**Прикладная онтология по теплофизике.**

**Проблема сведения**

Зеленчук А.М.a, Прядухин И.Ю.b, Молородов Ю.И.c, Фазлиев А.З.d

a Новосибирский государственный университет, Новосибирск

b Томский государственный университет, Томск

c Институт вычислительных технологий СО РАН, Новосибирск

d Институт оптики атмосферы СО РАН, Томск

*andrey.zelenchuk@gmail.com**,* *yumo@ict.sbras.ru**,* *faz@iao.ru*

*В работе рассмотрена проблема построения индивидов в прикладной OWL-онтологии по теплофизике. Представлены примеры описания индивидов,* *описывающих теплофизические свойства воды в жидком состоянии, простой источник данных о скачке плотности в литии и набор значений скачков плотности в литии.*

**Ключевые слова**: проблема сведения, онтология по теплофизике, источники данных.

*A problem of constructing individuals of OWL-ontology of thermophysics is discussed. The individual instances describing the thermodynamic properties of water in liquid state, a simple data source on change in density and a set presented values of change in density in lithium are represented.*

1. **Введение**

Основное отличие прикладной онтологии по той или иной части физики от онтологии физики или онтологий верхнего уровня состоит в том, что ключевой проблемой такой онтологии является решение проблемы сведения. Проблема сведения в логическом анализе научных знаний [1] сводится к нахождению представления сложных научных терминов через простые термины. Решение этой проблемы (проблемы сведения) состоит в «указании способов построения научных терминов, так чтобы были выявлены все связи терминов по значению вплоть до простых по построению терминов» [1]. Указание способов состоит в первую очередь в выборе представления знаний, в рамках которого описывается такое построение.

Последние 15 лет в компьютерных науках развивается подход Semantic Web (SW) [2], развивающий стандарты представления данных, информации и знаний, аккумулируемых в информационных ресурсах сети Интернет. Например, развитие формального языка XML способствовало созданию языка ThermoML [3,4] в рамках которого формализованы термины, используемые для описания термодинамических свойств. Такая формализация оказалась полезна для обмена данными между научными сообществами. Авторы работы [5] указали на ограничения использования ThermoML, подчеркнув возможность расширения номенклатуры терминов, которая не предсказуема и может повлечь их переопределение. Они предложили для представления данных использовать концепцию полуструктурированных данных [6], но осталось неясным с помощью какого формального языка предлагается его делать.

Наряду с языком XML в подходе SW развит язык OWL для описания семантических аннотаций информационных ресурсов и построения OWL-онтологий для систематизации этих ресурсов. Описание информационных ресурсов в виде семантических аннотаций, доступное как для агентов, так и для исследователей. В информационной системе такая онтология характеризует те или иные аспекты предметной области.

В этой работе мы демонстрируем, каким способом можно решать проблему сведения при описании физико-химических свойств химических веществ.

1. **Концептуализация предметной области**

Представление знаний в предметной области представляет собой подход, основанный на трех дисциплинах: логике, концептуализации и вычислениях [7]. В этом параграфе кратко рассмотрена концептуализация предметных областей, необходимая для построения прикладной онтологии по теплофизике. Более подробное изложение концептуализации предметных областей «Химическое вещество» и «Физические величины и теплофизические свойства» дано, например, в работе [3]. Предметная область «Информационные ресурсы в теплофизике» обсуждается в [5]. В разделе 2.3 введен термины «источник данных» и «значение физической величины». Первый термин уточняет термин «набор данных», а второй уточняет термин «свойство», используемые в работе [5].

* 1. **Химическое вещество**

Существует несколько способов идентификации химического вещества. Одиночные атомы (молекулы, комплексы и т.д.) идентифицируют по атомарному составу (стехиометрическая формула), названию, идентификатору CAS, идентификаторам InChI, InChI-key, электронным состояниям и т.д. Простые и сложные вещества, состоящие из атомов (молекул, комплексов и т.д.), классифицируют по равновесным состояниям.

Простые и сложные вещества могут находиться в разных равновесных состояниях. Выделяют однофазовые равновесные состояния (газообразное, твердое, жидкое, плазменное и др.), двухфазовые равновесные состояния (межфазные границы: жидкость-газ (кривая насыщения), твердое-газ (кривая сублимации), твердое-жидкость (кривая плавления), и др.) и трехфазные равновесные состояния (точки равновесия: тройные и критическая точки).

Для описания простых и сложных веществ используют разные модели, например, идеальный газ, бездефектный кристалл и т.д.

* 1. **Физические свойства, используемые в теплофизике**

Физические свойства, используемые в теплофизике можно отнести к разным группам: базовые свойства, объемные свойства, калорические (энергетические) свойства, механические, свойства связанные с переносом тепла и др. Общее число свойств более ста. Эти свойства могут применяться только для определенных фазовых состояний или для всех состояний, могут относиться ко всем веществам или только к части из них, и т.д. Более того с течением времени появляются новые свойства веществ и новые, неизвестные ранее, фазовые состояния вещества. Другими словами, предметная область «Теплофизические свойства веществ» является открытой предметной областью.

* 1. **Информационные ресурсы**

Полученные при измерениях и вычислениях результаты представляются в форме компьютерных информационных ресурсов, внутренняя структура которых создается с помощью формальных языков, построенных на основе логики. Ресурсы могут содержать в себе данные, информацию и знания [8].

Представление данных о термодинамических величинах на современном уровне описано в [3]. На сложности обусловленные открытостью предметной области «Теплофизические свойства» обратили внимание в работе [5]. В этой работе была представлена логическая структура физико-химических данных, основанная на концепции полу-структурированных данных. В ней авторы ввели в научный оборот термин «набор данных» и привели схемы блока данных «свойства» и представление метаданных в блоке данных «свойства». Такой подход характерен при построении информационных систем двухслойной архитектуры (слой данных и информационный слой).

Мы используем другие термины, уточняющие термин «набор данных». Уточнение, с одной стороны, связано со спецификацией этого термина в части происхождения этих данных (решением какой задачи являются эти данные, каким методом решена эта задача и какие входные данные использованы), числа веществ и публикаций, входящих в описание набора данных, а с другой стороны, с формальным описанием этого термина на языке OWL DL. Такая спецификация необходима для расширения круга информационных задач, решаемых на этом наборе данных.

При построении представления знания, имея концептуализацию и формальный язык для его описания, необходимо сформулировать задачу, для решения которой оно предназначено.

Для информационных ресурсов по теплофизике важной является задача построения экспертных источников данных, интенсионал (физические величины) которых определяется исследователями прикладных наук, и на экстенсионал (значения физических величин) наложены ограничения, следующие из качества данных необходимого для решения задач прикладных наук. Входными данными для решения этой задачи должны быть данные, обеспечивающие анализ качества данных в предметной области.

В силу того, что для разных прикладных наук требуются разные интенсионалы экспертных данных и для решения разных задач в рамках одной прикладной науки качество данных может быть разным, необходимо оценивать вариабельность структур данных и уровня качества данных.

Требование повышения уровня качества данных существенным образом влияет на структуры входных данных задачи построения экспертных источников данных.

Дадим определение новых терминов для использования в онтологии по теплофизике. Первичным источником данных называются источник, данные из которого характеризуют свойства вещества в равновесном фазовом состоянии определенного типа и являются решением вычислительной задачи или результатом измерения, выполненного одним методом. Все прочие источники данных назовем составными.



Рис.1. Общая схема источников данных.

На рис.1 показана общая схема источников данных. Фигурные скобки означают, что для описания составного источника данных могут использоваться несколько публикаций, несколько веществ и несколько вычислительных задач, решенных несколькими методами и для разных входных данных. Заметим, что использование полного набора первичных данных эффективно для оценки доверия к экспертным данным [9].

В силу того, что при составлении онтологий в физике значительная доля проектируемых свойств связана с физическими величинами обратимся к описанию свойств, характеризующих физические свойства вещества. Мы предлагаем использовать для описания значений физических величин в теплофизике структуру, показанную на рис.2. Она получена как производная от двух схем данных «физико-химические свойства» и «метаданные физико-химического свойства», описанных в работе [5]. Такое видоизменение обусловлено применением для построения языка OWL DL в котором «property» является конструктом этого языка, для которого исследователь не может создавать свойства.



Рис.2. Общая схема значения физической величины

Мы опустим детальное описание сущностей «количественное значение», «точность вычисления» и «тип точности», предполагая, их понимание читателем.

3. **Онтологическое описание предметной области**

Язык спецификации OWL 2 DL используется для построения онтологий, являющихся одним из видов представления знаний. Ключевыми конструктами языка являются понятия «Class», «Statement», «Property» и «Individual». Классы, свойства и индивиды определяются высказываниями, содержащими свойства, и связаны с отношениями и концептами, формирующими концептуализацию предметной области. Класс определяется с помощью заданных правил, позволяющих формировать ограничения, которым должны удовлетворять индивиды, входящие в его расширение. В широком смысле эти ограничения можно интерпретировать как вопрос, ответом на который является расширение, определяемое этими ограничениями класса.

Следуя этим рассуждениям можно сделать вывод, что для построения онтологии предметной области необходимо определить на какие вопросы в предметной области она должна предоставлять адекватные ответы. Поскольку понятийная часть OWL-онтологии включает в себя таксономию классов, то эту таксономию можно рассматривать как иерархию вопросов, на которые онтология содержит ответы. Поскольку свойства классов, объектных свойств (object property) и конкретных свойств (datatype property) определяются только спецификацией OWL DL, не исследователем, то наиболее значительный произвол в построении онтологий связан с конструированием индивидов. Все свойства в онтологии, в том числе и создаваемые исследователем, определяются заданием их областей определения и значения. Таким образом, проблема сведения является корневой при построении прикладных онтологий.

**4. Индивиды, характеризующие свойства веществ в определенном фазовом состоянии**

Индивиды представляют в OWL-онтологии независимые друг от друга сущности (концепты, объекты) которые могут существовать некоторый конечный промежуток времени. В предметной области «Количественная теплофизика», состоящей из трех предметных областей, можно ввести значительное число индивидов. Часть этих индивидов можно разбить на группы каждую из которых можно отнести к одной из трех предметных областей.

В этом параграфе рассмотрен пример формирования индивида, относящегося к предметной области «Химическое вещество».



Рис.3. Фазовая диаграмма состояний воды (https://en.wikipedia.org/wiki/Phase\_diagram)

Химическое вещество может находиться в разных фазовых состояниях, в том числе одновременно в разных (например, жидком, газообразном и твердом). Для всех этих случаев можно создавать соответствующие индивиды. Ниже приведем пример индивида для воды в жидком состоянии для двух случаев: однофазном состоянии и многофазном состоянии, когда каждая из фаз включает в себя жидкое состояние. На рис.3 представлена фазовая диаграмма воды, на которой в системе координат (Температура-Давление) указаны области существовании одно-, двух- и трехфазных состояний воды.

Для построения субъектно-предикатных структур, характеризующих индивиды в языке OWL DL, используем следующие обозначения. Индивиды будем представлять в виде прямоугольных четырехугольников, внутри которых первая строка является именем индивида. Последующие строки содержат пары «свойство-объект». В этих парах объекты, выделяются жирным шрифтом. Конкретные (datatype) свойства записываются наклонным шрифтом. Дуги соответствуют свойствам. Имена свойств легко восстанавливаются по значению свойства и индивиду (субъекту) к которому оно относится.

**4.1. Индивиды, описывающие фазовые состояния воды**

На рис.4 представлены два варианта индивида, описывающего свойства жидкой воды. Первый вариант характеризует однофазовое жидкое состояние воды, второй вариант содержит также двух и трех фазные состояния, включающие в себя жидкое состояние. Оба варианта соответствуют абстрактному описанию набора фазовых состояний, а не веществу в конкретном фазовом состоянии. Выбор варианта индивида определяется задачами, стоящими перед исследователем при формировании онтологии.

В структуре каждого из двух индивидов можно выделить три части относящиеся к предметным областям «Химические вещество», «Теплофизические свойства веществ» и «Информационные ресурсы». На рис.4 индивиды, характеризующие значения теплофизических свойств и области определения фазовых состояний, приведены без их дальнейшей детализации. Структура индивидов, характеризующая значения теплофизических свойств, на рисунке не представлена, но структуру каждого из них можно свести к виду, определенному для значения физической величины (см. рис.2). При этом вопрос о том, как представлять значение физической величины в виде числа или функции от температуры и давления, здесь не рассматривается. Индивид **SinglePhaseStateOfLiquidWater** описывает цифровую модель области однофазных жидких состояний воды, а индивид **SinglePhaseStateOfLiquidWaterClosure** описывает замыкание предыдущей области, т.е. область, которая наряду с однофазными состояниями содержит состояния равновесия фаз, включающие жидкую фазу воды: кривые плавления и насыщения, тройные точки и критическую точку. Эти индивиды являются составными источниками данных о теплофизических свойствах лития в жидкой фазе.



Рис.4. Индивид «Жидкое состояние воды»

Важным является выделение индивидов вида **SinglePhaseStateOfLiquidWater\_Field,** описывающих области (кривые и точки) равновесных состояний и физические свойства вещества в этих состояниях.

**4.2. Индивиды, описывающие источники данных**

Обратимся к примеру, позволяющему понять, как структура индивида связана с задачами, и, следовательно, вопросами на которые в онтологии можно получить ответы. Рассмотренные ниже примеры являются примитивными, но позволяют сделать ряд выводов, использование которых полезно при конструировании индивидов сложных онтологий.

Рассмотрим простую задачу поиска публикаций, в которых опубликованы значения теплофизических свойств. Для примера выбрана задача систематизации публикаций по скачку плотности (change in density) при переходе от твердого к жидкому состоянию для лития. Индивид **Li\_MeltingCurve** характеризует двухфазное состояние лития (твёрдое-жидкое), описываемое кривой плавления, которой соответствует цифровая модель PT-кривой (индивид **PT-Line2**). Индивиды **Primary, DP** и **Measurement** характеризуют тип источника данных, название метода и тип метода измерения использованного в эксперименте. Индивиды **Primary** и **Measurement** можно детализировать с необходимой степенью детализации. Этот индивид характеризует первичный источник данных в соответствии со схемой на рис.1.



Рис.5. Индивид «Описание скачка плотности при плавлении лития в статье 1914\_BeCa»

Набор таких индивидов представляет каталог публикаций, в которых представлены значения скачка плотности для лития. Онтологию, в которой собраны все подобные индивиды для лития, можно назвать онтологией информационных ресурсов по скачку плотности лития.

**4.3. Индивиды, описывающие наборы значений физических величин**

Второй пример построения индивида связан с задачей систематизации скачков плотности химических веществ. В этой онтологии каждый индивид связан с одним веществом, находящимся в состоянии, соответствующем кривой плавления, со всеми публикациями, в которых исследован скачок плотности для этого вещества и всеми значениями скачка плотности. Предполагается, что для каждого вещества существует только один индивид.



Рис.6. Индивид «Опубликованные описания скачка плотности при плавлении лития»

В индивиде **Li\_MeltingCurve\_Change\_in\_Density** состояние вещества описывается по аналогии с предыдущим примером (см. рис. 5). К индивидам, связанным с предметной областью «Теплофизические свойства веществ», относятся индивиды **1914\_BeCa\_Li\_MeltingCurve\_Change\_in\_Density** и **1935\_Losana\_Li\_MeltingCurve\_Change\_in\_Density**. Каждый из них (см. рис.6) характеризует публикацию из которых извлечены значение скачка плотности, метод с помощью которого измерено это значение и тип источника данных.

Набор таких индивидов представляет собой каталог информационных ресурсов, в котором представлены значения скачка плотности для ряда веществ. Онтологию, в которой собраны все подобные индивиды для разных веществ, можно назвать онтологией информационных ресурсов по скачку плотности химических веществ.

Дальнейшее усложнение структуры индивидов такого вида можно связать с задачами анализа качества опубликованных значений скачка плотности химических веществ. Таким образом, дополняя структуру индивида можно строить новые индивиды, которые будут представлять ответы на новые вопросы.

Поскольку построение индивидов может в разных научных сообществах происходить в зависимости от решаемых ими задач, то со временем накапливается значительное число онтологий по предметной области. Формально хорошим выбором является использование онтологии являющейся результатом онтологического соглашения [11]. Если же такое соглашение отсутствует, то пользователю приходится делать выбор на основании метрик онтологий и круга задач, которых ему необходимо решать.

**Заключение**

В работе рассмотрена проблема сведения в прикладной онтологии. При построении OWL-онтологий эта проблема является центральной и ее решение зависит от задач для решения, которых используются прикладная онтология. На примере индивидов, описывающих воду в жидком состоянии, простой источник данных о скачке плотности в литии и набор значений скачков плотности продемонстрировано решение проблемы сведения.

Дальнейшее направление работ связано с построением индивидов, характеризующих качество данных по теплофизике в рамках спецификации [10], определяющей словарь терминов, позволяющих адекватно описать качество наборов данных; построением таксономий классов прикладной онтологии по теплофизике, и информационной системы трехслойной архитектуры для представления информационных ресурсов по теплофизике.

**Литература**

1. Зиновьев А.А., Основы логической теории научных знаний, М., Из-во ЛКИ, 2010, Изд. 2-е доп. и испр., 264 С.
2. T. Berners-Lee, J. Hendler and O. Lassila, The Semantic Web, Scientific American, May 17, 2001.
3. XML-based IUPAC Standard for Experimental, Predicted, and Critically Evaluated Thermodynamic Property Storage and Capture (ThermoML). IUPAC project 2002-055-3-024.
4. Frenkel M*.* // Pure Appl. Chem. 2005. V. 77. № 8.P. 1349.
5. А.О.Еркимбаев, И.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, с.20-31.
6. Когаловский М.Р, Энциклопедия технологий баз данных, М.: Финансы и статистика. 2002.
7. J.Sowa, Knowledge Representation, Brooks/Cole, 2000, 594p.
8. D.De Roure, N.R. Jennings, N.R. Shadbolt, The Semantic Grid: A Future e‐Science Infrastructure, chapter 17, In book: Grid Computing: Making the Global Infrastructure a Reality, pp.437 – 470.
9. A. Fazliev, A. Privezentsev, D. Tsarkov, J. Tennyson, Ontology-Based Content Trust Support of Expert Information Resources in Quantitative Spectroscopy, In book: Knowledge Engineering and the Semantic Web, Communications in Computer and Information Science, V. 394, Springer, Berlin, Heidelberg, Eds: P. Klinov, D. Mouromtsev, pp.15-28.
10. Data Quality Vocabulary, W3C First Public Working Draft 25 June 2015, [http://www.w3.org/TR/2015/WD-vocab-dqv-20150625/](http://www.w3.org/TR/2015/WD-vocab-dqv-20150625/%D1%8E), Eds: R. Albertoni, C. Guéret, A. Isaac.
11. Data Quality Vocabulary, W3C First Public Working Draft 25 June 2015, [http://www.w3.org/TR/2015/WD-vocab-dqv-20150625/](http://www.w3.org/TR/2015/WD-vocab-dqv-20150625/%D1%8E), Eds: R. Albertoni, C. Guéret, A. Isaac.
12. Guarino N. Formal ontology in information systems / Guarino N. // Proc. of Conf. on Formal Ontology in Information Systems (FOIS'98). –Trento, 6–8 June 1998. Amsterdam: IOS Press, 1998. P. 3-15.