

## РАЗРАБОТКА СИСТЕМЫ ДОСТУПА К ДАННЫМ ИЗ ВНЕШНИХ ИСТОЧНИКОВ ДЛЯ ИНТЕЛЛЕКТУАЛЬНЫХ НАУЧНЫХ ИНТЕРНЕТ-РЕСУРСОВ<sup>1</sup>

**Загорулько Галина Борисовна**

Научный сотрудник, Институт систем информатики им. А.П. Ершова СО РАН,  
630090 г. Новосибирск, пр. Лаврентьева 6, e-mail: gal@iis.nsk.su

**Молородов Юрий Иванович**

К.ф.-м.н., с.н.с., Институт вычислительных технологий СО РАН,  
630090 г. Новосибирск, пр. Лаврентьева 6, e-mail: yumo@ict.sbras.ru

**Вишнев Кирилл Евгеньевич**

Магистрант, Новосибирский государственный университет,  
630090 г. Новосибирск, ул. Пирогова 1, e-mail: vishnev\_kirill@mail.ru

**Аннотация.** Описаны средства, разработанные для интеллектуальных научных интернет-ресурсов, которые предназначены для визуализации и обработки данных, хранящихся во внешних источниках. Представлена архитектура системы доступа к внешним источникам, схема её функционирования и используемые для её создания подходы, а также методика и примеры использования предложенных средств.

**Ключевые слова:** интеллектуальный научный Интернет-ресурс, онтология, сервис, внешнее хранилище данных, система доступа к данным.

**Цитирование:** Загорулько Г.Б., Молородов Ю.И., Вишнев К.Е. Разработка системы доступа к данным из внешних источников для интеллектуальных научных интернет-ресурсов // Информационные и математические технологии в науке и управлении. 2018. № 3 (11). С. 127–135. DOI:10.25729/2413-0133-2018-3-14

**Введение.** В современных научных сообществах стоит проблема разработки информационно-аналитических ресурсов, которые бы систематизировали накопленные знания и фактический материал в определенных областях знаний. Разработка средств, поддерживающих создание таких ресурсов, является актуальной задачей. Существует множество подходов, методов и средств её решения [1–3, 7]. Одним из таких средств, хорошо зарекомендовавшим себя во многих научных и практических областях, является технология разработки интеллектуальных научных интернет-ресурсов (ИНИР) [5]. Эта технология основывается на онтологическом подходе и предназначена, в первую очередь, для описания слабоструктурированных областей, для представления семантических зависимостей понятий и систематизации разнородной информации и средств её обработки. Однако потребность в систематизации информации возникает и в областях, традиционно считающихся хорошо формализованными. В таких областях, как правило, имеются аналитические модели и большие объемы используемых данных. В то же время в слабоструктурированных областях зачастую появляется потребность в представлении и обработке данных, хранящихся во внешних источниках.

<sup>1</sup> Исследования выполнены при частичной поддержке грантов РФФИ № 16-07-00569, № 18-07-01457, Инт.П СО РАН № АААА-А18-118022190008-8 и ПФНИ № АААА-А17-117120670141-7

В статье представлены средства, разработанные для типового ИНИР и предназначенные для доступа к данным из внешних источников. Кратко описаны функциональные возможности технологии разработки ИНИР. Основное внимание уделено системе доступа к внешним данным, её архитектуре, схеме функционирования. Описаны используемые для создания этой системы подходы. Методика и примеры использования предложенных средств продемонстрированы на примере ИНИР по теплофизическим свойствам химических веществ.

**1. Технология разработки интеллектуальных научных интернет-ресурсов.** ИНИР представляет собой систему с web-интерфейсом, которая содержит систематизированную информацию, относящуюся к определенной области знаний, и предоставляет содержательный доступ к этой информации, к методам ее обработки и методам решения задач, принятым в данной области знаний (ОЗ), а также к относящимся к ней интернет-ресурсам [5]. Основным компонентом ИНИР является онтология ОЗ. На её основе осуществляется систематизация информации данной ОЗ и функционирование ИНИР.

Технология разработки предоставляет методику построения онтологии, набор базовых онтологий, оболочку ИНИР, средства спецификации пользовательского интерфейса, редактор данных и набор сервисов, обеспечивающих функциональность ресурса.

Для разработки онтологии используются средства Semantic Web [9], редактор Protégé [10] и методика, предложенная авторами технологии разработки ИНИР. Данная методика ориентирована на использование базовых онтологий и паттернов онтологического проектирования [8].

При разработке оболочки ИНИР был использован сервис-ориентированный подход. Согласно этому подходу, вся функциональность ИНИР реализуется с помощью сервисов – локальных либо распределённых, слабо связанных, заменяемых компонентов, оснащённых стандартизированными интерфейсами для взаимодействия по стандартизированным протоколам. Такой подход позволяет разработчикам ресурса создавать различные сервисы для обработки информации, хранящейся как в контенте ИНИР, так и во внешних хранилищах, а также использовать сервисы сторонних разработчиков.

На рис. 1 представлен пользовательский Web-интерфейс ИНИР «Теплофизические свойства химических веществ» [6].

В левой области страницы показан верхний уровень иерархии понятий онтологии. В центральной части находится описание конкретного объекта, являющегося экземпляром выбранного в онтологии понятия. Этот объект представляет «Научный результат» – «Зависимость теплопроводности свинца от температуры». Помимо мета-информации, описывающей свойства данного объекта (год получения, авторов результата, описывающих его публикаций) и содержащейся в контенте данного ИНИР, рассматриваемый объект имеет еще один атрибут – «Значение», представленный гиперссылкой («>>>») на страницу с экспериментальными числовыми данными, которые хранятся во внешней базе данных. Дублирование подобной информации в контенте ИНИР было признано нецелесообразным. Вместе с тем, у пользователей ИНИР есть потребность не только просматривать внешние данные непосредственно в ИНИР, но и использовать их для решения задач ОЗ ресурса. Поэтому были созданы специальные средства, позволяющие работать с данными из внешних источников.

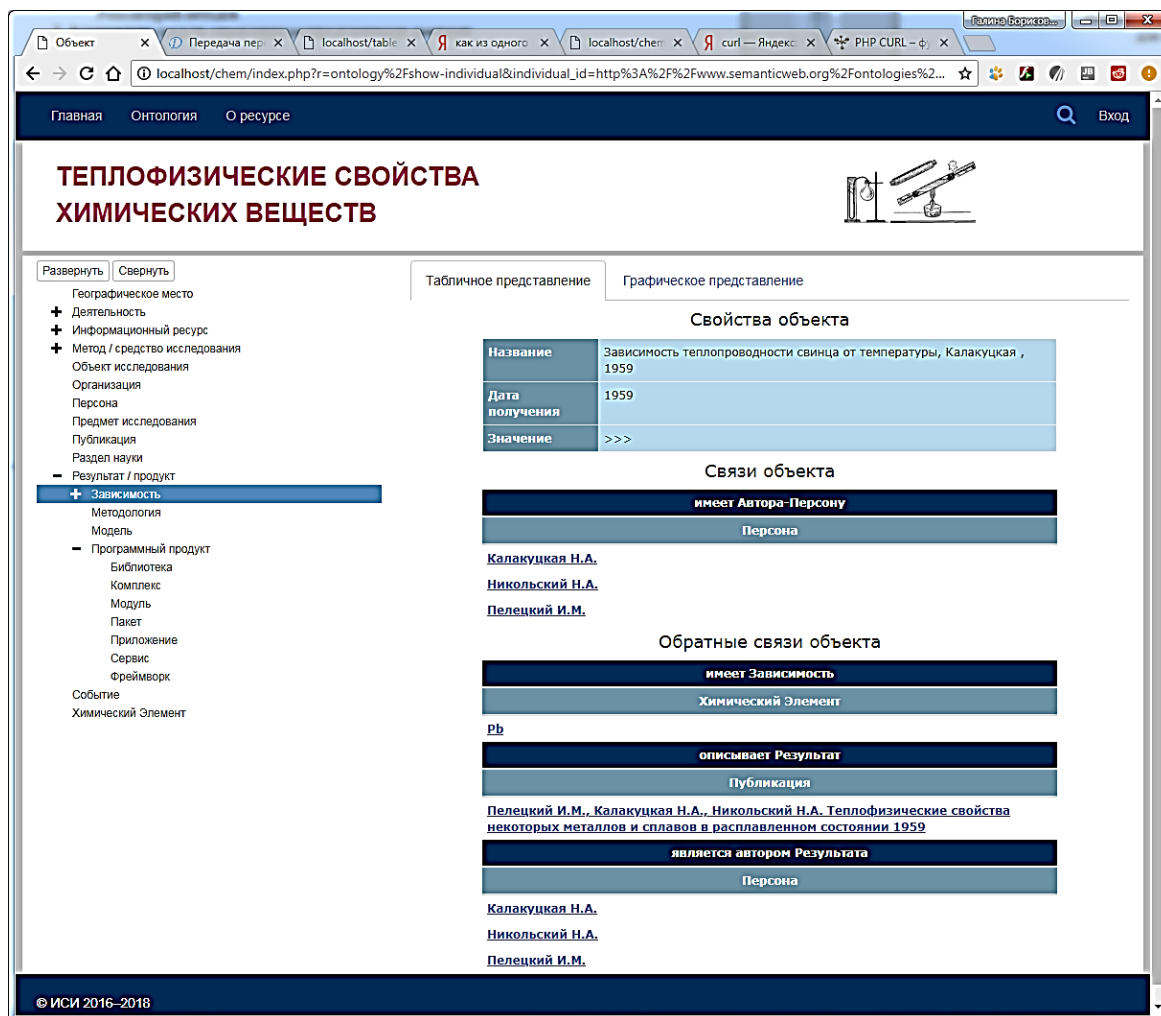


Рис. 1. ИНИР «Теплофизические свойства химических веществ»

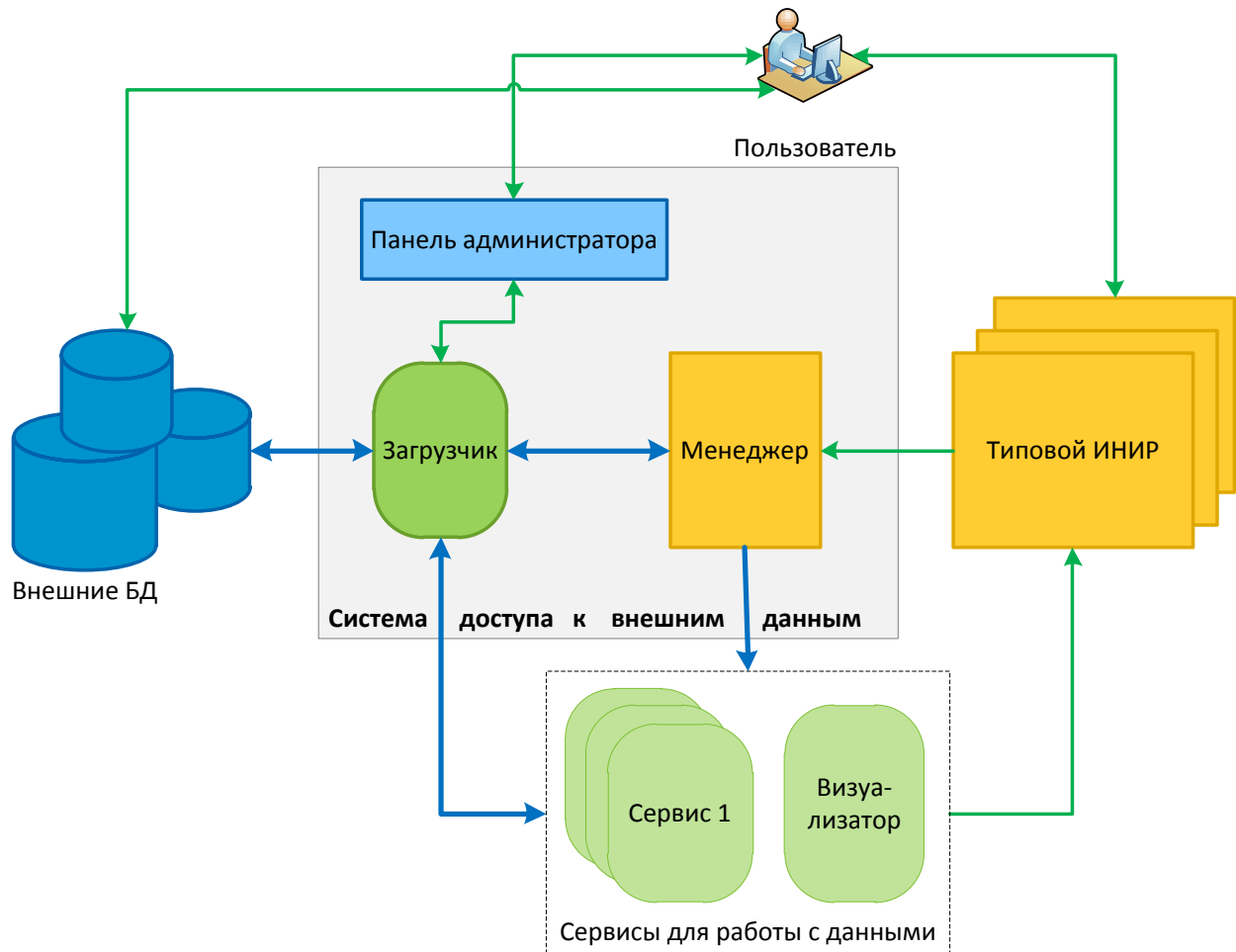
**2. Архитектура системы доступа к внешним данным и схема её функционирования.** Для доступа к внешним данным была разработана система, предоставляющая пользователям ИНИР следующие функциональные возможности:

1. Организация взаимодействия с внешними источниками данных. Это могут быть сторонние базы данных (БД), либо БД, созданные разработчиками конкретных ИНИР.
2. Описание информационных объектов со значениями из внешних БД.
3. Импорт значений свойств указанных объектов из внешних источников. Визуализация значений свойств объектов в виде таблиц или графиков.
4. Запуск сервисов анализа импортированных данных.
5. Использование импортированных данных для решения задач.

На рис. 2 представлены архитектурные компоненты системы доступа к данным и схема их взаимодействия.

Основным компонентом данной системы является сервис загрузки данных – *Загрузчик*, осуществляющий непосредственное взаимодействие с внешними хранилищами. Для подключения к системе конкретных БД служит *Панель администратора*, имеющая пользовательский Web-интерфейс, который позволяет зарегистрировать новые источники данных и сформировать шаблоны запросов для доступа к ним. *Загрузчик* имеет собственную БД, которая содержит адреса зарегистрированных в нем БД и информацию, требующуюся для построения запроса к конкретному ресурсу. Здесь могут использоваться как шаблоны

SQL запросов к реляционным БД, так и иные форматы запросов (REST API, SOAP, SPARQL и другие) к внешним ресурсам. Для построения конкретных запросов *Загрузчику* передаются необходимые параметры, которые извлекаются из онтологии ИНИР.



**Рис. 2.** Архитектура системы доступа к данным из внешних источников

*Сервисы* для работы с данными позволяют показывать их пользователю, выполнять их анализ или использовать для решения задач ОЗ ресурса.

Для организации взаимодействия ИНИР с *Загрузчиком* и *Сервисами* для работы с данными был разработан специальный плагин – *Менеджер*. Данный плагин предназначен для извлечения из онтологии ИНИР параметров, необходимых *Загрузчику* для построения запроса к внешней БД. *Менеджер* передает *Загрузчику* параметры, получает от него идентификатор запроса, который затем передает необходимому *Сервису*.

Рассмотрим схему функционирования системы доступа к внешним данным. Для того чтобы иметь возможность использовать в ИНИР внешние данные, необходимо предварительно через Панель администратора зарегистрировать в *Загрузчике* БД и шаблон запроса, выполняющего выборку необходимых данных. При этом каждой тройке (шаблон, строка подключения, тип БД) присваивается уникальный идентификатор, который сообщается инженеру знаний. Инженер знаний должен определить в онтологии ОЗ класс объектов, свойства которых будут принимать значений из внешней БД, и связать с этим классом полученный идентификатор. Кроме того, он должен позаботиться о том, чтобы связать свойства таких объектов с параметрами шаблона запроса к внешней БД (их именами

и порядком следования в шаблоне). Способ и средства осуществления этих предварительных действий описаны в следующем разделе.

При отображении в ИНИР свойств объектов, принадлежащих таким классам, появляется атрибут «Значение» с гиперссылкой «>>>>» (рис. 1). Переход по этой ссылке инициирует работу *Менеджера*, которому передается имя отображаемого объекта. По имени объекта *Менеджер* делает запрос к онтологии ИНИР, результатом которого будут связанный с классом данного объекта идентификатор и набор параметров шаблона запроса с их именами, значениями и порядком вхождения в запрос. Идентификатор и набор параметров передаются *Менеджером Загрузки*. *Загрузчик* записывает эту информацию во временную кэш-память и также присваивает каждому такому новому запросу уникальный идентификатор. В зависимости от того, в какой момент работы ИНИР и для каких целей вызывался *Менеджер*, он передает идентификатор запроса определенному *Сервису*, способному обратиться к *Загрузчику*, и, предъявив идентификатор запроса, получить и обработать данные, полученные от *Загрузчика*.

Для ресурса «Теплофизические свойства химических веществ», представленного на рис. 1, был разработан сервис визуализации данных (*Визуализатор*), который позволяет отображать зависимости некоторых теплофизических параметров, полученных в разное время разными авторами. Результат работы визуализатора показан на рис. 3

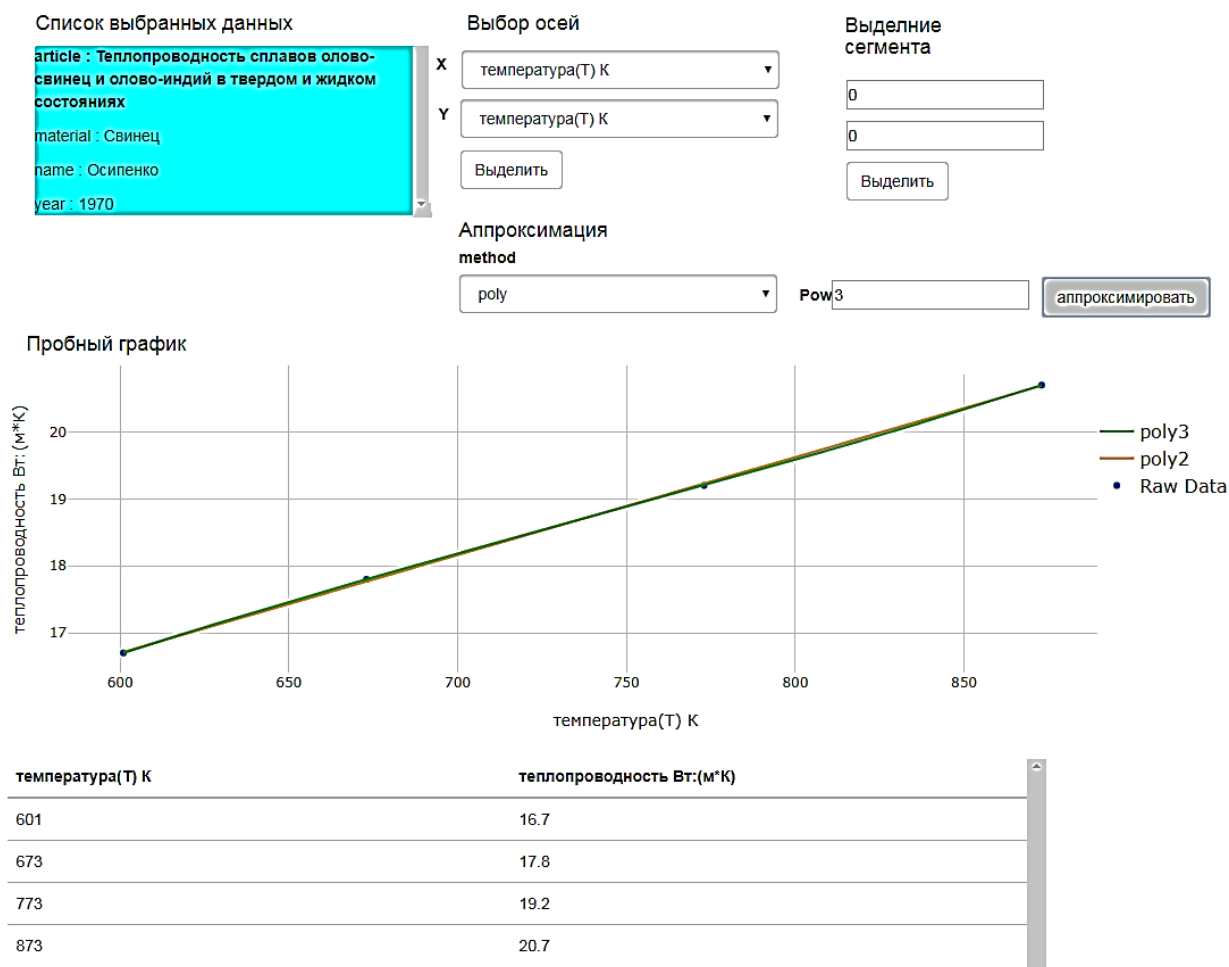
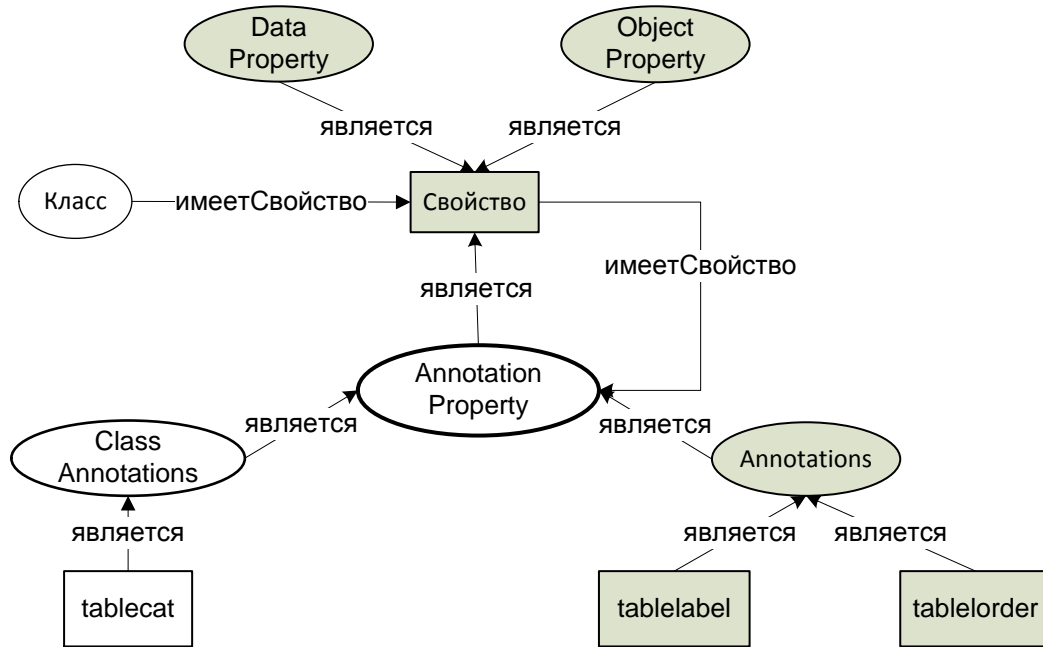


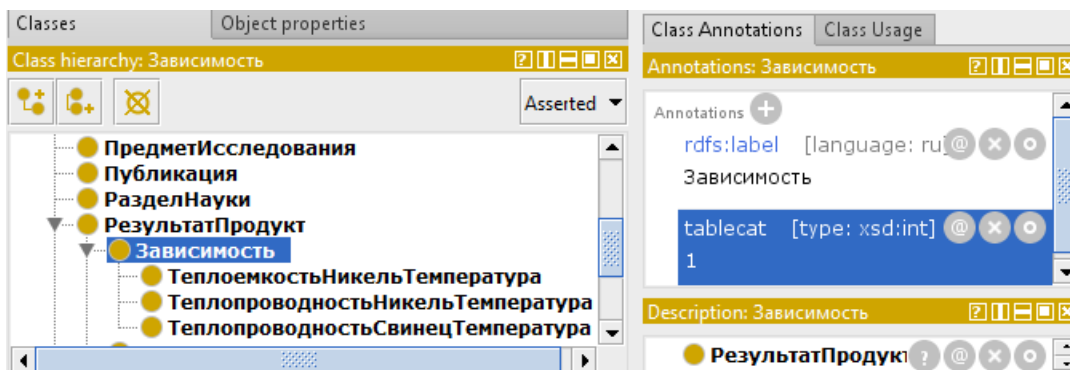
Рис. 3. Визуализация данных из внешних источников

**3. Особенности реализации взаимодействия ИНИР и системы доступа к внешним данным.** Поскольку онтология, являющаяся основным компонентом ИНИР, разработана средствами Semantic Web и редактора Protégé, их возможности были использованы для обеспечения связи класса объектов и их свойств с параметрами запроса к внешним БД. Был разработан паттерн онтологического проектирования (рис. 4), который относится к группе паттернов визуализации [4].



**Рис. 4.** Паттерн для установления связи класса и свойств объектов с параметрами запроса

Для задания связи объектов со значениями из внешней БД классу, к которому относятся эти объекты, приписывается аннотационное свойство `tablecat`, значением которого является уникальный идентификатор, сообщаемый инженеру знаний после регистрации в *Загрузчике* базы данных и шаблона запроса к ней. На рис. 5 показано, как в редакторе Protégé задается значение этого свойства для класса «Зависимость».



**Рис. 5.** Задание категории для класса объектов с данными из внешних источников

Тем свойствам объектов данного класса, которые входят в шаблон запроса, приписываются аннотационные свойства `tablelabel` и `tableorder`, задающие, соответственно, имя, под которым свойство объекта входит в шаблон запроса, и порядок вхождения (рис. 6). Значение этого свойства извлекается из онтологии для каждого конкретного объекта.

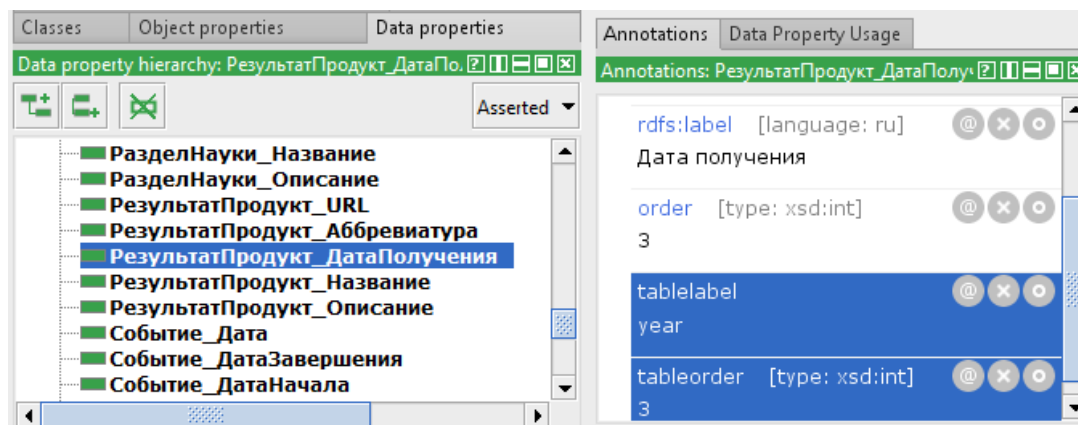


Рис. 6. Задание имен и порядка вхождения параметров шаблона запроса

Для класса «Зависимость» такими свойствами будут материал, для которого получена зависимость, фамилия получившего её автора, год получения и название публикации, в которой она была впервые описана. Так, год получения какой-либо зависимости теплофизических параметров будет входить в запрос под именем «year» и под номером 3 (рис. 6).

**Заключение.** В статье описана система доступа к данным из внешних источников, разработанная в рамках технологии создания интеллектуальных научных интернет-ресурсов. Рассмотрены архитектурные компоненты данной системы и схема их взаимодействия с ИНИР и друг другом.

Использование технологии и средств Semantic Web позволило максимально упростить установление связи между объектами контента ИНИР и значениями их свойств, хранящимися во внешних БД.

При реализации системы был использован сервис-ориентированный подход. Идея общения ИНИР и сервисов с помощью уникальных идентификаторов позволяет подключать к *Загрузчику* различные сервисы для работы с данными. Данный подход, а также наличие *Менеджера*, управляющего взаимодействием сервисов, позволяет легко масштабировать систему, наращивать функциональные возможности ИНИР, не производя изменений в его коде. При разработке сервисов особое внимание было уделено вопросам оптимизации работы предлагаемой архитектуры.

#### СПИСОК ЛИТЕРАТУРЫ

1. Брагинская Л.П., Григорюк А.П., Ковалевский В.В.. Научная информационная система «Активная сейсмология» для комплексных геофизических исследований // Вестник КРАУНЦ. Науки о земле. 2015. № 1. Выпуск № 25. С. 94–98.
2. Глоба Л.С., Новогрудская Р.Л. Модели и методы интеграции информационных и вычислительных ресурсов // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2012): материалы II Междунар. научн.-техн. конф. (Минск, 16-18 февраля 2012г.). Минск. БГУИР. 2012. С. 447–452.
3. Дударев В.А. Интегрированная информационная система по свойствам неорганических веществ и материалов // Труды XVII Международной конференции DAMDID/RCDL'2015, Обнинск, 13-16 октября 2015. С. 41–48.



4. Загорулько Ю.А., Боровикова О.И., Загорулько Г.Б. Применение паттернов онтологического проектирования при разработке онтологий научных предметных областей // Сборник научных трудов XIX Международной конференции DAMDID/RCDL'2017 (10-13 октября 2017 года, Москва, МГУ, Россия) / под. ред. Л.А. Калиниченко, Я. Манолопулоса, Н.А. Скворцова, В.А. Сухомлина. М.: Изд-во ФИЦ ИУ РАН. 2017. С. 332–340.
  5. Загорулько Ю.А., Загорулько Г.Б., Боровикова О.И. Технология создания тематических интеллектуальных научных интернет-ресурсов, базирующаяся на онтологии // Программная инженерия. 2016. № 2. С. 51–60.
  6. Загорулько Г.Б., Молородов Ю.И., Федотов А.М. Систематизация знаний по теплофизическим свойствам веществ // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2014. Т.12. № 3. С.48–56.
  7. Копайгородский А.Н. Применение онтологий в семантических информационных системах // Онтология проектирования. 2014. № 4 (14). С. 78–89.
  8. Gangemi A., Presutti V. *Ontology Design Patterns* // Handbook on Ontologies. Berlin: Springer. 2009. Pp. 221–243.
  9. Hitzler P., Krötzsch V., Rudolph S. *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC. 2009. 455 p.
  10. Protégé. A free, open-source ontology editor and framework for building intelligent systems . URL: <https://protege.stanford.edu/> (last visited 22.05.2018).
- 

**UDK 519.816, 004.827**

**DEVELOPMENT OF THE SYSTEM OF ACCESS TO DATA FROM EXTERNAL SOURCES FOR INTELLIGENT SCIENTIFIC INTERNET RESOURCES**

**Galina B. Zagorulko**

Scientific researcher of Laboratory "Artificial Intelligence"

A.P. Ershov Institute of Informatics Systems

Siberian Branch of the Russian Academy of Sciences

6, Acad. Lavrentjev pr., 630090, Novosibirsk, Russia, e-mail: gal@iis.nsk.su

**Yury I. Molorodov**

Scientific researcher of "Information Systems" Laboratory

Institute of Computational Technologies Systems

Siberian Branch of the Russian Academy of Sciences

6, Acad. Lavrentjev pr., 630090, Novosibirsk, Russia, e-mail: yumo@ict.sbras.ru

**Kirill E. Vishnev**

Student of Novosibirsk State University

1, Pirogov St., 630090, Novosibirsk, Russia, e-mail: vishnev\_kirill@mail.ru

**Abstract.** Described are the tools developed for intelligent scientific Internet resources that are designed to visualize and process data stored in external sources. The architecture of the access system to external sources, the scheme of its functioning and the approaches used for its creation, as well as the methodology and examples of the proposed means using are presented.



**Keywords:** intelligent scientific Internet resource, ontology, service, external data storage, data access system.

### References

1. Braginskaya L.P., Grigoryuk A.P., Kovalevskij V.V. Nauchnaya informacionnaya sistema «Aktivnaya seismologiya» dlya kompleksnyh geofizicheskikh issledovaniy [Scientific information system «Active seismology» for integrated geophysical studies] // Vestnik KRAUNC. Nauki o zemle = Bulletin of Kamchatka regional association «Educational-scientific center». Earth sciences. 2015. Issue № 25. № 1. Pp. 94–98. (in Russian).
2. Globa L.S., Novogrudskaya R.L. Modeli i metody integracii informacionnyh i vychislitel'nyh resursov [Models and methods of integration of information and computing resources] // Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nyh sistem: materialy II Mezhdunar. nauchn.-tekhn. konf. = Open Semantic Technologies for Intelligent Systems (OSTIS-2012): proceeding of the II intern. scientific.-tech. conf (Minsk, 16-18 February, 2012). Minsk. BGUIR. 2012. Pp. 447–452. (in Russian).
3. Dudarev, V.A. Integrirovannaya informacionnaya sistema po svojstvam neorganicheskikh veshchestv i materialov [Integrated information system on the properties of inorganic substances and materials] // Trudy XVII Mezhdunarodnoj konferencii DAMDID/RCDL'2015 = Proceeding of the XVII intern. conf DAMDID/RCDL'2015. Obninsk. October 13-16. 2015. Pp. 41–48. (in Russian).
4. Zagorulko Yu., Borovikova O., Zagorulko G. Application of Ontology Design Patterns in the Development of the Ontologies of Scientific Subject Domains // Selected Papers of the XIX International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2017) Moscow. Russia. October 9-13, 2017. Leonid Kalinichenko, Yannis Manolopoulos, Nikolay Skvortsov, Vladimir Sukhomlin (Eds.). Pp. 258–265. (in Russian).
5. Zagorulko Yu.A., Zagorulko G.B., Borovikova O.I. Tekhnologiya sozdaniya tematicheskikh intellektual'nyh nauchnyh internet-resursov, baziruyushchayasya na ontologii [Technology for building subject-based intelligent scientific internet resources based on ontology] // Programmnyaya inzheneriya = Software Engineering. 2016. no. 2. Pp. 51–60. (in Russian).
6. Zagorulko G.B., Molorodov YU.I., Fedotov A.M. Sistematizaciya znaniy po teplofizicheskim svojstvam veshchestv [Systematization of knowledge on thermophysical properties of substances] // Vestnik Novosibirskogo gosudarstvennogo universiteta. Seriya: Informacionnye tekhnologii = Novosibirsk State University Journal of Information Technologies. 2014. Volume 12. Issue No 3. Pp. 48–56. (in Russian).
7. Kopajgorodskij A.N. Primenenie ontologij v semanticheskikh informacionnyh sistemah [Ontologies application in semantic information systems] // Ontologiya proektirovaniya = Ontology of designing. 2014. № 4 (14). Pp. 78–89. (in Russian).
8. Gangemi A., Presutti V. Ontology Design Patterns // Handbook on Ontologies. Berlin: Springer. 2009. -Pp. 221–243.
9. Hitzler P., Krötzsch V., Rudolph S. Foundations of Semantic Web Technologies. Chapman & Hall/CRC. 2009. 455 p.
10. Protégé. A free, open-source ontology editor and framework for building intelligent systems. URL: <https://protege.stanford.edu/> (last visited 22.05.2018).