

ОНТОЛОГИИ И БАЗЫ ДАННЫХ – ВЗАИМНАЯ ДОПОЛНИТЕЛЬНОСТЬ ПРИ ИСПОЛЬЗОВАНИИ НАУЧНЫХ ДАННЫХ

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

Объединённый институт высоких температур РАН, Москва, Россия

e-mail: adilbek@ihed.ras.ru

Проанализирована роль онтологий в современных информационных системах, особенно применительно к научным данным. Проведен анализ преимуществ и недостатков онтологий при их сравнении с базами данных. Показано, что использование онтологий предметной области обеспечивает стандартизацию терминов и понятий, исключая их неоднозначность и ошибочную интерпретацию, унификацию семантики, согласованную в экспертном сообществе, распространение знаний с возможностью логического вывода. В качестве примеров рассмотрены онтологии в области наук о материалах и авторский проект теплофизической базы данных, основанной на онтологии.

Ключевые слова: базы данных, онтология, интеграция данных, Семантический Веб.

Ontology value for modern information technologies is studied especially with scientific data categorization and sharing. Advantages and disadvantages of ontologies in contrast with database conceptual scheme are defined. Subject field ontology may be successfully used for standardization of terms for excluding ambiguity and wrong interpretation, unification of semantics agreed among expert community, knowledge sharing with logical reasoning. Some domain ontologies for material science and author's project of thermophysical database developed with ontology are considered as examples.

Keywords: databases, ontology, data integration, Semantic Web.

Введение

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 данное Грубером [1]: «Онтология есть спецификация концептуализации», то есть 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 4 типа составляющих элементов: понятия/концепты, соотношения, экземпляры и аксиомы. Концепты – это классы сущностей или объектов в определенной области знания, например, ве-

щества, свойства, состояния и т.п. Отдельные классы связаны соотношениями либо типа класс-субкласс (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. 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. 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.

Стратегии объединения онтологий и БД

В последние годы возник ряд направлений, суть которых – попытка объединить в единой системе обе технологии (онтологий и БД), по возможности нивелируя ограничения и недостатки каждой. Принципиальные их различия детально рассмотрены в ряде публикаций [2-4], по-видимому, наиболее выпукло это сделано в лекции [5], представленной в сети slideshare.net.

Согласно [5], главное отличие состоит в том, что онтология записывает на формальном языке структуру предметной области (то есть вводит понятия и соотношения), в то время как концептуальная схема БД описывает на формальном языке определенную БД, не претендуя на распространение знаний. Поэтому БД фокусируется именно на данных (числовых, текстовых, и т.д.), в то время как

онтология ориентирована на толкование их смысла (семантики) при распространении согласованного знания. Поэтому главным ограничением БД считается потеря семантики, то есть недоступность смысла всех сущностей для лиц, незнакомых со схемой БД.

С другой стороны, имеется четкое соответствие между элементами БД и онтологии. Например, сущностям БД в онтологии соответствуют определенные классы, а атрибутам – свойства классов, ограничениям – аксиомы. Однако главный элемент онтологии – таксономия классов – не предусмотрен в БД. Наиболее существенные недостатки БД заключены в потере семантики при переходе от концептуальной к физической схеме и в невозможности поддерживать эволюционирующую схему при расширении предметной области за счет появления новых понятий и соотношений. К примеру, БД, ориентированную на хранение информации по свойствам веществ, определяемых химической формулой, практически невозможно перестроить на работу с материалами (в частности, с наноматериалами), идентификация которых требует детализации сведений о технологии производства. Общая сводка характерных признаков, фиксирующих сходства и различия онтологий и схем БД, приведена в таблице 1.

Хороший пример того, что значит семантика при распространении научных данных, дает единое соглашение об имено-

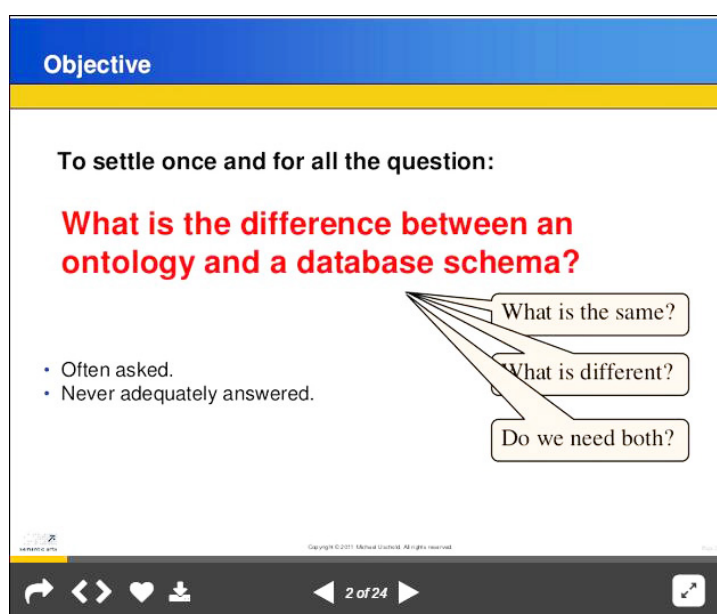


Рис. 1. Первый слайд из презентации [5]

Таблица 1

БД и онтологии – характерные признаки и отличия

DB schema	Ontology
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	Свойства
Ограничения	Аксиомы
Нет таксономии	Таксономия – ключевой элемент онтологии
Ограничения целостности	Ограничения смысла, согласованности и целостности
Данные обязательны	Экземпляры данных не обязательны (instances optional)
Словарь данных как дополнительный элемент	Комментарии – часть онтологии
Семантика только в концептуальной схеме и словаре данных, спроектирована для человека, не эволюционирует с изменением БД и приложений	Явное отражение семантики
Схему трудно изменять и поддерживать	Потенциально легче изменять и поддерживать

вании веществ, принятое согласно онтологии **ChEBI** (*Chemical Entities of Biological Interest Ontology*) [6], которая присваивает каждому из низкомолекулярных веществ уникальный идентификатор; 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. В итоге, формальный характер онтологии позволяет реализовать машинные выводы и рассуждения, что совершенно не предусмотрено в концептуальной схеме БД. Богатый потенциал онтологии состоит в поддержке семантики, и связанные с этим возможности интеграции разнородных ресурсов (БД, изданий и проч.) и возможности проведения логических рассуждений и выводов на основе

представленных данных.

Применительно к естественнонаучной тематике 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 [7-9], 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. Например, эволюция схемы данных может быть связана с постепенным включением наноструктур в химические БД, и расширением правил идентификации по мере включения в оборот новых объектов (см., например, [7,10]).

В то же время БД обеспечивают наиболее эффективный способ хранения данных, поскольку их отличает высокая производительность при поиске и реализации сложных запросов, недостижимая для других архитектур

тур. Поэтому в компьютерном сообществе возникло направление, цель которого – создание инструментов, использующих преимущества обеих концепций за счет соединения семантики с высокой производительностью при работе с данными. В самом общем виде в рамках этого направления реализуется так называемое *database-to-ontology mapping*.

Авторы одного из обзоров [4], исходя из того, какая из двух концепций (онтология или схема БД) может рассматриваться как ведущая, выделили две стратегии. В рамках первой БД рассматривается как ведущий элемент, усиленный связью с онтологией для реализации одной из нескольких целей: организация запросов с использованием согласованной семантики; эффективное проектирование БД; интеграция разнородных БД. Вторая стратегия, напротив, рассматривает БД как возможный инструмент для хранения или даже разработки онтологии.

Первой стратегии соответствует проектирование (или поддержка) БД, основанной на онтологии (**DBBO**, *database based on ontology*), когда все сущности, внесенные в БД (названия веществ, свойств, единиц измерения и т.п.), определены ссылкой на соответствующую онтологию. Пример, приведенный выше, со ссылкой на название вещества в онтологии **ChEBI** хорошо иллюстрирует эту процедуру.

В соответствии со второй стратегией предполагается проектирование структуры, получившей название онтологии, основанной на БД (**OBDB**, *Ontology Based on DB*). Задача такой структуры – разместить экземпляры онтологии в БД, чтобы обеспечить достаточную производительность при хранении больших объемов данных (то есть экземпляров онтологии). В основном эта стратегия нацелена на решение задач Semantic Web, где требуется организовать хранение и управление архивами документов, записанных на языке **RDF** (*resource definition framework*), что позволяет их связывать в так называемом пространстве связанных данных [11].

Для повышения эффективности инфраструктуры научных данных наиболее актуальна первая стратегия (**DBBO**), ориентированная на повышение функциональности БД. Онтология при этом лишь обеспечивает

семантику, но не содержит экземпляров – их роль выполняют записи БД.

Усиление функциональности за счет онтологии возможно также и на этапе проектирования БД. Подход, при котором в основу проекта БД положена онтология, заметно ускоряет процесс разработки и снимает ряд возникающих ограничений. В частности, концептуальные схемы, предлагаемые разными экспертами, унифицируются, поскольку базируются на единой информационной модели. При этом на этапе проектирования снимается проблема семантики, поскольку каждый элемент данных получается отображением онтологических понятий.

Наряду с проектированием, использование онтологий открывает возможности в интеграции БД, ранее созданных без ссылок на онтологии [12,13]. В ходе интеграции производится объединение схем и самих данных. Пользователь формулирует запросы в терминах онтологии, которые в свою очередь конвертируются в запросы к БД. Тем самым онтология играет роль эффективного посредника между пользователем и данными. Подобный подход получил название *доступ к данным, основанный на онтологиях* или **OBDA** (*ontology based data access*). В **OBDA** системе источник может рассматриваться как единая БД. Хотя в целом такой подход к интеграции БД является относительно новым, уже сейчас, по мнению авторов [3,4], можно указать несколько технологий, опробованных в разных проектах. В нашей работе [14] было показано, насколько эффективной оказывается онтология при проектировании БД по теплофизическим свойствам и интеграции данных из множества источников.

Примеры использования онтологий при проектировании БД

Есть 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 [15]. Стандарт предусматривает единую информационную модель для определения семантики и синтаксиса представления и единый словарь для определения смысла данных. Ранее практика использования материаловедческих онтологий была изучена в работах [14,16].

В наиболее общем виде онтология по свойствам материалов предложена в работе [17]. Онтология, покрывающая предметную область, построена в виде 7 суб-онтологий, распределенных по 3 группам, рис. 2.

Четыре базовые онтологии дают определения терминов, названий и словарей, причем каждая из онтологий основана на таксономии классов, представленных в словаре понятий. Базовые онтологии **Process** и **Environment** дают описания соответственно методов производства и измерения и характеристик среды (состав атмосферы, температура, pH и т.п.). Сверх четырех базовых, в общую онтологию включена суб-онтология **Materials Information**, детализирующая сведения по материалу. Используя базовые онтологии, эта онтология агрегирует все термины и концепции, характеризующие материал и конкретный образец, методы и условия измерения, критерии качества данных и проч. В частности, при построении онтологии **Unit Dimension** используется синтаксис языка MathML (версия, предназначенная для пере-

дачи формул), чтобы ввести операции, необходимые при согласовании различных единиц измерения.

Онтология [17], получившая название **Matinfo**, размещена на портале [18] и встроена в общую схему Semantic Web. Основная задача, решаемая при помощи разработанной онтологии, – обмен данными между тремя БД Японии: AIST (Advanced Industrial Science and Technology), NIMS (National Institute of Material Science) и широко используемой структурой **MatDB** [19], содержащей данные по тестированию материалов. Основным объемом информации в указанных БД относятся к тепловым и механическим свойствам. На рис. 3 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. В соответствии с классификацией, рассмотренной в предыдущем разделе, использование онтологии **Matinfo** полностью соответствует концепции **OBDA**, 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** [20] с главной целью интеграции структурно и тематически разнородных БД. Основу системы образует онто-

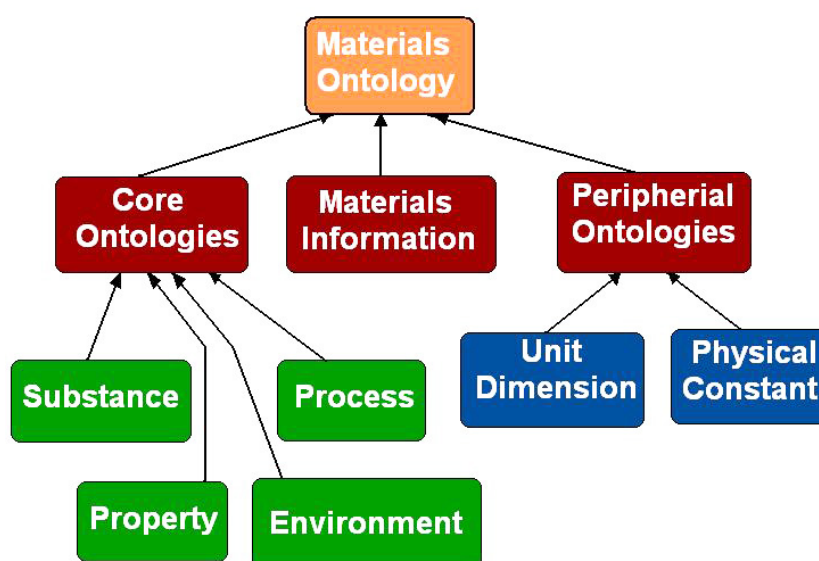


Рис. 2. Суб-онтологии, включенные в общую онтологию **Matinfo** по свойствам материалов [17]

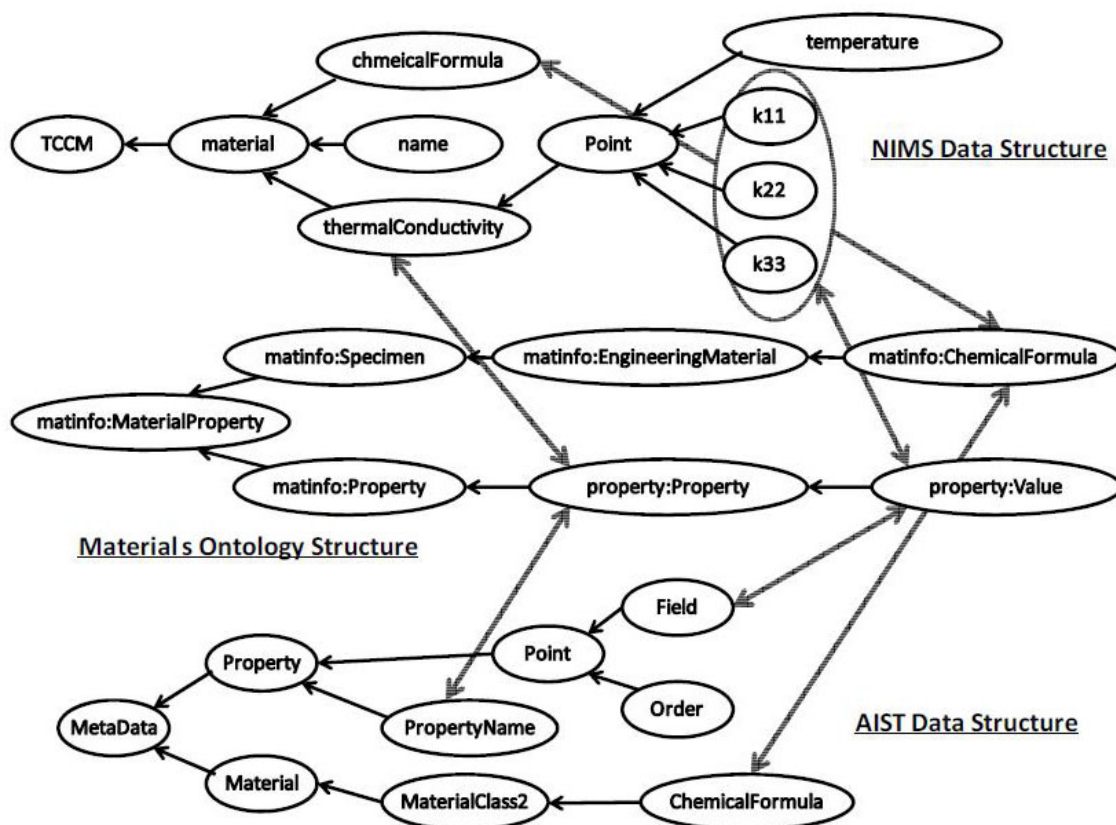


Рис. 3. Отображение схем данных для теплопроводности в БД AIST, NIMS и онтологии [17]

логия **MatOnto**, которая описывает базовые концепции и логические соотношения в науке о материалах. В ряде отношений, однако, эта онтология заметно отличается от **Matinfo**.

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). Последняя разработана Национальным институтом стандартов и технологий США (NIST). 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, гипотеза, эксперимент, активность и проч. 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. Онтология является базовым компонентом **MatSeek** и выполняет ряд задач: интеграция разнородных структурно и тематически БД; хранение сведений о генезисе данных (источник, метод получения данных, неопределенность); возможность формальных рассуждений с целью выделения нового знания. На рис. 4 показан верхний уровень онтологии с выделением основных классов и пояснением связей между концептами, относящимися к структуре и свойствам материалов.

Важным элементом, обогащающим выразительность онтологии **MatOnto**, является широкое привлечение концептов из других онтологий и словарей, например, онтологии по единицам измерений, академическому предметному классификатору и др.

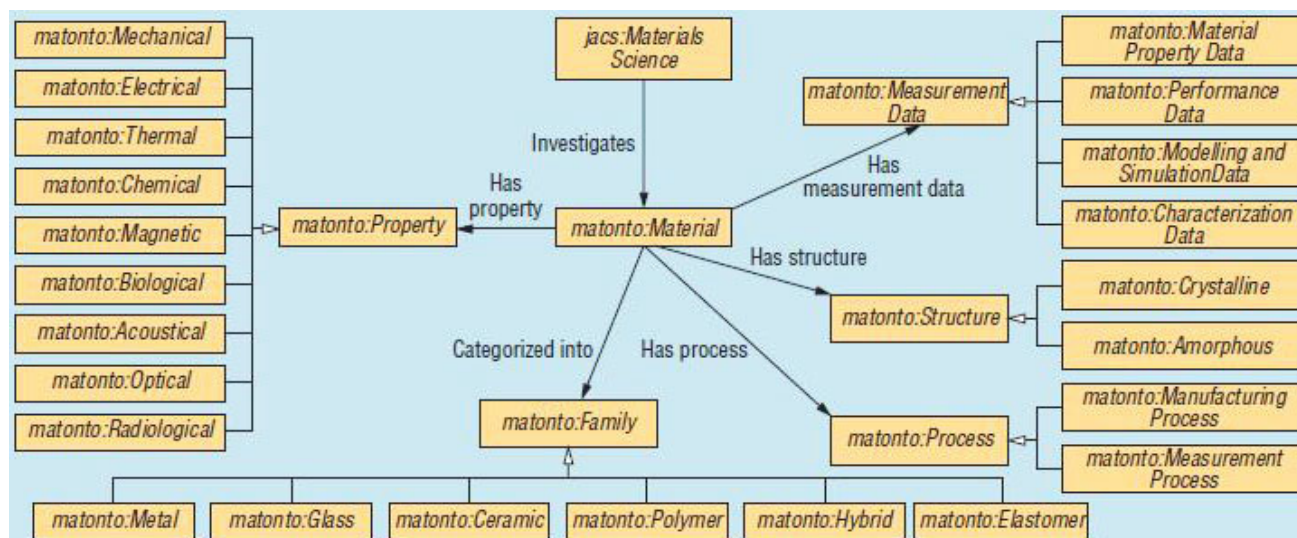


Рис. 4. Схематическое представление верхнего уровня онтологии MatOnto [20]

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.

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 онтология и основанная на ней система **MatSeek** соответствуют концепции **OBDA**, когда онтология выступает как посредник между пользователем и набором разнородных БД.

Концепция **DBVO** была опробована ранее при создании БД по теплофизическим свойствам веществ [21]. Наряду с унификацией семантики, здесь необходимо поддер-

живать 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, таблица 2. Среди них группа 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, определяемых требованиями предметной области, например, равенство энергии Гиббса для сосуществующих фаз (жидкость-газ или жидкость-твердое тело).

При разработке онтологии активно привлекались внешние источники (существующие онтологии и словари) для унификации семантики. В частности, для именования веществ использован словарь **ChemSpider** [22], который обеспечивает присвоение веществу

Таблица 2

Классы онтологии теплофизических свойств веществ [21]

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	Указание на агрегатное состояние вещества и ссылок 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 (суб-класс Functions)	Определяет перечень функций для вычисления значений свойств
ControlFunc (суб-класс Functions)	Определяет перечень проверочных функций, которые определяют, выполняется ли заданное соотношение при допустимой погрешности

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) [22].

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*). Математи-

ческие 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. Соответственно 12 классам онтологии создано 12 java классов, которые отображаются на таблицы реляционной БД данных, рис. 5.

Для системы загрузки данных разработан комплекс, позволяющий анализировать и загружать данные, с их проверкой на соответ-

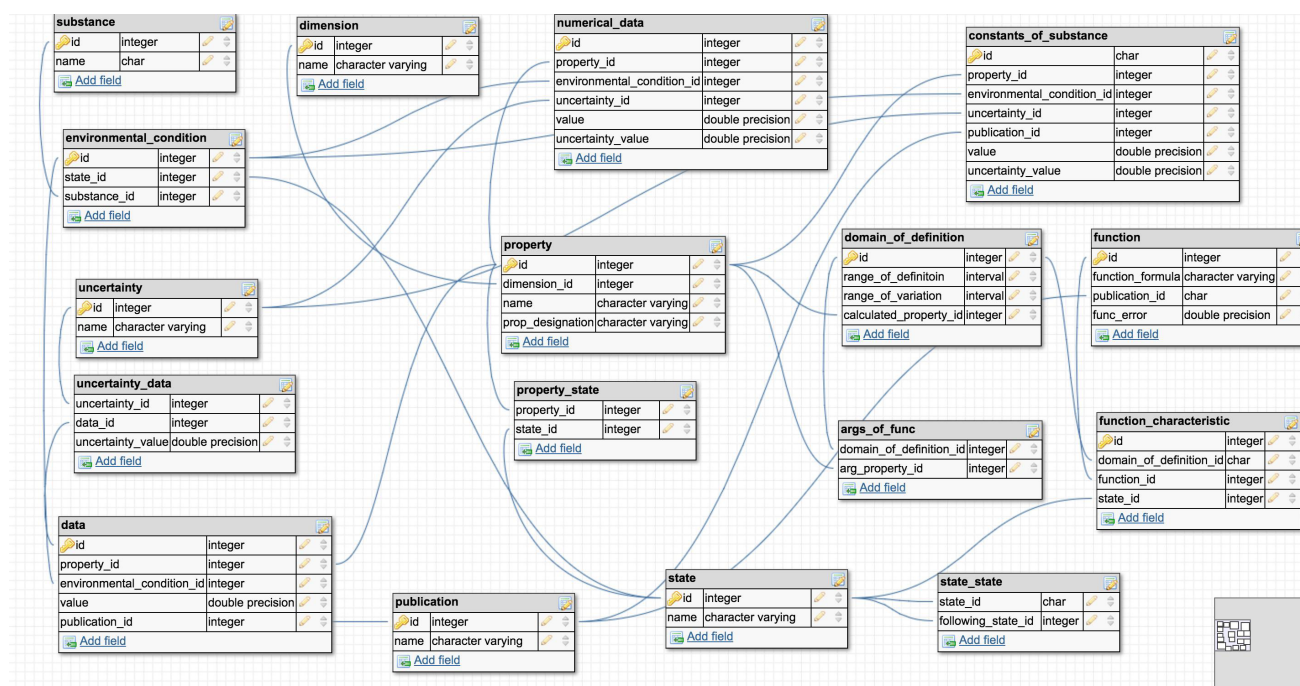


Рис. 5. Таблицы теплофизической БД [21]

ствие 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 температуре; при добавлении 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.

В работах [7-9] авторы обосновали 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 струк-

туры и физических моделей с появлением новых объектов – все это поставило экспертов перед необходимостью отхода от традиционных форм хранения в автономных БД.

Заключение

Здесь рассмотрен новый подход к работе с данными по свойствам, использующий различные технологии связывания БД с онтологиями [2-4]. 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.

Рассмотрено несколько возможных подходов к такому связыванию. Один из них, названный **DBVO**, предполагает проектирование БД путем отображения онтологии на концептуальную схему БД, то есть ее согласование с внешним стандартизованным ресурсом. Другая концепция (**OBDB**) имеет прямое отношение к методологии Semantic Web. В последнем 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 примеры показывают,

что привлечение онтологии позволяет не только унифицировать и сделать общедоступной семантику предметной области, но и обеспечить распространение знаний, закодированных в логических соотношениях между концептами.

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

Список литературы

1. Gruber T.R. A translational approach to portable ontology specification // Knowledge Acquisition. 1993. V. 5. No. 2. P. 199-220.
2. Тузовский А.Ф. Интеграция баз данных на основе онтологий. Слайды симпозиума «Онтологическое моделирование 2010». Сайт рабочей группы симпозиума «Онтологическое моделирование» // ontology.ipi.ac.ru
3. Когаловский М.Р. Системы доступа к данным, основанные на онтологиях // Программирование. 2012. №4. С. 55-77.
4. 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. V. 49. No 2. P. 166–187.
5. Uschold M. Ontologies and Database Schema: What's the Difference? // www.slideshare.net/UscholdM/
6. Degtyarenko K., de Matos P., Ennis M. et al. ChEBI: a database and ontology for chemical entities of biological interest // Nucleic Acids Research. 2008. V. 36. Database issue. D344–D350.
7. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Фокин Л.Р. Логическая структура физико-химических данных. Проблемы стандартизации и обмена численными данными // ЖФХ. 2008. Т. 82. №1. С. 20-31.
8. 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. Серия 1. 2004. №2. С. 7-14.
9. 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. Серия 1. 2008. №11. С. 13-19.
10. Елецкий А.В., Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Трахтенгерц М.С. Численные данные для объектов наномира // Мониторинг. Наука и технологии. 2012. № 2(11). С. 37-49.
11. 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. Серия 1. 2013. №12. С. 1-11.
12. Когаловский М.Р. Энциклопедия технологий баз данных. М.: Финансы и статистика, 2002.- 798 с.
13. 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.
14. 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. Серия 2. 2012. №8. С. 1-8.
15. Swindells N. The representation and exchange of materials and other engineering properties // Data Science Journal. 2009. V. 8, P. 190-200.
16. 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 // www.elbib.ru/index.phtml?page=elbib/rus/journal/2013/part6/EZKSS.
17. Ashino T. Materials ontology: an infrastructure for exchanging materials information and knowledge // Data Science Journal. 2010. V. 9. P. 54-61.
18. Materials Database Information Portal, <http://musigny.rds.toyo.ac.jp:8080>
19. 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. V. 11. P. ASMD11-ASMD16.
20. Kwok Cheung, Jane Hunter, John Drennan. MatSeek: An Ontology-Based Federated Search Interface for Materials Scientists // IEEE INTELLIGENT SYSTEMS. 2009. V. 24. No 01. P. 47-56.
21. 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. Инфраструктура научных информационных ресурсов и систем. Труды Четвертого Всероссийского симпозиума (С.-Петербург. 6-8 октября 2014 г.). Под ред. Е.В. Кудашева, В.А. Серебрякова. В 2- тт., Т. 1, стр. 87-111. М: ВЦ РАН
22. ChemSpider. Search and share chemistry // www.chemspider.com