

УДК 004.56,544.01,544.31

ОПЫТ СОЗДАНИЯ ТЕПЛОФИЗИЧЕСКИХ БАЗ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ СОВРЕМЕННЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ (обзор)

© 2020 г. Г. В. Белов^{1,2}, А. О. Еркимбаев¹, В. Ю. Зицерман¹,
Г. А. Кобзев¹, И. В. Морозов^{1,3,*}

¹ФГБУН Объединенный институт высоких температур РАН, Москва, Россия

²Московский государственный университет им. М.В. Ломоносова,
химический факультет, Москва, Россия

³Московский физико-технический институт, Долгопрудный, Россия

*E-mail: morozov@ihed.ras.ru

Поступил в редакцию 04.03.2020 г.

После доработки 23.03.2020 г.

Принят к публикации 30.03.2020 г.

Обзор суммирует результаты многолетних работ, проведенных в Термоцентре им. В.П. Глушко по созданию тематических баз данных и использованию новых информационных технологий, обеспечивающих интеграцию разнородных компьютерных ресурсов. Детально рассмотрены общие принципы формирования термодинамической базы данных ИВТАНТЕРМО и последние результаты по расширению ее фонда и функциональных возможностей. Дано описание теплофизической базы данных ТЕРМАЛЬ, включающей библиографические данные по широкому кругу теплофизических, оптических, электрических и других физических свойств. Рассмотрена реализация проекта ее обновления и расширения тематики. Изучены возможности новых информационных технологий для решения актуальных проблем интеграции разнородных ресурсов (баз данных, текстовых документов, электронных таблиц, графиков и данных в специализированных форматах) в рамках единой инфраструктуры. Показано, что в качестве наиболее эффективного средства систематизации и поиска может быть использовано онтологическое моделирование как средство организации гибкой структуры данных, характерных для веществ и материалов, свойства которых зависят от вида образца, технологии изготовления, влияния среды и т.п.

DOI: 10.31857/S0040364420040018

ОГЛАВЛЕНИЕ

Введение

1. Термодинамическая база данных ИВТАНТЕРМО

2. Документальная база данных по теплофизическим свойствам ТЕРМАЛЬ

3. Технология проектирования базы данных широкого профиля

4. Онтология теплофизических данных

5. Интеграция данных с использованием онтологий

Заключение

Список литературы

ВВЕДЕНИЕ

Систематизация данных — традиционное направление работ в теплофизике. Уже с конца XIX—начала XX веков термодинамические и теплофизические свойства вещества широко представлены в знаменитых справочных сериях:

Landolt–Börnstein (данные для физики и инженерии), Beilstein's Handbook of Organic Chemistry, Gmelin Handbook of Inorganic and Organometallic Chemistry [1]. Современный этап исследований характеризуется экспоненциальным ростом масштабов справочно-аналитической деятельности и выделением ее в самостоятельное направление исследований. Возникло даже специальное понятие о data-intensive science (дисциплина с интенсивным использованием данных) [2], в практике которой доминируют все стадии работы с численными данными: компиляция “сырых” данных, их долгосрочное хранение в базах данных (БД) с последующими обработкой, анализом и распространением рекомендованных данных.

Полностью отвечая основным критериям интенсивного использования данных, теплофизика как предметная область имеет некоторую специфику:

— исключительно высокий масштаб производства и потребления данных практически во всех отраслях промышленности;

– широкая и давняя распространенность БД и информационно-вычислительных средств при различных форматах и структуре данных;

– совместное использование БД и вычислительных приложений, например, для расчетов теплообмена, химического и фазового равновесия.

Теплофизические данные, характеризующие сжимаемость, энергетические и транспортные свойства вещества, представлены в мировых БД для широчайшего круга объектов: чистых соединений и растворов, органических и неорганических веществ, наноструктур и материалов, характеризующихся технологией изготовления. Для каждого из классов существуют свои физические модели, связанные с ними словари понятий и логические структуры данных. Существует обширная литература с описанием многочисленных БД по теплофизическим свойствам веществ, см., например, [3–6], а также специальный поисковый инструмент [7] со сведениями о доступных ресурсах по свойствам неорганических веществ.

В то же время резкий рост объемов данных при многообразии их форматов и моделей привел к тому, что действующая инфраструктура оказалась неспособной обеспечить эффективную организацию рабочего процесса. Среди наиболее серьезных препятствий чаще всего выступают ограниченность реляционной модели¹ и проблемы интеграции разнородных ресурсов. Другая проблема связана с неоднородностью источников данных, использующих разные форматы, модели и семантику (смысловое содержание данных), что затрудняет или даже исключает совместное использование разных ресурсов в задачах обработки данных. Осознание в научном сообществе перечисленных проблем стимулировало постановку масштабных исследований [9–12] с целью перестройки всей инфраструктуры научных данных с ориентацией на тенденции роста их объема при усложнении структуры и задач обработки. Интересно отметить, что потребность в интеграции термодинамических данных возникла, по видимому, задолго до того времени, когда эта проблема приобрела актуальность для информационной технологии и стимулировала создание ряда стандартов обмена данными [13].

В обзоре суммированы результаты многолетних исследований, проведенных сотрудниками Термоцентра им. В.П. Глушко и Теплофизического центра (ТФЦ) ОИВТ РАН, объединенных к настоящему времени в лабораторию теплофизических баз данных. Рассматривается опыт создания информационных систем по свойствам веществ и материалов, а также разработки и применения новых технологий, ориентированных в

основном на создание интегрированных систем хранения разнородных данных, отличающихся форматом и структурой данных. Сформулированы главные принципы формирования ее контента, обеспечивающие высокий уровень достоверности и согласованности данных. Представлены также последние результаты по обновлению фонда и расширению функциональных возможностей информационной системы.

Дано описание теплофизической БД ТЕРМАЛЬ, одного из старейших информационных ресурсов, функционирующих в ОИВТ РАН с 1974 г. [14]. Наряду с теплофизическими, БД ТЕРМАЛЬ способна хранить и классифицировать библиографические данные по широкому комплексу физических характеристик вещества: электрических, оптических, микросвойств и пр. Обсуждается реализация проекта обновления БД, связанного с преодолением наложенных при ее создании тематических ограничений: преимущественная ориентация на свойства неорганических веществ, исключение объектов, зависящих от технологии производства и факторов влияния (облучение, примеси, внешние поля), отказ от включения данных по свойствам растворов, сплавов и природных минералов.

В обзоре рассматривается новая технология систематизации данных по свойствам веществ, получившая название баз данных “широкого профиля” [15], пригодных для хранения, поиска и обработки данных по свойствам разнородных объектов: веществ, материалов, атомно-молекулярных и наноструктур. Их задача – обеспечить способность к настройке на предметную область с характерной для нее спецификой свойств, в частности способность учитывать такие факторы, как влияние технологии, предыстории и конфигурации образца, среды и т.п. Использование подобной технологии рассмотрено на примере разработки БД по свойствам и применению углеродных наноструктур [16, 17].

Изучаются возможности онтологического моделирования, одного из наиболее перспективных средств организации данных и знаний, широко применяемого в последнее время в естественно-научных и инженерных дисциплинах. Излагаются результаты работ по созданию первого образца онтологической модели предметной области “Теплофизические свойства веществ” [18, 19], выполненных совместно с ВЦ РАН. Описывается последующее развитие этой деятельности, которое позволило создать эффективную технологию интеграции разнородных теплофизических (и родственных) данных.

¹ Описание терминов, относящихся к теории баз данных, представлено в энциклопедии [8].

1. ТЕРМОДИНАМИЧЕСКАЯ БАЗА ДАННЫХ ИВТАНТЕРМО

В 1960–1980-е гг. два больших проекта в СССР и США по созданию системы справочных данных термодинамических величин привели к разработке справочников [20, 21]. Созданные в рамках этих проектов системы справочных данных послужили надежной основой для развития ракетно-космической отрасли в обеих странах. В дальнейшем информация о термодинамических свойствах индивидуальных веществ использовалась для исследования плазмохимических и металлургических процессов, процессов горения и взрыва, для создания моделей бинарных и тройных термодинамических систем и т.д.

История создания справочников по термодинамическим и термохимическим свойствам веществ в США описана в работах [22, 23], а созданию аналогичных справочников в СССР посвящена статья [24]. В этой же работе отмечается, что наиболее эффективным способом хранения, обновления и получения (генерации) информации о термодинамических свойствах веществ является создание программного комплекса, сопряженного с БД по термодинамическим свойствам веществ, включающей первичные экспериментальные данные, расчетные величины и библиографическую информацию. В [25] перечислены основные требования к БД по термодинамическим свойствам веществ. В частности, Л.В. Гурвич считал обязательной процедуру критического анализа всех первичных данных, позволяющих получить наиболее достоверные сведения, причем полученные значения в БД должны быть согласованы в рамках законов термодинамики и содержать оценки погрешностей. Описание процедуры анализа первичных данных, а при возможности и сами первичные данные вместе с описанием их источников также целесообразно хранить в БД.

Базы данных, удовлетворяющие указанным требованиям, принято называть “компетентно-подтвержденными”, что отличает их от “компильтивных” БД, информация в которых собрана без необходимой верификации, не содержит описания источников, методов обработки данных и оценки погрешностей.

Работы по созданию и эксплуатации компетентно-подтвержденной БД по термодинамическим свойствам индивидуальных веществ, которая получила название “ИВТАНТЕРМО”, ведутся в Термоцентре им. В.П. Глушко с 1980-х гг. параллельно с работами по накоплению и обновлению данных. На первом этапе этих работ был создан программный комплекс, включающий БД и предназначенный для работы на компьютерной системе HP-3000. Позже был разработан программный комплекс для персональных компью-

теров с операционной системой Windows для моделирования равновесных состояний сложных термодинамических систем, сопряженный с БД ИВТАНТЕРМО [26].

Кроме ИВТАНТЕРМО, существует еще несколько БД, которые нельзя отнести к чисто компильтивным. К их числу относится одна из наиболее известных в США баз данных по термодинамическим свойствам веществ NASA [27, 28]. К сожалению, эта БД не сопровождается подробным описанием способов получения информации и содержит лишь источники информации.

В Аргоннской национальной лаборатории США в рамках проекта “Active Thermochemical Tables” (ATCT) создана БД, задачей которой является получение наиболее точных согласованных данных о термохимических свойствах веществ [29, 30]. Кроме рекомендованных значений энтальпий образования и их погрешностей, в БД содержатся первичные экспериментальные и расчетные данные, а также сведения об источниках информации. Авторы проекта разработали и реализовали интересный метод согласования термохимических величин [31, 32], предполагающий построение матрицы, в которой отображены все доступные термохимические взаимосвязи между веществами (энтальпии и энергии Гиббса реакций, константы равновесия, энергии ионизации и т.д.). Все величины перед включением в БД критически оценивались.

Кроме того, в мире существует несколько исследовательских центров, в которых ведется разработка термодинамических БД и сопутствующего программного обеспечения [33–37]. Некоторые из них объединены в международные организации, такие как Scientific Group Thermodata Europe (SGTE) [38]. Основная часть информации в данных БД не является открытой и предоставляется за плату. Кроме БД, охватывающих широкий спектр веществ, существуют специализированные БД. Например, БД ThermoChimie [39] содержит термодинамические свойства около 2300 веществ, имеющих отношение к проблемам утилизации и хранения ядерных отходов.

Анализ публикаций свидетельствует о том, что наиболее интенсивно сегодня развиваются БД, с использованием которых можно анализировать фазовые диаграммы бинарных, тройных систем, а также фазовые диаграммы большей размерности [36, 40]. Однако в основе термодинамической модели системы все равно лежит информация о свойствах индивидуальных веществ, поэтому задачи уточнения, пополнения, хранения и распространения надежных справочных данных для индивидуальных веществ и сегодня остаются крайне важными. Актуальность этих задач подтверждается еще и тем, что регулярно появляются новые экспериментальные и теоретические данные о

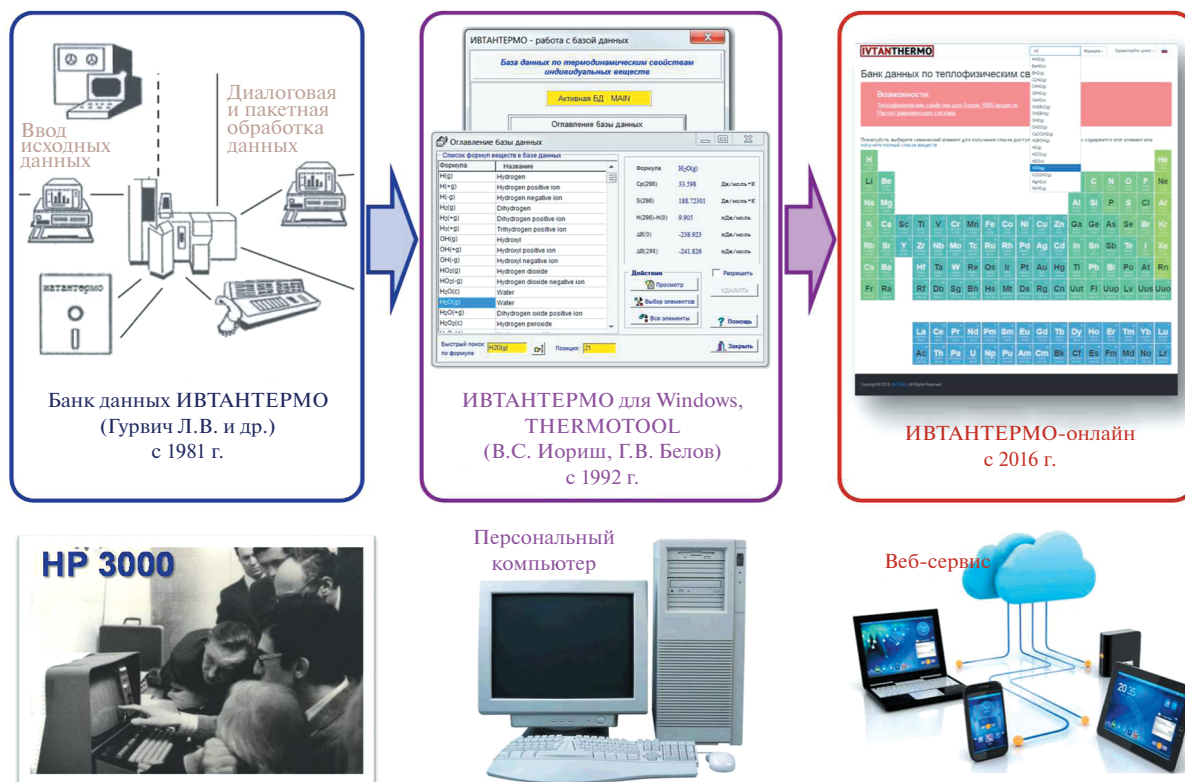


Рис. 1. Иллюстрация развития информационной системы "ИВТАНТЕРМО".

свойствах индивидуальных веществ, а также возникает потребность в расширении диапазона параметров.

С 2016 г. в ОИВТ РАН начата работа по созданию обновленной версии "ИВТАНТЕРМО", получившей название "ИВТАНТЕРМО-онлайн" [41, 42]. Эта система обладает новым расширяемым дизайном БД, клиент-серверной архитектурой, удобным веб-интерфейсом и рядом новых функций для онлайн-сервисов и автономной обработки данных (рис. 1). Отличительными особенностями "ИВТАНТЕРМО-онлайн" являются:

- сопровождение наборов данных унифицированными подробными описаниями источников и методов получения этих данных, включая ссылки на литературу, а в некоторых случаях и копии источников;
- хранение исходных экспериментальных или расчетных данных, использованных значений констант с оценками их погрешностей;
- регулярные обновления данных с сохранением истории всех изменений;
- двухступенчатая система внесения изменений с подтверждением со стороны экспертного совета;

– современный интерактивный веб-интерфейс с возможностью удаленной работы пользователей и экспертов;

– дополнение БД программами расчета и анализа термодинамических функций в конденсированной и газовой фазах, имеющими веб-интерфейс, совместимый с основным интерфейсом системы.

Для обработки экспериментальных данных с целью обновления и расширения БД "ИВТАНТЕРМО" необходима также разработка специализированного программного обеспечения. Информация о термодинамических свойствах веществ в БД хранится, как правило, в виде набора коэффициентов аппроксимирующих функций, что упрощает использование их в алгоритмах термодинамического моделирования [43, 44]. При этом существуют различия в обработке данных с получением аппроксимаций для конденсированных и газообразных фаз. В то время как термодинамические свойства газов обычно рассчитываются из молекулярных констант, инкремент энтальпии или теплоемкость конденсированного вещества измеряются непосредственно в эксперименте или определяются с помощью компьютерного моделирования. Из-за фазовых переходов в твердых телах соответствующий температурный интервал естественным образом разбивается на относи-

тельно небольшие интервалы, где термодинамические функции непрерывны и могут быть аппроксимированы одним полиномом. Напротив, для газовой фазы диапазон температур может достигать десятков тысяч кельвинов, а термодинамические функции имеют довольно сложный вид.

В процессе развития БД “ИВТАНТЕРМО” был разработан специальный алгоритм автоматического разбиения диапазона температур на интервалы для аппроксимации термодинамических свойств веществ в газовой фазе, выполняющий согласование коэффициентов полиномов на границах интервалов и обеспечивающий непрерывность и гладкость термодинамических функций при заданной максимальной погрешности аппроксимации [45].

Наряду с классическим методом расчета термодинамических свойств веществ в газовой фазе на основе молекулярных постоянных, использованном, в частности, при создании справочника [20], в последнее время активно развиваются методы квантовой химии, позволяющие определить межатомные потенциалы взаимодействия из “первых принципов” и на их основе рассчитать термодинамические функции. Например, в работах [46–48] развита методика определения термодинамических свойств двухатомных молекул в газовой фазе, основанная на численном решении стационарного уравнения Шредингера для полученного с помощью методов квантовой химии межатомного потенциала взаимодействия. Далее на основе колебательно-вращательного спектра молекулы определяется статистическая сумма. При этом рассматривается основное и несколько возбужденных состояний молекулы, вносящих наиболее существенный вклад в статистическую сумму в выбранном диапазоне температур. Эта методика имеет наибольшее значение для веществ, по свойствам которых существует недостаточное количество достоверных экспериментальных данных, а также для тех случаев, когда приближение “гармонический осциллятор—жесткий ротатор” или модель потенциала Морзе неприменимы.

Одновременно с этим развиваются и алгоритмы обработки экспериментальных данных по свойствам веществ в конденсированной фазе. Такие данные представлены в литературе, как правило, в виде таблиц, содержащих информацию о зависимости изобарной теплоемкости или инкремента энтальпии от температуры. В связи с этим возникает задача совместной обработки двух типов данных и объединения результатов различных экспериментов с целью получения гладкой температурной зависимости термодинамических функций для всей исследуемой области. Здесь также находят применение методы ро-

бастной (устойчивой к влиянию выбросов) оценки для исключения из рассмотрения явно недостоверных данных. Для решения этих задач в ОИВТ РАН была разработана компьютерная программа CondensedThermoFit с развитым пользовательским интерфейсом [49], с помощью которой удалось уточнить информацию в БД ИВТАНТЕРМО для ряда практически значимых химических соединений [50, 51].

В целом переход на веб-систему ИВТАНТЕРМО-онлайн и дополнение ее описанными выше алгоритмами обработки экспериментальных данных позволит создать информационную систему нового уровня, пригодную для совместной работы экспертов по уточнению и дополнению термодинамических данных по свойствам индивидуальных веществ. Ожидается, что это значительно ускорит процесс обновления рекомендованных данных в БД ИВТАНТЕРМО и повысит востребованность этой системы в России и за рубежом.

2. ДОКУМЕНТАЛЬНАЯ БАЗА ДАННЫХ ПО ТЕПЛОФИЗИЧЕСКИМ СВОЙСТВАМ ТЕРМАЛЬ

БД ТЕРМАЛЬ разработана и введена в эксплуатацию в середине 70-х годов прошлого века в ходе создания в ИВТАН организационной структуры под названием “Научно-информационный центр по теплофизическим свойствам чистых веществ” (Теплофизический центр, ТФЦ). Решение о его создании было принято президиумом АН при поддержке национального комитета по численным данным (CODATA), а с апреля 1974 г. ТФЦ приобретает статус государственного центра по данной тематике, входящего в состав Государственной службы стандартных справочных данных (ГСССД). Основные идеи и направления работы Центра, как и требования к БД ТЕРМАЛЬ, выдвинуты членом-корр. РАН проф. Э.Э. Шпильрайном [14].

В соответствии с исходной концепцией ТЕРМАЛЬ является документальной БД, предназначенной для хранения сведений о публикациях по теплофизическим и связанным с ними механическим, электрическим, оптическим свойствам неорганических и простейших органических веществ. С самого начала предполагалось, что фонды ТЕРМАЛЬ должны преимущественно содержать ссылки на данные об элементах, неорганических соединениях, включая нестехиометрические соединения и фазы переменного состава, а также простейшие и наиболее распространенные органические соединения (табл. 1).

Наиболее серьезное ограничение тематики справочного фонда состояло в исключении почти всех источников, содержащих данные по смесям, растворам, сплавам и природным минералам. Сделано несколько исключений из этого правила

Таблица 1. Органические вещества, включаемые в БД ТЕРМАЛЬ

Класс соединений		Стехиометрическая формула и название
Углеводороды	алканы	CH ₄ (метан), C ₂ H ₆ (этан), C ₃ H ₈ (пропан), C ₄ H ₁₀ (бутан)
	алкены	C ₂ H ₄ (этилен), C ₄ H ₈ (бутилен)
	алкины	C ₂ H ₂ (ацетилен)
Галогензамещенные углеводороды		Фреоны*
Оксиды углерода		CO ₂ (двуокись), CO (моноокись), C ₃ O ₂ (недоокись)
Сульфиды углерода		CS (моносulfид), CS ₂ (дисulfид), COS (карбонилсульфид), C ₃ S ₂ (субсульфид)
Карбиды элементов		
Карбонилы металлов		M _n (CO) _m , пример Ni(CO) ₄ , Fe ₂ (CO) ₉ и т.п.
Цианистые соединения		(CN) ₂ (цианоген), HCN (цианистоводородная кислота и ее соли с общей формулой M _n (CN) _m)
Кислоты		H ₂ CO ₃ (угольная), H ₂ C ₂ O ₄ (щавелевая), HCHO ₂ (муравьиная), HC ₂ H ₃ O ₂ (уксусная) и все их соли
		CH ₃ OH (метилловый спирт), C ₂ H ₅ OH (этиловый спирт), (CH ₃) ₃ CO (ацетон)

* https://en.wikipedia.org/wiki/List_of_refrigerants.

в отношении активно изучаемых композиций, а именно, сплавов щелочных металлов, системы Li–LiH (включая изотопные модификации), воздуха и смесей его компонентов (N₂, O₂, CO₂, Ar).

Желая ограничить фонд справочных данных результатами академических исследований, разработчики ТЕРМАЛЬ [14] полностью отказались от включения веществ со свойствами, зависящими не только от параметров состояния, но и от множества факторов влияния, среди которых: технология получения (характер термообработки, легирование), конфигурация образца и состояние поверхности, влияние среды и ряд других. Тем самым тематика БД была ориентирована только на вещества, свойства которых однозначно определяет стехиометрическая формула. Прочие объекты были отнесены к производственным материалам, данные для которых из-за множества плохо контролируемых факторов сочли неоправданным включать в фонд справочных данных.

Дальнейшая (почти полувековая) практика эксплуатации БД ТЕРМАЛЬ выявила недостатки столь жесткого разделения. Для множества чистых веществ выявлена четкая зависимость многих свойств, например излучательной способности, поверхностного натяжения или теплопроводности от состояния образца, включая его поверхность, состав окружающей среды, пористость, дефекты кристаллической структуры и пр. Помимо этого, в современных исследованиях достаточно часто приходится рассматривать веще-

ства в наноразмерном состоянии, когда свойства определяются размерным эффектом, топологией наноструктуры, условиями синтеза и другими факторами [16, 17].

Необходимость хранения и систематизации данных сложной структуры, не охваченных в исходной версии БД ТЕРМАЛЬ, послужила основанием для новой концепции – так называемых “БД широкого профиля” [15], пригодных для хранения, поиска и обработки данных по свойствам разнородных объектов и в условиях действия разнородных факторов влияния, о чем идет речь в следующем разделе. Ее разработка позволила приступить к решению задачи о реконструкции БД ТЕРМАЛЬ с целью снять или ослабить наложенные ограничения.

ТЕРМАЛЬ сформирована и поддерживается как документальная БД, т.е. содержащая данные об источнике (статья, автореферат диссертации, таблицы рекомендованных данных и т.д.), а не о самих свойствах. Этот принцип полностью оправдан с учетом неограниченного объема веществ, включаемых в тематику БД, а также наличием фонда самих источников, указатель на каждый из которых содержится в записи. За период эксплуатации в фондах ТЕРМАЛЬ накоплено около 80 тыс. документов, отражающих публикации по теплофизическим свойствам. В текущей версии данной базы реализована удобная для пользователей возможность получения полного текста до-

Таблица 2. Основные поля, используемые для индексирования источника

Тэг	Название поля	Тэг	Название поля
001	Авторы	015	Химическая формула
002	Название русское	016	Класс вещества
003	Название на языке оригинала	017	Номер документа/источника
004	Источник	018	Вид источника (статья, отчет и т.п.)
005	Конференция	019	Язык источника
006	Реферат	020	Год публикации
007	Место нахождения источника	021	Референт
008	Свойства	022	Нижняя температура изученного диапазона
009	Носитель	023	Верхняя температура изученного диапазона
010	Фаза	024	Нижнее давление изученного диапазона
011	Фазовый переход	025	Верхнее давление изученного диапазона
012	Тип свойства	026	Ссылка на электронный источник (файл) или сетевой документ
013	Физическое поле	027	Организация
014	Вид исследования (эксперимент, теория, моделирование)		

кумента в электронном виде непосредственно в ходе поиска.

Каждая запись или документ, из которых состоит фонд БД, представляет собой большую группу именованных полей (табл. 2).

Поля включают формальные сведения о документе (поля 017, 018, 021, 026), исчерпывающие библиографические данные и, наконец, элементы фактографических данных из источника, непосредственно относящиеся к теплофизическим свойствам. Из последней группы полей наиболее значимы при поиске и отборе документов поля 015 и 016 (формула и класс веществ), 008 и 012 (свойства и их тип) и 010 и 011 (сведения о фазе и фазовом переходе).

Управление БД ТЕРМАЛЬ осуществляется посредством программной оболочки CDS/ISIS, разработанной ЮНЕСКО для хранения документальной информации [52]. В то время как реляционные системы управления БД (СУБД) ориентированы на работу с БД, имеющими табличные формы, ISIS в своей основе имеет свободно-текстовый формат, наиболее адекватно представляющий структурированные нечисловые данные в форме записей произвольной длины. Библиографическая информация и описания произвольных документов могут быть представлены наилучшим образом именно в такой форме, что соответствует рекомендациям стандарта ISO-2709, предназначенного для обмена библиографической информацией. В своей основе СУБД ISIS построена на использовании ISO-2709: внутренняя организация основного файла записей, режимы импорта и экспорта. При этом использована версия ISIS для

операционной системы Windows [53], для адаптации которой потребовалось усовершенствовать локализацию программы на русский язык и разработать адекватную систему полей [54].

При поиске веществ помимо химической формулы используется специальная система кодирования, которая присваивает каждому классу веществ один или несколько кодов в соответствии с элементным составом и наличием функциональных групп. Например, такое соединение как NaCl должно быть отнесено к двум классам: соединения Na (код 101) и хлориды (код 1260). Для идентификации соединения оба кода должны быть занесены в поле 016 (табл. 2).

Все свойства, отмечаемые в документах БД, разбиты на десять типов, которые должны быть указаны в поле 012 (табл. 2): термические, калорические, механические, переносные/транспортные, электрические, магнитные, оптические, перекрестные, микросвойства и константы. К микросвойствам отнесены характеристики атомно-молекулярного уровня, необходимые для расчета термодинамических и транспортных свойств, например, потенциал взаимодействия, интеграл столкновений, поляризуемость молекулы и т.д.

Совокупность терминов (так называемых дескрипторов), определяющих физические свойства каждого из десяти типов, наряду с дескрипторами, обозначающими силовые поля, носители свойств, фазовые переходы и состояния, образует тезаурус, используемый как для отражения в документе основных понятий, затронутых в источнике, так и при организации машинного поиска. Общее число дескрипторов, отражающих свой-

ства и состояния, достигает примерно 300, что заметно превышает объем многих из традиционно используемых словарей, например принятого в стандарте обмена ThermoML [13] или в перечне терминов, определяющих свойства металлов и сплавов [55].

Тезаурус построен как последовательность дескрипторных статей, состоящих из заглавного дескриптора, списка дескрипторов и недескрипторов, семантически связанных с ним. Установлены три вида семантических отношений: синонимии (включая условную синонимию); подчинения; ассоциации. Подобная лингвистическая структура, по сути, является прообразом онтологии, хотя в период создания ТЕРМАЛЬ не было соответствующего опыта использования онтологий в научной информатике. Детальная разработка тезауруса с элементами семантических отношений облегчила поставленную задачу по реконструкции БД ТЕРМАЛЬ, используя новую технологию управления данными, основанную на онтологии (см. раздел 5). Конечная цель этой реконструкции – расширить справочный фонд, сняв указанные выше ограничения тематики: только индивидуальные вещества с исключением данных, отражающих проявление дополнительных факторов (технология, среда, состояние образца и т.п.).

3. ТЕХНОЛОГИЯ ПРОЕКТИРОВАНИЯ БАЗЫ ДАННЫХ ШИРОКОГО ПРОФИЛЯ

Концепция “базы данных широкого профиля” выдвинута авторами в статье [15]. Под этим термином понимается БД, решающая задачи хранения и обработки данных о свойствах веществ и материалов произвольного состава и структуры, равно как и атомно-молекулярных единиц (молекулы, кластеры, наноразмерные структуры типа углеродных нанотрубок и пр.), а также адекватного учета совокупности факторов, определяющих свойства объекта (структура и конфигурация образца, технология изготовления, факторы влияния и т.п.). Это существенным образом выделяет ее на фоне БД, описанных в двух предыдущих разделах. Отличительные черты подобных БД: отсутствие традиционных ограничений, т.е. жесткая ориентация на класс веществ; простота и однозначность идентификации веществ; фиксированная структура записи. Потребность расширения функциональных возможностей БД обусловлена:

- вовлечением в практику исследований таких объектов, как композиты, наноструктуры, фуллерены, вещества в метастабильном состоянии и т.п.;

- использованием физических моделей, включающих, наряду с теплофизическими, данные другой природы (реологические, эксплуатационные);

- проявлением в свойствах веществ влияния среды, геометрии, “предыстории” образца и т.п.

Изменчивость структуры БД многообразна и проявляется в различиях объема и типа данных; в появлении новых и/или утрате смысла прежних характеристик; в изменении правил идентификации вещества. Столь же многообразны и причины, определяющие вариации структуры теплофизических данных: различия физических свойств отдельных веществ; изменения в структуре данных при смене модели (например, формы уравнения состояния); зависимость описания от диапазона параметров; различия в способах представления данных; расширение требований к идентификации соединения и ряд других. В итоге для адекватной передачи особенностей отдельных веществ наборы данных должны варьироваться как по объему, так и по внутренней структуре.

Данные с варьируемой структурой, потребность в которых возникает в различных сферах деятельности, получили название полуструктурированных данных [8]. Первой и важнейшей из причин изменчивости структуры данных является изменчивость свойств, когда какое-то из веществ настолько выделяется из общего ряда, что требуется сменить модель, тип описания и выбор характерных параметров. Простейший и самый известный пример – пары щелочных металлов, содержащие мономеры и димеры, что заметно усложняет описание. Например, вязкость становится функцией давления, а не только температуры, как следует из кинетической теории газов для не реагирующих газов. Поэтому даже при малых давлениях объем данных должен быть расширен включением сведений о константе димеризации или в другом подходе вторым групповым интегралом. Такое же изменение свойств и, как следствие, структуры данных характерно для паров C, S, Se, P, As и других веществ со сложным составом пара [20]. При описании в БД неидеальных свойств вещества ассоциация молекул сказывается на логической структуре за счет появления новых данных в виде констант равновесия, непостоянства молекулярной массы и т.п.

Многообразны причины непостоянства структуры данных для твердой фазы: нестехиометрия состава; диаграмма состояния сложного типа, в том числе и с фрагментами неравновесных составов; полупрозрачность вещества и т.п. Например, из ряда оксидов урана UO_{2+x} только UO_2 характеризуется обычными точками плавления и кипения, линией сосуществования жидкость–пар и др. [56]. Уже для $UO_{2.25}$ или $UO_{2.67}$ (U_4O_9 , U_3O_8) характеристика ограничивается набором свойств для твердой фазы. В целом же характеристика любого из индивидуальных веществ является неполной без дополнительной информации по фазовой диаграмме, что резко расширяет требуемый объ-

ем данных. Из простых соединений SiO_2 дает яркий пример сложной структуры данных за счет множества полиморфных модификаций, для которых на диаграмме P – T должны быть заданы границы, а данные по калорическим свойствам должны дополняться и согласоваться с объемными характеристиками (коэффициент объемного расширения, модуль сжимаемости и др.). Публикация [57], содержащая согласованные термодинамические данные для SiO_2 , показывает, насколько заметно увеличивается их объем при сопоставлении с веществами, не имеющими столь богатого полиморфизма.

Изменчивость структуры данных для твердого состояния может иметь и другие причины: гистерезис свойств в процессах нагревания–охлаждения, размытость фазовых переходов, неравновесные процессы и метастабильные состояния, реологические особенности и т.п. Даже оптические свойства могут сказаться на структуре данных. Так, эффективная теплопроводность полупрозрачных веществ включает радиационную составляющую [58], что требует дополнительной информации о коэффициенте поглощения и геометрии образца.

В исходной публикации [15] приведено достаточно много примеров необходимости перехода к варьируемой структуре данных. Среди них изменение объема и структуры данных при высоких температурах или давлениях, проявление неравновесности (метастабильные фазы, гистерезис, стеклообразные модификации), что вводит зависимость данных от скорости процесса. При замедленной релаксации, например, вблизи критической точки, дисперсией по отношению к частоте внешнего воздействия обладают обычные транспортные свойства.

Можно выделить три главных фактора, определяющих необходимость в использовании гибкой структуры данных (т.е. полуструктурированных данных): 1) различия в номенклатуре свойств для различных типов веществ; 2) проявление в физических свойствах особенностей образца, а именно, его формы, размеров, технологии изготовления, наличия примесей и пр.; 3) проявление в физических свойствах так называемых факторов влияния, среди которых состояние внешней среды (ее состав, параметры состояния), механическая нагрузка, радиационные и химические факторы и т.п. Впервые необходимость учета многообразных факторов влияния как отдельной категории в данных о свойствах вещества признана в классификаторе [59], изданном еще в 1980 г. ГСССД.

Для широкого круга веществ (или широкого диапазона параметров) таблично-ориентированная структура данных является скорее исключением, чем правилом. Ее использование в реляци-

онных БД сопряжено с серьезными издержками: неоправданное сужение круга веществ, неадекватная передача их индивидуальных особенностей, проблемы с интеграцией данных из разных информационных систем.

Для эффективной работы с данными сложной и варьируемой структуры, в особенности необходимых при систематизации данных по наноструктурам, оказалось возможным использовать реляционные БД со специальной надстройкой, позволяющей по мере надобности усложнять логическую структуру [16, 60]. Эффективные возможности для создания такой надстройки заложены в БД PostgreSQL [61], предусматривающей возможность объявления собственных типов данных с достаточно сложной структурой. Одной из реализаций этой стратегии является введение типа COMPOSITE, понимаемого как массив, каждый элемент которого может иметь любой из допустимых типов, включая и вновь объявленный тип COMPOSITE. Это открывает принципиальную возможность заменить в ячейке атомарную единицу данных (число, строку) на целую структуру, сложность которой определяется декларированием переменной с приписанным типом COMPOSITE (например, массив, числовая таблица, таблица БД и т.п.). Преимущество такого подхода – тесная связь на уровне процедур подготовки и обработки данных со сложившейся в научном сообществе практикой работы с реляционными БД. На основе предложенной технологии создана тематическая БД по свойствам углеродных наноструктур [16, 62].

К настоящему времени концепция полуструктурированных данных [8, 15] нашла достаточно широкое применение. Пример ее использования дан в разработке платформы Citrination², где собрана информация для материалов всех классов. Специально предложенный для платформы формат PIF (Physical information file) [63] наилучшим образом приспособлен к данным сложной структуры и обеспечивает возможность подстройки хранилища данных к вариациям их структуры за счет общности, гибкости и легкости структурирования данных на стадии их загрузки. Другим примером является предложенная в NIST [55] система сбора данных по свойствам металлов и сплавов с детальным отражением особенностей образца и истории его подготовки к измерениям (примеси, атмосфера, тепловая и механическая обработка и др.).

4. ОНТОЛОГИЯ ТЕПЛОФИЗИЧЕСКИХ ДАННЫХ

Расширение функций, намеченное в концепции БД широкого профиля, естественным обра-

² <https://citrination.com/>

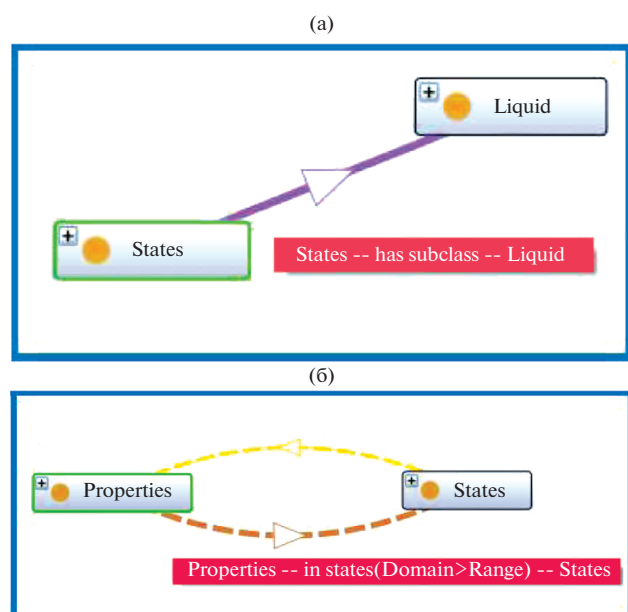


Рис. 2. Схематичное представление триплетов на примерах из онтологии “Теплофизические свойства веществ”.

зом заставило обратиться к новым средствам и технологиям работы с данными, в частности с технологией так называемой Semantic Web (семантической паутины), расширяющей возможности глобальной сети документов переходом к сети данных (Web of Data) и обеспечивающей связи между данными в документе, которые могут идентифицировать любые объекты, персоны или концепции. Выдвинутая создателем сети данных Тимом Бернерс-Ли концепция [64] предусматривала, что документы, представленные в сети, должны публиковаться в специальном формате, позволяющем программным агентам раскрывать их семантику, т.е. смысл или содержание. Роль явной и формальной спецификации семантики авторы [64] возлагали на онтологию, рассматривая ее как подлинное ядро нового подхода.

В общем смысле, онтология – это система понятий предметной области, которая представлена как набор сущностей, соединенных различными отношениями. Наиболее простое и точное ее определение дано Грубером [65]: “Онтология есть спецификация концептуализации”, т.е. стандартизованное и закодированное представление знаний в предметной области. При этом под концептуализацией предметной области понимается спецификация базовых понятий, отношений между ними и ограничений, а также создание соответствующих словарей. Под спецификацией подразумевается ее кодирование на одном из специальных языков, например OWL (Web Ontology Language).

Сам процесс кодирования – это запись логических утверждений о предметной области в форме так называемых триплетов: Субъект–Предикат–Объект, где все три элемента трактуются как ресурсы, и, следовательно, обладают уникальным адресом в информационной среде (URL – Uniform Resource Locator), например “<http://dataprop-subst.thermophysics.ru/dps24g.owl#Properties>”, а последний элемент (Объект) может быть представлен и “литералом” (т.е., символом, цифрой или текстом). Набор таких утверждений – триплетов о предметной области приводит к формированию “графа” – математической конструкции, состоящей из “узлов” и “ребер”. В качестве примеров показаны схемы триплетов для классов “States” и “Liquid” (рис. 2а), а также “Properties” и “States” (рис. 2б) из онтологии по теплофизическим свойствам веществ. В данном случае имеются следующие утверждения (запись приведена в пояснительной форме без употребления синтаксиса языка OWL):

1. “States – has subclass – Liquid” (класс “Агрегатное состояние” имеет подкласс “Жидкое состояние”);

2. “Properties – inStates – States” (экземпляр класса “Свойства” описывает свойства веществ или материалов в агрегатном состоянии, указанном в экземпляре класса “Агрегатное состояние”);

3. “States – inProperties – Properties” (объект, описываемый классом “Агрегатное состояние”, обладает свойствами, описываемыми экземпляром класса “Свойства”).

В приведенном примере утверждения 2 и 3 являются обратными (инверсными) по отношению друг к другу.

Совокупность триплетов, выражающих доступные для машинной обработки утверждения, представляет некоторый формат описания ресурсов под названием RDF (Resource Definition Framework), см. подробнее в [66].

По своим возможностям онтология значительно шире обычной таксономии (т.е. иерархически организованной системы классификации), так как поддерживает логические связи между понятиями, обусловленные спецификой предметной области. Добавляя к понятиям, представленным в виде таксономического словаря, семантику (смысл, правильное толкование) и логические соотношения, онтология описывает “знание” с возможностью его машинной интерпретации. В практическом плане онтология позволяет установить общепринятую в некотором сообществе трактовку понятий, исключить синонимии и неоднозначности, поддерживать эволюцию предметной области, связанную с расширением числа объектов и их характеристик, по сути решая на новом уровне задачи, поставленные при

создании БД широкого профиля (см. предыдущий раздел).

В целом же технология Semantic Web позволяет формировать из большого количества разнородных источников структурированной информации наборы научных данных и оперативно обновлять эти наборы, связывать данные из сформированного набора с данными из других открытых наборов, ориентируясь на их тематическое родство. Получить хорошее представление о концепциях Semantic Web и используемых при этом средствах можно, в частности, из отдельных глав обширного руководства [67] или из обзора, подводящего итоги внедрения этого подхода на 2018 г. [68]. Краткую справку читатель может получить в сетевой энциклопедии “Techopedia” [69].

Хотя задуманная идея Semantic Web относится к ресурсам любой тематики, естественнонаучные дисциплины образуют наиболее подходящую “площадку” для отработки ее возможностей. Понятийный багаж таких наук, как физика, химия, материаловедение и др. исходно уже достаточно формализован, чтобы лечь в основу онтологического описания. Тем самым открывается принципиальная возможность интеграции научных данных за счет спецификации содержимого разнородных источников [10, 66, 70]. Возможности новой технологии применительно к теплофизическим данным изучены в рамках совместного проекта ОИВТ РАН и Вычислительного центра (ВЦ) РАН³.

В работах [18, 19, 71, 72] впервые предложена упрощенная версия онтологии для предметной области “Теплофизические свойства веществ”. Эта версия формировалась при осознанном сужении круга объектов: исключались производственные “материалы”, смеси и растворы, игнорировалась зависимость свойств от давления. Последнее связано с возможностью пренебречь барической зависимостью конденсированной фазы, а для газовой фазы ограничиться оценкой вириальных коэффициентов, которые сами являются функциями температуры. На основании анализа литературы выбрана типовая модель предметной области, которая сводится к выделению тройки базовых понятий (вещество, свойство, состояние) и ряда дополнительных, которые характеризуют набор данных. Принимается, что каждый из наборов включает для одного вещества численные данные по нескольким свойствам, информацию по их неопределенности, источнику и т.п. Представление набора данных позволяет различать свойства в виде констант и в

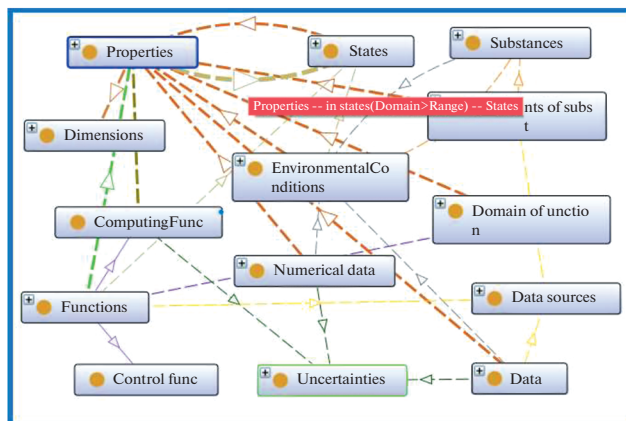


Рис. 3. Визуализация онтологии предметной области “Теплофизические свойства веществ”.

виде функций, заданных с помощью одномерных таблиц. Набор данных определяется указанием на имя вещества и физическое состояние, среди которых: три агрегатных состояния, три линии равновесия (кривые сублимации, плавления и насыщения), две характерные точки (тройная и критическая). Для твердого состояния возможно выделение отдельных фаз с указанием типа кристаллической решетки (кубической, тетрагональной и т.п.). Предусмотрена также возможность различных типов задания неопределенности: стандартная неопределенность, расширенная неопределенность (при указании уровня значимости), комбинированная неопределенность (с включением погрешности аргумента) и т.д. Подобная альтернативность принята и при включении сведений об источнике данных: один источник на весь набор, различные источники для каждого из свойств, различные источники для отдельных точек из набора.

Рассмотрим спецификацию концептов онтологии теплофизических данных. В качестве основных приняты такие классы, как Substances (вещества), States (состояния), Properties (свойства веществ), NumericalData (числовые значения свойств). Взаимосвязь этих классов показана на рис. 3, который представляет собой визуализацию онтологической модели. Показано лишь ограниченное количество классов онтологии и связей между ними.

Также на рис. 3 представлены классы Dimensions (размерности), Uncertainties (погрешности), EnvironmentConditions (внешние условия). Последний из классов определяет условия, которым соответствуют измеренные или оцененные данные: прежде всего, это варьируемый параметр состояния, а также некоторые параметры, характеризующие состояние среды – давление воздуха, влажность, содержание примесей и пр.

³ Проект РФФИ № 13-07-00218а. “Разработка онтологической модели предметной области “Теплофизические свойства веществ” и ее применение для целей интеграции научных информационных ресурсов”, рук. А.О. Еркимбаев. Проект выполнялся совместно с ВЦ РАН (В.А. Серебряков).

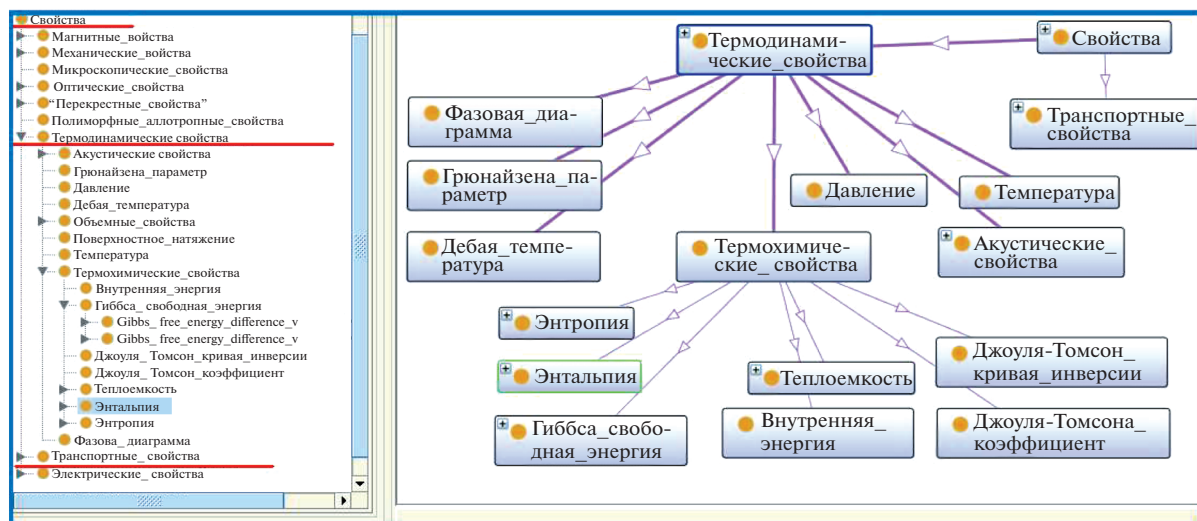


Рис. 4. Фрагмент иерархии классов по теплофизическим свойствам веществ, представленный в специализированном редакторе онтологий Protégé; левая часть – список классов в виде древовидной структуры, в правом окне – в виде графа.

Включенные в онтологию два класса (Publication, Data) расширяют представление о внешних источниках, описанное при концептуализации. Первый соответствует обычным источникам/библиографии, ссылки на которые должны быть включены в набор данных. Второй предполагает использование расположенных в сети справочников и словарей, например, с данными по молекулярному весу. Последняя группа классов детализирует и расширяет возможности представления свойств посредством аналитических функций. Два класса в этой группе (Functions, DomainOfFunctions) решают задачу вычисления функции и контроля области ее определения и изменения. Два других класса (ComputingFunc и ControlFunc) являются дочерними по отношению к классу Functions. Первый из них объединяет функции, используемые для вычисления свойства, второй – функции для контроля соотношений между свойствами.

Предложенная онтология формализует предметную область до уровня, позволяющего охватить практически все виды данных по свойствам, представленных как в виде таблиц, так и в виде математических выражений. В качестве примера на рис. 4 показан фрагмент онтологии с набором классов для описания теплофизических свойств веществ. Класс верхнего уровня “Свойства” соответствует классу “Properties”, показанному на рис. 3. Рисунок 4 – это снимок экрана при работе со специализированным редактором онтологий, позволяющим отображать классы в виде древовидной структуры (слева) и в виде графа (справа), при этом названия классов могут отображаться на русском языке. Более детальное описание онто-

логии для теплофизических свойств веществ приведено в [18, 19, 72].

Проект онтологии теплофизических данных предполагал ее использование как элемента общей технологии Semantic Web, позволяющей программным агентам расшифровывать семантику данных, что обеспечивает их сетевую интеграцию [73, 74]. Практическим воплощением концепции Semantic Web является проект Linked Open Data (LOD) – открытых связанных данных [75]. Целью этого проекта является наполнение сети Интернет данными в стандартных форматах Semantic Web, а также установление связей между данными из различных источников. Таким образом, формируется единое пространство открытых связанных данных (ОСД). Публикация данных в пространстве ОСД позволяет увеличить степень повторного использования данных, понизить степень их дублирования и ценность за счет связывания с другими данными и, наконец, облегчить потребление данных заинтересованными сторонами. В этом случае меняются роль и место публикации, которая перестает быть изолированной единицей, а становится элементом глобальной БД за счет связывания ее с тематически родственными публикациями и БД.

В ходе совместного проекта ОИВТ РАН и ВЦ РАН создана инфраструктура в виде онтологий и программного обеспечения для распространения данных и знаний в глобальной сети, точнее в рамках пространства ОСД. В ее основе лежит применение так называемой RDF-модели (Resource Definition Framework) для публикации структурированных данных в Интернете и установление явных ссылок-связей между сущностями в различных наборах данных. Преимущества ОСД как ин-

тегрирующей среды: связывание документов и данных различного формата, поддержка семантики за счет обращения к общедоступным словарям и онтологиям, автоматический характер процесса связывания тематически родственных документов.

Поскольку значительный объем научных данных, особенно данных по свойствам, имеют структурированный характер и находятся в реляционных БД (relational database, RDB), одна из основных технологий Semantic Web состоит в использовании отображения RDB → RDF [75]. Для реализации этого представления в сети Интернет использовались инструменты для публикации данных в ОСД, аннотирования и связывания с родственными документами [66].

При отработке технологии проводилось связывание авторских наборов термодинамических данных с двумя ресурсами, признанными наиболее релевантными решаемой задаче: ChemSpider [76] и ChEBI (Chemical Entities of Biological Interest) [77]. Они используются как платформы для аннотирования и сопровождения существующих данных, а также как источники дополнительных сведений о веществе, отсутствующих в авторском наборе данных. Таким образом, удалось обеспечить доступ внешним пользователям, включая сетевых программных агентов, к содержимому авторского набора термодинамических данных. В целом работы по интеграции теплофизических данных в ОСД носили поисковый характер [75, 78].

5. ИНТЕГРАЦИЯ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ ОНТОЛОГИЙ

Интеграция данных заключается в их соединении из различных источников и предоставлении пользователю унифицированного представления этих данных. Интеграция позволяет преодолеть: разнородность источников, использующих различные модели; автономность источников, спроектированных независимо под решение различных задач различными методами; физическую распределенность источников, доступных только через сетевые протоколы удаленного доступа. Создание технологий интеграции БД и в целом разнородных ресурсов уже давно находится в центре внимания специалистов по информационным технологиям.

При решении этой задачи, как правило, различают интеграцию на физическом и логическом уровнях [79–82]. В первом случае проводится консолидация данных с их извлечением из источников и помещением в единое хранилище, во втором — федерализация данных посредством отображения, реализуемого медиатором между глобальным и локальным представлениями данных, при том что физически данные остаются в локальных источниках. Оба способа неизбежно

связаны с перестройкой источников и вмешательством в их работу. Альтернативой служит семантическая интеграция данных на базе онтологий, которые охватывают предметную область, отражаемую в интегрируемых ресурсах [83–85]. В этом случае на основе нескольких ресурсов и онтологий создается единая инфраструктура, но без формирования ОСД с наборами данных.

Сравнение возможностей онтологии и БД в задачах хранения и распространения данных было предметом специального исследования [86]. 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. Распространение информации посредством онтологии проводится согласованным образом, т.е. передаваемая ею структура данных является общедоступной и одинаково трактуемой в определенном сообществе. В то же время концептуальная схема БД, определяя все понятия и структуру данных, служит только для тех целей, что реализует конкретная БД.

Одновременно БД обеспечивают эффективный способ хранения данных и высокую производительность при поиске и реализации сложных запросов. Поэтому, рассматривая роль онтологии в информационной инфраструктуре, исследователи выбирают в качестве основного пути не “вытеснение” БД, а создание системы хранения и управления данными из БД на основе связывания их с понятиями онтологии (database-to-ontology mapping). Задача этого метода — использовать преимущества обеих конструкций за счет соединения семантики с высокой производительностью при работе с данными.

Осознание огромного потенциала, заложенного в онтологиях для систематизации и распространения научного знания, стимулировало активную деятельность по разработке предметно-ориентированных онтологий для формализации отдельных дисциплин: химии [87], материаловедения [88–90], нанотехнологий [91]. Возможности онтологий для интеграции теплофизических данных изучены авторами в работах [70, 92, 93], а также в исследованиях СО РАН [94, 95]. Так, в [94] предложена архитектура системы, интегрирующей теплофизические знания и ресурсы на базе онтологии для семантического структурирования информации, навигации и поиска данных. Авторы [95] на примере теплофизики рассмотрели одну из проблем концептуализации знания, а именно проблему “сведения”, т.е. представления сложных научных терминов через простые. Соответственно, проектирование онтологий, включая определение классов и их свойств, можно связать с конструированием их экземпляров/индивидов.

В качестве конкретных примеров рассмотрены экземпляры, представляющие фазовые состояния воды, источники данных и наборы значений физических величин. Как иллюстрация экземпляров, представляющих значения величин, выбраны наборы данных по скачкам плотности при плавлении лития. Предложенные в работе [95] правила построения экземпляров предусматривали достаточно полную детализацию, включение сведений об источнике, методе измерения, точности и т.п.

Специальная версия онтологии для реакторной теплофизики предложена в работе [96]. Рассмотрены возможные сценарии ее использования на этапах проектирования, эксплуатации и интеграции автономных ресурсов, преимущественно БД.

В работах [70, 93] возможности онтологического подхода к интеграции ресурсов расширены путем включения некоторых элементов технологий Big Data. Как и предыдущие версии, новая система ориентирована на интеграцию разнородных по типу и изменяющихся по структуре данных. В качестве базовой единицы хранения данных предложено использовать файлы в JSON (JavaScript Object Notation) формате [97], а описание данных (метаданные) возложено на методы семантического моделирования и хранения. Основная идея проекта заключается в связывании атрибутов (ключей), а также и по возможности значений (объектов) с концептами онтологий. Важно, что эта операция отражается как на содержании литеральных (текстовых) данных онтологической модели, так и в изменении наименования атрибутов/ключей наборов данных в формате JSON. При этом по существу возникают две связанные системы хранения данных:

- репозиторий онтологий и моделей метаданных для оперирования данными;
- локальные или распределенные хранилища файлов данных в формате JSON и, возможно, других форматах (текстовых, графических и т.п.).

Предлагаемое решение использует две важные особенности данных. Онтологии обеспечивают безболезненную модернизацию текущей структуры данных, при этом исходные источники данных остаются “изолированными” и не изменяемыми. Владельцы ресурсов периодически выгружают данные в JSON-файлы по шаблонам, связанным с онтологическими моделями. При этом состав, количество и актуальность выгружаемых данных для “внешнего” мира они определяют сами. Такой тип взаимодействия является пассивным, в отличие от активного, когда дается доступ к БД по различным стандартным протоколам с соответствующей нагрузкой на СУБД.

Связь с онтологией дает также возможность организовать интеграцию совершенно различных типов данных приведением их атрибутов к еди-

ным концептам предметно-ориентированной онтологической модели. Применение файлов в формате JSON позволяет организовать “выгрузку” данных с совершенно разной исходной структурой и практически из всех типов БД. Вся дальнейшая работа по управлению данными происходит с JSON-файлами, ничуть не затрагивая исходные ресурсы. Наконец, использование одной из наиболее популярных платформ высокопроизводительных вычислений Apache Spark [98] для работы с большими массивами разнородных данных позволяет организовать весь процесс хранения, обработки и поиска данных среди JSON-файлов, в общем случае размещенных на разных серверах сети. В то же время наличие репозитория онтологий позволяет проводить поиск, совершая навигацию по классам онтологии, т.е. переходя на верхний и нижний уровни иерархии классов [83–85, 92]. Зафиксировав требуемое понятие, пользователь использует возможности платформы для запроса к содержательной части хранилища (к JSON-файлам) на традиционном для реляционных БД языке SQL (язык структурированных запросов).

Все функциональное управление комплексом данных базируется на свободно-распространяемом программном обеспечении в веб-среде. На рисунке 5 представлена схема управления и работы платформы [93].

Один из ключевых моментов предложенной технологии – стандартизация хранения данных, полученных из множества источников, различающихся структурой, форматом и семантикой, в текстовом документе, записанном в JSON-формате, принятом для обмена данными и метаданными. Преимущество текстового документа – возможность простого чтения и редактирования, доступность для восприятия человеком, удобная форма хранения и обмена произвольной структурированной информацией.

Разработанная технология использовалась в проекте радикальной перестройки БД ТЕРМАЛЬ [99] с расширением ее тематики и функций. В соответствии с принятым в работах [93, 94] сценарием, проект подобной перестройки предполагает решение двух ключевых задач.

– Переход от автономной (изолированной) БД к интегрированной структуре, способной обеспечить работу с ресурсами разного формата и содержания (таблицы, тексты, коды, сетевые документы и т.п.). Одно это позволяет резко расширить объем хранилища за счет сочетания неоднородных данных: структурированных и неструктурированных, экспериментальных и расчетных, “сырых” и рекомендованных.

– Создание средств настройки системы на предметную область, т.е. определенный тип веществ с характерной для них номенклатурой

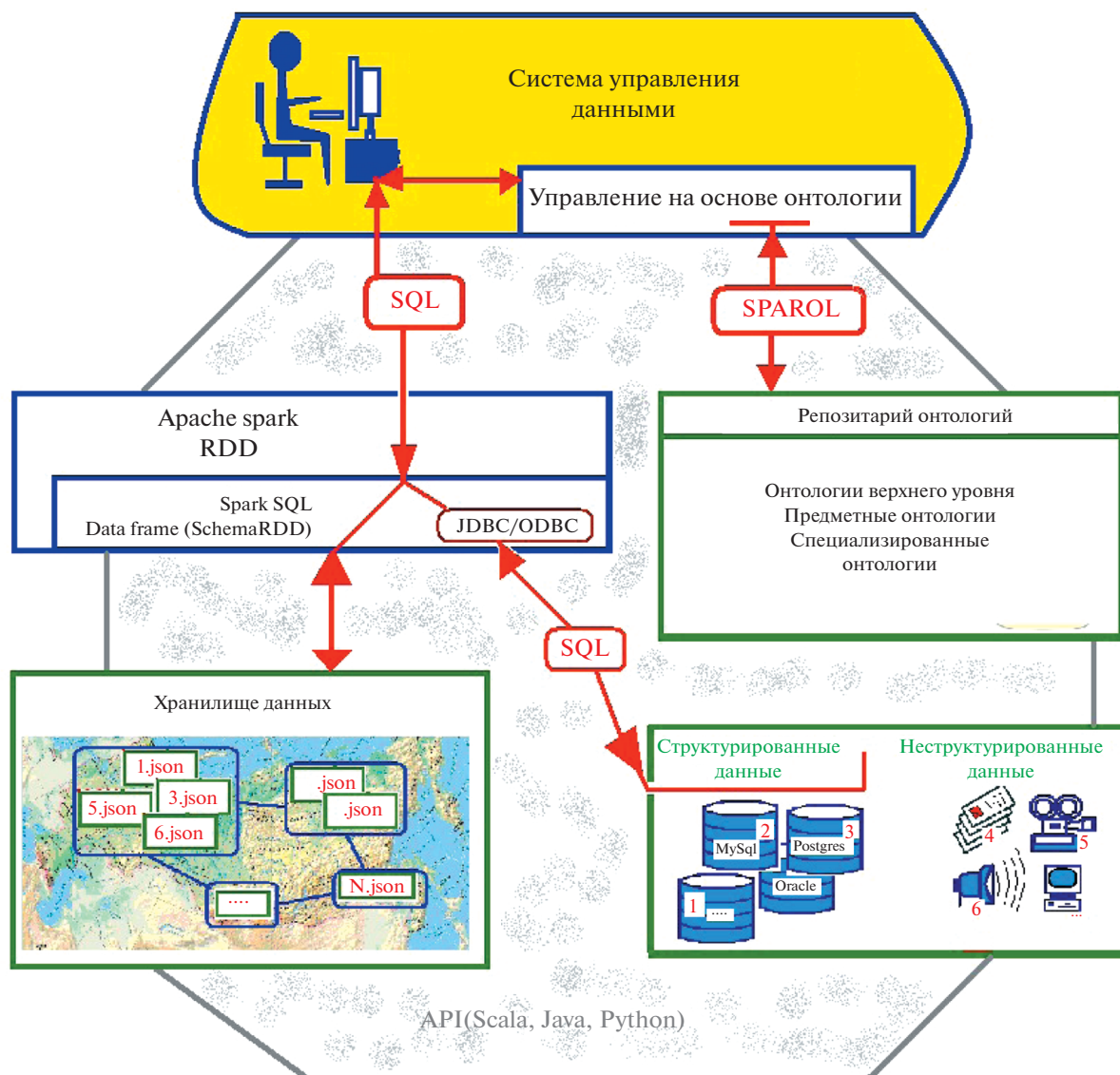


Рис. 5. Схема веб-среды для интеграции разнородных и распределенных данных по свойствам веществ.

свойств и идентификаторов. Именно ее решение позволит отказаться от обременительных ограничений, введенных в период создания БД ТЕРМАЛЬ.

Вторая задача предусматривает создание онтологического репозитория (ontology-based data manager), включающего предметно-ориентированные онтологии, которые охватывают терминологию и логические связи определенных предметных областей, например теплофизики, конструкционных материалов, наноструктур и т.п. Для них, как правило, уже доступны представленные в сети разработки, на основе которых можно составить и поддерживать собственную предметно-ориентированную онтологию. Обязательным также является включение в репозиторий некоторых онтологий верхнего уровня (upper-level ontol-

ogies), содержащих базовые типы и соотношения для описания научной активности, например онтология SIO (Semanticscience Integrated Ontology), представленной на портале Ontobee [100]. Предложенная структура предполагает двухуровневую систему хранения.

На рис. 6 продемонстрирована организация связи между данными, которые представлены в БД ТЕРМАЛЬ, и классами онтологии. Верхний фрагмент рисунка иллюстрирует поиск документов, включающих понятие “энтальпия”. Как видно, всего найдено 3335 записей. Нижний фрагмент демонстрирует визуализацию классов онтологии, соответствующих калорическим свойствам. Наличие связи БД–онтология позволяет вести навигацию по классам онтологии и содержатель-

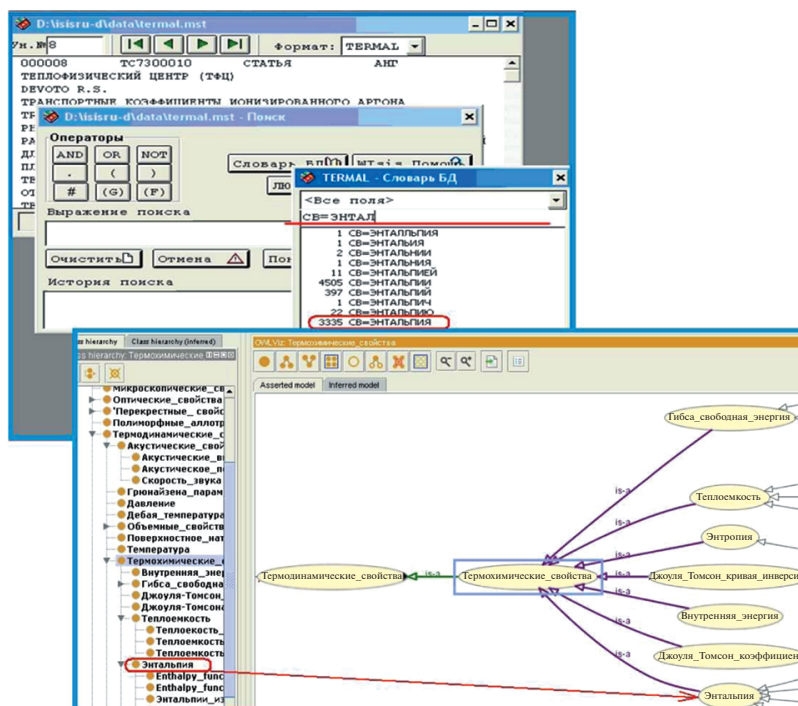


Рис. 6. Связывание содержания БД ТЕРМАЛ с классами онтологии.

ный поиск, не внедряясь в структуру самой БД, и при необходимости вводить новые понятия.

Подобная двухуровневая система хранения данных, основанная на онтологии, обеспечивает решение обеих задач, составляющих содержание проекта перестройки БД ТЕРМАЛ: интеграция разнородных ресурсов и вариация структуры данных.

ЗАКЛЮЧЕНИЕ

С первых лет существования получение и систематизация справочных данных по теплофизическим свойствам составляли одно из основных направлений в работе ОИВТ РАН. В обзоре рассмотрены результаты, полученные коллективом Термоцентра им. В.П. Глушко и Теплофизического центра по разработке теплофизических БД, и прежде всего история и современные возможности термодинамической БД ИВТАНТЕРМО и документальной БД ТЕРМАЛ. Создание и поддержка подобных БД являются результатом обширных исследований, включающих анализ и обработку эксперимента, а также постановку сложных термодинамических и квантово-химических расчетов для вещества в различных фазовых состояниях.

С начала XXI века возникает принципиально новый этап в работе с численными данными в большинстве естественнонаучных дисциплин, начиная с наук о Земле и вплоть до химии и мате-

риаловедения [2, 4, 70, 82]. Характерные для этого этапа тенденции к формированию наук с интенсивным использованием данных привели к внедрению в практику научных коллективов новых информационных технологий, таких как Semantic Web и Big Data. Обе технологии призваны преодолеть фундаментальные проблемы, связанные как с экспоненциальным ростом данных, так и с безграничным многообразием самих ресурсов (БД, веб-страницы, коллекции файлов и т.д.), подлежащих интеграции. Естественно, эта общая тенденция не могла обойти теплофизику, в которой работа с первичными данными, их анализ и обобщение всегда занимали важное место.

В обзоре описаны результаты нескольких проектов, имеющих своей целью расширить возможности традиционных БД путем перехода к интегрированным структурам, способным к представлению разнородных данных при непрерывной эволюции их логической структуры, связанной с необходимостью охвата веществ с ранее неизвестными свойствами. Итогом является отработанная технология хранения, систематизации и распространения теплофизических и родственных данных при крайнем многообразии как самих веществ, так и способов их получения, методов измерений, влияний внешней среды и других факторов. Развитие этой технологии и применение ее в различных областях теплофизики и науки о материалах имеет важное практическое значение.

СПИСОК ЛИТЕРАТУРЫ

1. *Потапов В.М., Кочетова Э.К.* Химическая информация. Где и как искать химику нужные сведения. М.: Химия, 1988. 224 с.
2. *Еркимбаев А.О., Цицерман В.Ю., Кобзев Г.А.* Интенсивное использование цифровых данных в современном естествознании // Научно-техническая информация. Сер. 2. Информационные процессы и системы. 2017. № 9. С. 9.
3. *Вассерман А.А., Мальчевский В.П.* Базы данных и автоматизированные информационные системы по теплофизическим свойствам газов и жидкостей // Технические газы. 2009. № 5. С. 59.
4. *Киселева Н.Н., Дударев В.А., Земсков В.С.* Компьютерные информационные ресурсы неорганической химии и материаловедения // Успехи химии. 2010. Т. 79. № 2. С. 162.
5. *Diky V., Bazyleva A., Paulechka E., Magee J.W., Martinez V., Riccardi D., Kroenlein, K.* Validation of Thermophysical Data for Scientific and Engineering Applications // J. Chem. Thermodyn. 2019. V. 133. P. 208.
6. *Saxena S.K.* Thermodynamic Databases and Phase Diagrams. In: Informatics for Materials Science and Engineering / Ed. Rajan K. Butterworth-Heinemann, 2013. Pt. 11. P. 245.
7. Information Resources on Inorganic Chemistry, IRIC. <http://iric.imet-db.ru/DB.asp>
8. *Когаловский М.Р.* Энциклопедия технологий баз данных. М.: Финансы и статистика, 2002. 800 с.
9. *Hey T.* The Fourth Paradigm: Data Intensive Scientific Discovery / Eds. Hey T., Tansley S., Tolle K. Redmond, WA: Microsoft Research, 2009. 286 p.
10. *Bizer C.* Interlinking Scientific Data on a Global Scale // Data Sci. J. 2013. V. 12. P. GRDI6.
11. *Wilkinson M.D., Dumontier M., Aalbersberg I.J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.W., da Silva S.L.B., Bourne P.E., Bouwman J.* The FAIR Guiding Principles for Scientific Data Management and Stewardship // Sci. Data. 2016. V. 3. P. 160018.
12. *Rumble J.R. Jr.* Accessing Materials Data: Challenges and Directions in the Digital Era // Integr. Mater. Manuf. Innov. 2017. V. 6. P. 172.
13. *Frenkel M.* Global Information Systems in Science: Application to the Field of Thermodynamics // J. Chem. Eng. Data. 2009. V. 54. P. 2411.
14. *Горгораки Е.А., Краевский В.Л., Трахтенгерц М.С., Швальб В.Г., Шпильрайн Э.Э., Якимович К.А.* Автоматизированная информационно-поисковая система Теплофизического Центра ИВТАН // Обзоры по теплофизическим свойствам веществ. М.: ИВТАН, 1977. № 4.
15. *Еркимбаев А.О., Цицерман 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.
16. *Еркимбаев А.О., Цицерман В.Ю., Кобзев Г.А.* Систематизация данных по физико-химическим свойствам и применению углеродных наноструктур // ТВТ. 2010. Т. 48. № 6. С. 869.
17. *Елецкий А.В., Еркимбаев А.О., Цицерман В.Ю., Кобзев Г.А., Трахтенгерц М.С.* Теплофизические свойства наноразмерных объектов: систематизация и оценка достоверности данных // ТВТ. 2012. Т. 50. № 4. С. 524.
18. *Серебряков В.А., Теймуразов К.Б., Хайруллин Р.И., Еркимбаев А.О., Цицерман В.Ю., Кобзев Г.А., Трахтенгерц М.С.* Практическая реализация системы интеграции теплофизических данных на основе онтологической модели предметной области // Тр. IV всерос. симп. "Инфраструктура научных информационных ресурсов и систем". СПб., 6–8 окт. 2014. М.: ВЦ РАН, 2014. Т. 1. С. 87.
19. *Еркимбаев А.О., Цицерман В.Ю., Ключкина Е.К., Кобзев Г.А., Серебряков В.А., Теймуразов К.Б., Хайруллин Р.И.* Система хранения и интеграции численных данных по теплофизическим свойствам веществ // Тр. XIX Всерос. науч. конф. "Научный сервис в сети Интернет", 18–23 сент. 2017. Новороссийск. М.: ИПМ им. М.В. Келдыша, 2017. С. 137.
20. *Гурвич Л.В., Вейц И.В., Медведев В.А. и др.* Термодинамические свойства индивидуальных веществ. Спр. изд. / Под ред. Глушко В.П. М.: Наука, 1981. 624 с.
21. *Chase M.W., Davies C.A., Downey J.R. Jr., Frurip D.J., Mc Donald R.A., Syverud A.N.* JANAF Thermochemical Tables, third ed. // J. Phys. Chem. Ref. Data. 1985. V. 14. Suppl. 1.
22. *Wagman D.D.* Data Bases: Past, Present and Future // Pure Appl. Chem. 1992. V. 64. № 1. P. 37.
23. *Ruscic B., Bross D.H.* Thermochemistry // Computer Aided Chem. Eng. 2019. V. 45. P. 3.
24. *Гурвич Л.В.* ИВТАНТЕРМО – автоматизированная система данных о термодинамических свойствах веществ // Вестн. АН СССР. 1983. № 3.
25. *Gurvich L.V.* Reference Books and Data Banks on the Thermodynamic Properties of Individual Substances // Pure Appl. Chem. 1989. V. 61. P. 1027.
26. *Belov G.V., Iorish V.S., Yungman V.S.* IVTANTHERMO for Windows – Database on Thermodynamic Properties and Related Software // Calphad. 1999. V. 23. № 2. P. 173.
27. *Burcat A.* Thermochemical Data for Combustion Calculations // Combustion Chemistry. N.Y.: Springer, 1984. P. 455.
28. *Burcat A., Ruscic B.* Third Millennium Ideal Gas and Condensed Phase Thermochemical Database for Combustion (with Update from Active Thermochemical Tables). No. ANL-05/20. Argonne, IL, USA: Argonne National Lab., 2005.
29. *Ruscic B., Pinzon R.E., Von Laszewski G., Kodeboyina D., Burcat A., Leahy D., Montoy D., Wagner A.F.* Active Thermochemical Tables: Thermochemistry for the 21st Century // J. Phys. Conf. Ser. 2005. V. 16. № 1. P. 078.
30. База данных "Active Thermochemical Tables". <https://atct.anl.gov/>
31. *Ruscic B., Pinzon R.E., Morton M.L., von Laszewski G., Bittner S.J., Nijssure S.G., Amin K.A., Minkoff M., Wagner A.F.* Introduction to Active Thermochemical Tables: Several "key" Enthalpies of Formation Revisited // J. Phys. Chem. A. 2004. V. 108. P. 9979.
32. *Ruscic B.* Uncertainty Quantification in Thermochemistry, Benchmarking Electronic Structure Computations, and Active Thermochemical Tables // Int. J. Quantum Chem. 2014. V. 114. P. 1097.

33. *Linstrom P.J., Mallard W.C.* The NIST Chemistry Web-Book: A Chemical Data Resource on the Internet // *J. Chem. Eng. Data.* 2001. V. 46. P. 1059.
34. Facility for the Analysis of Chemical Thermodynamics (FACT). <http://www.crct.polymtl.ca/fact/>
35. Thermodata. <http://thermodata.online.fr/>
36. Thermo-Calc Software. <http://www.thermocalc.com/>
37. Materials-oriented Little Thermodynamic Database (MALT). <http://www.kagaku.com/malt/>
38. *Dinsdale A.T.* SGTE Data for Pure Elements // *Calphad.* 1991. V. 15. P. 317.
39. *Giffaut E. et al.* Andra Thermodynamic Database for Performance Assessment: ThermoChimie // *Appl. Geochem.* 2014. V. 49. P. 225.
40. *Bale C.W., Bélisle E., Chartrand P. et al.* FactSage Thermochemical Software and Databases, 2010–2016 // *Calphad.* 2016. V. 54. P. 35.
41. *Belov G.V., Dyachkov S.A., Levashov P.R., Lomonosov I.V., Minakov D.V., Morozov I.V., Sineva M.A., Smirnov V.N.* The IVTANTHERMO-Online Database for Thermodynamic Properties of Individual Substances with Web Interface // *J. Phys. Conf. Ser.* 2018. V. 946. № 1. P. 012120.
42. *Morozov I.V.* Thermodynamic Database for Pure Substances IVTANTHERMO-Online // *CEUR Workshop Proc.* 2019. V. 2523. P. 325.
43. *Белов Г.В., Иорш В.С., Юнган В.С.* Моделирование равновесных состояний термодинамических систем с использованием ИВТАНТЕРМО для Windows // *ТВТ.* 2000. Т. 38. № 2. С. 191.
44. *Белов Г.В.* Термодинамическое моделирование: методы, алгоритмы, программы. М.: Научный мир, 2002. 184 с.
45. *Belov G.V., Aristova N.M., Morozov I.V., Sineva M.A.* On Approximation of the Heat Capacity of Substances in the Gaseous State // *J. Math. Chem.* 2017. V. 55. № 8. P. 1683.
46. *Мальцев М.А., Морозов И.В., Осина Е.Л.* Термодинамические свойства димеров аргона Ar_2^+ и Ar_2 // *ТВТ.* 2019. Т. 57. № 1. С. 42.
47. *Мальцев М.А., Морозов И.В., Осина Е.Л.* Термодинамические свойства ArH^+ и ArH // *ТВТ.* 2019. Т. 57. № 3. С. 367.
48. *Мальцев М.А., Морозов И.В., Осина Е.Л.* Термодинамические свойства ArO^+ и ArO // *ТВТ.* 2020. Т. 58. № 2. С. 202.
49. *Sineva M.A., Morozov I.V., Belov G.V., Aristova N.M., Lavrinenko Ya.* Simultaneous Analysis of the Enthalpy Increment and Heat Capacity Data Measurements for Updating the IVTANTHERMO Database // *J. Phys. Conf. Ser.* 2019. V. 1385. P. 012025.
50. *Аристова Н.М., Белов Г.В., Морозов И.В., Синева М.А.* Термодинамические свойства диоксида урана в конденсированном состоянии // *ТВТ.* 2018. Т. 56. № 5. С. 677.
51. *Линева В.И., Синева М.А., Белов Г.В., Морозов И.В.* Термодинамические свойства ванадия в конденсированном состоянии // *ТВТ.* 2020. Т. 58. № 1. С. 41.
52. CDS ISIS, Wikipedia. https://en.wikipedia.org/wiki/CDS_ISIS
53. *Бакстон Э., Хопкинсон А.* Руководство по CDS/ISIS для Windows. Пер. с англ. под общ. рук. Шрайберга Я.Л. М.: Ассоциация ЭБНИТ, 2002.
54. *Трахтенгерц М.С.* Новый эффективный инструмент для текстовых баз данных 3/4 CDS/ISIS for Windows // *Научно-техническая информация. Сер. 2. Информационные процессы и системы.* 2006. № 6. С. 1.
55. *Wilthan B., Pfeif E.A., Diky V.V., Chirico R.D., Kattner U.R., Kroenlein K.* Data Resources for Thermophysical Properties of Metals and Alloys. Part 1: Structured Data Capture from the Archival Literature // *Calphad.* 2017. V. 56. P. 126.
56. *Labroch D., Dugne O., Chatillon C.* Thermodynamics of the O–U System. II. Critical Assessment of the Stability and Composition Range of the Oxides UO_{2+x} , U_4O_{9-y} and U_3O_{8-z} // *J. Nucl. Mat.* 2003. V. 312. P. 50.
57. *Mao H., Sundman B.B., Wang Z., Saxena S.K.* Volumetric Properties and Phase Relations of Silica – Thermodynamic Assessment // *J. Alloys Comp.* 2001. V. 327. P. 253.
58. *Сергеев О.А., Мень А.А.* Теплофизические свойства полупрозрачных материалов. М.: Изд-во стандартов, 1977.
59. *Гейвандов Э.А., Колосова И.И., Коматова Л.В. и др.* Классификатор свойств веществ и материалов. М.: Изд-во стандартов, 1980. 132 с.
60. *Еркимбаев А.О., Зицерман В.Ю.* Построение информационных систем по свойствам веществ с использованием технологии слабоструктурированных данных. Теплофизические свойства веществ и материалов // *Матер. докл. и сообщ. XI Рос конф. по теплофизическим свойствам веществ.* Санкт-Петербург, Россия. 4–7 окт. 2005. СПб., 2005. С. 93.
61. PostgreSQL: The World’s Most Advanced Open Source Relational Database. <http://www.postgresql.org>
62. *Елецкий А.В., Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Трахтенгерц М.С.* База данных по физико-химическим и эксплуатационным свойствам углеродных наноструктур. Свид. о гос. рег. базы данных № 2014621608 от 26 ноября 2014 г.
63. *Michel K., Meredig B.* Beyond Bulk Single Crystals: A Data Format for all Materials Structure–property–processing Relationships // *MRS Bulletin.* 2016. V. 41. № 8. P. 617.
64. *Berners-Lee T., Hendler J., Lassila O.* The Semantic Web // *Sci. Amer.* 2001. V. 284. № 5. P. 35.
65. *Gruber T.R.* A Translation Approach to Portable Ontologies // *Knowledge Acquisition.* 1993. V. 5. № 2. P. 199.
66. *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.
67. *Harth A., Janik M., Staab S.* Handbook of Semantic Web Technologies. Berlin–Heidelberg: Springer, 2011. 1035 p.
68. *Gandon F.* A Survey of the First 20 Years of Research on Semantic Web and Linked Data // *Ingénierie des Systèmes d’Information.* 2018. V. 23. № 3–4. P. 11.
69. Электронный справочник “Techopedia”, описание технологии “Semantic Web”. <https://www.techopedia.com/definition/27961/semantic-web>

70. *Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Косинов А.В.* Интеграция информационных ресурсов с данными по свойствам веществ и материалов. Практическая реализация и доступные средства // Научно-техническая информация. Сер. 2. 2018. № 10. С. 11.
71. *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.
72. *Атаева О.М., Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Серебряков В.А., Теймуразов К.Б., Хайруллин Р.И.* Интеграция данных по теплофизическим свойствам веществ методами онтологического моделирования // Тр. XV Всерос. науч. конф. “Электронные библиотеки: перспективные методы и технологии, электронные коллекции”. Ярославль, Россия. 14–17 окт. 2013. Ярославль: ЯрГУ, 2013. 422 с.
73. *Нестеренко А.К., Сысоев Т.М., Бездушный А.А., Бездушный А.Н., Серебряков В.А.* Интеграция распределенных данных на основе технологий Semantic Web и рабочих процессов // Электронные библиотеки. 2004. Т. 7. № 4.
74. *Harth A., Janik M., Staab S.* Semantic Web Architecture. In: Handbook of Semantic Web Technologies. Berlin–Heidelberg: Springer, 2011. P. 43.
75. *Кузнецов К.А.* Система интеграции научных данных в пространстве Linked Open Data // Программирование. 2013. № 1. С. 36.
76. ChemSpider. Search and Share Chemistry. <http://www.chemspider.com>
77. Chemical Entities of Biological Interest (ChEBI), a Database and Ontology of Molecular Entities Focused on Small Chemical Compounds. <http://www.ebi.ac.uk/chebi/>
78. *Устинова Е.С.* Интеграция данных по свойствам веществ в специализированное пространство связанных данных. Дипломная работа. М.: ВМК МГУ, 2014.
79. *Калиниченко Л.А.* Методы и средства интеграции неоднородных баз данных. М.: Наука, Гл. ред. физ.-мат. лит., 1983.
80. *Ландэ Д.В.* Основы интеграции информационных потоков. Киев: Инжиниринг, 2006. 240 с.
81. *Doan A.H., Halevy A., Ives Z.* Principles of Data Integration. Elsevier, Inc., 2012.
82. *Дударев В.А.* Интеграция информационных систем в области неорганической химии и материаловедения. М.: Красанд, 2016. 320 с.
83. *Тузовский А.Ф.* Интеграция баз данных на основе онтологий. Слайды симпозиума “Онтологическое моделирование 2010”. <http://www.myshared.ru/slide/381212/>
84. *Коголовский М.Р.* Системы доступа к данным, основанные на онтологиях // Программирование. 2012. № 4. С. 55.
85. *Calvanese D., Calvanese D., De Giacomo G., Lembo D., Lenzerini M., Rosati R., Ruberti G.A.* Ontology-Based Data Access and Integration. In: Encyclopedia of Database Systems / Eds. Liu L., Özsu M.T. N.Y.: Springer, 2018.
86. *Uschold M.* Ontology and Database Schema: What’s the Difference? // Appl. Ontology. 2015. V. 10. № 3–4. P. 243.
87. *Degtyarenko K., de Matos P., Ennis M. et al.* ChEBI: a Database and Ontology for Chemical Entities of Biological Interest // Nucl. Acids Res. 2008. V. 36. Database Iss. P. D344.
88. *Ashino T.* Materials Ontology: an Infrastructure for Exchanging Materials Information and Knowledge // Data Sci. J. 2010. V. 9. P. 54.
89. *Cheung K., Hunter J., Drennan J.* MatSeek: An Ontology-Based Federated Search Interface for Materials Scientists // IEEE Intelligent Systems. 2009. V. 24. № 1. P. 47.
90. *Zhang X., Hu C., Li H.* Semantic Query on Materials Data Based on Mapping MatML to an OWL Ontology // Data Sci. J. 2009. V. 8. P. 1.
91. *Erkimbaev A.O., Zitserman V.Yu., Kobzev G.A., Kosinov A.V.* Ontological Concepts and Taxonomies for Nano World // J. Inform. Knowledge Management. 2019. V. 18. № 2. 1950014.
92. *Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Косинов А.В.* Связывание онтологий с базами данных по свойствам веществ и материалов // Научно-техническая информация. Сер. 2. 2015. № 12. С. 1.
93. *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 // CEUR Workshop Proc. 2017. V. 2022. P. 219.
94. *Федотов А.М., Молородов Ю.И., Зеленчук А.М.* Концепция и архитектура информационной системы для исследования теплофизических свойств материалов // Вестн. Новосиб. гос. ун-та. Сер. Информационные технологии. 2015. Т. 13. № 4. С. 43.
95. *Зеленчук А.М., Прядухин И.Ю., Молородов Ю.И., Фазлиев А.З.* Прикладная онтология по теплофизике. Проблема. Сведения // Тр. V Всерос. симп. “Инфраструктура научных информационных ресурсов и систем”. СПб., 6–8 окт. 2015. Т. 1. С. 164.
96. *Чусов И.А., Кириллов П.Л., Проняев В.Г., Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Фокин Л.Р.* Онтологии и базы данных по теплофизическим свойствам реакторных материалов // Изв. вузов. Ядерная энергетика. 2019. № 1. С. 5.
97. The JSON Data Interchange Format. Standard ECMA-404, 1th ed., October 2013. <http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf>
98. *Karau X., Konvinski Э., Венделл П., Захария М.* Изучаем Spark: молниеносный анализ данных. Пер.с англ. М.: ДМК Пресс, 2015. 304 с.
99. *Kosinov A.V., Erkimbaev A.O., Zitserman V.Yu., Kobzev G.A.* Ontology-based Methods of Thermophysical Data Integration // J. Phys. Conf. Ser. 2019. V. 1385. P. 012033.
100. Портал проекта “Ontobee”. <https://www.ontobee.org>