

# ОНТОЛОГИИ И БАЗЫ ДАННЫХ ПО ТЕПЛОФИЗИЧЕСКИМ СВОЙСТВАМ РЕАКТОРНЫХ МАТЕРИАЛОВ

**И.А. Чусов\*, П.Л. Кириллов\*, В.Г. Проняев\*\*, А.О. Еркимбаев\*\*\*,  
В.Ю. Зицерман\*\*\*, Г.А. Кобзев\*\*\*, Л.Р. Фокин\*\*\***

*\* Институт атомной энергетики НИЯУ МИФИ*

*249040, Калужская обл., г. Обнинск, Студгородок, 1*

*\*\* Частное учреждение «Атомстандарт»*

*117342, Москва, ул. Бутлерова, 17*

*\*\*\* Объединенный институт высоких температур РАН*

*125412, Москва, ул. Ижорская, 13, строение 2*



Анализируется информационная технология хранения, систематизации и распространения теплофизических данных для атомной энергетики. Общая тенденция в областях, связанных с широким использованием научных данных, состоит в переходе от традиционных баз данных к созданию единой инфраструктуры, способной преодолеть резкое нарастание объема данных при усложнении их структуры, связанной с постоянным расширением классов материалов. Подобная инфраструктура обеспечивает интероперабельность, включая обмен и распространение данных. Предложен общий принцип управления данными для реакторной теплофизики, основанный на предметно-ориентированной онтологии ReactorThermoOntology (RTO). Онтология включает в себя единый словарь всех понятий, расширенный за счет логических связей и аксиом. Предложенная в работе онтология объединяет термины, характерные для реакторных материалов, их характеристик, а также информационных сущностей, определяющих текстовые, математические и компьютерные структуры. В закодированном виде онтология становится управляющей надстройкой, способной интегрировать разнородные данные. Важнейшая ее особенность – возможность перманентного расширения, необходимая по мере введения в практику новых материалов. Рассмотрены возможные сценарии ее использования на этапах проектирования, эксплуатации и интеграции автономных ресурсов, преимущественно БД. В качестве эффективного средства интеграции предложено использовать технологию больших данных при многообразных вариациях их логической структуры.

**Ключевые слова:** теплофизические свойства, реакторные материалы, ядерное топливо, онтология, база данных, интеграция данных, JSON-формат.

## ВВЕДЕНИЕ

Развитие атомной энергетики неразрывно связано с обширной инфраструктурой, обеспечивающей хранение данных и знаний по технологиям, рабочим процессам, материалам и другим аспектам. Ее сердцевину образуют базы данных (БД), причем рост

© *И.А. Чусов, П.Л. Кириллов, В.Г. Проняев, А.О. Еркимбаев, В.Ю. Зицерман, Г.А. Кобзев, Л.Р. Фокин, 2019*

объема данных, связанный с расширением круга объектов, требует вовлечения в практику множества неоднородных БД, резко отличающихся форматами, семантикой и т.д. Это затрудняет их совместное использование при обработке и использовании в вычислениях. В результате возникает проблема интеграции разнородных ресурсов, созданных при существенных различиях в модели и форме представления данных [1]. Ее осознание стимулировало масштабные исследования для перестройки всей инфраструктуры данных с ориентацией на рост их объема при множестве структур и моделей.

Ключевая идея будущей инфраструктуры – формирование пространства, соединяющего данные разной структуры, исходно локализованные в автономных БД [2, 3]. Одной из определяющих идей при организации такого пространства является использование онтологий, семантически точных и машинно-обрабатываемых определений сущностей и их соотношений [4]. Онтология формализует предметную область, обеспечивая семантическое единство для отдельных БД. Ее функция значительно шире обычной таксономии, так как она поддерживает логические связи между понятиями, обусловленные спецификой предметной области. Добавляя к данным семантику (смысл) и логические соотношения, онтология описывает «знание» с возможностью его машинной интерпретации и тем самым становится необходимой надстройкой при создании интегрирующей инфраструктуры.

Широкие возможности онтологий стимулировали активность по их разработке для формализации многих дисциплин – химии, материаловедения, наук о Земле и др. [5, 6]. В атомной энергетике достаточно давно разработана семантическая система управления знаниями [7], по идеям близкая онтологиям, поскольку опирается на действующий в сети набор словарей и таксономий, обеспечивающих семантическое единство ресурсов.

Статья посвящена созданию онтологии для теплофизики реакторных материалов. Теплофизические свойства материалов с учетом их значимости при моделировании предъявляют повышенные требования к ресурсам, обеспечивающим их накопление и систематизацию. Ранее для интеграции теплофизических данных была предложена упрощенная онтология, выделяющая базовые концепции (вещество, свойство, данные) и сопутствующие понятия (размерность, состояние, неопределенность, источник) [8]. В ее рамках удалось стандартизировать набор понятий и ввести требуемые по смыслу ограничения, например, соответствия между агрегатным состоянием вещества и его свойствами. Построить ее аналог для реакторных материалов возможно лишь при существенном расширении понятийного аппарата с охватом ранее игнорируемых факторов – состав композиции, характеристики образца, сведения о технологии, режиме эксплуатации и т.п. В [9] был обрисован прототип подобной онтологии (NuclThermo Ontology), фиксирующий ее структурные элементы: таксономию материалов и свойств; модели представления данных и их неопределенности; экстра-факторы типа пористости, степени выгорания, облучения и т.д.; аксиоматику использования свойств применительно к определенным классам материалов.

Предлагается реализация намеченного проекта, детализирующая иерархию классов и ассоциативные соотношения, определяющие логические связи между концепциями. Следующий раздел посвящен описанию онтологии при возможности перманентного наращивания новых понятий, например, в перечне материалов и номенклатуре свойств. Вспомогательные материалы в виде громоздких диаграмм, иллюстрирующих таксономию объектов и их характеристик, а также OWL-файл, кодирующий содержание онтологии, из-за большого объема представлены на сайте [10].

Третий раздел посвящен наиболее важному в практическом отношении вопросу – как, базирясь на действующей онтологии, обеспечить интеграцию нескольких БД по теплофизическим свойствам, различающихся форматом, моделью данных и их семантикой. Под интеграцией понимается их виртуальное соединение с общим интерфейсом и еди-

ной системой запросов. Большой опыт в применении онтологий для интеграции материаловедческих БД [5] позволяет считать вполне реализуемой и задачу интеграции данных по реакторной теплофизике.

### ПРОЕКТ ОНТОЛОГИИ RTO

Проектирование онтологий [4], как правило, начинается с выделения фундаментальных понятий, заимствованных из предметно-независимой онтологии **BFO** (Basic Formal Ontology) (рис. 1).

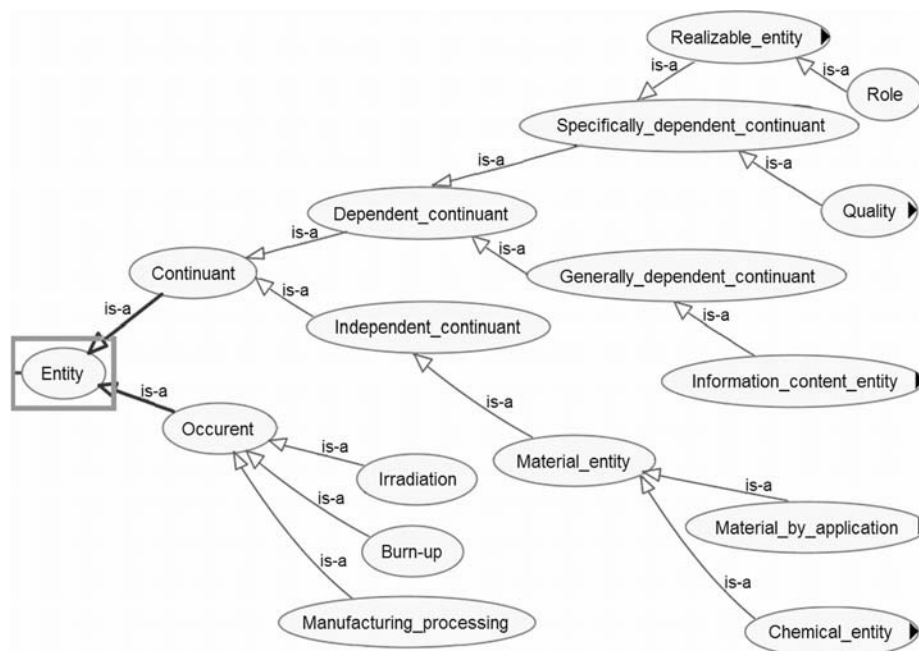


Рис. 1. Классы верхнего уровня онтологии

Как видно, исходное понятие «класс **Entity**» имеет двух наследников: **Continuant** для сущностей, сохраняющих со временем свою идентичность, и **Occurent** для сущностей, имеющих временные части и развивающиеся со временем. Класс **Continuant** имеет в качестве наследников **Independent Continuant** и **Dependent Continuant** для объектов предметной области и их атрибутов. Класс **Dependent Continuant** также имеет два дочерних класса: **Specifically Dependent Continuant** (объединяющий характеристики объектов) и **Generically Dependent Continuant** (суммирующий виды данных и соответствующие концепции) с одним дочерним классом **Information\_content\_entity**. Класс **Specifically dependent continuant** имеет два дочерних класса: **Quality** (вся совокупность свойств объектов, процессов и документов), а также **Realizable entity** для определения объектов в терминах их возможных ролей или способностей. В итоге формируется строгая и универсальная структура, адаптируемая к любой предметной области. На том же рисунке показаны следующие уровни иерархии, включающие в себя базовые понятия реакторной теплофизики. Все виды веществ и материалов представлены классом **Material\_entity**, класс **Quality** имеет в качестве дочерних классы **Material Quality**, **Sample**, **State**, **Information\_content\_quality**, **Process\_quality**. Первый из них суммирует все виды физических характеристик типа плотность, теплопроводность и т.д. Класс **Sample** определяет характеристики конкретного образца (марка производителя, форма и структура образца и т.д.), класс **State** выделяет агрегатное состояние (газ, жидкость и т.п.). Два последних класса характеризуют информационные сущности (доку-

мент, идентификатор) и процессы.

Специфика реакторной теплофизики определяется прежде всего иерархией материалов, наследующих класс **Material\_entity** (рис. 2). Его дочерний класс **Material\_by\_application** в соответствии с принятой классификацией [11] объединяет материалы пяти категорий: топливо, теплоноситель, замедлитель, поглощающий и конструкционный материалы. Другой дочерний класс **Chemical\_entity** включает в себя не относящиеся к этим категориям вещества или многокомпонентные системы.

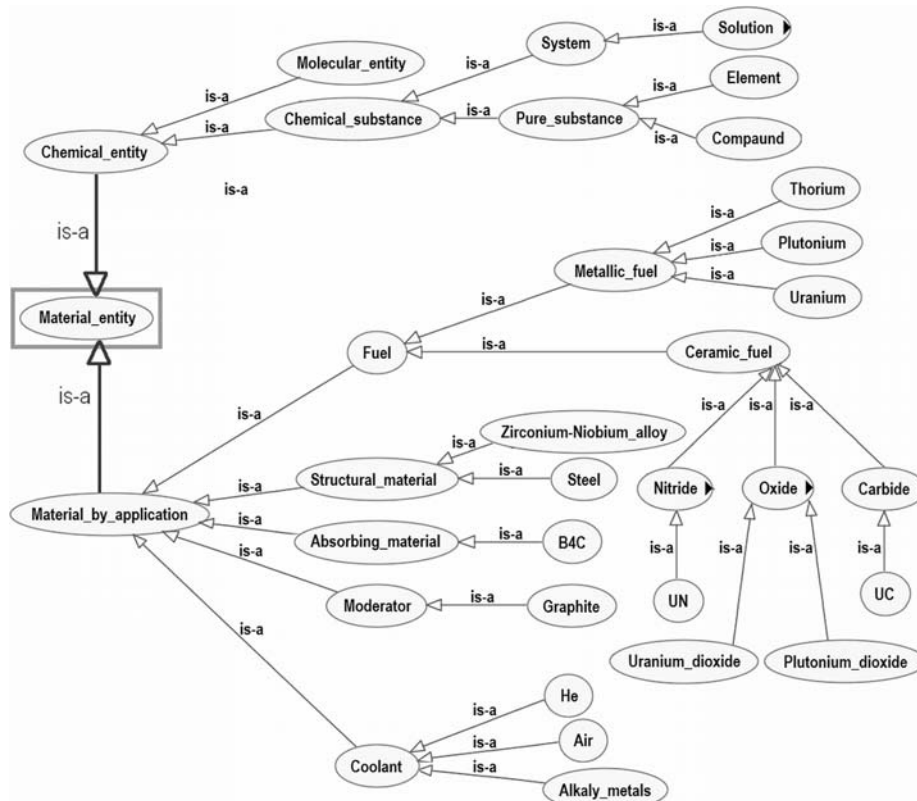


Рис. 2. Классы, определяющие виды материалов – сокращенная версия иерархии

На рисунке 2 показаны самые типичные классы, более детальная диаграмма [10] содержит классы, представляющие распространенные виды реакторных материалов и их многочисленные экземпляры. Там указаны композитные и солевые топлива с их экземплярами, например, металлокерамика  $UO_2-Zr$  или солевой расплав  $LiF-BeF-ThF_4-UF_4$ . В качестве теплоносителей показаны металлы и сплавы р-блока периодической таблицы (Pb, Pb-Bi), в качестве замедлителей – оксидные системы  $Dy_2O_3-TiO_2$ ,  $Dy_2O_3-HfO_2$  и т.д.

Исследования атомных материалов часто охватывают объекты, не подпадающие под выделенные пять категорий. К ним относятся системы типа U-O, включающие в себя оксиды  $UO$ ,  $UO_2$ ,  $UO_3$ ,  $U_3O_7$ ,  $U_3O_8$ ,  $U_4O_9$  в различных фазовых состояниях, растворы и соединения Np, Am, Cm (так называемые minor actinides), продукты деления (fission products) в смесях с другими материалами, а также многочисленные соединения и частицы (атомы, молекулы, изотопы). Все они относятся к иерархии, порождаемой классом **Chemical\_entity**, отражая при этом возможную роль или назначение в классе верхнего уровня **Role** (см. рис. 1).

Приведенная на рис. 2 таксономия отчетливо иллюстрирует возможность ее расширения за счет новых материалов или их категорий. Например, можно ввести класс

**CerMet\_fuel** (композитных топлив) типа  $UO_2$ -Zr,  $UO_2$ -Al и далее расширить его за счет систем U-PuO<sub>2</sub>, U-PuN, U-PuC [12], включающих в себя матрицу из U, вместо неделимых веществ, например, Zr или Al. Другой важный момент – наличие класса **Sample** (см. рис. 1), детализирующего специфику конкретного образца в дополнение к общим свойствам материала. Как правило, он суммирует такие детали, как заводская марка (grade), форма, размер, структура, отклонение от стехиометрии, например, для оксидов  $MO_{2\pm x}$ . Особенности реакторных материалов отражает также класс **Occurent** (см. рис. 1), включающий в себя технологию изготовления (подготовка образца), облучение и выгорание топлива. Например, для идентификации поглотителей типа  $Dy_2O_3$ -TiO<sub>2</sub>,  $Dy_2O_3$ -HfO<sub>2</sub> [13] потребовалось, кроме состава, уточнить форму образца (в виде таблетки), ее размеры (диаметр и высота), а также технологию – спекание или плавление. Обе возможности онтологии, расширения иерархии и детализации, образца решают проблему непрерывной подстройки структуры данных к особенностям новых объектов и понятий [14].

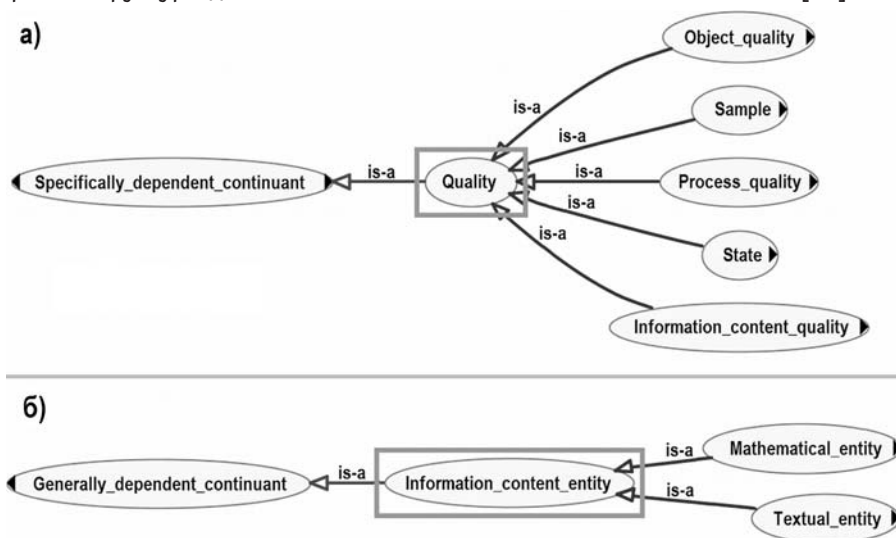


Рис. 3. Классы, определяющие характеристики а) – материальных объектов; б) – информационных сущностей

Наряду с классом **Material\_entity** полноту представления данных обеспечивают два класса верхнего уровня: **Quality** и **Information\_content\_entity** (рис. 3). Первый из них включает в себя сущности, определяющие свойства материалов, процессов, информационных сущностей, а также образца и состояния вещества. Второй класс **Information\_content\_entity** включает в себя сущности, отражающие представление данных в публикациях или БД. Они обеспечивают идентификацию объектов и информационных ресурсов, представление документов, формирование численных данных. Первые две задачи решает класс **Textual\_entity** с дочерними классами **Identifier**, **Document**, **Document\_part**. В числе идентификаторов предусмотрены ключевые слова, названия, ID и т.д. Все концепции, необходимые для представления численных данных, собраны в классах, дочерних по отношению к **Mathematical\_entity**. Среди них типовые понятия – **number**, **quantity**, **variable** (с выделением **independent** and **dependent**) и др. Ведущим является понятие **Data set**, т.е. *контейнер* определенной структуры, содержащий константы и таблицу значений нескольких физических величин (плотность, теплоемкость, энтальпия и т.д.). Указание на используемые единицы измерения выполняет класс **Unit\_of\_measurement**. Заметим, что на рис. 3 показаны лишь начала тех иерархий, что порождены классами **Quality** и **Information\_content\_entity**, а более полную картину можно найти на сайте [10]. Там можно увидеть всю совокупность физических свойств, например, данные по диаграммам состояния и кристалличес-

ким структурам, включенные в иерархию **Material\_quality**. Классы **Sample** и **State** объединяют данные по образцу и фазовому состоянию, класс **Property\_quality** – характеристики процессов, например, степень выгорания.

Особую ценность данным придает иерархия **Information\_content\_quality**. Она включает в себя классы **Uncertainty**, **Status\_descriptor** (разделяющий типы данных – экспериментальные, расчетные, справочные, рекомендованные) и **Quality\_descriptor**, т.е. суммарную оценку качества данных с учетом их неопределенности, полноты описания образца и метода, степени воспроизводимости и т.д.

Реализованный здесь принцип [4] предполагает активное использование классов и отдельных концепций из других онтологий, ранее разработанных и представленных в сети. Этот принцип обеспечивает согласование разнородных словарей, исключающее неидентичность в определениях одного и того же понятия.

Наряду с онтологией BFO для импорта понятий использованы две онтологии общенаучного содержания из репозитория **Ontobee** ([www.ontobee.org](http://www.ontobee.org)): **Semanticscience Integrated Ontology** (SIO) и **Chemical Information ontology** (ChemInf). Онтология SIO предоставляет обширный набор типов и соотношений для описания научной активности. Онтология ChemInf охватывает типовые концепции в химии и науке о материалах – молекула, раствор, вещество и т.д. Термины из внешних онтологий почти полностью покрывают набор понятий, дочерних по отношению к таким классам, как **Chemical\_entity**, **Information\_content\_entity**, **Information\_content\_quality** и **Material\_Quality** (частично). Все новые понятия и соответствующие им классы предназначены, в основном, для сущностей, характерных именно для реакторной теплофизики, прежде всего тех, что включены в иерархию **Material\_by\_application** (см. рис. 2 и в более полном виде на сайте [10]).

Таблица 1

**Ограниченный набор типичных объектных свойств (ассоциативных соотношений), импортированных из онтологии SIO**

№	ObjectctProperties	ID
1	<i>Is_related_to</i>	SIO_000001
2	<i>Denotes</i>	SIO_000020
3	<i>Has_part</i>	SIO_000028
4	<i>Is_denoted_by</i>	SIO_000060
5	<i>Is_part_of</i>	SIO_000068
6	<i>Is_contained_in</i>	SIO_000128
7	<i>Contains</i>	SIO_000202
8	<i>Is_connected_to</i>	SIO_000203
9	<i>Has_quality</i>	SIO_000217
10	<i>Is_quality_of</i>	SIO_000218
11	<i>Is_source_of</i>	SIO_000219
12	<i>Has_role</i>	SIO_000228
13	<i>Is_component_part_of</i>	SIO_000313
14	<i>Is_covalently_connected_to</i>	SIO_000334
15	<i>Is_weakly_interacting_with</i>	SIO_000335
16	<i>Has_component_part</i>	SIO_000362
17	<i>Is_described_by</i>	SIO_000557
18	<i>Describes</i>	SIO_000563
19	<i>Has_identifier</i>	SIO_000671
20	<i>Has_data_item</i>	SIO_001277
21	<i>Is_data_item_in</i>	SIO_001278

В полном объеме эффективность заимствования из внешних онтологий проявляется при выборе так называемых объектных свойств или ассоциативных соотношений.

Сама по себе иерархия классов, описанная выше, представляет собой лишь словарь в виде таксономии понятий для объектов, их характеристик, процессов, документов и т.д. Объектные свойства устанавливают логические связи между объектом и его характеристикой, объектом и его частью, документом и предметом его описания и т.д. В качестве подходящего источника объектных свойств здесь принята упомянутая **Semanticscience Integrated Ontology** ([www.ontobee.org/ontology/SIO](http://www.ontobee.org/ontology/SIO)). Она включает в себя набор из 207-ми объектных свойств, способных адекватно представить почти все мыслимые связи между объектами и их атрибутами. В таблице 1 приведена ограниченная выборка из этого набора с наиболее характерными связями.

<b>has part [SIO:000028]</b>	
<b>label</b>	has part
<b>identifier</b>	SIO:000028
<b>description</b>	has part is transitive, reflexive and antisymmetric relation between a whole and itself or a whole and its part.
<b>term type</b>	ReflexiveProperty TransitiveProperty ObjectProperty
<b>subproperty of</b>	is location of
<b>superproperty of</b>	has proper part
<b>inverse of</b>	is part of
<b>is defined by</b>	<a href="http://semanticscience.org/ontology/sio.owl">http://semanticscience.org/ontology/sio.owl</a>

Рис. 4. Образец представления объектного класса *has\_part* в онтологии SIO

Каждое из свойств определяется ID, включенным в уникальный сетевой адрес (URL). Так, индекс SIO\_000028 определяет URL [http://semanticscience.org/resource/SIO\\_000028.rdf](http://semanticscience.org/resource/SIO_000028.rdf), где приведено детальное описание свойства *has\_part* (рис.4). Пример его использования дает утверждение **molecule has\_part some atom**; для обратного утверждения используется свойство *is\_part\_of* (SIO\_000068).

В итоге онтология RTO определяет для реакторной теплофизики согласованный словарь терминов и спецификацию их смысла, что позволяет вести поиск и логические выводы. Существенно, что можно непрерывно наращивать классы, связанные с появлением новых материалов или новых факторов, например, наноструктурой топлива. Расширяемый перечень классов, дочерних по отношению к **Sample**, позволяет включить исходно непредусмотренные особенности образца – конфигурацию, примеси, пористость и т.д. Наконец, огромный набор текстовых и математических терминов обеспечивает требуемую подстройку типа данных, например, замену единичного числа на массив или интервал значений.

## ОНТОЛОГИЯ КАК СРЕДСТВО ПРОЕКТИРОВАНИЯ И ИНТЕГРАЦИИ БД

**Взаимная дополнительность онтологий и БД.** Возникло направление, предполагающее синтез БД с онтологиями с целью использования преимуществ обоих инструментов. В его основе лежит четкое соответствие между элементами БД и онтологии. Так, сущностям БД в онтологии соответствуют определенные классы, атрибутам – свойства классов, ограничениям – аксиомы. Их главные отличия в том, что онтология записывает на формальном языке структуру предметной области, а концептуальная схема БД описывает определенную БД, не претендуя на распространение знаний. Иначе говоря, БД фокусируется именно на данных (числовых, текстовых и т.д.), в то время как онтология ориентирована на толкование их смысла и реализацию машинных выводов. Потеря семантики является главным ограничением БД, так как смысл сущностей не доступен для лиц, не знакомых со схемой БД.

Другой принципиальный недостаток БД – невозможность поддерживать эво-

люцию схемы данных. Практика систематизации данных по свойствам показывает непригодность жесткой структуры для широкой совокупности веществ с учетом специфики образцов [14]. Например, эволюция схемы данных может быть связана со спецификой эксплуатации ядерных реакторов, отражающейся на свойствах материалов.

В то же время БД обеспечивают наиболее высокую производительность при реализации сложных запросов, не достижимую для других архитектур. Именно поэтому оправданно внедрение систем, сочетающих доступность семантики с высокой производительностью при работе с данными. В задачах проектирования или интеграции БД реализуется *database-to-ontology mapping*, когда возможности БД усилены связью с онтологией для запросов с использованием общей семантики или интеграции разнородных БД. Онтология обеспечивает семантику, т.е. словарь, соотношения между понятиями и структуру данных, но не содержит экземпляров – их роль выполняют записи БД. В ходе интеграции производится объединение схем и самих данных, а пользователь формулирует в терминах онтологии запросы, которые конвертируются в запросы к БД. Тем самым онтология играет роль эффективного посредника между пользователем и данными. В работе [8] было показано, насколько эффективной оказывается онтология при проектировании и интеграции теплофизических и материаловедческих БД.

#### Примеры эффективного использования онтологий по свойствам веществ.

Из БД по свойствам веществ, интегрированных с онтологией, наибольшей известностью пользуется химическая БД **ChEBI** ([www.ebi.ac.uk/chebi/](http://www.ebi.ac.uk/chebi/)). Каждой из включенных в БД молекулярных сущностей (а также их группам и классам) присвоен уникальный идентификатор **ChEBI ID** (например, **ChEBI:15377** для H<sub>2</sub>O), доступный для цитирования пользователем сети или программным агентом (чтобы выйти на требуемую запись, достаточно ввести ее **ChEBI ID** в Google). Набор данных включает в себя элементарные сведения о молекуле (формула, масса, заряд), структурные данные, гиперссылки на другие БД, в частности, БД NIST ([webbook.nist.gov/chemistry/](http://webbook.nist.gov/chemistry/)) с обширной термодинамической информацией.

Каждая запись содержит также фрагмент онтологии с возможностью навигации по родительским и дочерним классам. Ассоциативные соотношения типа тех, что представлены в табл.1, указывают на объекты, функционально связанные с исходным, например, соотношение *has\_role* связывает запись для H<sub>2</sub>O с записью для термина «парниковый газ», а соотношение *has\_part* указывает на гидрат, для которого H<sub>2</sub>O является составной частью. Начиная навигацию с определенной сущности, можно выйти на множество других, определив их логические и ролевые связи, встроенные в многоуровневые таксономии, что позволяет реализовать сложные запросы с поиском веществ по структуре, функциям, роли и т.д. В целом, это означает, что наряду с данными для конкретного вещества **ChEBI** предоставляет фрагмент «знаний» из предметной области.

Для интеграции нескольких БД используется преобразование внешнего запроса, составленного в терминах онтологии, в набор подзапросов к изолированным источникам данных без их объединения в едином хранилище. Центральный компонент такой виртуальной системы – *посредник* с единым интерфейсом доступа и единой моделью, основанной на онтологии или наборе онтологий. Включенный в посредник процессор выполняет разделение запроса на подзапросы к источникам данных. К источникам данных прикрепляются так называемые *адаптеры* для разрешения проблем, возникающих из-за разнородности моделей. Таким образом, в основе интеграции лежит использование архитектуры «посредник-адаптер» глобальной модели в виде онтологии и *федерирование* данных, т.е. их слияние при



выполнении запроса к системе из автономных БД. Детали технологии при различных подходах к интеграции обсуждаются в [15, 16]

Возможности интеграции на основе онтологии **Matinfo** рассмотрены в [17]. Онтология для предметной области «Конструкционные материалы» построена в виде семи субонтологий, из которых основные концепции дают четыре базовые: **Substance** для веществ, смесей и материалов; **Property** для химических, тепловых и механических свойств; **Process** для методов производства и измерения; **Environment** для характеристик среды (состав атмосферы, температура, pH и т.д.). Кроме четырех базовых в общую онтологию включена субонтология **Materials Information**, которая агрегирует все термины и концепции, характеризующие материал и конкретный образец, методы и условия измерения, критерии качества данных и т.д. Две периферические онтологии **Unit Dimension** и **Physical Constant** покрывают потребность в представлении единиц измерения и констант.

Онтология решает задачу обмена данными между тремя БД Японии: AIST (Advanced Industrial Science and Technology), NIMS (National Institute of Material Science) и широко используемой структурой **MatDB** (<https://odin.jrc.ec.europa.eu>), содержащей данные по тестированию материалов. Основной объем информации в указанных БД относится к тепловым и механическим свойствам. Обмен данными реализован с использованием промежуточного уровня, занятого онтологией **Matinfo**.

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 можно отразить как скалярную величину, так и произвольную матрицу.

В атомной энергетике примером подобной интеграции служит система обмена данными между упомянутой **MatDB** в составе сети Online Data & Information Network for Energy и подготовленным Окриджской Национальной лабораторией США справочником **Gen IV Materials Handbook** (<https://gen4www.ornl.gov/>). Интеграция проведена без физического объединения ресурсов, путем конверсии двух различных форматов к единому формату данных [18]. Техника импорта и экспорта данных в обоих ресурсах использовала XML-представление, которое не унифицирует семантику, что требует детального знания обеих концептуальных схем. Заметное усложнение проблемы с увеличением числа и многообразия ресурсов является дополнительным аргументом в пользу онтологического подхода к интеграции БД.

**Технология Big Data в задачах интеграции ресурсов.** Новые возможности в интеграции научных ресурсов открывает использование подходов, предназначенных для работы с большими данными. Применительно к свойствам материалов источником больших данных является публикационный поток, причем их объем определяет не столько число изучаемых объектов, сколько многообразие условий синтеза, измерений, микроструктурных особенностей и т.д. Из трех признаков отнесения к категории «больших данных» (3V – Volume, Velocity, Variety) применительно к данным о свойствах именно последний играет решающую роль. В [19] показано, что этот подход позволяет с минимальными затратами преодолеть две

главные проблемы на пути интеграции ресурсов: многообразие схем, терминологии, типов и форматов данных и необходимость перманентной подстройки создаваемой структуры к вариациям номенклатуры объектов и понятий. Предлагаемые решения базировались на совместном использовании ранее применяемых в отдельности технологий: стандарт обмена в виде текстовых структурированных документов; онтологически основанное управление данными; платформа работы с большими данными **Apache Spark** (<http://spark.apache.org/docs/>). Их применение позволяет конвертировать множество первичных ресурсов (БД, файловые архивы и т.д.) к стандартному текстовому формату JSON с последующим использованием онтологий для семантической интеграции. Задачи хранения разнородных данных, доступа к ним и анализ возлагаются при этом на платформу **Apache Spark**.

JSON-формат оказался одним из наиболее удобных для обмена данными за счет простого чтения и редактирования, доступности для восприятия человеком, хранения иерархических структур. JSON-формат является рабочей единицей для ряда платформ, в частности, для **Apache Spark**, позволяющей организовывать обмен, хранение и организацию запросов для распределенных данных.

Разработанный в [19] сценарий предусматривает конверсию каждого из ресурсов к JSON-документам, связанными с репозитарием, включающим в себя как предметно-ориентированные, так и онтологии верхнего уровня. Роль онтологий состоит в привнесении семантики в документы, а также возможности корректировать структуру данных. Связывание документов с онтологиями позволяет вести семантический поиск, выявляя информацию верхнего и нижнего уровней (родительские и дочерние классы) и боковые связи (*related terms*), не зная схем данных источников.

Платформа **Apache Spark** позволяет справиться с нарастающим объемом и распределенным характером данных по свойствам за счет высокой производительности и выраженной ориентации на работу с данными, включая их хранение, обработку, анализ в распределенной среде. Среди прочих технологических особенностей ее отличает наличие встроенных библиотек для аналитической обработки, в том числе для организации SQL-запросов, посредством которых можно получить доступ к содержимому структурированных JSON-документов. Именно возможность SQL-запросов к обширным массивам данных играет ключевую роль в решаемой задаче их интеграции. Другая особенность, определяющая эффективность Spark при хранении и обработке данных, – способность поддерживать взаимодействие с множеством типов хранилищ, от **HDFS** (*Hadoop Distributed Files System*) до традиционных БД на локальных компьютерах.

Преимущества **Apache Spark** заключаются в большей скорости вычислений, работе с данными различной природы (текстовыми, полуструктурированными и структурированными) из различных источников (файлы различных форматов, СУБД и потоковые данные). В совокупности предложенные средства открывают неограниченные по производительности и многообразию возможности работы со сложными данными, к которым относятся данные по свойствам сложных композиций, с которыми имеет дело реакторная теплофизика.

## ЗАКЛЮЧЕНИЕ

В работе обоснована актуальность внедрения нового подхода к работе с данными по свойствам материалов, основанного на онтологиях как средствах, обеспечивающих интеграцию разнородных данных. Богатые возможности онтологий в областях, связанных с использованием обширных массивов данных, подтверждаются их широким внедрением для целей систематизации, поиска, и доверенных компьютеру логических построений. Предложен вариант онтологии для теплофизи-

зических данных с их высокими требованиями к представлению неопределенности, деталям эксперимента и образца и т.д. Показано, что за счет редактирования онтологий можно преодолеть две главные проблемы в работе с данными: эволюцию работающих БД, связанную с расширением номенклатуры объектов и понятий; учет экстра-факторов, влияющих на свойства материала (технология изготовления, размеры и структура образца, влияние среды, условий эксплуатации и др.). Рассмотрены примеры использования онтологий в химии и материаловедении как средств интеграции автономных ресурсов с различной структурой и форматом данных. В качестве наиболее эффективного средства предложено использовать некоторые технологии так называемых «больших данных», изначально приспособленных к работе с множеством гетерогенных источников.

### Литература

1. Дударев В.А. Интеграция информационных систем в области неорганической химии и материаловедения. – М.: КРАСАНД. – 2016. – 320 с.
2. Bizer C. Interlinking scientific data on a global scale. // *Data Science Journal*. – 2013. – Vol. 12. – PP. GRDI6-GRDI12.
3. Frenkel M. Global Information Systems in Science. Application to the Field of Thermodynamics. // *J. Chem. Engn. Data*. – 2009. – Vol. 54. – No 9. – PP. 2411-2428.
4. Uschold M., Gruninger M. Ontologies: principles, methods and applications. // *Knowledge Engineering Review*. – 1996. – Vol. 11. – No 2. – PP. 93-136.
5. Zhang X., Zhao C., Wang X. A survey on knowledge representation in materials science and engineering: An ontological perspective. // *Computers in Industry*. – 2015. – Vol. 73. – PP. 8-22.
6. Brodaric B., Gahegan M. Ontology use for semantic e-Science. // *Semantic Web*. – 2010. – Vol. 1. – PP. 149-153.
7. Knowledge Organization Systems and Semantic Technology. International Atomic Energy Agency. Nuclear Knowledge Management (NKM). Электронный ресурс: [www.iaea.org/nuclearenergy/nuclearknowledge/](http://www.iaea.org/nuclearenergy/nuclearknowledge/) (дата доступа 10.08.2018).
8. Еркимбаев А.О., Цицерман В.Ю., Кобзев Г.А., Косинов А.В. Связывание онтологий с базами данных по свойствам веществ и материалов // *Научно-техническая информация. Серия 2*. – 2015. – № 12. – С. 1-16.
9. Chusov I.A., Kirillov P.L., Bogoslovskaya G.P. et al. Development of the system of reactor thermophysical data on the basis of ontological modeling. // *Journal of Physics Conference Series*. – 2017. – Vol. 891. – #012172.
10. Портал THERMOPHYSICS.RU. Электронный ресурс: <http://www.thermophysics.ru/RTO/index.html> (дата доступа 10.08.2018)
11. Thermophysical Properties of Materials for Nuclear Engineering: Tutorial for students of specialty «Nuclear Power Plants». Ed. by Prof. P.L. Kirillov. – Obninsk: IATE, 2006. – 182 p.
12. Mishra S., Kaity S., Banerjee J., Nandi C., Dey G.K., Khan K.B. U-PuO<sub>2</sub>, U-PuC, U-PuN cermet fuel for fast reactor. // *Journal of Nuclear Materials*. – 2018. – Vol. 499. – PP. 272-283.
13. Рисованный В.Д., Захаров А.В., Муралева Е.М. Новые перспективные поглощающие материалы для ядерных реакторов на тепловых нейтронах. // *ВАНТ. Серия: Физика радиационных повреждений и радиационное материаловедение*. – 2005. – № 3. – С. 87-93.
14. Erkimbaev A.O., Zitserman V.Y., Kobzev G.A., Fokin L.R. The logical structure of physicochemical data: Problems of numerical data standardization and exchange // *Russian Journal of Physical Chemistry A*. – 2008 – Vol. 82. – No 1. – PP. 15-25.
15. 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. – No 2. – PP. 166-187.
16. Kogalovsky M.R. Ontology-based data access systems. // *Programming and Computer Software*. – 2012. – Vol. 38. – No 4. – PP. 167-182.
17. Ashino T. Materials ontology: an infrastructure for exchanging materials information and knowledge. // *Data Science Journal*. – 2010. – Vol. 9. – PP. 54-61.

18. *Lin L, Austin T, Ren W.* Interoperability of materials database systems in support of nuclear energy development and potential applications for fuel cell material selection. // *Materials Performance and Characterization.* – 2015 – Vol. 4. – No 1. – PP. 115-130.

19. *Erkimbaev A.O., Zitserman V.Yu., Kobzev G.A., Kosinov A.V.* Standardization of Storage and Retrieval of Semi-structured Thermophysical Data in JSON-documents Associated with the Ontology. Proceedings of the XIX International Conference «Data Analytics and Management in Data Intensive Domains» (DAMDID/RCDL'2017), Moscow, Russia, October 10–13, 2017. Электронный ресурс: <http://ceur-ws.org/Vol-2022/paper36.pdf> (дата доступа 10.08.2018).

Поступила в редакцию 22.08.2018 г.

#### Авторы

Чусов Игорь Александрович, зав. кафедрой, д.т.н.  
E-mail: [igrch@mail.ru](mailto:igrch@mail.ru)

Кириллов Павел Леонидович, д.т.н., профессор, ИАТЭ НИЯУ МИФИ  
E-mail: [plkirillov@gmail.com](mailto:plkirillov@gmail.com)

Проняев Владимир Григорьевич, в.н.с. к.ф.-м.н.  
E-mail: [vgpronyaev@yandex.ru](mailto:vgpronyaev@yandex.ru)

Еркимбаев Адильбек Омирбекович, зав. лабораторией, к.т.н.  
E-mail: [adilbek@ihed.ras.ru](mailto:adilbek@ihed.ras.ru)

Зицерман Владимир Юрьевич, вед. н.с., к.ф.-м.н.  
E-mail: [vz1941@mail.ru](mailto:vz1941@mail.ru)

Кобзев Георгий Анатольевич, гл. н.с., д.ф.-м.н., проф.  
E-mail: [kobzev@oivtran.ru](mailto:kobzev@oivtran.ru)

Фокин Лев Рувимович, гл. н.с., д.т.н.  
E-mail: [lfokin@mail.ru](mailto:lfokin@mail.ru)

UDC 621.039.4; 04.652

### ONTOLOGIES AND DATABASES ON THERMOPHYSICAL PROPERTIES OF NUCLEAR REACTOR MATERIALS

Chusov I.A.\*, Kirillov P.L.\*, Pronyaev V.G.\*\*, Erkimbaev A.O.\*\*\*, Zitserman V.Yu.\*\*\*, Kobzev G.A.\*\*\*, Fokin L.R.\*\*\*

\* Institute for Nuclear Power Engineering, NRNU «MEPhI»,  
1 Studgorodok, Kaluga reg., Obninsk, Russia 249040

\*\* Private Institution «Atomstandart»,  
17 Butlerova st., Moscow, Russia 117342

\*\*\* Joint Institute for High Temperatures of Russian Academy of Sciences  
13, bd.2 Izhorskaya st., Moscow, Russia 125412

#### ABSTRACT

The study is dedicated to the information technologies for storage, systematization and distribution of thermophysical data for nuclear power engineering. The general trend existing in the areas involving wide use of scientific data is the shifting from conventional databases to the development of a consolidated infrastructure capable of overcoming sharply growing volumes of scientific data with continuously increasing complexity of the data structure due to the expansion of the range of materials. The above infrastructure ensures interoperability, including data exchange and dissemination. The general principle of data management for thermophysical

properties of the nuclear reactor materials based on the subject-oriented ReactorThermoOntology (RTO) is suggested in the present paper. The ontology includes a unified glossary of all concepts, expanded through logical connections and axioms. The suggested RTO ontology combines the terms typical for reactor materials, their characteristics, as well as all types of information entities, determining textual, mathematical and computer structures. In the coded form, the ontology becomes the control add-in that can integrate heterogeneous data. Its most important feature is the possibility of its permanent expansion, which is necessary with introduction of new materials and terms related to them, e.g. nanostructures characteristics. Beside the ontology, description of the reactor materials, the possible scenarios for the use of the ontology during the phases of design, operation and integration of autonomous resources, primarily databases, are examined in the paper. The use of Big Data technology with diverse variations of logical structures of the data is suggested as the most efficient tool for data integration. Joint use of the technologies which were applied separately before, such as exchange standard in the form of the structured text documents, data control based on the ontology and platform for the work with big data, allows the conversion of multiple existing primary resources (databases, files, archives, etc.) to the standard JSON text format for the subsequent semantic integration.

**Key words:** thermophysical properties, reactor materials, nuclear fuel, ontology, database, data integration, JSON-format.

#### REFERENCES

1. Dudarev V.A. *Integration of information systems on inorganic chemistry and material Science*. Moscow. KRASAND Publ., 2016, 320 p. (in Russian).
2. Bizer C. Interlinking scientific data on a global scale. *Data Science Journal*, 2013, v. 12, pp. GRDI6-GRDI12.
3. Frenkel M. Global Information Systems in Science. Application to the Field of Thermodynamics. *J. Chem. Engr. Data*, 2009, v. 54, no. 9, pp. 2411-2428.
4. Uschold M., Gruninger M. Ontologies: principles, methods and applications. *Knowledge Engineering Review*, 1996, v. 11, no. 2, pp. 93-136.
5. Zhang X., Zhao C., Wang X. A survey on knowledge representation in materials science and engineering: An ontological perspective. *Computers in Industry*, 2015, v. 73, pp. 8-22.
6. Brodaric B., Gahegan M. Ontology use for semantic e-Science. *Semantic Web*, 2010, v. 1, pp. 149-153.
7. Knowledge Organization Systems and Semantic Technology. International Atomic Energy Agency. Nuclear Knowledge Management (NKM). Available at: [www.iaea.org/nuclearenergy/nuclearknowledge/](http://www.iaea.org/nuclearenergy/nuclearknowledge/) (accessed Aug 10, 2018).
8. Erkimbaev A.O., Zitserman V.Y., Kobzev G.A., Kosinov A.V. Associating the ontologies with databases by properties of substances and materials. *Nauchno-Tekhnicheskaya Informatsiya. Ser. 2*, 2015, no. 12, pp. 1-6 (in Russian).
9. Chusov I.A., Kirillov P.L., Bogoslovskaya G.P., Yunusov L.K., Obysov N.A., Novikov G.E., Pronyaev V.G., Erkimbaev A.O., Zitserman V.Yu., Kobzev G.A., Trachtengerts M.S, Fokin L.R. Development of the system of reactor thermophysical data on the basis of ontological modeling. *Journal of Physics Conference Series*, 2017, v. 891, #012172.
10. Web Portal THERMOPHYSICS.RU. Available at: <http://www.thermophysics.ru/RTO/index.html> (accessed Aug 10, 2018) (in Russian).
11. *Thermophysical Properties of Materials for Nuclear Engineering*. Tutorial for students of specialty «Nuclear Power Plants». Ed. by Prof. P.L. Kirillov. Obninsk. IATE Publ., 2006, 182 p.
12. Mishra S., Kaity S., Banerjee J., Nandi C., Dey G.K., Khan K.B. U-PuO<sub>2</sub>, U-PuC, U-PuN

- cermet fuel for fast reactor. *Journal of Nuclear Materials*, 2018, v. 499, pp. 272-283.
13. Risovany V.D., Zakharov A.V., Muraleva E.M. *New advanced absorbing materials for thermal reactors. VANT. Ser.: Fizika radiatsionnyh povrezhdeniy i radiatsionnoe materialovedenie*, 2005, no. 3, pp. 87-93 (in Russian).
14. Erkimbaev A.O., Zitserman V.Y., Kobzev G.A, Fokin L.R. The logical structure of physicochemical data: Problems of numerical data standardization and exchange. *Russian Journal of Physical Chemistry A*, 2008, v. 82, no. 1, pp. 15-25.
15. 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, pp. 166-187.
16. Kogalovsky M.R. Ontology-based data access systems // *Programming and Computer Software*, 2012, v. 38, no 4, pp. 167-182.
17. Ashino T. Materials ontology: an infrastructure for exchanging materials information and knowledge. *Data Science Journal*, 2010, v. 9, pp. 54-61.
18. Lin L, Austin T, Ren W. Interoperability of materials database systems in support of nuclear energy development and potential applications for fuel cell material selection. *Materials Performance and Characterization*, 2015, v. 4, no. 1, pp. 115-130.
19. Erkimbaev A.O., Zitserman V.Yu., Kobzev G.A., Kosinov A.V. Standardization of Storage and Retrieval of Semi-structured Thermophysical Data in JSON-documents Associated with the Ontology. Proc. of the XIX International Conference «Data Analytics and Management in Data Intensive Domains» (DAMDID/RCDL'2017), Moscow, Russia, October 10–13, 2017. Available at: <http://ceur-ws.org/Vol-2022/paper36.pdf> (accessed Aug 10, 2018).

#### Authors

Chusov Igor Aleksandrovich, Head of Department, Professor, Dr. Sci. (Engineering)

E-mail: [igrch@mail.ru](mailto:igrch@mail.ru)

Kirillov Pavel Leonidovich, Principal Science Researcher, Professor, Dr. Sci. (Engineering)

E-mail: [plkirillov@gmail.com](mailto:plkirillov@gmail.com)

Pronyaev Vladimir Grigorevich, Chief Specialist, Cand. Sci. (Phys.-Math.)

E-mail: [vgpronyaev@yandex.ru](mailto:vgpronyaev@yandex.ru)

Erkimbaev Adilbek Omirbekovich, Head of Laboratory, Cand. Sci. (Engineering)

E-mail: [adilbek@ihed.ras.ru](mailto:adilbek@ihed.ras.ru)

Zitserman Vladimir Yurievich, Leading Researcher, Cand. Sci. (Phys.-Math.)

E-mail: [vz1941@mail.ru](mailto:vz1941@mail.ru)

Kobzev Georgy Anatolievich, Principal Science Researcher, Professor, Dr. Sci. (Phys.-Math.),

E-mail: [kobzev@oivtran.ru](mailto:kobzev@oivtran.ru)

Fokin Lev Ruvimovich, Principal Science Researcher, Dr. Sci. (Engineering)

E-mail: [lfokin@mail.ru](mailto:lfokin@mail.ru)