образом строятся паттерны многоместных отношений, паттерны для представления проектов, событий и др.

#### 4. Описание методов визуализации информации для пользователя средствами онтологии

Для пользователей очень важно удобное представление содержащейся в контенте ИНИР информации. В связи с этим нами был предложен способ отображения этой информации на экране монитора с помощью шаблонов визуализации, строящихся с использованием свойств, объявленных в онтологии ИНИР как *AnnotationProperty*. Всего было введено три таких свойства: *order*, *link* и *shortlink*.

Шаблоны визуализации задаются для каждого класса онтологии. Различают шаблоны визуализации объектов класса и шаблоны визуализации ссылок на объекты.



Шаблон визуализации объектов класса включает все его атрибуты и связанные с ним отношения. По умолчанию атрибуты класса и связанные с ним отношения, в том числе и их атрибуты, отображаются в том порядке, в котором они заданы в онтологии. По желанию пользователя порядок показа атрибутов может быть изменен (см. рис. 4). Для этого с помощью свойства *order* задаются порядковые номера атрибута.

Шаблон визуализации ссылки на объект какого-либо класса может включать как атрибуты данного класса, так и атрибуты связанных с ним классов и заданных между ним и другими классами отношений.

Вводится два вида ссылок на объект — полные и краткие. Полные ссылки используются при отображении списка объектов заданного класса, краткие — при отображении ссылки на объект при визуализации какого-либо другого объекта.

Для описания полных ссылок используется свойство *link*, для кратких — *shortlink*. Значением этих свойств является целое число, задающее порядок следования компонентов в ссылке.

Фамилия	Боровикова
Предыдущая фамилия	Россеева
Имя	Олеся
Отчество	Игнатьевна
<b>работает в Организации</b>	
<b>Организация</b>	
<a href="#">Институт систем информатики им. А.П. Ершова СО РАН ИСИ СО РАН</a>	
Должность	младший научный сотрудник
Дата начала	1999
<b>развивает Метод</b>	
<b>Метод исследования</b>	
<a href="#">Онтологическое моделирование</a>	
<b>участвует Персона в Деятельности</b>	
<b>Деятельность</b>	
<a href="#">Методы и технологии создания и управления интеллектуальными научными Интернет-ресурсами на основе онтологий и семантических сетей</a>	
Роль	исполнитель
Начало участия	2014
Завершение участия	2015

Рис. 4. Визуализация объекта класса «Персона».

Fig. 4. Visualization of an object of “Person” class

## 5. Заключение

В статье обсуждены проблемы и особенности построения онтологии для тематических ИНИР с использованием технологии Semantic Web.

Для упрощения процесса создания онтологии предложена методология, включающая набор методик и паттернов разработки онтологий, а также метод построения онтологии ИНИР на основе множества базовых онтологий.

Предложенные паттерны поддерживают стандартизацию представления сложных сущностей и благодаря этому обеспечивают лучшее понимание онтологий разными людьми, делая возможной коллективную разработку онтологий и повышая уровень их переиспользуемости.

Наличие довольно представительного набора базовых онтологий и метода построения всех онтологий ИНИР на их основе значительно снижает сложность и трудоемкость процесса создания ИНИР.

### Источники финансирования и выражение признательности

Исследования, посвященные разработке онтологий для тематических интеллектуальных научных интернет-ресурсов, были поддержаны Российским фондом фундаментальных исследований (гранты 13-07-00422 и 16-07-00569).

### Список литературы

1. Загоруйко Ю. А., Загоруйко Г. Б., Боровикова О. И. Технология создания тематических интеллектуальных научных интернет-ресурсов, базирующаяся на онтологии // Программная инженерия. 2016. Т. 7, № 2. С. 51—60.
2. Лапшин В. А. Онтологии в компьютерных системах. М.: Научный мир, 2010. 224 с.
3. Benjamins V. R., Fensel D. Community is Knowledge! in (KA)2 // Proc. of 11<sup>th</sup> Banff Knowledge Acquisition for Knowledge-based Systems workshop KAW'98 (Banff, Canada, April 1998). — Calgary : University of Calgary, 1998. С. KM.2-1—KM.2-18.
4. Hillmann D. Using Dublin Core [Electronic resource]. Режим доступа: <http://dublincore.org/documents/usageguide/> (дата обращения 25.10.2016).
5. Hitzler P., Krötzsch V., Rudolph S. Foundations of Semantic Web Technologies. Chapman & Hall/CRC, 2009. 455 с.
6. Antoniou G., Harmelen F. Web Ontology Language: OWL // Handbook on Ontologies. 2009. С. 91—110.
7. Dodds L., Davis I. Linked Data Patterns. 2012 [Electronic resource]. Режим доступа: <http://patterns.dataincubator.org/book> (дата обращения 25.10.2016).

### References

- [1] Y. A. Zagorulko, G. B. Zagorulko, and O. I. Borovikova, “Technology for Building Subject-Based Intelligent Scientific Internet Resources Based on Ontology,” *Softw. Eng.*, vol. 7, no. 2, pp. 51–60, 2016. (In Russ.). doi: 10.17587/prin.7.51-60

- [2] V. A. Lapshin, *Ontologii v kompyuternih sistemah* [Ontology in computer systems]. Moscow: Nauchnyi Mir, 2010. (In Russ.).
- [3] V. R. Benjamins and D. Fensel, "Community is Knowledge! in (KA)2," in *Proc. of 11<sup>th</sup> Banff Knowledge Acquisition for Knowledge-based Systems workshop KAW'98 (Banff, Canada, April 1998)*, 1998, p. KM.2-1–KM.2-18.
- [4] D. Hillmann, "Using Dublin Core." [Online]. Available: <http://dublincore.org/documents/usageguide/>
- [5] P. Hitzler and M. Krötzsch..., *Foundations of semantic web technologies*. Chapman & Hall, CRC Press, 2009.
- [6] G. Antoniou and F. Harmelen, "Web ontology language: OWL," in *Handbook on ontologies*, 2009, pp. 91–110. doi: 10.1007/978-3-540-92673-3\_4
- [7] L. Dodds and I. Davis, "Linked Data Patterns," 2012. [Online]. Available: <http://patterns.dataincubator.org/book>.

### **Acknowledgement**

The studies devoted to the development of ontologies for subject-based intelligent scientific Internet resources have been supported by the Russian Foundation for Basic Research (grants 13-07-00422 and 16-07-00569).