

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ
ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 12

Москва 2015

ОБЩИЙ РАЗДЕЛ

УДК 004.65 : [54.03/.04 : 620.22]

А.О. Еркимбаев, В.Ю. Зицерман, Г.А. Кобзев, А.В. Косинов

Связывание онтологий с базами данных по свойствам веществ и материалов*

Рассмотрены возможности нового подхода, интегрирующего потенциал онтологии и традиционной базы данных, применительно к задачам хранения и распространения данных по свойствам веществ и материалов. Представлен обзор нескольких технологий проектирования и интеграции научных баз данных, ориентированных на их объединение с онтологиями. Показано, что использование онтологии предметной области обеспечивает стандартизацию терминов и понятий, исключаящую их неоднозначность и ошибочную интерпретацию, унификацию семантики, согласованную в экспертном сообществе, распространение знаний, сопровождаемое возможностью логического вывода и автоматизацией поиска. В качестве примеров использования онтологии рассмотрены: одна из наиболее масштабных баз данных в области химии ChEBI; несколько систем хранения и интеграции данных по свойствам материалов; авторский проект базы данных по теплофизическим свойствам индивидуальных веществ.*

Ключевые слова: численные данные, базы данных, онтология, систематизация данных

* Работа выполнена при поддержке РФФИ – проект № 13-07-00218

ВВЕДЕНИЕ

Накопление и систематизация фактографических данных о свойствах веществ и материалов издавна является одним из основных направлений деятельности в естественнонаучных и инженерных дисциплинах: физической химии, кристаллографии, материаловедении и др. [1–3]. Широкое использование для этих задач электронных ресурсов (баз данных, вычислительных систем, Web-порталов и др.) при постоянно возрастающем объеме их использования позволяет отнести эту сферу научной деятельности к так называемой *e-Science*, которую отличает не столько компьютеризация исследования, сколько преимущественная ориентация на работу с данными, включая их получение, систематизацию, хранение и распространение для коллективной работы. Особенности *e-Science* позволяют говорить о новом этапе в научной методологии, *4-й парадигме* [4], дополняющей стандартную триаду: теория, эксперимент и моделирование.

Ключевым (хотя далеко не единственным) элементом при работе с данными о свойствах является база данных (БД). Если на определенном этапе переход к БД от многотомных справочников обеспечил заметные достижения в создании и поддержке информационных ресурсов, то со временем резкий рост объемов данных при многообразии их форматов и моделей привел к тому, что действующая инфраструктура оказалась неспособной обеспечить эффективную организацию рабочего процесса [5].

Автономность БД с жесткой фиксацией используемых терминов и логических структур практически исключает возможность беспрепятственного обмена данными без активного участия человека-эксперта. В информационном сообществе в качестве главного ограничителя возможностей БД называют так называемую *потерю семантики* в ходе проектирования БД, когда принятая концептуальная модель реализуется в виде физической модели. По сути, это означает, что каждая из БД имеет собственный понятийный аппарат, который (без специальных инструкций) остается неизвестным для пользователя. Чтобы построить правильный запрос, пользователь любой из БД должен знать названия сущностей и их атрибутов, как минимум – принятые названия веществ и их физических свойств. При этом любые тематически родственные БД с неизбежностью имеют различия в концептуальной схеме.

Применительно к свойствам вещества первым шагом к преодолению неоднородности БД была бы стандартизация названий путем использования общедоступных и одинаково трактуемых индексов, например **CAS Register Number**. По сути, это проблема повторяет на новом этапе ту же проблему стандартизации, которая еще до компьютерной эры заняла важное место в международном сотрудничестве в физике, химии, кристаллографии и др. Под эгидой IUPAC (Международный союз по чистой и прикладной химии) и CODATA (комиссия по численным данным в науке и технологии) разработаны многочисленные рекомендации для обозначений, терминологии, химической номенклатуры, единиц

измерения и пр. Повсеместный переход к компьютерным средствам заметно обостряет проблему, поскольку встает вопрос не только одинаковой трактовки человеком, но и совместимости данных с множеством приложений и компьютерных средств поиска. Поэтому даже такой простейший аспект, как разноречивость в названии веществ, обусловленный богатой синонимией, может стать непреодолимым барьером при поиске.

Проблема интеграции ресурсов и, прежде всего, БД достаточно давно обсуждается как в компьютерном сообществе, так и среди экспертов по систематизации научных данных [1, 3, 5, 6]. Один из очевидных подходов – выработка принятой в сообществе логической схемы, обеспечивающей стандартизацию в именовании веществ, свойств, состояний и прочих аспектов. В качестве интегрирующей модели стала использоваться модель, основанная на стандартах XML, долгое время служившего наиболее используемым форматом при обмене структурированной информацией – между программами, между людьми, между компьютером и человеком [7]. XML, или расширяемый язык разметки (*eXtensible Markup Language*) создан для описания данных и концентрирует свое внимание на их **сущности**, что принципиально отличает его от языка HTML, созданного для отображения, т.е. показа данных. Существенно, что XML – это свободный и расширяемый язык, где разработчик имеет возможность предлагать свои теги, используя которые можно создать структуру документа в полной аналогии с таблицами реляционной БД. Тот факт, что данные хранятся в обычных текстовых файлах, позволяет в технологии XML использовать программно и аппаратно независимые решения. Обмен данными путем их преобразования в XML-формат позволяет упростить связь несовместимых сред и делает данные доступными для различных программ, благодаря чему он стал языком, применяемым для обмена информацией через Интернет, а сами XML-документы своеобразными хранилищами данных. Стандарт под названием XML Schema предоставляет значительную часть информации, которая хранится обычно в БД, и может быть принят как стандарт структуры данных в определенной профессиональной среде.

Разумеется, XML-технология была подхвачена в естественнонаучных дисциплинах для целей обмена и интеграции численных данных. Во множестве дисциплин появились собственные версии XML [8] со своими словарями, средствами поддержки в виде настраиваемых браузеров и программ, реализующих графические представления, вычислительные сервисы и пр. Они позволяют работать с данными самой разной структуры и содержания, например, полученными в молекулярной динамике (MODML), вычислительной нейрофизиологии (NeuroML), биоинформатике (BSML) и даже *спелеологии* (Cave Survey Markup Language). Широкую известность и распространение получили языки для описания химической информации (CML) [9] и математических выражений (MathML) [10].

Для стандартизации обмена данными в термодинамике и материаловедении были разработаны и ши-

роко анонсированы языки ThermoML и MatML. Проект ThermoML [11], поддержанный IUPAC, в качестве основной цели имел стандартизацию форм хранения и обмена теплофизическими данными для более чем 120 свойств при различных формах представления и различном статусе данных (экспериментальные, расчетные, справочные). Близкие по сути задачи призван решать язык MatML [12, 13], с той разницей, что объектами являются не вещества с известной стехиометрией, а материалы, свойства которых существенно зависят от технологии изготовления, факторов внешнего воздействия и т.п. Возможности и ограничения в применении обоих языков достаточно подробно описаны в нашей работе [6].

К этой же тематике, связанной с физическими свойствами, относится стандарт XSAMS: XML Schema for Atomic, Molecular and Solid Data [14], созданный для совместимости БД, содержащих параметры атомных и молекулярных процессов, а также характеристик взаимодействия частиц с поверхностью твердого тела. Стандарт предлагает физикам универсальную структуру, но только для обмена данными, без контроля их достоверности. Процесс описывается путем ссылки на начальное и конечное состояние, а сами данные представлены либо в табличной форме, либо в виде параметров аппроксимирующих выражений. Поскольку считаются важными источник и происхождение данных, схема накладывает жесткие требования на информацию по источникам данных и методам их генерации. Полное описание структуры XSAMS-документов можно найти в руководстве [15].

Постепенно, однако, в информационном сообществе пришло осознание недостатков, присущих XML-технологиям, особенно при стандартизации данных в науке. XML-схема обеспечивает синтаксис и структуру данных, но не их семантику (т.е. смысл), необходимую при гибком отображении словарей и схем данных. В итоге для тех же целей роль XML-схем заняли онтологии, формализующие в закодированном виде структуру знаний. Подробнее их возможности в организации научного знания описаны в следующем разделе. Здесь достаточно отметить, что размещение в сети онтологии, закодированной на языке OWL (Web Ontology Language), позволяет каждому из понятий предметно-ориентированного словаря присвоить уникальный идентификатор, так называемый URI (Uniform Resource Identifier), на который могут ссылаться автономные ресурсы, обеспечивая его единую трактовку.

Ориентация на онтологии легла в основу концепции **онтологически основанных БД**, которая предусматривает семантическое отображение структур данных, потенциально сводящее локальные БД в единое пространство данных при согласовании их терминологии и логической структуры. Доступность семантики при компьютерной обработке данных – не единственное преимущество онтологий при их сопоставлении с традиционными БД или жестким XML-стандартом. К достоинствам онтологического подхода относятся: легкость запроса информации при общедоступной семантике; возможности инте-

грации ресурсов; эффективные средства проектирования БД.

Осознание огромного потенциала, заложенного в онтологиях для систематизации и распространения научного знания, стимулировало активную деятельность по разработке предметно-ориентированных онтологий для формализации отдельных сегментов многих научных дисциплин: медико-биологических, химии, наук о Земле и др. Появились онтологии для наук о материалах [16–18], а также наноструктур [19]. Параллельно в сфере информационных технологий разработаны новые продукты, полученные гибридизацией БД и онтологий. Механизм работы всех этих продуктов использует некоторое взаимное отображение между концептуальной схемой БД и онтологией.

В целом традиционные БД и онтологии по своим возможностям для систематизации данных оказались взаимно дополнительны. БД способны к структурированию и хранению больших объемов данных при высокой эффективности их поиска. С одной стороны, онтологии позволяют устранить их главные дефекты – отсутствие общепонятной семантики, жесткую, неспособную к эволюции структуру. С другой стороны, сама онтология совместно с экземплярами данных не способна заменить БД из-за падения эффективности поиска с ростом объема данных.

Цель настоящей работы – выявить возможности и преимущества новых технологий, использующих связку «онтология–БД» при организации научных данных по свойствам вещества. Онтология обеспечит для предметной области, определяемой кругом объектов и набором свойств, терминологический словарь, соотношения между понятиями и структуру данных, а ее связка с БД – возможности хранения при эффективном поиске больших объемов данных.

ОНТОЛОГИИ И БД – ВОЗМОЖНОСТИ СОВМЕСТНОГО ИСПОЛЬЗОВАНИЯ

В последние годы, концепция **онтологии**, исходно возникшая как философская категория, перешла в сферу информатики, где она определяет структуру для формальной спецификации понятий и отношений в определенной области знаний. Наиболее простым и точным ее определением считается то, которое дал Грубер [20]: «Онтология есть спецификация концептуализации», т.е. стандартизованное и закодированное представление знаний в определенной предметной области. Кодирование или запись на машинно-интерпретируемом языке (например, OWL) обеспечивает принципиальную возможность восприятия и поиска данных программными агентами. Главное назначение онтологии в том, что они добавляют к данным семантику (смысл, правильное толкование) и соотношения, что в совокупности описывает «знание» с возможностью его машинной интерпретации.

При разработке онтологии используют четыре типа составляющих элементов: понятия/концепты, соотношения, экземпляры и аксиомы. Концепты – это классы сущностей или объектов в определенной области знания, например, вещества, свойства, состояния и т.п. Отдельные классы связаны соотношениями – либо типа класс–subclass (таксономия), либо ассоциативными соотношениями, которые определяют

свойства или роли отдельных понятий. Общие понятия конкретизируют экземпляры, так что онтология вместе с экземплярами, представленная на OWL, выполняет функцию записи данных, примерно так же, как XML-документ (см. Введение). Наконец, аксиомы устанавливают присущие данной области ограничения значений отдельных понятий или экземпляров. Например, при определении свойств молекул аксиомы могут выделить класс радикалов, по отношению к которым лишено смысла их отнесение к твердой фазе. Тем самым аксиомы позволяют достаточно формально представить многочисленные «реалии» предметной области.

Оценить богатый потенциал, заложенный в онтологиях для хранения и распространения данных, можно, сопоставляя их возможности с БД [21]. 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 время как 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, т.е. передаваемая ею структура данных является общедоступной и одинаково трактуемой в определенном сообществе. Все члены сообщества могут использовать онтологию и имеют доступ к информации.

Простой пример дает единое соглашение об именовании веществ, принятое согласно онтологии **ChEBI** (Chemical Entities of Biological Interest Ontology) [19], которая присваивает каждому из низкомолекулярных веществ уникальный идентификатор, и для согласованной номенклатуры необходима ссылка на эту онтологию и соответствующий веществу ID. Так, ID, присвоенный метану, имеет вид **CHEBI:16183**, а ссылка в любой из БД на уникальный URI www.ebi.ac.uk/chebi/

searchId.do?chebiId=CHEBI:16183 обеспечивает единую (и доступную для компьютера) трактовку этого термина. Аналогично можно обеспечить ссылки и на другие понятия, представленные в онтологии: названия свойств, физических состояний, единиц измерения и т.п.

Более полный перечень ключевых различий между БД и онтологией приведен в табл. 1 [21]. Наряду с семантикой, которая теряется в ходе проектирования БД, в таблице указаны некоторые дополнительные признаки. В частности, БД немыслима без дополнения в виде данных, в то время как в онтологии наличие экземпляров не обязательно (*instances optional*). С другой стороны, онтология немыслима без таксономии классов (хотя и не сводится к ней), а в БД таксономия как структурный компонент отсутствует. В итоге формальный характер онтологии позволяет реализовать машинные выводы и рассуждения, что совершенно не предусмотрено в концептуальной схеме БД.

Применительно к тематике «свойства веществ и материалов» особую роль играет возможность на уровне онтологий поддерживать эволюцию схемы данных, связанную с расширением круга объектов и появлением новых, ранее неизвестных понятий (последний пункт в табл. 1). Как отмечалось в наших работах [1, 2, 22], вся практика систематизации данных по свойствам показывает невозможность поддерживать их логическую схему для широкой совокупности веществ, не вступая в противоречие с объективной потребностью передать их специфику и описания, принятые в разных коллективах. Хорошим примером служит возникшая необходимость перестройки стандарта ThermoML при переходе к задачам представления данных для свойств биоматериалов (ферментов и протеинов) и равновесий в растворах [23]. Другой пример эволюции схемы данных связан с постепенным включением наноструктур в химические БД, что с необходимостью ведет к изменению правил идентификации [1, 19].

Таблица 1

БД и онтологии – характерные признаки и отличия

Концептуальная схема БД	Онтология
Определяет структуру БД на формальном языке	Определяет набор концепций и соотношений, которые представляют содержание и структуру предметной области на формальном языке
Фокусируется на данных	Фокусируется на смысле
Сущности	Классы
Атрибуты	Соотношения
Ограничения	Аксиомы
Нет таксономии	Таксономия – ключевой элемент
Данные – ключевой элемент	Экземпляры данных не обязательны
Семантика только в концептуальной схеме, спроектирована для человека, не эволюционирует с изменением БД и приложений	Семантика – ключевой элемент, доступный программной обработке
Схему трудно изменять и поддерживать	Потенциально легче изменять и поддерживать

Вывод о безусловном преимуществе онтологий был бы, однако, совершенно преждевременным. Помимо того, что БД обеспечивают наиболее популярный и эффективный способ хранения данных, их отличает высокая производительность при поиске и реализации сложных запросов, недостижимая для других архитектур. Поэтому, рассматривая роль онтологий в информационной инфраструктуре, исследователи выбрали в качестве основной стратегии не «вытеснение» БД, а создание своеобразной связки путем так называемого *database-to-ontology mapping*. Ее задача использовать преимущества обеих конструкций, в основном за счет соединения семантики с высокой производительностью при работе с данными. Анализ предложенных концепций и продуктов дан в ряде обзоров [24–26]. Авторы [26], исходя из того, какая из двух концепций (онтология или схема БД) может рассматриваться как ведущая, выделили две стратегии. Первая из них ориентирована на использование онтологий для усиления функциональности БД. Тогда БД рассматривается как ведущий элемент, усиленный связью с онтологией для целей: реализации запросов с использованием согласованной в сообществе специалистов семантики; облегчения проектирования БД; обеспечения интеграции нескольких БД. Вторая стратегия, напротив, рассматривает БД как возможный инструмент усиления или даже разработки онтологии.

В соответствии с двумя стратегиями можно выделить две концепции взаимного отображения БД и онтологии и два типа продуктов. Первой концепции соответствует БД, основанная на онтологии (**DBBO** – *database based on ontology*), когда смысл сущностей, записанных в БД, определен ссылкой на соответствующую онтологию. Для каждой таблицы или атрибута устанавливается связь с концепцией, включенной в онтологию, – см. приведенный выше пример со ссылкой на название вещества в онтологии **ChEBI**. Вторая концепция предполагает проектирование структуры, получившей название онтологии, основанной на БД (**OBDB**, *Ontology Based on DB*). Задача такой структуры разместить экземпляры онтологии в БД, чтобы обеспечить достаточную эффективность при управлении, т.е. загрузке, поиске и реализации сложных запросов. В основном эта концепция нацелена на решение задач Semantic Web, где требуется организовать хранение и управление архивами документов, записанных на языке **RDF** (*resource definition framework*), который используется для связывания тематически родственных документов в глобальной сети [27].

В настоящей статье нас будет интересовать первая стратегия, ориентированная на технологию БД. Система, построенная по принципу **DBBO**, включает: исходно принятую БД с заполняющими ее записями; онтологию для семантического индексирования БД; возможные ссылки на другие онтологии с целью расширения словаря; соотношения между каждым элементом БД и онтологическим понятием. Онтология в такой структуре лишь обеспечивает семантику, но не содержит экземпляров – их роль выполняют записи БД. По сравнению с традиционной архитектурой БД, система **DBBO** включает два дополни-

тельных уровня. Первый из них (онтологический) включает одну или несколько онтологий для определения понятий предметной области, причем эти определения никак не связаны с конкретными приложениями, для которых спроектирована БД. Второй уровень (концептуальный) представлен соотношениями, связывающими онтологический уровень с логической моделью данных, причем выбор концептов может быть использован для автоматической генерации логической модели БД.

Следует заметить, что выделение концепции **DBBO** (как и сама аббревиатура), по-видимому, принадлежит авторам [26], в то время как другие авторы [24, 25] неправомерно используют аббревиатуру **OBDB** вне зависимости от конкретной схемы взаимного отображения онтологии и БД. Работ, посвященных методам усиления БД за счет связи с онтологией, пока немного, хотя уже предложены конкретные алгоритмы отображения и языки запросов, расширяющие возможности традиционного языка SQL. Особый интерес представляет алгоритм **Round-Trip Engineering** («поездка в оба конца») [28], способный обеспечить согласование реляционной схемы БД и концептуальной схемы или онтологии в тех случаях, когда возникает потребность в поддержке их эволюции при расширении информационного содержания. Алгоритм позволяет синхронизировать концептуально-реляционное отображение с появлением новой семантики.

Усиление функциональности за счет онтологии возможно не только для уже существующей БД, но и на этапе ее проектирования. От проектировщика требуются хорошие знания о специфике предметной области, которые он должен формализовать в виде концептуальной схемы, используя различные инструменты типа **UML**. Для одной и той же предметной области схемы разных экспертов будут радикально отличаться, а сами БД становятся гетерогенными, по сути, недоступными для машинно-реализуемой интеграции. Подход, при котором в основу проекта БД положена определенная онтология, заметно ускоряет процесс разработки и снимает ряд возникающих ограничений. В частности, концептуальные схемы, предлагаемые разными экспертами, в известной степени унифицируются, поскольку базируются на единой модели области. Кроме того, уже на этапе проектирования снимается проблема семантики, поскольку каждый элемент данных получается отображением онтологического понятий, имеющих точное определение и смысл.

Наряду с проектированием БД, использование онтологий открывает возможности в интеграции БД, ранее созданных без ссылок на онтологии. Методы интеграции издавна занимают важное место в технологии БД [3, 6, 7]. Под интеграцией в общем случае понимается их физическое или виртуальное соединение в одном месте, что позволяет пользователю использовать общий интерфейс и единую систему запросов. В ходе интеграции производится объединение схем и самих данных. Онтология предоставляет естественный и достаточно эффективный способ объединения схем, позволяя преодолеть проблемы с отсутствием семантики в исходных БД. Пользователь

может формулировать в терминах онтологии запросы, которые будут обращены к «подключенным» к онтологии БД. Тем самым онтология будет играть роль эффективного посредника между пользователем и данными.

Подобный подход получил название *доступ к данным, основанный на онтологиях*, или **OBDA** (*ontology based data access*). В **ODBA**-системе источник может рассматриваться как единая БД. Хотя в целом такой подход к интеграции БД является относительно новым, уже сейчас, по мнению авторов [25, 26], можно указать несколько технологий, опробованных в разных проектах. Простейшая из них предполагает, что для набора БД принимается одна онтология как посредник между пользователем и данными. При этом сама онтология может включать в качестве модулей несколько специализированных онтологий. Существенно при этом, чтобы все объединяемые ресурсы примерно одинаково отражали специфику предметной области. Затруднения возникают, когда один из ресурсов отражает какой-то определенный аспект, не характерный для других. Кроме того, одна онтология может не справиться с

задачей интеграции ресурсов, когда их содержание расширяется так, что выходит за грань ранее принятой концептуализации. Альтернативой этой технологии служит использование множества онтологий, каждая из которых соответствует одной из подлежащих интеграции БД. В этом случае любые изменения или расширения, относящиеся к одному ресурсу, затрагивают только одну онтологию. Ключевым элементом технологии является процедура взаимного отображения онтологий (*alignment of ontologies*). Наконец, был предложен и промежуточный подход, когда каждому ресурсу ставится в соответствие одна онтология, но над ними располагается общий контролируемый словарь базисных терминов общей предметной области. Схематично различия трех технологий представлены на рис. 1.

В нашей работе [29] было показано, насколько эффективным оказывается онтология при проектировании БД по теплофизическим свойствам и интеграции данных из множества источников. Другие реализованные к настоящему времени примеры интеграции БД при посредстве онтологий по свойствам веществ рассмотрены в следующем разделе.

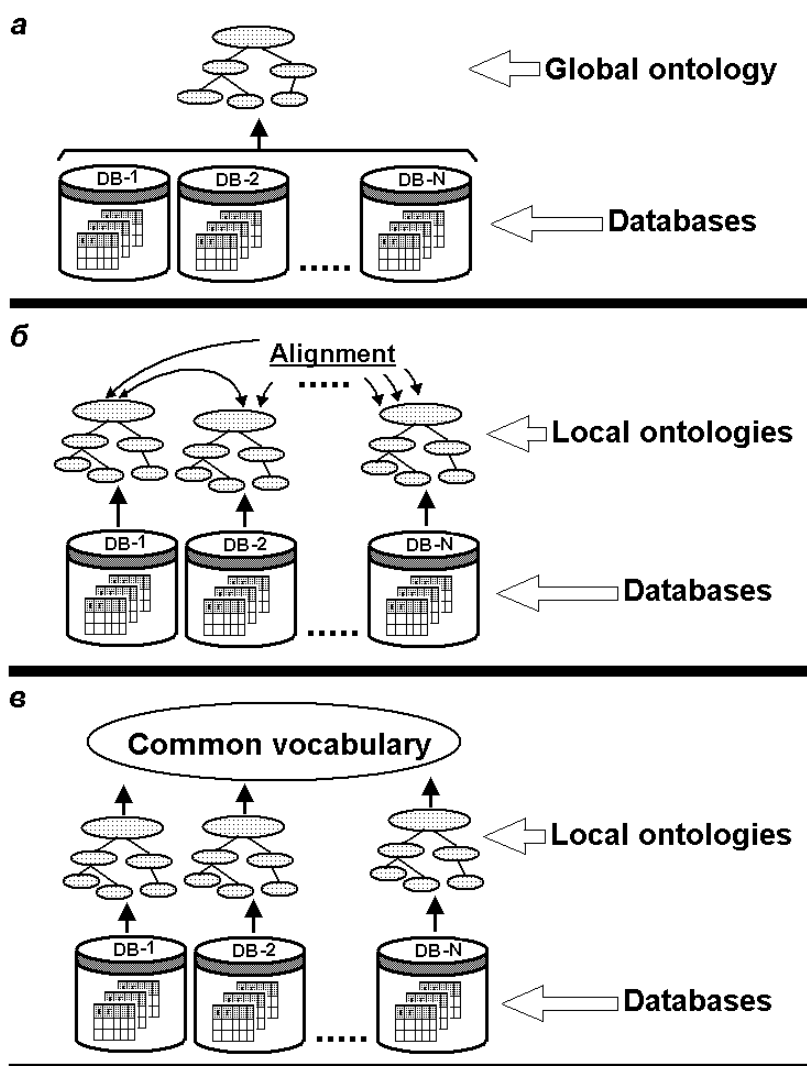


Рис. 1. Иллюстрация различий в технологиях **OBDA** (*ontology based data access*) [26]

ПРАКТИКА СВЯЗЫВАНИЯ ОНТОЛОГИЙ С БД ПО СВОЙСТВАМ ВЕЩЕСТВ И МАТЕРИАЛОВ

Один из наиболее впечатляющих примеров расширения функциональности БД за счет связывания с онтологией дает упомянутая выше БД химического содержания **ChEBI** [19], преимущественно ориентированная на низкомолекулярные соединения (исключены белки, нуклеиновые кислоты и пептиды). БД охватывает молекулярные сущности, их группы и классы. Первое из понятий относится к любым идентифицируемым по составу атомам, молекулам, ионам, наноструктурам и т.п., второе – к группам связанных атомов (или один атом) в составе молекулы (methyl, CH₃), а третье (класс) к совокупности молекул или групп, охваченных классификационным признаком, например, alkane (R-CH₃). Каждую сущность, а также группу и класс идентифицирует уникальный идентификатор **ChEBI ID** (например, **CHEBI:15377** для воды), свободный для цитирования в сети пользователем или программным агентом.

В записи для молекулярной сущности (рис. 2) указаны химические данные (определение, формула, масса, заряд), названия, как принятое в **ChEBI**, так и синонимы из других источников, структурная информация, кодированная в линейных нотациях (**InChI** и

SMILES), а также детализированная в **Molfile** (координаты атомов и матрица связности). Наряду с этим в записи представлены регистрационные номера в химических классификаторах (CAS, Beilstein и др.) и ссылки на другие БД с обширной информацией, например, термодинамической в БД NIST (webbook.nist.gov/chemistry/), или токсикологической в БД ChemIDplus (chem.sis.nlm.nih.gov/chemidplus/chemidlite.jsp).

Главным элементом онтологии, интегрированной с БД, является таксономия в виде ациклического графа, когда любое понятие может происходить от нескольких родительских понятий. Внизу рис. 2 показаны дочерние (incoming) и родительские (outgoing) классы, непосредственно связанные с веществом в записи (в данном случае **paracetamol**, **C8H9NO2**). На рис. 3 приведен фрагмент древовидной структуры, охватывающий всю таксономию, связанную с веществом, причем пользователь имеет возможность навигации с вызовом данных для любой сущности, представленной на «дереве». Обращение к URI с указанием онтологии и уникального ID непосредственно связывает внешний ресурс (или пользователя) с соответствующей записью, например URI www.ebi.ac.uk/chebi/searchId.do?chebiId=46195 связывает его с записью в БД для парацетамола.

The image shows a screenshot of a ChEBI record for paracetamol. The record is displayed in a web interface with several tabs: 'Main', 'ChEBI Ontology', and 'Automatic Xrefs'. The 'Main' tab is active. On the left, there is a chemical structure of paracetamol (4-acetaminophenol). To the right of the structure, there are several fields: 'ChEBI Name' (paracetamol), 'ChEBI ID' (CHEBI:46195), 'Definition' (A member of the class of phenols that is 4-aminophenol in which one of the hydrogens attached to the amino group has been replaced by an acetyl group.), 'Stars' (been manually annotated by the ChEBI Team.), 'Secondary ChEBI IDs' (CHEBI:46191, CHEBI:2386), and 'Supplier Information' (eMolecules:474380, eMolecules:27677450, ZINC18274777). Below these fields, there is a 'Download Molfile' button and a section for 'Chemical structure searches' with three links: 'Find compounds which contain this structure', 'Find compounds which resemble this structure', and 'Take structure to the Advanced Search'. At the bottom, there is a section for 'Additional chemical data' with fields for 'Formula' (C8H9NO2), 'Net Charge' (0), 'Average Mass' (151.16260), 'InChI' (InChI=1S/C8H9NO2/c1-6(10)9-7-2-4-8(11)5-3-7/h2-5,11H,1H3,(H,9,10)), 'InChIKey' (RZVAJINKPMORJF-UHFFFAOYSA-N), and 'SMILES' (CC(=O)Nc1ccc(O)cc1). Annotations with callout boxes point to the 'Recommended ChEBI name' (paracetamol), the 'Illustration' (chemical structure), and the 'Additional chemical data' (Formula field).

Рис. 2. Внешний вид типовой записи в БД ChEBI

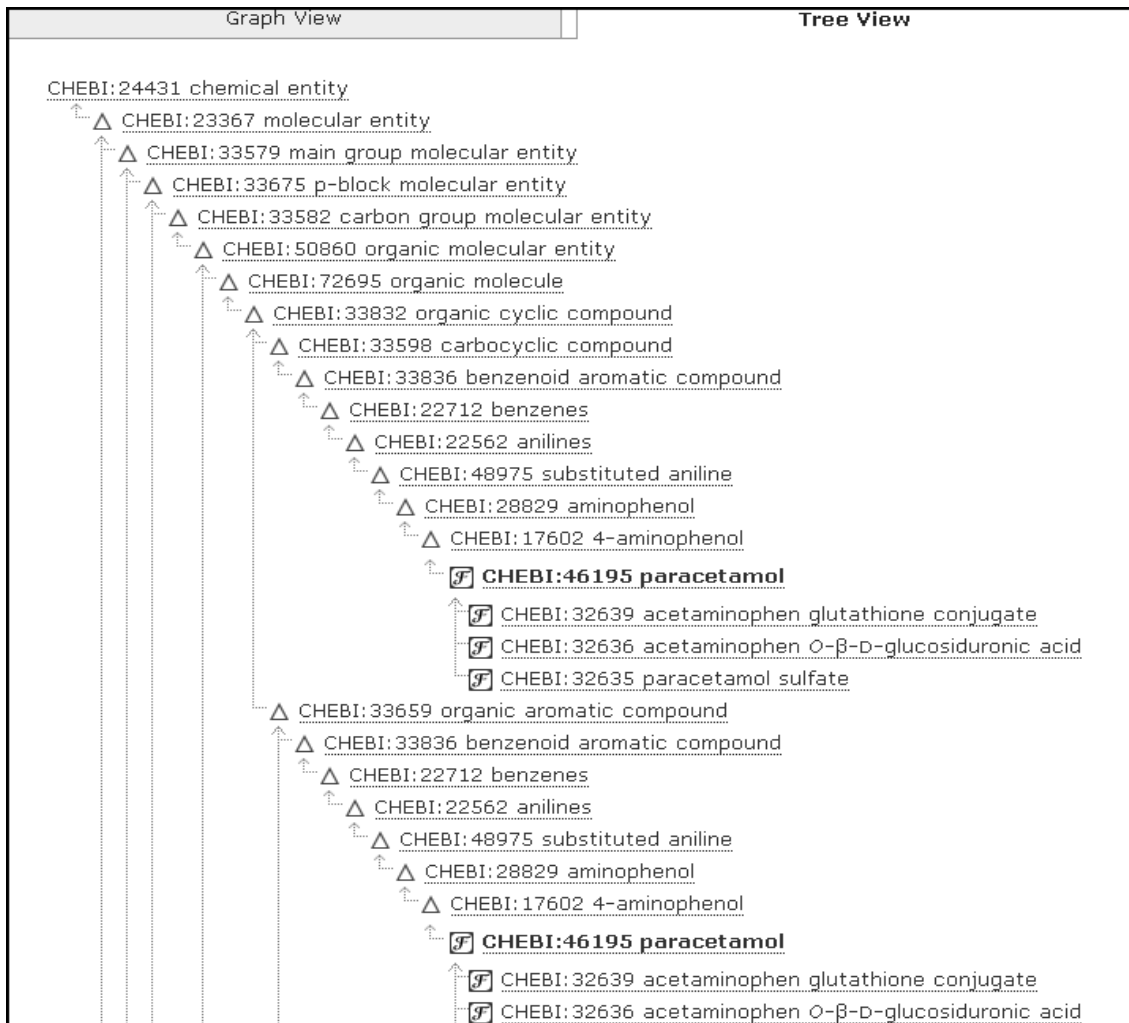


Рис. 3. Фрагмент «дерева», включающего сущность **CHEBI:46195** в общей таксономии

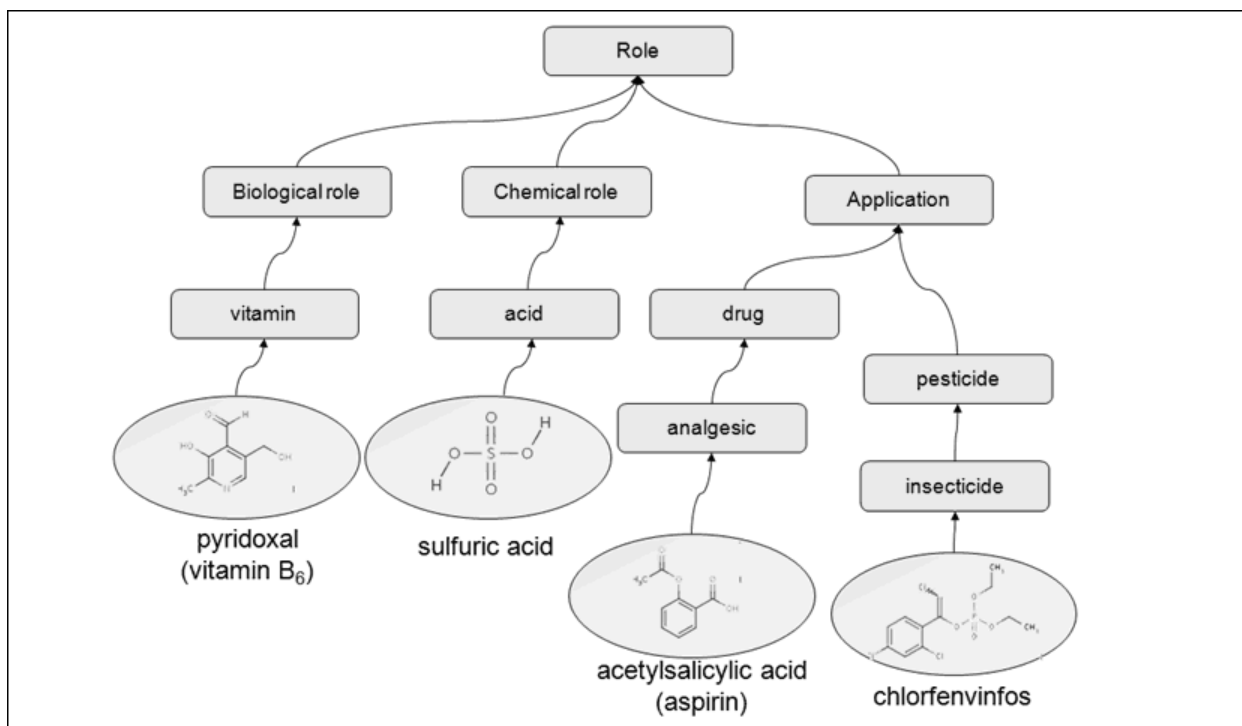


Рис. 4. Схематичное представление суб-онтологии **Role**

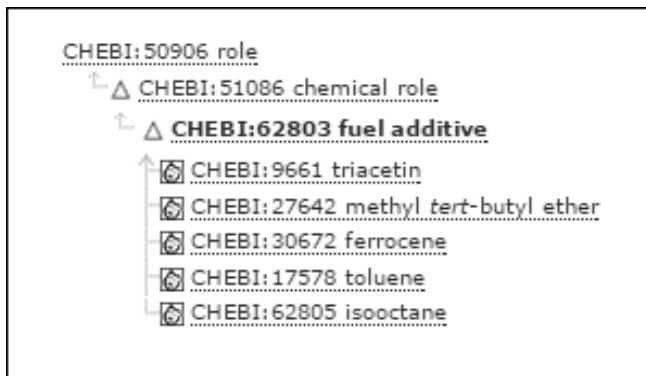


Рис. 5. Сущности, относящиеся к суб-классу **fuel additive** из суб-онтологии **chemical role**

Таким образом, наличие онтологии открывает для пользователя или внешнего ресурса возможности, исходно отсутствующие в БД. В частности, по идентификатору **ChEBI** можно сделать ссылки на любые сущности БД, определить их логические и ролевые связи, встроенные в многоуровневые таксономии. В результате можно возложить на программного агента реализацию сложных запросов с поиском веществ по структуре, функциям, биологической роли и т.п. В целом это означает, что наряду с данными для конкретного вещества **ChEBI** предоставляет фрагмент «знаний», характеризующих предметную область.

Есть также несколько удачных примеров использования онтологий для интеграции данных в материаловедении, где многообразие типов данных и богатство словарей проявляются наиболее ярко. Среди таких примеров база знаний **PLINIUS**, оперирующая результатами исследований по свойствам керамик, онтологическое описание свойств ползучести конструкционных материалов, стандарт ISO 10303-235: “**Engineering properties for product design and verification**” с охватом информации о промышленных изделиях, наряду с традиционными материалами [6]. Стандарт предусматривает единую информационную модель для определения семантики и синтаксиса представления и единый словарь для определения смысла данных. В сравнении с ранее внедренным языком разметки **MatML** [13, 14], онтология способна обеспечить более высокий уровень стандартизации материаловедческих данных, формализующей определение свойств, методов обработки и использования.

В наиболее общем виде онтология по свойствам материалов предложена в работе [16]. Онтология, покрывающая предметную область, построена в виде семи субонтологий, распределенных по трем группам, **рис. 6**. Четыре базовые онтологии дают определения терминов, названий и словарей, представляющих основные концепции для каждой из областей. Каждая из онтологий основана на таксономии классов, представленных в словаре понятий. Базовые онтологии **Process** и **Environment** на **рис. 6** дают описания соответственно методов производства и измерения и характеристик среды (состав атмосферы, температура, pH и т.п.). Сверх четырех базовых, в общую онтологию включена субонтология **Materials Information**, детализирующая сведения по конкрет-

ному объекту путем агрегирования других классов (вещество, свойство и т.п.). Используя базовые онтологии, эта онтология агрегирует все термины и концепции, характеризующие материал и конкретный образец, методы и условия измерения, критерии качества данных и пр. В частности, при построении онтологии **Unit Dimension** используется синтаксис **MathML** (версия, предназначенная для передачи формул), чтобы ввести операции, необходимые при согласовании различных единиц измерения.

Онтология, получившая название **Matinfo**, была размещена на портале [30] и встроена в общую схему **Semantic Web**. Основная задача, решаемая при посредстве разработанной онтологии [16], – обмен данными между тремя БД Японии: **AIST** (Advanced Industrial Science and Technology), **NIMS** (National Institute of Material Science) и широко используемой структурой **MatDB** [31], содержащей данные по тестированию материалов. Основной объем информации в указанных БД относится к тепловым и механическим свойствам. На **рис. 7** показана часть структуры тепловых данных, представленных в двух БД (**AIST**, **NIMS**) и онтологии. Обмен данными реализован с использованием промежуточного уровня, занятого онтологией **Matinfo**. Стрелки на рисунке показывают взаимное отображение концепций онтологии и сущностей в БД.

Структура данных **NIMS** приспособлена к хранению экспериментальных данных, в соответствующих полях хранятся основные метаданные, **thermalConductivity** и **chemicalFormula**. Напротив, в БД **AIST** метаданные конкретизируются пользователем, в поле с именем **property** вводится текстовая строка, что позволяет дополнять или модифицировать имена свойств при фиксированной схеме. Дополнительную сложность вносит то, что БД **AIST** способна хранить как скалярные, так и тензорные значения теплопроводности, характерные для монокристаллов или материалов с выраженной анизотропией строения. Эта же возможность закреплена и в онтологии, где возможно отразить как скалярную величину, так и произвольную матрицу. Отображение полей каждой из БД производится посредством **XSLT** (*Extensible Stylesheet Language Transformations*) шаблонов, причем первичная подстройка осуществлялась в полуавтоматическом режиме.

В соответствии с классификацией, рассмотренной в предыдущем разделе, использование онтологии **Matinfo** полностью соответствует концепции **OBDA** (доступ к данным, основанный на онтологии), причем здесь использована одна онтология, включающая в себя все понятия в науке о материалах, которые могут встретиться в интегрируемых источниках.

Аналогичные цели решает разработанная система **MatSeek** [17] с главной целью интеграции структурно и тематически разнородных БД. Концептуальную основу этой системы образует онтология **MatOnto**, которая аналогично **Matinfo** описывает базовые концепции и логические соотношения в науке о материалах. В ряде отношений, однако, эта онтология заметно отличается.

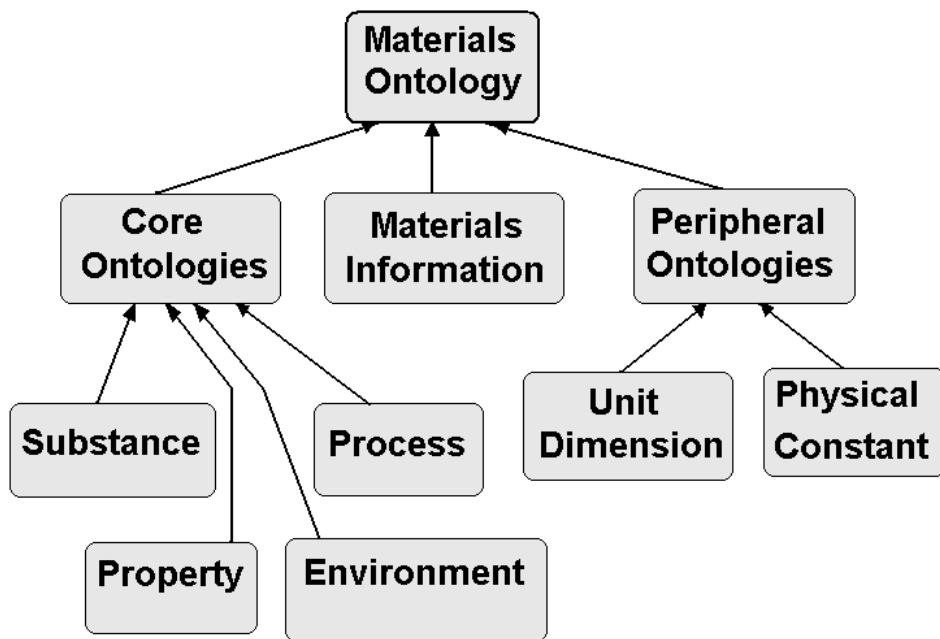


Рис. 6. Суб-онтологии, включенные в общую онтологию **Matinfo** по свойствам материалов [16]

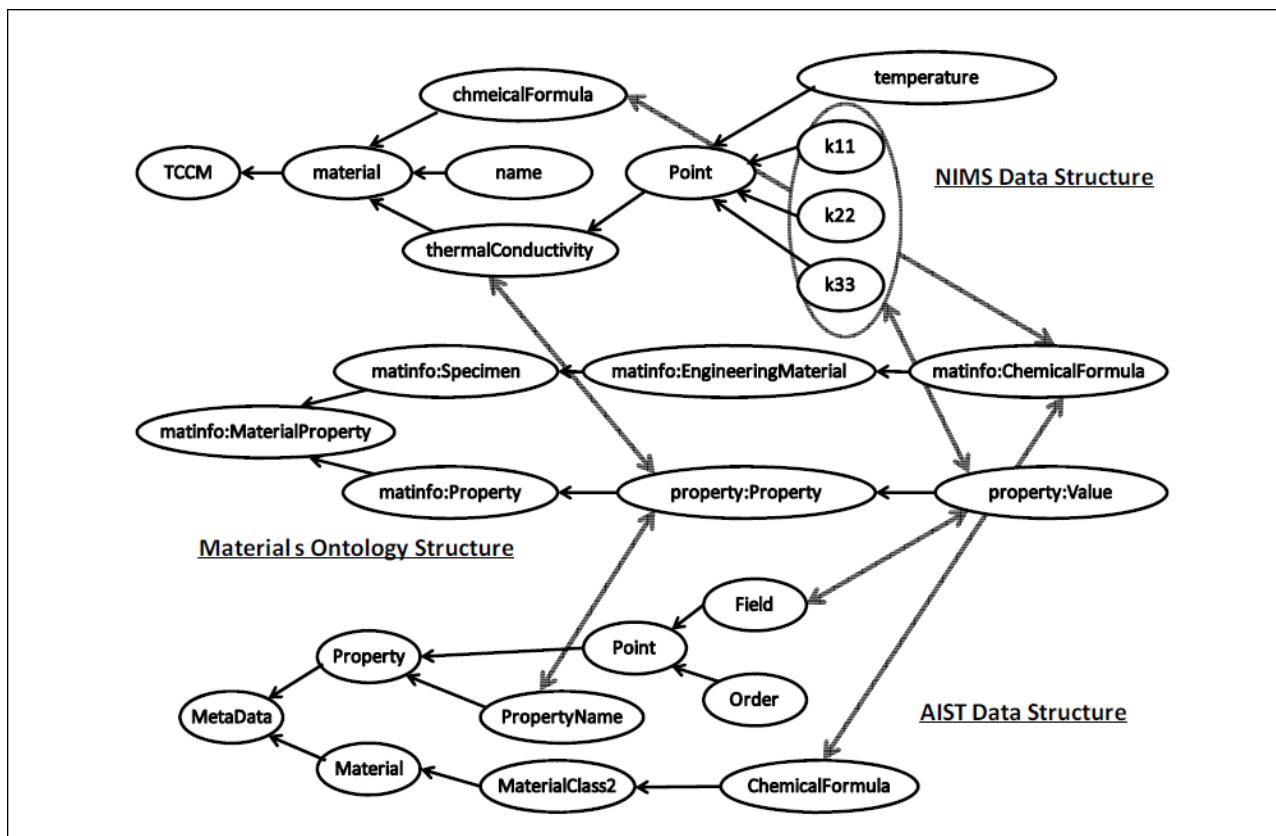


Рис. 7. Отображение схем данных для теплопроводности в БД AIST, NIMS и онтологии [16]

Прежде всего, она нацелена на интеграцию данных из достаточно далеких предметных областей, которые объединены лишь тем, что используются на равных основаниях при исследовании. Например, при разработке топливных элементов в качестве интегрируемых ресурсов выбраны три БД: Inorganic Crystal Structure Database (ICSD), Ionic Radii database, и Phase Equilibria Diagrams (PED). Онтология, установленная в системе **MatSeek** на стороне сервера, записана на языке OWL и выступает посредником между пользователем и группой неоднородных БД.

Второе важное отличие, связанное с проведением логических рассуждений, состоит в активном привлечении внешних онтологий, в том числе онтологий так называемого верхнего уровня, аккумулирующих базовые научные концепции: процесс, гипотезу, эксперимент, активность и пр. Наконец, онтология встроена в достаточно сложную информационную структуру **MatSeek**, функционирующую как приложение в среде Semantic Web. **MatSeek** предоставляет Web-платформу для работы с БД и аналитическими средствами, позволяет формулировать основанные на онтологии динамические запросы к интегрируемым ресурсам, предлагает пользователю поисковый интерфейс в стиле Google. Ядром этой системы является онтология, выполняющая ряд задач: интеграцию разнородных структурно и тематически БД; представление информации о происхождении данных, включая описание источника, методов получения, неопределенности и т.п.; проведение формальных рассуждений с целью выделения нового знания. На рис. 8 показан верхний уровень онтологии с выделением основных классов и пояснением связей между концептами, относящимися к структуре и свойствам материалов. Как видно из рисунка, онтология имеет 5 классов верхнего уровня:

- **Property** – таксономия с 9 субклассами (механические, электрические и т.д.);
- **Family** – таксономия материалов с 6 субклассами (металлы, стекла, керамика, полимеры, гибридные материалы, эластомеры);

- **Process** – два субкласса (**Manufacturing** и **Measurement**);
- **Structure** – два субкласса (**Crystalline** и **Amorphous**);
- **Measurement Data** – таксономия результатов измерений, интегрирует 4 субкласса, охватывающих данные из разных источников (**Material Property-Data, PerformanceData, Modelling and Simulation-Data, CharacterizationData**)

Как уже говорилось, важным элементом, обогащающим выразительность онтологии **MatOnto**, является широкое привлечение концептов из других онтологий и словарей, например, онтологий по единицам измерений, академическому предметному классификатору и др. Сверх того, понятия, актуальные для материаловедения, связаны с более абстрактными лингвистическими понятиями, присущими естественному языку. Для их внедрения использована онтология верхнего уровня **DOLCHE** (*Descriptive Ontology for Linguistic and Cognitive Engineering*), формулирующая такие абстракции как **Endurant** (thing, object), **Perdurant** (process, event), **Abstract, Quality**. Другая из привлеченных онтологий **EXPO** ориентирована на формализацию научного контента в любой из естественнонаучных дисциплин, ориентированных на эксперимент как основную методологию. Ключевые понятия **EXPO** определяют цели, объект эксперимента, выдвинутые гипотезы, результаты и т.п. Привлечение онтологий верхнего уровня обеспечивает расширяемость **MatOnto** с распространением на другие сегменты материаловедения.

Таким образом, потенциал онтологии **MatOnto** раскрывается за счет ее способности к интеграции неоднородных ресурсов, в том числе тематически удаленных, как, например, БД по ионным радиусам и фазовым диаграммам. По логике построения и решаемым задачам онтология, как и вся система **MatSeek**, соответствует концепции **OBDA**, когда онтология выступает как посредник между пользователем и набором разнородных БД.

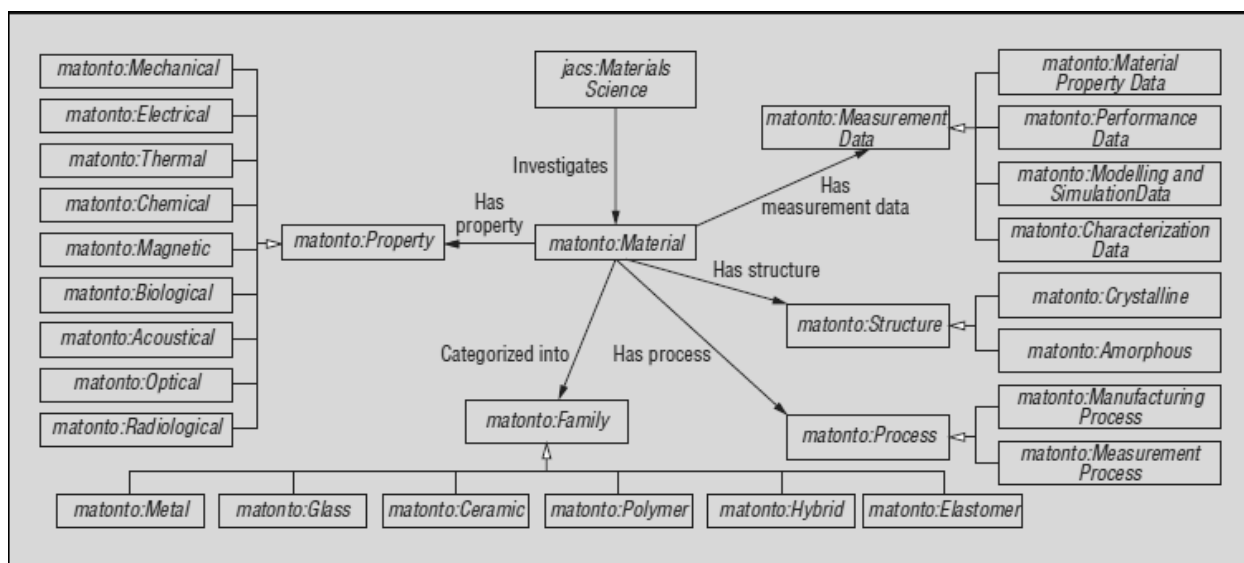


Рис. 8. Схематическое представление верхнего уровня онтологии **MatOnto** [17]

Заслуживает внимания еще одна попытка использования онтологий в сфере материаловедения для доступа к опытным данным [18]. В отличие от рассмотренных проектов в духе **OBDA**, здесь предложена иная концепция, использующая трансформацию языка **MatML** в соответствующую онтологию. Эта трансформация преодолевает принципиальные ограничения **MatML** при обмене данными. Прежде всего, это древовидная структура документа, противоречащая практике запросов, использующих базовые термины предметной области. Кроме того, как отмечалось во Введении, **XML** схемы не описывают семантику, что и послужило основанием для отказа от **MatML** как средства интеграции данных в пользу онтологического моделирования. В то же время при переходе к онтологии в работе [18] предложено сохранить и интегрировать содержание **MatML**-документов – в какой-то степени по аналогии с реляционными БД. Этот подход предполагает набор правил для выделения из данной **MatML**-схемы соответствующей онтологии, включающей классы и соотношения, а также специальный алгоритм, преобразующий данные исходного документа в экземпляры онтологии. Построенной путем такой конверсии онтологии авторы [18] присвоили название **MatOWL**. При этом, поскольку **MatOWL** есть все-таки **OWL**-версия **MatML**-схемы, эта конверсия ограничена в части семантики. Поэтому предусмотрено расширение онтологии за счет дополнительных концепций, соотношений и аксиом – либо путем их прямого внедрения, либо путем отображения **MatOWL** на какую-либо из существующих онтологий в области науки о материалах.

Для конверсии самих данных, содержащихся в **MatML**-документах, в экземпляры онтологии предложена довольно сложная процедура, использующая промежуточную конструкцию – объектную модель **MatOO**. Модель имеет древовидную структуру как исходный **MatML**-документ, а также представлена в виде классов и соотношений, как должно быть в онтологии. **Рис. 9** схематично представляет трансформацию данных в экземпляры онтологии в **RDF**-формате.

Таким образом, исходный материал в виде **MatML**-документов удается конвертировать в типовой формат, предусмотренный **Semantic Web** для целей распространения в сети. Далее при большом объеме данных становится актуальной апелляция к технологиям **OBDB**, чья задача разместить экземпляры онтологии в БД, чтобы обеспечить достаточную эффективность при управлении, т.е. загрузке, поиске и реализации сложных запросов.

ПРОЕКТИРОВАНИЕ ТЕПЛОФИЗИЧЕСКОЙ БД НА ОСНОВЕ ОНТОЛОГИИ

Общая концепция **DBVO** была опробована ранее при создании БД по теплофизическим свойствам веществ [32]. Роль онтологии здесь особенно существенна, поскольку, наряду с унификацией семантики, необходимо поддерживать эволюцию схемы данных и аксиомы, отражающие логические и математические ограничения, присущие данной области. Ключевое понятие – набор данных, включающий для одного вещества несколько констант и температурных функций, а также сведения о фазовом состоянии вещества, единицах измерений, неопределенности и источнике данных. Основные списки – веществ, свойств, фазовых состояний, единиц измерений и т.д. считаются открытыми, что позволяет в рамках онтологии поддерживать эволюцию схемы данных.

Концептуализация предметной области привела к выбору 12 базовых понятий, послуживших основой для построения соответствующих классов (**табл. 2**). Среди них группа из 4 базовых классов (вещества, состояния, свойства, численные данные), 6 вспомогательных классов и 2 класса, определяющих вычисляемые функции и аргументы. При этом класс **Functions** порождает 2 subclasses, определяющих значения свойств и выполнение математических ограничений, устанавливаемых требованиями предметной области, например, равенство энергии Гиббса для сосуществующих фаз (жидкость–газ или жидкость–твердое тело).

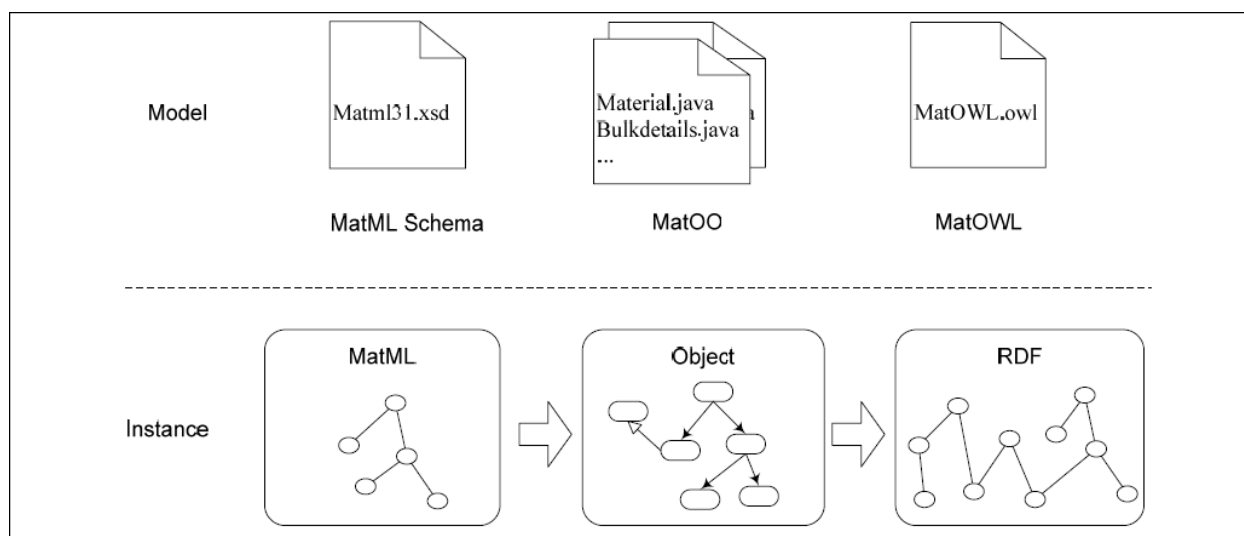


Рис. 9. Схематичное представление конверсии **MatML**-документа в экземпляры онтологии [18]

Классы онтологии теплофизических свойств веществ [32]

Core classes	
Substances	Определяет вещества, для которых приводятся данные
States	Определяет агрегатные состояния вещества
Properties	Определяет свойства вещества
NumericalData	Определяет набор численных данных для функциональной зависимости свойств вещества в определенных условиях от аргументов
Subsidiary classes	
ConstantsOfSubstance	Определяет набор численных значений констант в определенных условиях среды.
EnvironmentalConditions	Указывает агрегатное состояние вещества и ссылки на набор классов
Uncertainties	Определяет тип погрешности физической величины
Dimensions	Определяет размерности физических величин
Data	Определяет перечень данных из справочников физических величин
DataSource	Определяет источники данных для классов NumericalData , Data
Functions and arguments	
DomainOfFunctionDefinition	Определяет перечень аргументов и ограничений физических свойств для функций
Functions	Определяет функции для вычисления и проверки корректности значений физических величин
ComputingFunc (субкласс Functions)	Определяет перечень функций для вычисления значений свойств
ControlFunc (субкласс Functions)	Определяет перечень проверочных функций, которые контролируют, выполняется ли заданное соотношение при допустимой погрешности

При разработке онтологии активно привлекались внешние источники (существующие онтологии и словари) для унификации семантики. В частности, для именования веществ использован словарь **ChemSpider**, который обеспечивает присвоение веществу уникального идентификатора, например, для водорода **CSID:762** и соответствующего URI www.chemspider.com/Chemical-Structure.762.html. Отдельные термины, связанные с фазами, свойствами, размерностями приняты из онтологий **ChemAxiom** и **QUDT** (*Quantities, Units, Dimensions and Data Types*) [32].

Физические принципы, определяющие свойства веществ, накладывают целую совокупность ограничений на использование понятий. Логические ограничения записаны с использованием конструкций языка OWL. Среди логических ограничений: разбиение класса свойств на два непересекающихся класса (свойства-функции и свойства-константы); обязательность определения аргумента для свойства-функции; согласованность ссылок на состояния вещества с видом свойств-функций (например, запрет на свойство **viscosity** в состоянии **solid**). Математические ограничения задаются отдельно для каждого экземпляра-свойства. Они относятся к свойствам-функциям: требования к области определения, области существования, характеру монотонности и, возможно, другим характеристикам (например, связи двух и более функций). Математические ограничения касаются не классов, а экземпляров свойств, поскольку области определения и существования функций определяются для каждого свойства отдельно и список свойств допускает расширение. На основании построенной онтологии сгенерирована реляционная БД в СУБД PostgreSQL. В соответствии с 12 классами онтологии

создано 12 java-классов, которые отображаются на таблицы реляционной БД данных, рис. 10.

На том же рис. 3 (справа) приведен список ассоциативных свойств (соотношений), определяющих каждую сущность. Первые два (**is_a**, **has_part**) определяют простейшие виды логической связи, восемь остальных – фиксируют химический контекст в отношениях двух веществ, например, соотношение **has_functional_parent** (как видно из рис. 2) указывает, чем является *paracetamol* по отношению к другому веществу *paracetamol sulfate*. Особое значение в списке имеет последнее соотношение **has_role**, позволяя раскрыть множество аспектов в свойствах и приложении веществ. Оно связывает базовую субонтологию **Molecular structure**, в рамках которой построена таксономия веществ, с другой субонтологией **Role**, которая включает три класса верхнего уровня (**biological role**, **chemical role**, **application**), рис. 4. При этом **chemical role** классифицирует сущности по их «химически значимой» роли (кислота, основание, лиганд), **biological role** определяет их «биологически значимую» роль, а **application** дает классификацию по целевому использованию, например **drug**, **pesticide** и др. (рис. 4). Все сущности, относящиеся к этой субонтологии, так же как и вещества, имеют уникальный ID, что позволяет связывать внешний ресурс с любой из сущностей в онтологии **role**. Пользователь может совершать навигацию по дереву, соответствующему ролевой онтологии, точно так же, как и для веществ, включенных в БД. Например, выделив дочерний субкласс **fuel additive** в классе **chemical role**, можно найти все вещества, отвечающие этому понятию (рис. 5) и выявить их характеристики и связи.

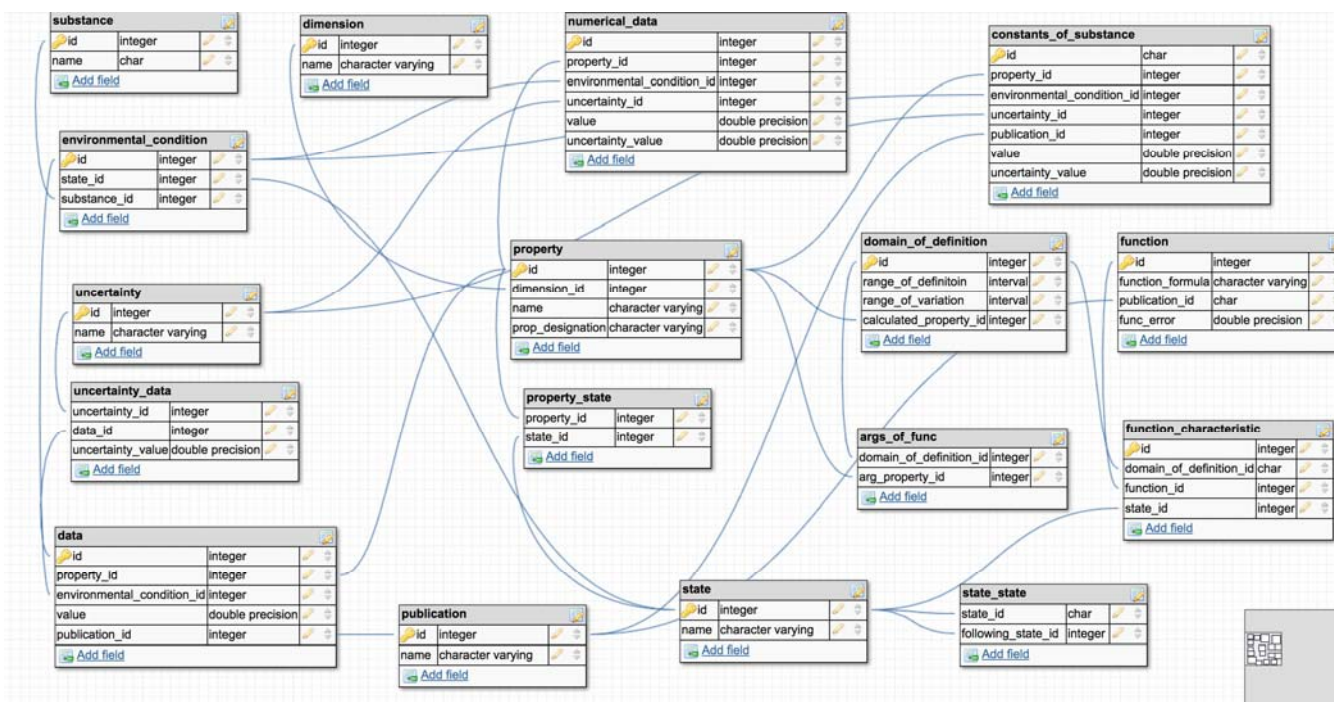


Рис. 10. Таблицы теплофизической БД [32]

Для системы загрузки данных разработан комплекс, позволяющий анализировать и загружать данные с их проверкой на соответствие онтологии и выполнение ограничений. Задача системы не ограничивается хранением введенных данных, предполагая возможность ряда вычислительных операций, в том числе: для контроля корректности численных данных, основанного на физических принципах; расчета физического свойства в произвольной точке, например, при произвольно заданной пользователем температуры; для случая добавления пользователем нового свойства в виде формулы или программного кода.

Разработанная система заметно облегчила проектирование реляционной БД за счет связи с онтологией. Помимо унификации семантики, онтология обеспечила выполнение логических и математических связей между понятиями и сохранение за пользователем права наращивать списки веществ, свойств, единиц измерения и прочих элементов набора данных при эволюции концептуальной схемы.

ЗАКЛЮЧЕНИЕ

Более десятилетия назад авторы опубликовали статью [2], в которой была обоснована необходимость активного внедрения новых технологий работы с данными для поддержки и эффективного развития общенаучной тематики «Свойства веществ и материалов». Ее специфика, включая многообразие форм представления, требования к совместному хранению «сырых» и обработанных данных, перманентное обновление логической структуры и физических моделей с появлением новых объектов – все это поставило экспертов перед необходимостью отхода от традиционных форм хранения в автономных БД. За

прошедшие годы доминирование работы с данными стало тенденцией во многих дисциплинах, называемых теперь *e-science* [4, 5], что, естественно, не обошло и отмеченную нами тематику.

В настоящей работе изучен один из новых подходов к работе с данными по свойствам, использующий различные технологии связывания БД с онтологиями [24–27]. Сами по себе БД, будучи созданы в разных коллективах, с неизбежностью порождают разноречивой терминологии, логических схемах и форматах данных. Онтология оказалась идеальным средством унификации семантики и подчинения множества концептуальных схем единой структуре, к тому же согласованной в научном сообществе. В результате открывается путь к широкой интеграции структурно или даже тематически разнородных ресурсов с возможностью обмена данными и использования в совместной работе. Нами проанализирован опыт связывания онтологий с БД для ряда областей знания: химии, материаловедения, теплофизики.

В компьютерном сообществе предложено несколько возможных подходов к такому связыванию. Один из них, названный **DBVO**, предполагает проектирование БД путем отображения онтологии на концептуальную схему БД, т.е. ее согласование с внешним стандартизированным ресурсом. Другая концепция, **OBDB**, имеет прямое отношение к методологии Semantic Web. В этом случае БД используется как хранилище онтологий совместно с экземплярами данных, которые могут быть включены в пространство связанных данных. Наконец, предложена концепция **OBDA** для интеграции разнородных БД за счет использования онтологии как семантического посредника при формулировании запроса между пользователем и БД.

Рассмотренные в работе примеры показывают, что привлечение онтологии, формализующей знания в предметной области, позволяют не только унифицировать и сделать общедоступной семантику предметной области, но и обеспечить распространение знаний, закодированных в логических соотношениях между концептами. Применительно к тематике «Свойства веществ и материалов» формализация знания в онтологии позволяет решить старую проблему [1] – постоянно варьировать форму идентификации новых объектов и номенклатуру свойств и характеристик, что особенно актуально для таких областей, как биоинформатика или нанотехнологии.

СПИСОК ЛИТЕРАТУРЫ

1. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Фокин Л.Р. Логическая структура физико-химических данных. Проблемы стандартизации и обмена численными данными // Журнал физической химии. – 2008. – Т. 82, № 1. – С. 20–31.
2. Зицерман В.Ю., Кобзев Г.А., Фокин Л.Р. Возможности и перспективы информационных технологий в подготовке и распространении справочных данных: свойства веществ и материалов // Научно-техническая информация. Сер. 1. – 2004. – № 2. – С. 7–14.
3. Киселева Н.Н., Дударев В.А., Земсков В.С. Компьютерные информационные ресурсы неорганической химии и материаловедения // Успехи химии. – 2010. – Т. 79, № 2. – С. 162–188.
4. Hey T. The Fourth Paradigm: Data Intensive Scientific Discovery / eds. T. Hey, S. Tansley, K. Tolle. – Redmond, WA: Microsoft Research, 2009. – 286 p.
5. Castelli D., Manghi P., Thanos C. A vision towards Scientific Communication Infrastructures. On bridging the realms of Research Digital Libraries and Scientific Data Centers // Int. J. Digit. Libraries. – 2013. – Vol. 13. – P. 155–169.
6. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Сон Э.Е., Сотников А.Н. Интеграция баз данных по свойствам вещества. Подходы и технологии // Научно-техническая информация. Сер. 2. – 2012. – № 8. – С. 1–8.
7. Когаловский М.Р. Энциклопедия технологий баз данных. – М.: Финансы и статистика, 2002. – 798 с.
8. O'Reily. XML from the inside out. Resource Guide. Science. – URL: www.xml.com/pub/rg/Science.
9. Murray-Rust P., Rzepa H.S. Markup Languages – How to Structure Chemistry-Related Documents // Chemistry International. – 2002. – Vol. 24, № 4. – P. 9–13.
10. Елизаров А.М., Липачев Е.К., Малахальцев М.А. Веб-технологии для математика: основы MathML. – М.: Физматлит, 2010. – 192 с.
11. Frenkel M., Chirico R.D., Diky V. et al. XML-based IUPAC standard for experimental, predicted, and critically evaluated thermodynamic property storage and capture (ThermoML). IUPAC Recommendations 2006 // Pure Appl. Chem. – 2006. – Vol. 78, № 3. – P. 541–612.
12. Ojala T., Over H.-H. Approaches in using MatML as a common language for materials data exchange // Data Science Journal. – 2008. – Vol. 7. – P. 179–195.
13. Kaufman J.G., Begley E.F. MatML. A Data Interchange Markup Language // Advanced Materials & Processes/November. – 2003. – P. 35–36.
14. Park Jun-Hyoung, Song Mi-Young, Yoon Jung-Sik. Design of System for Atomic, Molecular and Plasma-Material Interaction Data Exchange with XSAMS // International Journal of Software Engineering and Its Applications. – 2014. – Vol. 8, № 6. – P. 149–156.
15. VAMDC-XSAMS reference guide [v 1.0] [r 12.07]. – URL: <http://vamdc.eu/documents/standards/dataModel/vamdcxsams/index.html>.
16. Ashino T. Materials ontology: an infrastructure for exchanging materials information and knowledge // Data Science Journal. – 2010. – Vol. 9. – P. 54–61.
17. Cheung Kwok, Hunter Jane, Drennan John. MatSeek: An Ontology-Based Federated Search Interface for Materials Scientists // IEEE INTELLIGENT SYSTEMS. – 2009. – Vol. 24, № 01. – P. 47–56.
18. Zhang Xiaoming, Hu Changjun, Li Huayu. Semantic Query on materials data based on mapping MatML to an OWL ontology // Data Science Journal. – 2009. – Vol. 8. – P. 1–17.
19. Degtyarenko K., de Matos P., Ennis M. et al. ChEBI: a database and ontology for chemical entities of biological interest // Nucleic Acids Research. – 2008. – Vol. 36. – Database issue. – D344–D350.
20. Gruber T.R. A translational approach to portable ontology specification // Knowledge Acquisition. – 1993. – Vol. 5, № 2. – P. 199–220.
21. Uschold M. Ontologies and Database Schema: What's the Difference? – URL: www.slideshare.net/UscholdM/.
22. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А. Роль метаданных в создании и использовании информационных ресурсов о свойствах веществ и материалов // Научно-техническая информация. Сер. 1. – 2008. – № 11. – С. 13–19.
23. Frenkel M., Chirico R. D., Diky V. et al. Extension of ThermoML: The IUPAC standard for thermodynamic data communications. (IUPAC Recommendations 2011) // Pure Appl. Chem. – 2011. – Vol. 83, № 10. – P. 1937–1969.
24. Тузовский А.Ф. Интеграция баз данных на основе онтологий. Слайды симпозиума «Онтологическое моделирование 2010» // Сайт рабочей группы симпозиума «Онтологическое моделирование». – URL: ontology.ipi.ac.ru.
25. Когаловский М.Р. Системы доступа к данным, основанные на онтологиях // Программирование. – 2012. – № 4. – С. 55–77.
26. Laallam F.Z., Kherfi M.L., Benslimane S.M. A survey on the complementarity between database and ontologies: principles and research areas // Int. J.

- Computer Applications in Technology. – 2014. – Vol. 49, № 2. – P. 166–187.
27. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Серебряков В.А., Теймуразов К.Б. Технология научных публикаций в среде «открытых связанных данных» // Научно-техническая информация. Сер. 1. – 2013. – №12. – С. 1–11.
28. An Y., Hu X., Song I.-Y. Round-Trip Engineering for Maintaining Conceptual-Relational Mappings // Proc. Of 20th International Conference on Advanced Information Systems Engineering (CAiSE'08), LNCS 5074, P. 296–311. Springer, 2008.
29. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Серебряков В.А., Шиолашвили Л.Н. Интеграция данных по свойствам веществ и материалов на основе онтологического моделирования предметной области // Журнал «Электронные библиотеки». – 2013. – Т. 16, № 6. – URL: www.elbib.ru/index.phtml?page=elbib/rus/journal/2013/part6/EZKSS.
30. Materials Database Information Portal. – URL: <http://musigny.rds.toyo.ac.jp:8080>.
31. Austin T.S.P., Over H.H. MatDB online – a standards-based system for preserving, managing, and exchanging engineering materials test // Data Science Journal. – 2012. – Vol. 11. – P. ASMD11–ASMD16.
32. Серебряков В.А., Теймуразов К.Б., Хайруллин Р.И., Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Трахтенгерц М.С. Практическая реализация системы интеграции теплофизических данных на основе онтологической модели предметной области // Труды Четвертого Всероссийского симпозиума «Инфраструктура научных информационных ресурсов и систем». (С.-Петербург. 6–8 октября 2014 г.) / под ред. Е.В. Кудашева, В.А. Серебрякова. – М: ВЦ РАН. – Т. 1. – С. 87–111.

Материал поступил в редакцию 18.08.15.

Сведения об авторах

ЕРКИМБАЕВ АДильБЕК ОМИРБЕКОВИЧ – кандидат технических наук, зав. лабораторией баз данных объединенного института высоких температур РАН (ОИВТ РАН), Москва
e-mail: adilbek@ihed.ras.ru

ЗИЦЕРМАН ВЛАДИМИР ЮРЬЕВИЧ – кандидат физико-математических наук, ведущий научный сотрудник лаборатории баз данных ОИВТ РАН
e-mail: vz1941@mail.ru

КОБЗЕВ ГЕОРГИЙ АНАТОЛЬЕВИЧ – доктор физико-математических наук, главный научный сотрудник, советник НИЦ электрофизики и тепловых процессов ОИВТ РАН
e-mail: gkbz@mail.ru

КОСИНОВ Андрей Владимирович – инженер-программист ОКБ ОИВТ РАН
kosinov@gmail.ru