А.И. Прохоров, Д.А. Варламов,
Е.С. Амосова, П.К. Берзигияров,
В.Е. Туманов

**Внедрение предметных онтологий в
систему научной аналитики по
физической химии радикальных
реакций**

Рекомендуемая форма библиографической ссылки

Прохоров А.И., Варламов Д.А., Амосова Е.С., Берзигияров П.К., Туманов В.Е. Внедрение предметных онтологий в систему научной аналитики по физической химии радикальных реакций // Научный сервис в сети Интернет: труды XVIII Всероссийской научной конференции (19-24 сентября 2016 г., г. Новороссийск). — М.: ИПМ им. М.В.Келдыша, 2016. — С. 298-302. — doi:[10.20948/abrau-2016-26](https://doi.org/10.20948/abrau-2016-26)

Размещена также [презентация к докладу](#)

Внедрение предметных онтологий в систему научной аналитики по физической химии радикальных реакций

А.И. Прохоров¹, Д.А. Варламов², Е.С. Амосова¹,
П.К. Берзигияров¹, В.Е. Туманов¹

¹ *Институт проблем химической физики РАН, г. Черноголовка, РФ*

² *Институт экспериментальной минералогии РАН, г. Черноголовка, РФ*

Аннотация. Представлен процесс внедрения предметных онтологий по химии радикальных реакций в хранилище данных предметно-ориентированной системы научной аналитики в области физической химии.

Ключевые слова: онтология, хранилище данных, базы данных, интеграция, радикальная реакция, физическая химия

1. Введение

В работе [1] представлено описание предметно-ориентированной системы научной аналитики в области физической химии радикальных реакций, разработанной на основе технологии хранилищ данных. В работах [2,3] разработаны онтологические модели предметных областей для данной системы.

Внедрение разработанных онтологий в указанную систему и превращение ее в семантическую систему в Интернет является актуальной научно-практической задачей. Целью настоящей работы является представить процесс внедрения предметных онтологий в хранилище данных системы.

Взаимоотношение онтологий и хранилищ данных подробно исследовались в работе [4]. Существует два общих подхода к использованию онтологий в хранилищах данных: проектирование хранилищ данных на основе онтологий (см., например, [5]) и внедрение онтологических моделей в существующие хранилища данных (см., например, [6]). В настоящей работе мы следуем второму подходу, когда разработанные онтологические модели интегрируются в существующее хранилище данных.

2. Постановка задачи

Достижение поставленной цели состоит из двух основных шагов: представление разработанных онтологий в виде реляционной таблицы данных

(RDF-тройки) и разработки интерпретатора семантического запроса в SQL-запрос к хранилищу данных с последующим возвращением результата.

Как правило, пользователь не обладает всей точной информацией относительно реагентов реакции, а с другой стороны в большинстве случаев пользователь ищет аналогичные реакции для оценки реакционной способности реакций изучаемого им химического процесса.

Поэтому использование предметных онтологий дает возможность получить список реакций определенного класса для последующего принятия решений относительно реакционной способности интересующей группы реакций. С этой точки зрения онтология предметной области обладает неоспоримым преимуществом в поиске подходящих реакций.

3. Решение проблемы

Для того чтобы интегрировать в систему [1] разработанные онтологии [2,3], было решено сохранить их в той же реляционной базе данных (РБД), где содержатся данные по физической химии радикальных реакций.

Выделяют два базовых подхода к организации хранения онтологий в РБД:

- 1) использование «единственной» таблицы для хранения всех RDF-триплетов (подход «вертикальная таблица») (рис.1);
- 2) отображение иерархии онтологических сущностей (классов, свойств, экземпляров) в схему РБД.

<i>graph</i>	<i>subject</i>	<i>predicate</i>	<i>object</i>
<u>Chem</u>	<u>Molecules</u>	<u>rdf:type</u>	<u>Compounds</u>
<u>Chem</u>	<u>has_BDE_Value</u>	<u>rdf:type</u>	<u>rdfs:Property</u>
<u>Chem</u>	<u>has_BDE_Bond</u>	<u>rdf:type</u>	<u>rdfs:Property</u>
<u>Chem</u>	<u>has_BDE_Error</u>	<u>rdf:type</u>	<u>rdfs:Property</u>
<u>Chem</u>	<u>has_brutto_formula</u>	<u>rdf:type</u>	<u>rdfs:Property</u>
<u>Chem</u>	<u>has_CAS_number</u>	<u>rdf:type</u>	<u>rdfs:Property</u>
<u>Chem</u>	<u>Benzenamine_Aniline</u>	<u>rdf:type</u>	<u>Molecules</u>
<u>Chem</u>	<u>Benzenamine_Aniline</u>	<u>rdfs:label</u>	<u>'Benzenamine; Aniline'</u>
<u>Chem</u>	<u>Benzenamine_Aniline</u>	<u>has_brutto_formula</u>	<u>'C6H7N1'</u>
<u>Chem</u>	<u>Benzenamine_Aniline</u>	<u>has_CAS_number</u>	<u>'62-53-3'</u>
<u>Chem</u>	<u>Benzenamine_Aniline</u>	<u>has_BDE_Value</u>	<u>'375.3'</u>
<u>Chem</u>	<u>Benzenamine_Aniline</u>	<u>has_BDE_Bond</u>	<u>'N-H'</u>
<u>Chem</u>	<u>Cyclopropane</u>	<u>has_BDE_Bond</u>	<u>'C-H'</u>
<u>Chem</u>	<u>Cyclopropane</u>	<u>has_BDE_Value</u>	<u>'427'</u>
<u>Chem</u>	<u>Cyclopropane</u>	<u>has_CAS_number</u>	<u>'75-19-4'</u>
<u>Chem</u>	<u>Cyclopropane</u>	<u>has_brutto_formula</u>	<u>'C3H6'</u>
<u>Chem</u>	<u>Cyclopropane</u>	<u>rdf:type</u>	<u>Molecules</u>
<u>Chem</u>	<u>Cyclobutane</u>	<u>has_CAS_number</u>	<u>'287-23-0'</u>
<u>Chem</u>	<u>Cyclobutane</u>	<u>rdf:type</u>	<u>Molecules</u>
<u>Chem</u>	<u>Cyclobutane</u>	<u>has_BDE_Value</u>	<u>'418.5'</u>
<u>Chem</u>	<u>Cyclobutane</u>	<u>has_BDE_Bond</u>	<u>'C-H'</u>
<u>Chem</u>	<u>Cyclobutane</u>	<u>has_BDE_Error</u>	<u>'3.5'</u>
<u>Chem</u>	<u>Cyclobutane</u>	<u>has_brutto_formula</u>	<u>'C4H8'</u>

Рис. 1. Фрагмент «единственной» таблицы TRIPLES для хранения RDF-триплетов.

В нашем случае мы используем первый подход, когда все RDF-триплеты хранятся в унифицированной таблице БД, содержащей в общем случае четыре колонки: «граф», «субъект», «объект» и «предикат» (рис.1). Данный подход реализован, в частности, в Jena SDB и 3store. Он характеризуется достаточно высокой временной сложностью выборки RDF-триплетов, однако, мы полагаем, в нашем случае этот подход вполне пригоден.

Поскольку разработанные онтологии хранятся в RDF хранилище и реализовано оно в виде «одной» таблицы в реляционной базе данных, то любой SPARQL запрос можно быть представлен в виде SQL запроса с self-join к таблице RDF хранилища (рис.2).

Рис. 2. Представление SPARQL запроса в виде SELF JOIN SQL запроса в случае использования «единственной» таблицы для RDF-триплетов.

Для загрузки онтологий, разработанных в редакторе Protégé, в RDF хранилище использовано собственное решение. Оно заключается в следующем:

1. выгрузка онтологий непосредственно редактором Protégé в файлы в формате Turtle (Terse RDF Triple Language), так как это наиболее компактный вариант удобный для последующей обработки содержимого файлов;
2. загрузка файлов онтологий в RDF хранилище с помощью дополнительного сервиса системы.

Поскольку, основным скриптовым языком разработки на стороне сервера является JavaScript, то для реализации сервисов были использованы библиотеки N3.js, SPARQL.js [7,8]. На базе этих библиотек были разработаны и реализованы: сервис загрузки файлов онтологий в реляционную базу в варианте «единой» таблицы для всех RDF-триплетов, специальный сервис для обслуживания RDF хранилища и интерпретатор семантического запроса к предметно-ориентированному хранилищу данных с возможностью получения релевантного ответа. Общая схема реализации сервисов представлена на рис. 3.

Рис. 3. Общая схема реализации загрузки онтологий и обработки семантических запросов к данным, хранящимся в RDF хранилище (Triple Store)

На данном этапе в системе научной аналитики по физической химии радикальных реакций данные RDF-представлений онтологий связаны с данными исходных экспериментальных коллекций (реляционные данные) посредством общих индексов, которыми являются:

1. Уникальные численные идентификаторы химических соединений, т.е. CAS-номера химических реагентов;
2. Эмпирическая классификация радикальных реакций, разработанная под руководством профессора Е.Т. Денисова [9].

4. Заключение

Предложен подход интеграции предметных онтологий в предметно-ориентированную систему научной аналитики по физической химии радикальных реакций. В процесс интеграции были вовлечены две предметные онтологии: онтологическая модель радикальных реакций в жидкой фазе и онтологическая модель по термохимии радикальных реакций в жидкой фазе.

Разработаны сервис загрузки файлов онтологий в реляционную базу знаний в варианте «единой» таблицы для всех RDF-триплетов, сервис обслуживания RDF хранилища, интерпретатор семантического запроса к предметно-ориентированному хранилищу данных, который выдаст релевантный ответ.

Проведение работ по данной тематике поддержано грантом РФФИ №15-07-08645-а.

Литература

1. Туманов В.Е., Прохоров А.И., Лазарев Д.Ю., Соловьева М.Е. Система научной осведомленности по физической химии радикальных реакций. // Информационные ресурсы России. — 2010. — № 5. — С. 16-21
2. Амосова Е.С., Туманов В.Е. Преставление химических реакций, реагентов и их термохимических свойств в интеллектуальной системе по физической

- химии радикальных реакций в жидкой фазе с использованием онтологической модели предметной области. // Бутлеровские сообщения. — 2014. — Т. 39. — № 7. — С. 39-46
3. Амосова Е.С., Берзигияров П.К. Разработка представления онтологической модели по физической химии радикальных реакций отношениями реляционной базы данных и реализация его в предметно-ориентированной системе научной осведомленности. // Бутлеровские сообщения. — 2016. — Т. 45. — № 1. — С. 152-158
 4. Ontologies and Data Warehousing. // NETWORK INFERENCE. INC., 2004. — URL: www.networkinference.com
 5. Pardillo J., Mazon J.-N. Using ontologies for the design of data warehouses. // International Journal of Database Management Systems. — 2011. — V. 3 — №2. — P. 73-87
 6. Choi N., Song I-Y, Han H. A survey on ontology mapping. // SIGMOD Record. — 2006. — V. 35. — № 3. — P. 34-41
 7. Verborgh R. Lightning-fast RDF in JavaScript [Электронный ресурс] // URL: <https://ruben.verborgh.org/blog/2013/04/30/lightning-fast-rdf-in-javascript/> (дата обращения: 25.08.2016)
 8. Verborgh R. Writing a SPARQL parser in JavaScript [Электронный ресурс] // URL: <https://ruben.verborgh.org/blog/2014/08/22/writing-a-sparql-parser-in-javascript/> (дата обращения: 25.08.2016)
 9. Денисов Е.Т. Новые эмпирические модели реакций радикального отрыва. // Успехи химии. — 1997. — Т. 66. — № 10. — С. 953