

Труды XVII Всероссийской научной конференции

Научный сервис в сети Интернет

Новороссийск, 21-26 сентября 2015 г.

**ИПМ им.М.В.Келдыша
2015**

УДК 519.7

ББК 73

Н 34

- Н 34 Научный сервис в сети Интернет: труды XVII Всероссийской научной конференции (21-26 сентября 2015 г., г. Новороссийск). — М.: ИПМ им. М.В.Келдыша, 2015. — 336 с.
ISBN 978-5-98354-015-6

Сборник содержит доклады, сделанные участниками XVII Всероссийской научной конференции «Научный сервис в сети Интернет». Конференция проходила с 21 по 26 сентября 2015 года. Подробную информацию о конференции можно найти по адресам <http://agora.guru.ru/abrau2015/> и <http://keldysh.ru/abrau/2015/>.

Конференция поддержана Российским фондом фундаментальных исследований, проект 15-07-20607-г.

ИНФРАСТРУКТУРА ВЕБ-СЕРВИСА ИНТЕЛЛЕКТУАЛЬНОЙ ПРОБЛЕМНО-ОРИЕНТИРОВАННОЙ СИСТЕМЫ НАУЧНОЙ АНАЛИТИКИ ПО ФИЗИЧЕСКОЙ ХИМИИ РАДИКАЛЬНЫХ ЖИДКОФАЗНЫХ РЕАКЦИЙ В ИНТЕРНЕТ

А.И. Прохоров¹, Д.А. Варламов^{1,2}, Е.С. Амосова¹,
М.Е. Соловьева¹, В.Е. Туманов¹, П.К. Берзигияров¹

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

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

Разработка проблемно-ориентированных веб-сервисов в рамках создания интеллектуальных информационных систем по прогнозированию физико-химических свойств соединений методами интеллектуального анализа экспериментальных кинетических и термохимических данных на основе онтологических моделей является актуальной междисциплинарной задачей.

Решению задачи извлечения опубликованных данных о химических реакциях и константах скорости реакций на основе онтологических моделей в биологии посвящен ряд работ зарубежных исследователей [1-3]. Методология и общие подходы к разработке инфраструктуры проблемно-ориентированных веб-сервисов, направленных на интеллектуальный анализ больших объемов информации, обсуждались в работах [5-8].

Целью настоящей работы является описание прототипа программно-технологической инфраструктуры проблемно-ориентированного веб-сервиса для автоматического сбора и семантического анализа кинетических и термохимических данных из открытых оцифрованных источников в Интернет с последующей их разбраковкой, сохранением и использованием в интеллектуальной проблемно-ориентированной системе научной аналитики по физической химии радикальных реакций, которая является развитием проблемно-ориентированной системы по физической химии радикальных реакций [9] за счет внедрения в нее семантики на основе онтологических моделей предметной области.

В настоящее время информационная система [9] имеет многослойную программно-технологическую архитектуру как показано на рис. 1.

Верхний уровень системы представляет собой предметно-ориентированное веб-приложение, которое предоставляет пользователю интерфейс доступа к программным компонентам системы: информационной подсистеме, аналитической подсистеме, подсистеме дистанционного обучения, встроенной подсистеме объяснений и подсистеме производства новых профессиональных знаний.

Ниже доступного для пользователей уровня скрыт уровень программных агентов, распределенных на различных узлах локальной сети. В системе

реализованы интеллектуальные агенты, реактивные агенты, агенты обученных искусственных нейронных сетей, которые реализуют работу встроенных в портал экспертных систем и выполняют функции поиска информации. Набор интеллектуальных агентов системы является активным компонентом системы. Их можно представить в виде специализированных сервисов. При этом сами агенты ориентированы на обработку научных данных в узкоспециализированном разделе предметной области.

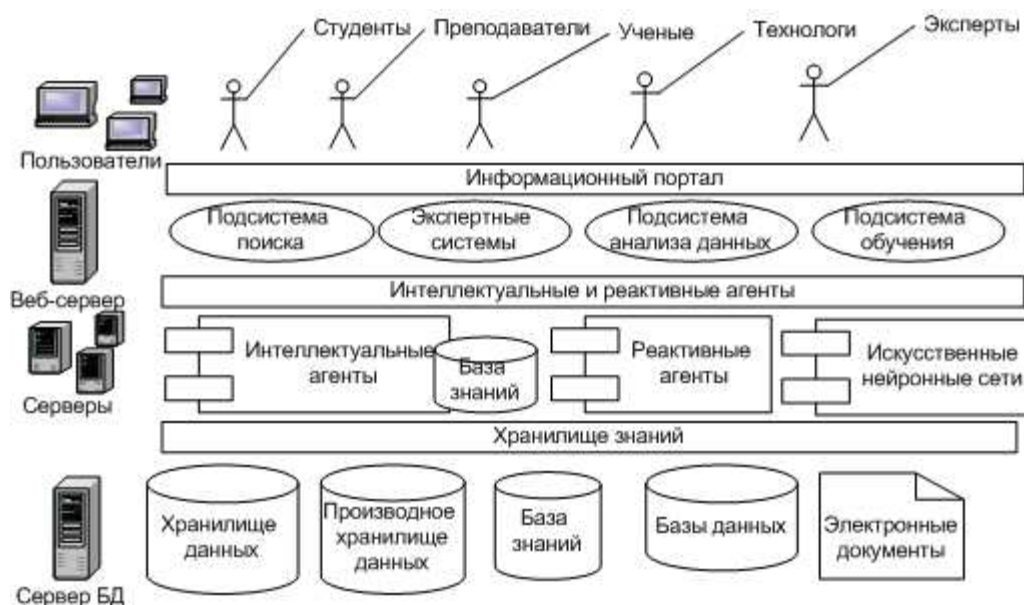


Рис. 1. Программно-технологическая архитектура системы научной осведомленности.

Вся система опирается на уровень хранилищ системы. Фактологическая часть хранилищ включает в себя:

- хранилище данных по константам скоростей радикальных жидкофазных реакций (накоплено более 30000 значений);
- киоск данных по энергиям диссоциации связей молекул (более 1000 объектов);
- база данных по энтальпиям образования радикалов (около 980 объектов)
- база данных по энтальпиям образования органических молекул (экспериментальные данные, до 3000 информационных объектов).

Хранилище знаний настоящей системы содержит эмпирические и рассчитанные факты, продукционные правила и процедуры расчета, а в совокупности с экспертными системами образует виртуальную подсистему производства новых профессиональных знаний. Дополнительно в качестве механизмов производства новых знаний в системе используются обученные искусственные нейронные сети (для реакций $R\cdot + RH$, $C_6H_5\cdot + RH$ и $H + RH$ в водных растворах). Например, для хранилища по константам скорости в системе реализованы две экспертные системы (ЭС). Одна – классическая на

продукционных правилах (многоагентная), другая – на основе искусственных нейронных сетей. Решение об использовании конкретной ЭС определяет специализированный интеллектуальный агент. Если решение задачи пользователя можно определить расчетным путем на основе уже имеющихся данных, то управление по расчету нужной величины передается соответствующему резистивному агенту.

Дальнейшее развитие системы мы видим в увеличении интеллектуальной составляющей следующим образом – преобразование уровня хранилища системы в уровень хранилища знаний посредством реализации для системы ряда онтологических моделей (онтологий).

В [10] разработана одна из онтологических моделей представления радикальных жидкофазных реакций для внедрения ее в создаваемую систему (рис. 2).

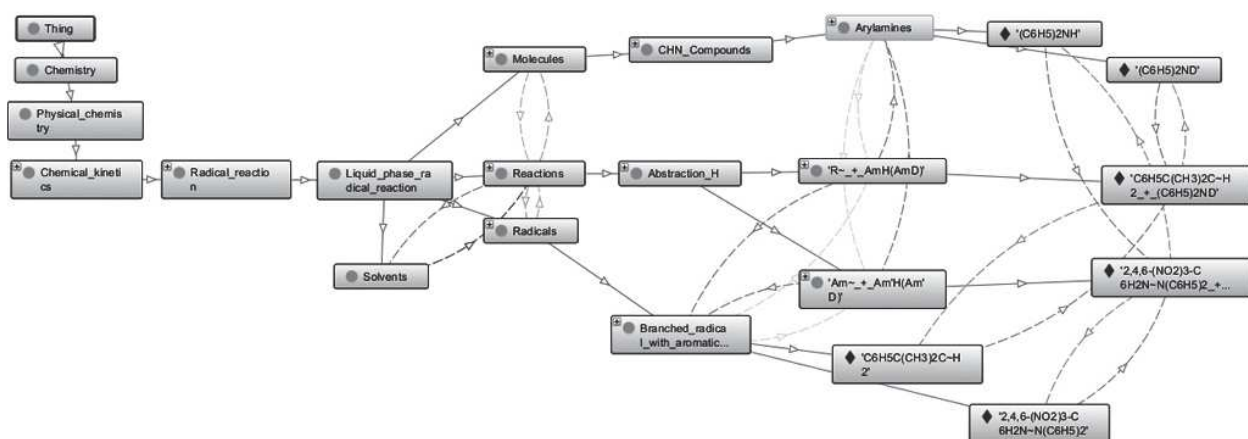


Рис. 2. Графическое представление фрагмента онтологии предметной области «радикальные реакции в жидкой фазе», отображающее классы и подклассы, а также связи между классами и экземплярами классов.

Добавляя в систему онтологические модели, с их помощью мы связываем все имеющиеся данные хранилищ, витрин, баз на уровне их измерений в единое хранилище знаний. При этом имеющиеся в системе продукционные правила необходимо согласовать с правилами, получаемыми из онтологий. Поэтому часть «старых» продукционных правил может перейти в разряд аксиом или фактов.

К этому моменту наша система практически будет готова к семантическому поиску в ней. Но в данный момент подобный сервис в системе отсутствует, поэтому предполагается его реализовать, поставив ему еще и дополнительную задачу: искать не только внутри системы, но еще и во вне неё. Добывая дополнительно данные извне, мы расширяем количество фактического материала и его актуальность, и, как следствие, достоверность предсказаний.



Рис. 3. Программно-технологическая архитектура веб-сервиса для автоматического сбора кинетических и термохимических данных радикальных жидкофазных реакций из открытых оцифрованных источников в Интернете.

Предполагается, что основными задачами прототипа веб-сервиса для автоматизации сбора кинетических и термохимических данных из открытых оцифрованных источников в Интернете будут:

1. поиск информации (поисковый профиль «ключевые слова», список сайтов (журналов), в перспективе семантический поиск);
2. формирование блока документов (форматы pdf, doc, htm, xml), либо автономно (из offline-коллекций), либо через Интернет;
3. парсинг документов (извлечение данных из текстовых документов: семантическая модель документа, кластеризация, классификатор (байесовы сети), искусственные нейронные сети, нечеткие нейросети);
4. верификация и сохранение полученных в хранилище данных системы.

Таким образом, действующая информационная система [9] будет обновлена до интеллектуальной проблемно-ориентированной системы научной аналитики по физической химии радикальных жидкофазных реакций. Её «опорная» часть «Хранилище знаний» пополнится дополнительными фактологическими составляющими: онтологическими моделями по кинетике и термохимии радикальных жидкофазных реакций и, связанным с ними, хранилищем оцифрованных документов, которые предполагается собирать, в том числе и из открытых источников в Интернет. Кроме того, всё хранилище знаний системы организуется вокруг разрабатываемых онтологических моделей в единое целое. Это и наличие в онтологиях предметных понятий и связей между ними позволяет реализовать в системе подсистему семантического поиска как в материалах хранилища знаний системы, так и в Интернет. Семантический поиск и сбор документов в Интернет будет осуществляться автоматизированной частью веб-сервиса с последующим

интеллектуальным анализом и обработкой полученных документов. Основной задачей системы останется возможность предсказания значений кинетических и термодинамических характеристик радикальных жидкофазных реакций.

Проведение работ по данной тематике поддержано грантом РФФИ № 15-07-08645-а, рук. Прохоров А.И.

ЛИТЕРАТУРА

1. Weber L., Irmer M., Bobach C. Extraction of chemical reactions from full text documents: From n-tuples of value attribute pairs toward the automated construction of reaction databases [Электронный ресурс] / http://www.chemaxon.com/wp-content/uploads/2014/07/B05_Weber_OntoChem_ACS_2014.pdf (дата обращения 20.05.2015 г.)
2. Hakenberg J., Schmeier S., Kowald A., Klipp E., Leser U. // OMICS A Journal of Integrative Biology, 2004. Vol. 8, 2. P.131-152.
3. Jessop D.M., Adams S.E., Murray-Rust P. Mining chemical information from open patents // Journal of Cheminformatics. Vol. 3. № 40. P. 1-17.
4. Bingjun Sun, Qingzhao Tan, Prasenjit Mitra, C. Lee Giles Extraction and Search of Chemical Formulae in Text Documents on the Web / International World Wide Web Conference Committee (IW3C2). WWW 2007, May 8.12, 2007, Banff, Alberta, Canada. [Электронный ресурс] / <http://www.sciweavers.org/read/extraction-and-search-of-chemical-formulae-in-text-documents-on-the-web-37295> (дата обращения 20.05.2015 г.)
5. Wild D.J. A Web Service Infrastructure for Chemoinformatics // 4th Joint Sheffield Conference on Chemoinformatics, June 2007. [Электронный ресурс] / <http://cisrg.shef.ac.uk/shef2007/talks/wild.pdf> (дата обращения 20.05.2015 г.)
6. Wild D.J. Integrating text and literature sources with traditional chemoinformatics tools. [Электронный ресурс] / <http://acscinf.org/docs/meetings/233nm/presentations/233nm07.pdf> (дата обращения 20.05.2015 г.)
7. Viroli M., Casadei M., Nardini E., Omicini A. Towards a Pervasive Infrastructure for Chemical-Inspired Self-organising Services // Self-Organizing Architectures. Lecture Notes in Computer Science. 2010. Vol. 6090. P. 152-176.
8. Ai C. S., Blower P.E., Jr., Ledwith R. H. Extraction of chemical reaction information from primary journal text. // J. Chem. Inf. Comput. Sci. 1990. Vol. 30, № 2. P. 163–169.
9. 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.
10. Амосова Е.С., Туманов В.Е. Представление химических реакций, реагентов и их термодинамических 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.

СИСТЕМАТИЗАЦИЯ ФИЗИКО-ХИМИЧЕСКИХ ДАННЫХ. ВОЗМОЖНОСТИ ОНТОЛОГИЙ И БАЗ ДАННЫХ, ПРИНЦИП ДОПОЛНИТЕЛЬНОСТИ

В.А. Серебряков¹, А.О. Еркимбаев², В.Ю. Зицерман², Г.А. Кобзев²,
К.Б. Теймуразов¹, Р.И. Хайрулин²

¹ *Вычислительный центр им. А.А.Дородницына РАН*

² *ОИВТ РАН*

Накопление и систематизация физико-химических данных и прежде всего данных о свойствах вещества является одним из основных направлений деятельности в естественнонаучных дисциплинах. До недавнего времени основным информационным ресурсом этой области была традиционная БД. Однако резкий рост их числа при многообразии форматов и моделей привел к тому, что действующая инфраструктура оказалась неспособной обеспечить эффективную организацию рабочего процесса. Автономность БД с жесткой фиксацией используемых терминов и логических структур практически исключает возможность беспрепятственного обмена данными без активного участия человека-эксперта. В итоге интерес в научном сообществе сместился в сторону принятия онтологий как более эффективных средств формализации и распространения научных данных. Главное назначение онтологий в том, что они добавляют к данным семантику (смысл, правильное толкование) и соотношения, что в совокупности описывает «знание» с возможностью его машинной интерпретации. Оценить богатый потенциал, заложенный в онтологиях для хранения и распространения данных, можно, сопоставляя их возможности с БД [1-3]. При кажущемся сходстве решаемых задач между ними имеются глубокие различия. Онтология служит для распространения информации, определяя на формальном языке концепции и соотношения, которые представляют содержание и структуру предметной области. В то же время концептуальная схема БД, определяя все понятия и структуру данных, служит только для тех целей, что реализует конкретная БД. Распространение информации посредством онтологии проводится согласованным образом, то есть передаваемая ею структура данных является общедоступной и одинаково трактуемой в определенном сообществе. Все члены сообщества могут использовать онтологию и имеют доступ к информации.

Авторы ряда публикаций специально выделили все факторы, определяющие сходства и различия обеих конструкций. Их сводка, зафиксированная в лекции [2], приведена в таблице 1.