

ПРЕДМЕТНО-ОРИЕНТИРОВАННАЯ СИСТЕМА НАУЧНОЙ ОСВЕДОМЛЕННОСТИ ПО ФИЗИЧЕСКОЙ ХИМИИ РАДИКАЛЬНЫХ РЕАКЦИЙ

А.И. Прохоров, В.Е. Туманов, Д.Ю. Лазарев, М.Е. Соловьева

Введение

Успешное использование и развитие систем деловой осведомленности или бизнес - аналитики (Business Intelligence System, BI System) привел к идее разработки и создания систем научной осведомленности (Science Intelligence System, SI System). В статье [1] *система научной осведомленности* определена «как информационная инфраструктура, которая обеспечивает принятие решений и совместную работу научного сообщества в рамках выделенной предметной области знаний». Там же рассмотрена общая архитектура таких систем в разрезе категорий основных пользователей и использования современных информационных технологий.

В основу создания таких систем положены технологии складирования данных (Data Warehousing) и анализа и извлечения знаний (Data Mining) [2]. В системах научной осведомленности научные решения, методология и методы исследований интегрируются в общую библиотеку решений, а данные из разнородных источников интегрируются в общее хранилище данных, которое через предметно-ориентированные информационные ресурсы поставляют информацию пользователям: ученым, аспирантам, студентам, технологам, представителям промышленности. Особенностью систем научной осведомленности является предоставление пользователям, помимо собственно профессиональной информации, инструментария для анализа данных.

В [3] был рассмотрен общий подход к построению систем управления фундаментальными знаниями на примере физико-химических данных. В [4] был использован аналогичный подход для создания интеллектуальной информационной системы по физической химии радикальных реакций, при этом модель системы рассматривалась с точки зрения внедрения в нее элементов прикладного искусственного интеллекта для производства новых предметных знаний.

В [5] было дано определение *предметно-ориентированных систем научной осведомленности*, как узкоспециализированных систем научной осведомленности, которые кроме возможности решения задач интеллектуального анализа данных наделены способностью производства новых профессиональных знаний.

Целью настоящей работы является описание программно-технологической архитектуры предметно-ориентированной системы научной осведомленности по физической химии радикальных реакций и ее отдельных программных компонентов.

Предметно-ориентированная система научной осведомленности по физической химии радикальных реакций

Разрабатываемая предметно-ориентированная система научной осведомленности по физической химии радикальных реакций рассматривается как интеллектуальная система в Интернет, назначением которой является сбор, хранение, верификация, извлечение, распространение и производство новых предметно-ориентированных знаний по физической химии радикальных реакций.

Предметную область системы составляют следующие объекты и их основные характеристики:

- органическая молекула (химическая формула, наименование, характеристика реакционного центра, CAS Number, тип разрываемой связи, энергия диссоциации связи, энтальпия образования молекулы),
- радикал (химическая формула, наименование, характеристика реакционного центра, CAS Number, энтальпия образования радикала),
- радикальная реакция (реагенты, тип реакции, реакционная способность, фаза, растворитель, условия протекания реакции),
- библиографическая ссылка на источник фактографических данных,
- комплект лекций и задач для дистанционного обучения.

Для разработки и создания системы были использован подход на основе порталных технологий, технологии интеллектуальных агентов, технологии прикладного искусственного интеллекта, технологии баз и хранилищ данных. Активным компонентом системы является интеллектуальный агент, который можно представить в виде веб - приложения, наделенного искусственным интеллектом, и расположенного за внешним информационным порталом. При этом сами агенты ориентированы на обработку научных данных в узкоспециализированном разделе предметной области. Общая программно-технологическая архитектура системы показана на Рис. 1.

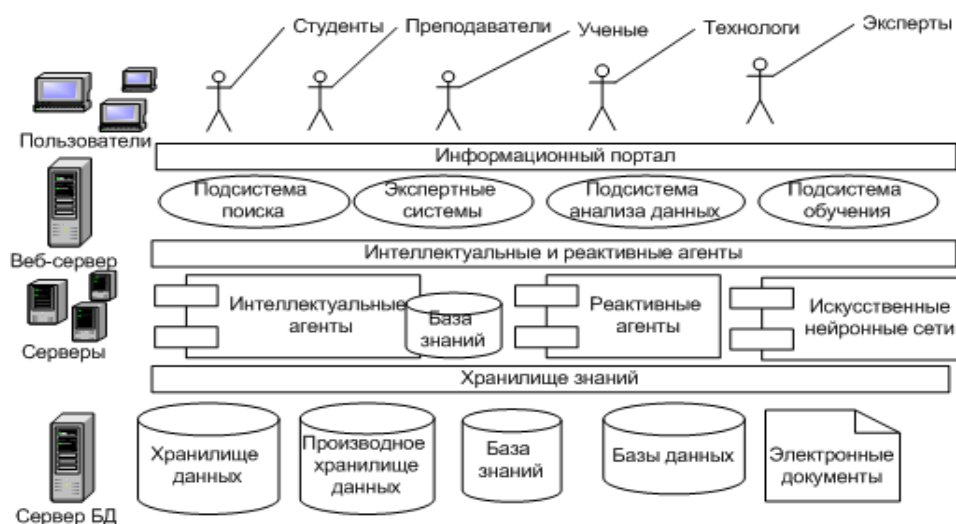


Рис. 1. Программно-технологическая архитектура предметно-ориентированной системы научной осведомленности по физической химии радикальных реакций.

Как показано на Рис.1, система состоит из нескольких программных слоев. Первый слой реализован как предметно-ориентированное веб-приложение, которое предоставляет пользователю интерфейс и принимает управляющие решения на основе таблиц решений. Веб-приложение предоставляет доступ к следующим программным компонентам системы: информационному и аналитическому компонентам, компоненту дистанционного обучения и компоненту производства новых профессиональных знаний.

Второй слой состоит из интеллектуальных агентов, реактивных агентов и обученных искусственных нейронных сетей, которые реализуют работу встроенных в портал экспертных систем и функций поиска информации. Агенты распределены в различных узлах локальной сети.

Третий слой представляет собой хранилище знаний, которое состоит из:

- хранилищ экспериментальных данных по константам скорости радикальных жидкофазных реакций, энергиям диссоциации связей молекул, энтальпиям образования молекул, реализованных как связанные киоски данных;
- производных хранилищ данных по константам скорости радикальных реакций в жидкой и газовой фазах, энергиям диссоциации связей молекул, энтальпиям образования молекул, в которые пользователи имеют возможность заносить полученные ими в результате работы с системой новые значения;
- Базы знаний, которая состоит из правил продукций, таблиц решений, процедур выполнения расчетов и алгоритмов кластерного анализа данных, и общих фактов, используемых экспертными системами;
- Электронных документов, которые в частности представляют собой материалы учебных лекций по физической химии радикальных реакций, а также тезаурусы терминов.

Система предназначена для решения следующих задач:

- Поиск данных о реакционной способности реагентов в бимолекулярных радикальных реакциях в жидкой фазе;
- Поиск значений энергий диссоциации связей органических молекул, а также энтальпий образования радикалов и молекул;
- Поиск в базе данных библиографических ссылок;
- Оценка реакционной способности реагентов радикальных бимолекулярных реакциях в жидкой и газовой фазах по термохимическим и кинетическим данным;
- Оценка энергий диссоциации связей органических молекул по кинетическим данным.

Хранилище знаний

Хранилище знаний в совокупности с экспертными системами образуют виртуальную подсистему производства новых профессиональных знаний (констант скоростей радикальных реакций, энергий диссоциации связей молекул).

Хранилище знаний есть предметно-ориентированная, интегрированная, поддерживающая временные ряды данных электронная коллекция, которая содержит данные, знания, процедуры генерирования знаний, и используется для анализа и исследования данных, производства новых знаний и поддержки принятия решений.

Хранилище знаний как компонент производства новых профессиональных знаний включает в себя (Рис. 2):

- хранилище данных для исследования (Exploration Data Warehouse), содержащее экспериментальные данные по реакционной способности радикальных реакций в жидкой фазе;

- встроенную экспертную систему для управления оценкой реакционной способности радикальных реакций;
- интеллектуальный агент для оценки константы скорости и энергии активации реакции в жидкой и газовой фазе;
- веб-сервис, посредством которого осуществляется вызов обученных искусственных нейронных сетей для прогнозирования значений констант скорости и энергии активации жидкофазных радикальных реакций определенных классов.
- производное хранилище данных, содержащее рассчитанные данные по реакционной способности радикальных реакций в жидкой и газовой фазах;
- встроенную экспертную систему для оценки энергии диссоциации связи молекул по кинетическим данным радикальных реакций;
- хранилище данных энергий диссоциации связей органических молекул, которое может быть пополнено новыми данными в результате работы экспертной системы оценки энергии диссоциации связей молекул по кинетическим данным радикальных реакций;
- тезаурус основных терминов и понятий предметной области;
- тезаурус описаний алгоритмов и процедур прогнозирования физико-химических характеристик молекул;
- базу знаний, содержащую производные правила и факты, которые используются встроенными экспертными системами.

В системе все хранилища данных реализованы как взаимосвязанные киоски данных.

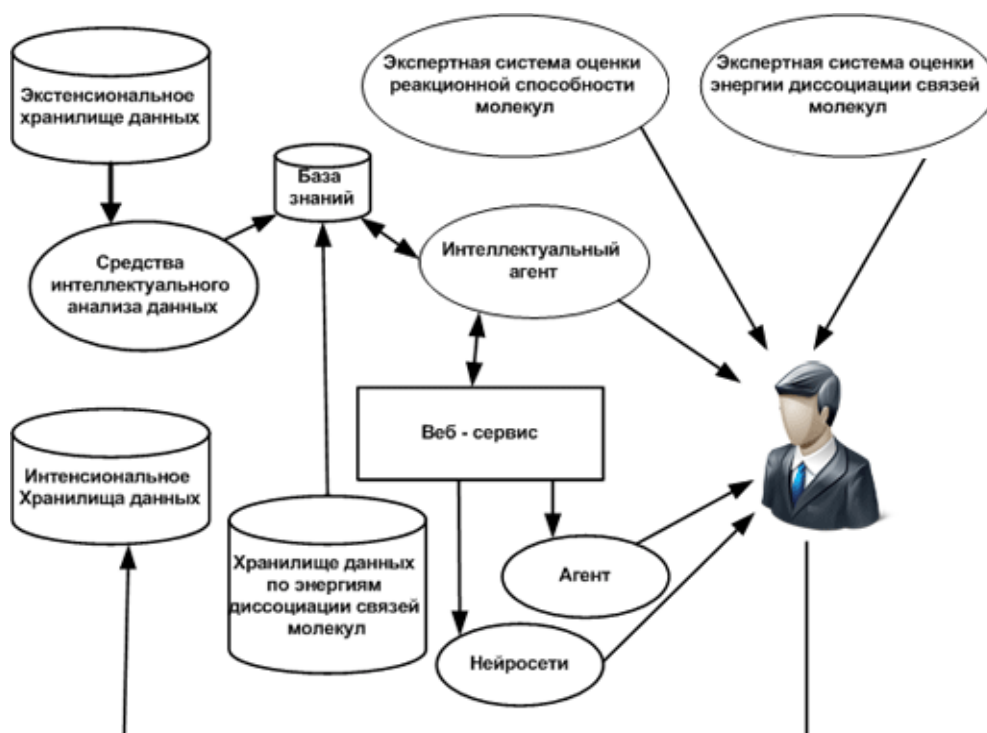


Рис. 2. Хранилище знаний предметно-ориентированной системы научной осведомленности по физической химии радикальных реакций.

В результате работы пользователей системы хранилище знаний пополняется новыми профессиональными знаниями. В качестве механизмов производства новых знаний в настоящей системе используются встроенные экспертные системы, обученные искусственные нейронные сети и интеллектуальные агенты. Как показала опытная эксплуатация системы использование встроенных экспертных систем в хранилище знаний оправдано в тех случаях, когда возможно использовать знания экспертов или получить такие знания в процессе интеллектуального анализа данных, что является трудоемким по времени процессом. Более перспективным оказалось использование искусственных нейронных систем, которые лучше аппроксимируют в данном случае зависимости в данных и дают более точный прогноз реакционной способности. Однако искусственные нейронные сети не всегда можно обучить из-за отсутствия достаточного набора данных.

Структуры данных для хранилищ и баз данных

Информационный компонент системы состоит из баз и хранилищ данных. Хранилища данных системы реализованы как взаимосвязанные киоски данных, построенных методом многомерного моделирования. Информационный компонент системы включает в себя:

- Киоск данных констант скоростей радикальных жидкофазных реакций (КД КСРЖФР);
- Киоск данных энергий диссоциации связей органических молекул;

- Киоск данных энтальпий образования органических молекул;
- Базу данных по энтальпиям образования радикалов;
- Базу данных библиографических ссылок.

Многомерная модель для КД КСРЖФР приведена на диаграмме Рис. 3, а перечень атрибутов представлен в Табл. 1. Поскольку киоски данных являются англоязычными, то сохранены английские наименования колонок таблиц.

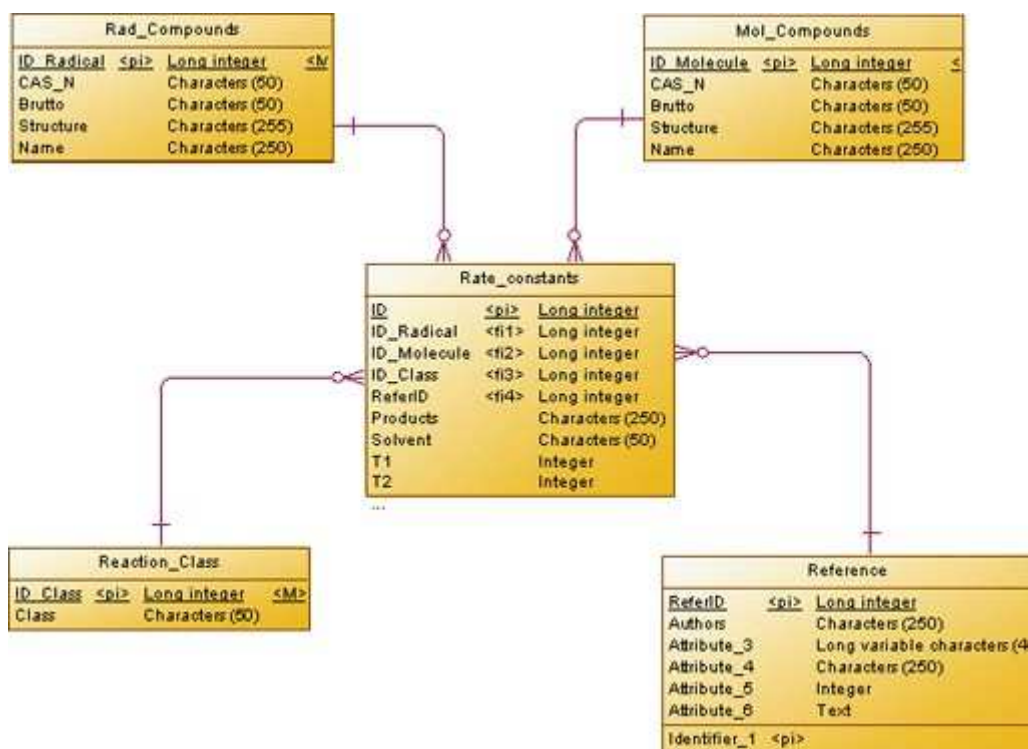


Рисунок 3. Логическая схема многомерной модели данных Киоска данных констант скоростей радикальных жидкофазных реакций

Как видно из диаграммы, таблица фактов *Rate_constants* содержит фактографические данные о константе скорости радикальной реакции и перечень ключей таблиц измерений. Факты не являются аддитивными и полуаддитивными по всем измерениям. Однако по комбинации измерений к некоторым фактам могут быть применены статистические функции. Будем называть такие факты полуаддитивным по статистической функции. Таким фактом является логарифм предэкспонента, к которому можно применить статистической среднее для определенной комбинации измерений, фиксирующей набор реакций с одинаковыми реакционными центрами.

Таблица 1. Перечень атрибутов Киоска данных по константам скорости радикальных жидкофазных реакций

Поле	Описание поля
Измерение <i>Mol_Compounds</i> - содержит перечень реагентов молекул	
ID_Molecule	Идентификатор молекулы
CAS_N	Регистрационный номер молекулы
Brutto	Атомарная формула молекулы
Structure	Полулинейная структурная формула
Name	Наименование молекулы
Измерение <i>Rad_Compounds</i> - содержит перечень реагентов радикалов или атомов.	
ID_Radical	Идентификатор радикала или атома
CAS_N	Регистрационный номер радикала
Brutto	Атомарная формула радикала
Structure	Полулинейная структурная формула
Name	Наименование радикала

Измерение <i>Reaction_Class</i> - содержит перечень классов реакций	
ID_Class	Идентификатор класса реакции
Class	Класс реакции
Измерение <i>Reference</i> - содержит ссылки на литературные источники	
ReferID	Идентификатор первоисточника

Authors	Авторы
Title	Наименование работы
Source	Описание первоисточника
Year_S	Год издания
Abstract	Аннотация работы
Таблица фактов <i>Rate_constants</i> содержит данные о константе скорости	
ID	Уникальный суррогатный ключ
ID_Class	Идентификатор класса реакции
ID_Radical	Идентификатор радикала или атома
ID_Molecule	Идентификатор молекулы

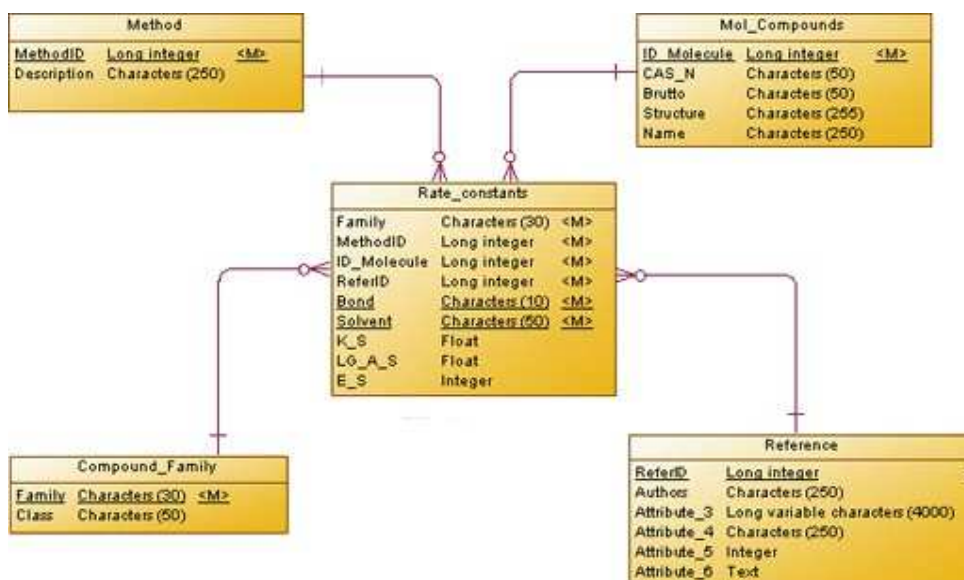


Рис. 4. Логическая схема многомерной модели данных Киоска данных энергии диссоциации органических молекул.

ReferID	Идентификатор первоисточника
Products	Продукты реакции
Solvent	Растворитель
T_T1_T2	Температура
K_S	Константа скорости реакции
LG_A_S	Логарифм предэкспонента
E_S	Энергия активации

Многомерная модель киоска данных энергии диссоциации связей органических молекул приведена на Рис. 4, а перечень атрибутов представлен в Табл. 2.

Таблица 2. Перечень атрибутов Киоска данных энергии диссоциации связей органических молекул

Поле	Описание поля
Измерение <i>Mol_Compounds</i> - содержит перечень реагентов молекул	
ID_Molecule	Идентификатор молекулы
CAS_N	Регистрационный номер молекулы
Brutto	Атомарная формула молекулы
Structure	Полулинейная структурная формула
Name	Наименование молекулы
Измерение <i>Compound_Family</i> - содержит классификацию органических молекул	
Family	Класс молекул
SubFamily	Подкласс молекул

Измерение <i>Reference</i> - содержит ссылки на литературные источники	
ReferID	Идентификатор первоисточника
Authors	Авторы
Title	Наименование работы
Source	Описание первоисточника
Year_S	Год издания
Abstract	Аннотация работы
Измерение <i>Method</i> — содержит описание методов определения энергии диссоциации связи	
MethodID	Идентификатор метода измерения или оценки
Description	Описание метода
Таблица фактов <i>BDE_value</i> содержит данные о энергии диссоциации связи	
ID_Molecule	Идентификатор молекулы
Bond	Тип связи
Fragment	Фрагмент молекулы
ReferID	Идентификатор первоисточника
MethodID	Идентификатор метода измерения или оценки
Dn	Значение энергии диссоциации связи
dDn	Величина погрешности
NOB	Число связей данного типа в молекуле

Логическая структура киоска данных энтальпий образования молекул аналогична многомерной модели данных киоска данных энергий диссоциации связей органических молекул, с той лишь разницей, что таблицы измерений находятся в логической связи с таблицей, содержащей значения энтальпии образования молекул. Логическая структура базы данных по энтальпиям образования радикалов включает в себя таблицу измерения *Rad_Compounds* и таблицу, содержащую значения энтальпий образования радикала.

Логическая структура базы данных библиографических ссылок представляет собой объединение в виртуальной таблице таблиц измерений *Reference* киосков данных.

Интерфейс системы

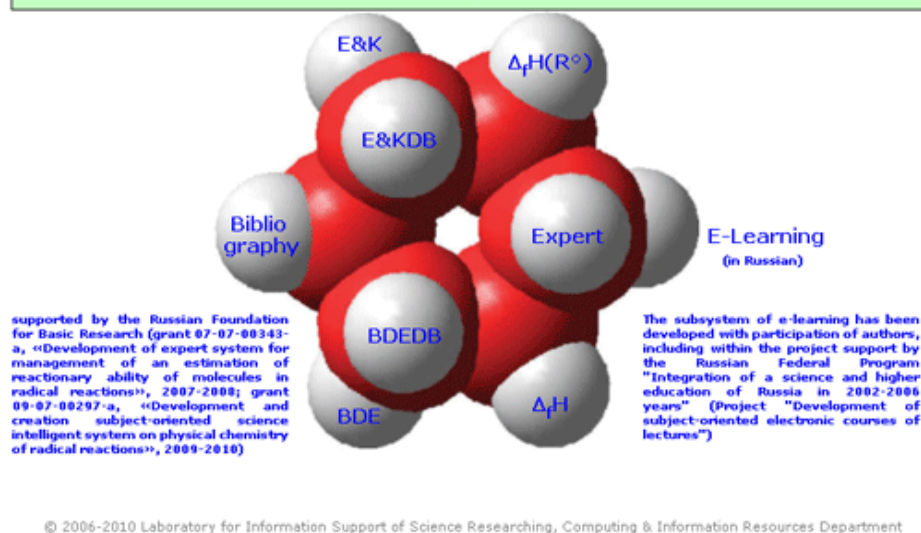


Рис. 5. Главное меню предметно-ориентированной системы научной осведомленности по физической химии радикальных реакций

На Рис. 5 приведена электронная форма с главным меню предметно-ориентированной системы научной осведомленности по физической химии радикальных реакций.

Каждый пункт меню (URL-ссылка) соответствует варианту использования системы. Пункт меню «E&k» отсылает пользователя к подсистеме оценки реакционной способности элементарной реакции, которая включает в себя встроенную экспертную систему и набор искусственных нейронных сетей для производства новых значений констант скорости и энергий активации радикальной химической реакции.

Пункт меню «E&kDB» предоставляет пользователю интерфейс для поиска данных о реакционной способности молекул в радикальных реакциях в жидкой и газовой фазах. Киоск экспериментальных данных по константам скорости молекул в радикальных реакциях в жидкой фазе насчитывает более 31000 записей. Данные по газовой фазе включают в себя только расчетные значения констант и энергий активаций. База данных по константам скорости молекул в газовой фазе собирается NIST и опубликована на сайте <http://kinetics.nist.gov/kinetics/index.jsp>.

Пункт меню « $\Delta_f H(R^\circ)$ » предоставляет пользователям доступ к базе данных по энтальпиям образования радикалов. База данных в настоящее время содержит более 1000 значений. Пункт меню « $\Delta_f H$ » предоставляет пользователю доступ к базе данных по энтальпиям образования молекул. База данных в настоящее время также содержит более 1000 значений. Содержание обеих баз данных основывается на результатах полученных авторами [6]. Аналогичные базы данных представлены на сайтах NIST (<http://webbook.nist.gov/chemistry/>) и МГУ им. М.В. Ломоносова (<http://www.chem.msu.ru/rus/handbook/ivtan/welcome.html> и <http://www.chem.msu.ru/cgi-bin/tkv.pl?show=welcom.html/welcome.html>).

Пункт меню «BDEDB» предоставляет пользователям доступ к киоску данных по энергиям диссоциации связей органических молекул, который в настоящее время содержит более 1000 записей.

Пункт меню «BDE» отсылает пользователя к экспертной системе для оценки энергий диссоциации связей органических молекул по кинетическим данным радикальных реакций отрыва.

Пункт меню «Bibliography» предоставляет пользователю доступ к базе данных библиографических ссылок.

Пункт меню «Expert» предоставляет доступ к интерфейсу эксперта, который может редактировать данные в базах данных и хранилище знаний.

Пункт меню «e-learning» отправляет пользователя в подсистему дистанционного обучения, которая включает в себя электронные курсы лекций, электронные задачки и тесты контроля знаний в удаленном режиме.

Активными элементами (агентами) производства новых знаний в системе являются встроенные экспертные системы (по прогнозированию констант скорости и энергии диссоциации связей), а также экспертная система контекстного управления навигацией пользователя. Полученные в результате использования пользователем экспертных систем константы скоростей реакций или энергии диссоциации связей органических молекул могут быть сохранены в хранилище знаний. Такая возможность делает систему активной и позволяет заинтересованному научному сообществу накапливать в ней новые знания.

Возможность пополнения системы предметными знаниями накладывает на ее функционал определенные ограничения. Такие ограничения связаны с необходимостью обеспечить достоверность заносимых в нее данных. Поэтому хранилище знаний системы состоит из двух разделов: базового раздела, составленного экспертами по данным научных публикаций, и раздела, произведенного пользователями системы.

При попытке занесения новых данных экспертная система контекстного управления на основе нечетких рассуждений делает ряд проверок и выводов о достоверности этих данных, а затем принимает решение либо о занесении данных с определенным показателем их надежности, либо об отказе в запоминании данных. При сохранении данных система просит пользователя заполнить анкету.

Заключение

Разработка и публикация в Интернет предметно-ориентированных систем научной осведомленности на основе использования хранилищ знаний позволит научному сообществу создавать распределенные сети для сбора, хранения, извлечения, интеллектуального анализа, распространения и производства знаний в узкоспециализированных областях исследований и технологий.

Включение в такие системы подсистемы дистанционного обучения предметно-ориентированным знаниям значительно расширяет круг ее потенциальных пользователей (студентов и аспирантов), что способствует самостоятельному формированию у них профессиональных знаний, а преподавательскому составу высших учебных заведений предоставляет дополнительный учебный материал и электронный ресурс-справочник.

Работа выполнена при поддержке гранта РФФИ 09-07-00297-а «Разработка и создание предметно-ориентированной системы научной осведомленности по физической химии радикальных реакций».

ЛИТЕРАТУРА:

1. R. Hackathorn Science Intelligence. Can a Business Intelligence Approach Enable “Smart” Science?. DM Review. 2005. (<http://www.DMReview.com>)
2. R.J. Thierauf Effective Business Intelligence Systems. Westport. Quorum Books. 2001. 392 p.
3. Q. Dong, X. Yan, R.D. Chirico, R.C. Wilhoit, M. Frenkel Database Infrastructure to Support Knowledge Management in Physicochemical Data. // 18-th CODATA Conference. 2002, Sep 29 – 3 Oct., Montreal, Canada, 36 P.
4. V.E. Tumanov Data Warehousing and Data Mining in Thermochemistry of Free Radical Reactions. // Fourth Winter Symposium on Chemometrics "Modern Methods of Data Analysis". Russia. Chernogolovka. February. 15-18. 2005. P. 28-29.
5. В.Е. Туманов Предметно-ориентированные системы научной осведомленности. // Информационные технологии. 2009. № 5. С. 12-18.
6. Е.Т. Денисов, В.Е. Туманов Оценка энергий диссоциации связей по кинетическим данным радикальных жидкофазных реакций. // Успехи химии. 2005. Т. 74. № 9 С. 905-938.