

Интеграция данных по теплофизическим свойствам веществ методами онтологического моделирования

О.М. Атаева

ВЦ им. А.А. Дородницына РАН

oli@ultimeta.ru

А.О. Еркимбаев

ОИВТ РАН

adilbek@ihed.ras.ru

В.Ю. Зицерман

vz1941@mail.ru

Г.А. Кобзев

В.А.Серебряков

К.Б. Теймуразов

ОИВТ РАН

gkbz@oivtran.ru

ВЦ им. А.А. Дородницына РАН

serebr@ccas.ru

kbt@intrimg.ru

Р.И. Хайруллин

МФТИ

rena-2006@yandex.ru

Аннотация

Предложена схема интеграции данных по теплофизическим свойствам веществ. Изучена специфика предметной области и построена онтология, допускающая использование методов Semantic Web. Показано, как исходно разнородные данные по свойствам индивидуальных веществ могут быть связаны в рамках сетевого пространства Linked Open Data.

1. Предметная область – специфика данных

Теплофизика относится к дисциплинам, в которых работа с численными данными занимает центральное место. При том же предмете исследования, что и в молекулярной физике и физике конденсированного состояния, в теплофизике проявляют повышенное внимание к первичным данным с изучением их достоверности, воспроизводимости, согласовании разнородных свойств. Развитие теплофизики сопровождается нарастающим производством новых данных, публикуемых в десятках журналов различного профиля. Современный этап характеризуется выделением систематизации данных в самостоятельное направление, наряду с экспериментом и теорией, а также повсеместным переходом от печатной формы к базам данных (БД). Потребность в методах обмена неоднородными данными возникла в теплофизике задолго до того момента, когда проблема интеграции данных приобрела актуальность для информационного сообщества. Так один из первых стандартов обмена термодинамическими данными, получивший название COSTAT (Codata STandard

Thermodynamics), был разработан Термодинамическим Исследовательским Центром США в течение 1985-1987 г.г. под эгидой Международной комиссии по численным данным (CODATA) [1].

Большой проблемой на пути интеграции теплофизических данных является многообразие форматов представления в отсутствие общепринятых рецептов и стандартов записи [2]. Другим фактором является резкое расширение видов вновь изученных объектов при множестве факторов, определяющих их свойства. Проблемой является также необходимость удовлетворить нарастающие требования к аттестации неопределенности, представляя целый набор характеристик: стандартная неопределенность, доверительный интервал и т.п. Наконец, при работе с численными данными приходится учитывать, что в публикациях и БД используют несколько типовых форм, а именно: табличную, графическую и математическую (в виде хранимых формул или программных кодов). Графическая форма иллюстрирует характер зависимостей, рассеяние опытных точек и т.п. Табличная форма наиболее надежна в передаче данных, легко контролируема в отношении пропусков, ошибок в знаке или порядке величин и т.п. Математическая форма, избавляя от интерполяции, требует повышенной тщательности в обнаружении ошибок, легко выявляемых в табличной форме. Доминирующей формой в экспериментальных работах и справочниках является именно табличная, а математическая форма дополняет табличную, обеспечивая удобство для пользователя и работу вычислительных приложений.

Важно также заранее представлять характер зависимостей, число независимых переменных, функций и проч. В принципе, число переменных зависит от состава смеси и параметров состояния.

Состав n-компонентной смеси определяется n-1 параметром с учетом экстенсивности свойств, а типичными параметрами являются температура и давление. Практические задачи облегчает то, что подавляющее большинство данных определяет один параметр, как правило, температура, поскольку свойства твердой и жидкой фаз в достаточно широком диапазоне проявляют слабую барическую зависимость. Применительно к газовой фазе, практика работы в химической промышленности показывает, что для учета неидеальности допустимо ограничиться первыми поправками по плотности (или давлению), которые сами являются функциями температуры. Не предполагая универсальность этого тезиса, можно принять, что подавляющее большинство публикуемых теплофизических данных сводится к представлению температурных функций, как например, теплоемкость, энтальпия, вязкость и т.п., а также численных констант, например, критических постоянных вещества. Если рассматривать чистые вещества, игнорируя барическую зависимость, мы приходим к задаче интеграции однотипных данных, представленных константами и функциями одной переменной. При таком сужении предметной области интеграция данных делается относительно обозримой, хотя и здесь приходится учитывать крайнее многообразие вариантов, связанное с идентификацией вещества, принимаемой номенклатурой свойств, фазовым многообразием и т. п.

Целью данного исследования является стандартизация процессов обмена и интеграции данных по теплофизическим свойствам в рамках **онтологического моделирования**. В работах авторов [3, 4] рассмотрено несколько удачных примеров использования онтологий применительно к свойствам вещества, в основном из области материаловедения, где многообразие типов данных и богатство словарей проявляются наиболее ярко. Рассматривая онтологию как элемент *Semantic WEB*, можно обеспечить доступ к многочисленным словарям и онтологиям общенаучного содержания, таким как онтология QUIDT [5] по физическим величинам и единицам измерения или к международному словарю химической терминологии [6]. Для описания источников данных может быть использована разработанная в связи с проектом ЕНИП (Единое научное информационное пространство) онтология научных данных [7], с подробным описанием класса «Публикация».

Тем самым, онтология обеспечивает единую информационную модель для определения

семантики и синтаксиса представления и единый словарь для определения смысла данных. Формализуя понятия и концепции предметной области, онтология позволяет адаптировать описание к различным схемам, используемым в БД или документах. Соответствующий пример отображения онтологии по свойствам материалов на схемы разнородных БД по теплофизическим свойствам приведен в работе [8]. Наконец, важный элемент, присущий онтологическому моделированию – возможность перманентного *наращивания* новых понятий, и прежде всего, за счет расширения перечня веществ и номенклатуры свойств.

2. Типовая структура данных и общие принципы концептуализации

Наиболее адекватное определение, данное Т. Грубером [9] трактует онтологию как **спецификацию концептуализации**. Второй из этих терминов означает строгое описание понятий предметной области, их связей и отношений средствами естественного языка, в то время как первый (**спецификация**) означает формализацию понятий и связей, например, в терминах OWL. Описание типовой структуры теплофизических данных, приведенное ниже, и означает концептуализацию предметной области, предваряющую построение онтологии.

Упрощенная схема данных базируется на трех элементах: (1) перечень/словарь веществ; (2) перечень/словарь свойств и (3) формат/шаблон для представления набора фактографических данных. С каждым элементом из списка веществ связано одно или несколько названий и формульное обозначение. С каждым элементом из списка свойств связано его название, обозначение и единица измерения. При построении шаблона фиксируются: список свойств, рассматриваемых как константы; список свойств, рассматриваемых как функции; способы задания неопределенности и представления источников данных. Таблица на рис. 1 из справочника [10] иллюстрирует характерный набор данных: текст с названием вещества и формулой, две константы и 5 свойств в виде столбцов таблицы, из которых первый отвечает аргументу, а четверка остальных представляет функции. Набор данных упрощен за счет исключения сведений о неопределенности и источнике.

Water H₂O(g)

$\Delta_f H_0^*$	-238.923 kJ/mol	$\Delta_f H_{298}^*$	-241.826 kJ/mol
------------------	-----------------	----------------------	-----------------

T	Cp	F	S	H
K	J/K*mol	J/K*mol	J/K*mol	kJ/mol
298.15	33.596	155.502	188.722	9.905
300.00	33.605	155.708	188.930	9.967
400.00	34.292	165.287	198.683	13.358
500.00	35.254	172.766	206.434	16.834
600.00	36.360	178.935	212.958	20.414
700.00	37.556	184.210	218.652	24.109
800.00	38.808	188.839	223.748	27.927
900.00	40.091	192.980	228.393	31.872
1000.00	41.385	196.738	232.684	35.946
1100.00	42.670	200.190	236.689	40.148
1200.00	43.927	203.390	240.456	44.479
1300.00	45.138	206.380	244.020	48.932
1400.00	46.287	209.190	247.407	53.504
1500.00	47.354	211.846	250.638	58.187
1500.00	47.273	211.846	250.634	58.181
1600.00	48.369	214.368	253.721	62.965
1700.00	49.344	216.770	256.683	67.851

Рис. 1

На рис. 2 показана другая форма таблицы из [11], где приведены рекомендованные данные по второму вириальному коэффициенту CO. Шаблон, отличается в нескольких моментах: (1) приведена

только одна функция; (2) неопределенность приведена в каждой точке и представлена двумя характеристиками; (3) ссылки на источники приведены в каждой точке.

Argon (cont.)

Table 2. (cont.)

$\frac{T}{K}$	$\frac{B_{\text{exp}} \pm \delta B}{\text{cm}^3 \cdot \text{mol}^{-1}}$	$\frac{B_{\text{exp}} - B_{\text{calc}}}{\text{cm}^3 \cdot \text{mol}^{-1}}$	Ref. (Symbol in Fig. 1)	$\frac{T}{K}$	$\frac{B_{\text{exp}} \pm \delta B}{\text{cm}^3 \cdot \text{mol}^{-1}}$	$\frac{B_{\text{exp}} - B_{\text{calc}}}{\text{cm}^3 \cdot \text{mol}^{-1}}$	Ref. (Symbol in Fig. 1)
398.15	-1.2 ± 1.0	0.2	49-mic/wij(□)	573.15	10.8 ± 0.1	0.1	67-kal/mil(■)
400.00	-0.9 ± 0.1	0.3	92-ewi/tru-1(▲)	573.15	10.6 ± 0.1	0.0	67-kal/mil(■)
400.00	-1.1 ± 0.1	0.1	96-est/tru(●)	600.00	11.8 ± 0.1	0.0	92-ewi/tru-1(▲)
423.15	1.4 ± 1.0	0.4	49-mic/wij(□)	673.15	14.3 ± 0.1	-0.2	67-kal/mil(■)
423.15	1.4 ± 1.0	0.3	49-mic/wij(□)	673.15	14.2 ± 0.1	-0.3	67-kal/mil(■)
450.00	3.3 ± 0.1	0.0	96-est/tru(●)	700.00	15.1 ± 0.1	-0.2	92-ewi/tru-1(▲)
473.15	5.1 ± 0.1	0.1	67-kal/mil(■)	773.15	17.1 ± 0.1	-0.1	67-kal/mil(■)
473.15	5.2 ± 0.1	0.2	67-kal/mil(■)	773.15	17.0 ± 0.1	-0.2	67-kal/mil(■)
500.00	7.0 ± 0.1	0.2	92-ewi/tru-1(▲)				
Further references: [10-onn/cro, 25-hol/ott, 30-tan/mas, 53-wha/lup, 56-cot/ham, 60-lec, 62-fen/hal, 62-poo/sav, 66-cra/son, 67-wei/wyn, 68-byr/jon, 69-lic/sch, 70-bla/hal, 70-bos/col, 71-pro/can, 72-osb, 73-pop/cha, 74-bel/rei, 74-hah/sch, 74-sch/heb, 76-san/uri, 77-ren/sch, 77-sch/sch, 79-ewi/mar, 79-sch/leu-1, 80-per/sch, 80-sch/geh, 80-woo/kro, 82-ker, 84-ker/hae, 88-pat/jof, 89-ewi/owu, 91-lop/roz]							

Рис. 2

При работе с данными должен быть проведен контроль корректности, основанный на физических принципах. Простейшая форма контроля - проверка знака функции и аргумента, характера монотонности и проч. особенностей, следующих из физических закономерностей. Необходим также контроль взаимосвязи некоторых функций, в тех случаях, когда она следует из их определения. Например, для термодинамических функций на рис. 1 должно выполняться соотношение $F = S - H/T$. Будучи точным, подобное соотношение для опытных данных выполняется лишь с некоторой априорно задаваемой точностью, например, $|F - (S - H/T)| \leq \delta$.

Наряду с тремя базовыми элементами (вещества, свойства, шаблон данных), схема должна включать древовидный справочник состояний, к которым относятся теплофизические данные. Этот перечень включает три агрегатных состояния, линии равновесия типа liquid-gas и характерные точки (тройную и критическую). Для твердого состояния возможно выделение отдельных фаз (кубической, тетрагональной, гексагональной и др). Возникают связи и ограничения в назначении свойств в зависимости от агрегатного состояния. Например, свойство **viscosity** может использоваться лишь для веществ, находящихся в состояниях **liquid** или **gas**. Другие характеристики (например, энтальпия испарения) применимы только на фазовых границах.

Определенные усложнения связаны с многовариантностью в задании свойств и их неопределенности. Так с каждым свойством и аргументом ассоциируется единица измерения, а также способ представления: прямой, по отклонению от некоторой реперной величины, по отношению к некоторой реперной величине. В двух последних случаях набор данных должен включать значение реперной величины. Точно также для неопределенности должен быть предусмотрен идентификатор, определяющий способ ее задания, как и само значение неопределенности. В зависимости от условий эксперимента может быть принято, что неопределенность характеризует весь набор данных или должна быть отнесена к каждой из точек, как на рис.2.

4. Онтология теплофизических данных: классы и свойства

Как следует из определения [9], вторым этапом построения онтологии является **спецификация** с выделением классов и записью всех связей и отношений на языке OWL. Предложенная онтология включает 11 основных классов и 2 класса потомка. Диаграмма на рис. 3 показывает связи классов, после чего приведен перечень классов с указанием назначения и основными атрибутами каждого из классов.

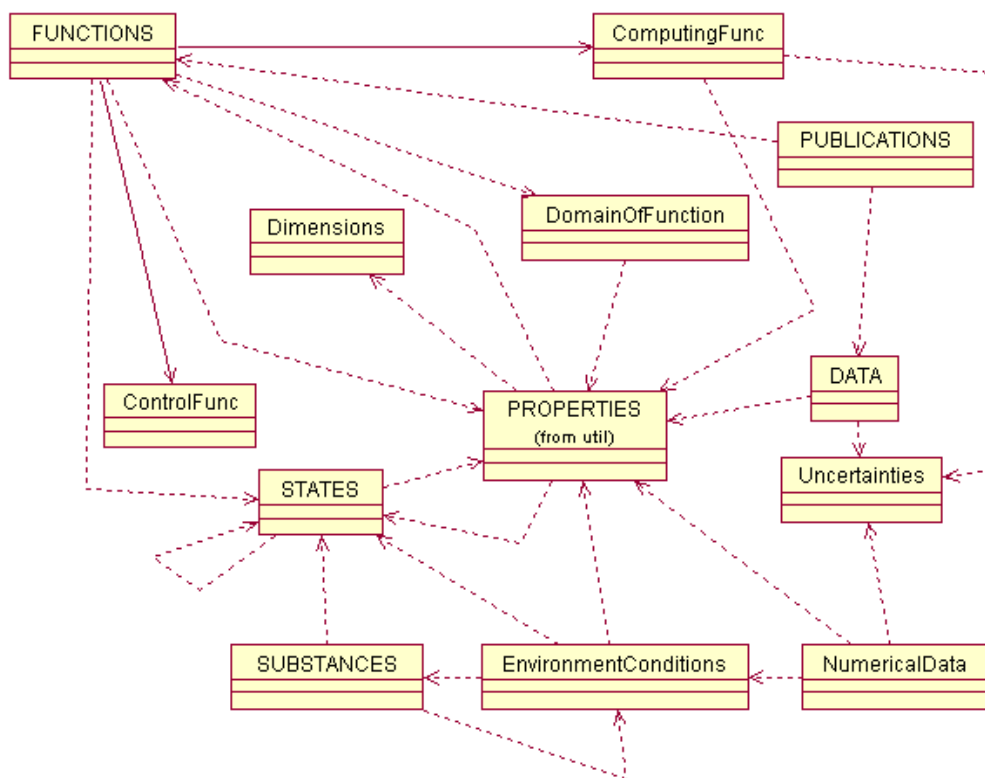


Рис. 3

Первая четверка классов определяет ключевые понятия, выделенные в разделе 3 для представления набора данных: вещества, свойства, состояния, численные данные. Смысл большей части атрибутов достаточно понятен из их названия. Так атрибуты **InConditoins** и **InState** в классе **Substances** отсылают к перечням состояний и внешних условий, перечни которых даны в соответствующих классах. В классе **Properties** введен атрибут **Formula** для возможности представления данных в математическом (не табличном) виде. Использован также ряд атрибутов для таких понятий как единицы измерения и неопределенность.

Класс **Substances**: Определяет вещества, для которых приводятся данные

Substances
1. String Name 2. String Subst_Formula 3. EnvironmentalConditons inConditioints 4. States inState

Класс **States**: Определяет агрегатные состояния вещества.

States
1. String Name 2. States ParentStates 3. Properties inProperty

Класс **Properties**: Определяет свойства вещества.

Properties
1. String Name 2. String(MathMl) PropDesignation 3. Function Formula 4. States inState 5. Dimensions Dimension

Класс **NumericalData**: Определяет набор численных данных по свойствам вещества в определенных условиях.

NumericalData
1. Decimal Value 2. Decimal UncertaintyValue 3. Uncertainties UncertaintyType 4. Properties Property 5. EnvironmentConditions inConditions 6. Publication Datasource

Следующая группа классов определяет базовые понятия, сопровождающие физические величины:

единицы измерения (размерности), неопределенность данных и внешние условия, например, давление среды. К условиям в данной работе относятся и константы, связанные с веществом. С точки зрения онтологии это не совсем корректно, но в целях упрощения схемы принято временное решение об отнесении констант к условиям. Соответственно, класс **EnvironmentConditions** содержит указание на агрегатное состояние вещества и ссылок на набор классов **NumericalData**, в которых указываются численные значения свойств с заданной погрешностью.

Класс **Dimensions**: Определяет размерности физических величин.

Dimensions
1. Decimal Name

Класс **Uncertainties**: Определяет тип погрешности физической величины.

Uncertainties
1. Decimal Name

Класс **EnvironmentConditions**: Задаёт перечень свойств, определяющих условия среды, в которых находится вещество, численные значения свойств вычисляются с помощью класса **NumericalData**.

EnvironmentConditions
1. Properties inProperty 2. Substances inSubstances 3. States State

Приводимые далее два класса (**Data**, **Publication**) вводят внешние источники, дополняющие набор данных, например, справочную информацию по молекулярным весам, а также сведения о публикациях, откуда приняты наборы данных.

Класс **Data**: Определяет перечень данных из справочников физических величин.

Data
1. Decimal Value 2. Decimal UncertaintyValue 3. Publication Datasource 4. Properties inProperty 5. Uncertainties Uncertainty

Класс **Publication**: Определяет источники данных для классов **NumericalData**, **Data** и **Functions**.

Publication
1. Decimal Name

Наконец, последняя группа состоит из двух основных классов (**Functions**, **DomainOfFunctions**) и двух потомков класса **Functions**: **ControlFunc** и **ComputingFunc**. В совокупности они решают задачу вычисления свойств по формулам при контроле допустимой области изменения аргумента и функций, а также заранее установленных соотношений между различными свойствами, которые в экспериментальных данных выполняются с точностью до некоторой погрешности. Функции делятся на два типа: «вычислительные функции» и «контрольные функции». Вычислительные функции дополнительно содержат указание на вычисляемое свойство, тип и величину погрешности. Результатом вычисления функции является значение свойства, помещаемое в БД. Контрольные функции являются булевскими и отвечают на вопрос, выполняется ли заданное соотношение с требуемой степенью погрешности или нет. Формулы записываются с помощью языка MathML[12].

Класс **Functions**: Определяет функции для вычисления и проверки корректности значений физических величин.

Functions
1. String FuncFormula
2. Publication Datasource
3. DomainOfFunction Domain
4. States inStates

Класс **DomainOfFunction**: Определяет перечень аргументов и ограничений физических свойств для функции вычислителя. (Область определения, содержание будет дорабатываться)

DomainOfFunction
1. Properties inProperty

Класс **ControlFunc** (Подкласс класса **Functions**): Определяет перечень проверочных функций, которые определяют, выполняется ли заданное соотношение с требуемой степенью погрешности.

ControlFunc
1. Decimal RequiredUncertainty

Класс **ComputingFunc** (Подкласс класса **Functions**): Определяет перечень функций для вычисления значений свойств.

ComputingFunc
1. Decimal UncertaintyValue
2. Uncertainties Uncertainty
3. Properties Property

В итоге, предложенная онтология формализует предметную область до уровня, позволяющего для «суженной» предметной области охватить практически все виды представляемых в литературе или компьютерных средах данных по свойствам как в виде таблиц, так и математических выражений. При этом возможно произвольное расширение на новые виды свойств, единицы измерения, способы задания неопределенности и прочие элементы, сопровождающие набор экспериментальных или справочных данных. Окончательная задача – разработка софта, обеспечивающего экспорт данных из БД или публикаций в форме, соответствующей разработанной онтологии.

4. Технология интеграции и связывания данных

Конечная цель онтологии – обеспечить возможности интеграции, то есть единообразного представления разнородных данных, отвечающих выбранной предметной области с учетом наложенных ограничений. Подход к решению подобной задачи в рамках концепций **Semantic Web** рассмотрен в публикациях [3, 4]. В данной работе предложена для целей интеграции технология **Linked Open Data (LOD)**, являющаяся порождением и развитием Semantic Web [13]. Публикация в среде LOD предоставляет издателям и потребителям информации значительно больше возможностей в сравнении с простым размещением данных в традиционном “web of documents”. В отличие от гипертекста, где *линки* связывают отдельные документы, записанные в HTML, технология связанных данных обеспечивает связи между произвольными *вещами*, отмеченными в документе посредством URI, которые могут идентифицировать любые объекты, персоны или концепции. Обширный перечень руководств по технологии LOD представлен на сайте [14].

Концепция LOD базируется на трех технологиях, каждая из которых поддерживается стандартами группы W3C: Hypertext Transfer Protocol (HTTP), URI (Uniform Resource Identifier) и RDF (Resource Definition Framework). В качестве единой модели связанных данных предлагается RDF, некоторая модель для представления данных и метаданных, состоящая из пригодных для машинной обработки утверждений, каждое из которых имеет формальный вид «субъект — предикат — объект» и называется **триплетом**. RDF обеспечивает построение моделей данных, не касаясь самой семантики, отсылая за интерпретацией смысла данных к наличным в сети словарям и онтологиям. Реальный механизм, который позволяет использовать возможности словарей и приписать моделям RDF-данных семантику, состоит в аннотировании семантических метаданных посредством специального синтаксиса под названием RDF Schema или язык описания RDF словарей (RDF

Vocabulary Definition Language). Соответствующие описания словарей сами записаны на RDF, представляя определения словарей в форме графов и допуская их публикацию в среде Linked Open Data. В целом RDF Schema есть не что иное как семантическое обобщение RDF, обеспечивающее структуру для описания словарей, то есть, предметно-ориентированных классов и свойств.

Язык, определяемый спецификацией RDFS, состоит из коллекции RDF ресурсов, которые могут быть использованы, чтобы описать свойства других RDF ресурсов в предметно-ориентированных словарях. Соответствующий документ имеет статус рекомендаций, подготовленных группой W3C в 2004 г. [15].

Еще один из основных стандартов W3C – язык запросов SPARQL [16]. Это обозначение является рекурсивным акронимом от англ. *SPARQL Protocol and RDF Query Language*. SPARQL представляет собой язык запросов к данным, представленным по модели RDF, точно так же, как язык SQL обеспечивает запросы к таблицам реляционной БД, а также протокол для передачи этих запросов и ответов на них. Предоставление SPARQL-точек доступа (*SPARQL-endpoint*) является рекомендованной практикой при публикации данных во всемирной паутине.

Поскольку исходно многие документы, подлежащие интеграции посредством связывания в пространстве LOD, загружены в БД, возникла задача конверсии реляционных данных в RDF-формат. Идеи такого отображения высказаны еще в 1998 году Т. Бернерс-Ли, который обсуждал сходства и различия моделей RDF и «сущность-связь» [17]. В настоящее время разработан специальный инструмент для представления реляционных БД в качестве связанных данных на Web - D2R Сервер [18] поддерживающий большинство БД (Oracle, MySQL, PostgreSQL и др.). Сервер публикует связанные данные, позволяя клиенту запрашивать данные из БД посредством SPARQL протокола. Смысл всей операции по конверсии RDB→RDF не просто унификация формата, но и некоторое обогащение содержания за счет явного моделирования соотношений между сущностями, которые в реляционной БД были несущественны.

Исходя из доступности указанных технологий LOD, здесь предложена следующая технологическая цепочка. Сделанная запись онтологии на языке OWL обеспечивает построение БД и пользовательского интерфейса. Все внешние документы, вне зависимости от происхождения загружаются во вновь организованную БД, где формируется хранилище реляционных данных. В то же время, онтология обеспечивает описание на более высоком уровне, чем в обычной схеме данных. Без нее интеграция в пространство LOD невозможна, а SPARQL запросы делаются именно на основе OWL онтологии.

Посредством сервера D2R организуется конверсия этих данных в RDF формат при

возможности их автоматического связывания с родственными данными, доступными в сети.

Соединение графов модели данных, представленной RDF, с доступными словарями и онтологиями обеспечивает фундамент всей технологии публикации связанных данных. Технология успешна при работе с распределёнными знаниями за счет автоматического сцепления RDF-файлов, размещённых в сети любыми авторами с последующей возможностью найти в собранном документе сведения, которых не было ни в одной из его частей. Важным моментом в предварительной подготовке является поиск и подбор уже действующих словарей, например, с использованием <http://lov.okfn.org/dataset/lov>.

Потенциал технологий LOD применительно к теплофизическим данным в известной степени подтверждает накопленный успешный опыт ее применения в химии, где важнейшим аспектом является интеграция данных по структуре молекул и свойствам вещества [19].

Литература

1. R. C. Wilhoit, K.N. Marsh. CODATA Standard Thermodynamics. Rules for Preparing COSTAT Message for Transmitting Thermodynamic Data, Report to CODATA Task Group on Geothermodynamic Data and Chemical Thermodynamic Tables, Paris (1987).
2. А.О. Еркимбаев, В.Ю. Зицерман, Г.А. Кобзев, Л.Р. Фокин. Логическая структура физико-химических данных. Проблемы стандартизации и обмена численными данными. Журнал физической химии 2008. Т.82. №1. стр. 20-31.
3. А.О. Еркимбаев, А.Б. Жижченко, В.Ю. Зицерман, Г.А. Кобзев, Э.Е. Сон, А.Н. Сотников. Интеграция баз данных по свойствам вещества. Подходы и технологии. Научно-техническая информация. Сер. 2. Информационные процессы и системы. ВИНТИ РАН, 2012, №8, стр. 1-8.
4. А.О. Еркимбаев, В.Ю. Зицерман, Г.А. Кобзев, В.А. Серебряков, Л.Н. Шиолашвили Интеграция данных по свойствам веществ и материалов на основе онтологического моделирования предметной области. Труды XV Всероссийской объединенной конференции «Интернет и современное общество» (IMS-2012), Санкт-Петербург, Россия, 2012, стр. 38-47.
5. Hodgson R, Keller PJ: QUDT - Quantities, Units, Dimensions and Data Types in OWL and XML 2011. <http://qudt.org/>

6. International Union of Pure and Applied Chemistry. Compendium of Chemical Terminology. Gold Book, Version 2.3.2. 2012-08-19, <http://goldbook.iupac.org/>
7. А.А.Бездушный, А.Н.Бездушный, А.К.Нестеренко, В.А.Серебряков, Т.М.Сысоев, К.Б.Теймуразов, В.И.Филиппов. Информационная Web-система «Научный институт» на платформе ЕНИИП. ВЦ им. А.А. Дородницына РАН, Москва 2007.
8. T. Ashino. Materials ontology: an infrastructure for exchanging materials information and knowledge. Data Science Journal, 2010, V. 9, pp. 54-61.
9. Th. Gruber, <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>
10. Электронная библиотека учебных материалов по химии. База данных ИВТАНТЕРМО. www.chem.msu.su/rus/handbook/ivtan/welcome.html
11. J.H. Dymond, K.N. Marsh, R.C. Wilhoit et al. Virial Coefficients of Pure Gases and Mixtures. Springer-Verlag Berlin Heidelberg 2002
12. www.w3.org/Math/
13. T. Berners-Lee. Design Issues: Linked Data. Online at www.w3.org/DesignIssues/LinkedData.html
14. <http://linkeddata.org>
15. <http://www.w3.org/TR/rdf-schema/>
16. www.w3.org/TR/rdf-sparql-query
17. T. Berners-Lee. [Relational Databases on the Semantic Web](http://www.w3.org/DesignIssues/RDB-RDF.html), , Design Issue Note, 1998-2009, www.w3.org/DesignIssues/RDB-RDF.html
18. <http://d2rq.org/d2r-server>
19. J.G. Frey, C.L. Bird. Cheminformatics and the Semantic Web: adding value with linked data and enhanced provenance//. WIREs Comput. Mol. Sci. – 2013. - doi: 10.1002/wcms.1127; <http://onlinelibrary.wiley.com/doi/10.1002/wcms.1127/pdf>

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