

А. М. Федотов^{1,2}, Ю. И. Молородов¹, А. М. Зеленчук²

¹ *Институт вычислительных технологий СО РАН
пр. Акад. Лаврентьева, 6, Новосибирск, 630090, Россия*

² *Новосибирский государственный университет
ул. Пирогова, 2, Новосибирск, 630090, Россия*

fedotov@sbras.ru, yumo@ict.sbras.ru, andrey.zelenchuk@gmail.com

КОНЦЕПЦИЯ И АРХИТЕКТУРА ИНФОРМАЦИОННОЙ СИСТЕМЫ ДЛЯ ИССЛЕДОВАНИЯ ТЕПЛОФИЗИЧЕСКИХ СВОЙСТВ МАТЕРИАЛОВ

Статья посвящена созданию информационной системы (ИС), обеспечивающей интеллектуальный анализ и интеграцию знаний и информационных ресурсов по изучению теплофизических свойств материалов, в том числе органических и неорганических веществ, в широком диапазоне температур. ИС обеспечивает систематизацию и интеграцию доступных информационных ресурсов в едином информационном пространстве. Использование онтологии для описания предметной области информационной системы позволяет семантически структурировать его информационное наполнение и организовать в нем навигацию и содержательный поиск информации.

Информационная система включает систему импорта данных из файлов, в базу данных (БД), которая может находиться локально или на другом сетевом ресурсе. Для отображения данных в виде интерактивных графиков, с использованием алгоритмов сглаживания, интерполяции и аппроксимации используются ресурсы, разработанные на основе свободных программных продуктов.

Ключевые слова: информационная система, управляемая онтологией навигация, информационные ресурсы, онтология, содержательный поиск.

Введение

В последнее время в мире активно разрабатываются и используются специализированные пакеты программ (например, пакет программ REFPROP¹), которые позволяют получать данные по теплофизическим свойствам для некоторых классов веществ и материалов.

Опыт практического использования таких программ выявил их существенные недостатки. Они не дают доступа к таблицам первичных экспериментальных данных, не позволяют выбирать модели для их обработки. Используемые модели не являются физически обоснованными, что не позволяет проводить экстраполяцию данных, отсутствует также реальная возможность прогнозирования свойств и т. д. Все это делает их пригодными только для проведения инженерных расчетов, а не научных исследований. Информационная система, которая могла бы объединить достоинства существующих программных средств и устранить перечисленные выше недостатки, будет являться важным инструментом для развития теорий свойств веществ и фазовых превращений.

Интеграция информационных ресурсов в единую информационную среду и организация доступа к вычислительным ресурсам – это еще одно из важнейших направлений развития современных информационных технологий. Решение проблем создания и интеграции ин-

¹ <http://refprop.software.informer.com/>

формационных ресурсов и продуктов становится необходимым условием развития многих стран, в том числе и России.

Современное состояние проблемы

Научные базы данных содержат информацию о свойствах изучаемых объектов и являются либо чисто фактографическими, либо содержат еще библиографическую и текстовую информацию. Большая часть баз данных, особенно в России, являются локализованными в лабораториях и научных центрах и малодоступны извне. За рубежом с развитием сетевых технологий в последние десять-пятнадцать лет стали создаваться национальные и международные службы научных баз данных. Например, база данных по свойствам материалов университета Пердью (штат Индиана, США) [1] содержит фактографические базы данных по термофизическим, механическим, электрическим, оптическим и др. свойствам металлов и сплавов. Такого рода базы данных имеются и в США и Япония. Они на базе NIST (National Institute of Standards and Technology – Национальный институт стандартов и технологий, США) [2] и NIM3 (National Institute for Materials Science Technology – Национальный институт материаловедения, Япония) [3] создали и активно развивают обширные комплексы материаловедческих информационных систем, основанных на базах данных. Там же создана фактографическая база данных по аморфным материалам, которая ведется с помощью реляционной системы управления базами данных (СУБД). Она открыта для интерактивного доступа по компьютерным сетям, а новые поступления в базу данных отображаются в системе электронных телеконференций. Однако, как правило, эти базы данных недоступны для внешних пользователей. В России существует ряд специализированных баз данных (некоторые даже обладают всемирной известностью, например, ИВТАНТЕРМО, созданные в ОИВТ РАН [4] и ИМЕТ РАН [5], разработанных различными организациями, и никак не связанных друг с другом.

Предлагаемая в данной работе сетевая база данных по диаграммам состояния и физико-химическим свойствам материалов не имеет аналогов в России.

Обоснование необходимости проведения работы

Цели и задачи

Потребность многих исследователей в достоверной информации о диаграммах состояния материалов и их термодинамических свойствах чрезвычайно велика. Имеющийся в литературе фактический материал по этим вопросам разбросан по огромному числу периодических изданий, многие из которых недоступны по причине их отсутствия в библиотеках города, региона и даже страны. Имеющийся же справочный материал зачастую также плохо доступен либо устарел. Достаточно сказать, что в библиотеках такого крупного научного центра, как Екатеринбург, находятся справочники по диаграммам состояния, изданные, как правило, 30–40 лет назад. На момент написания статьи отсутствует даже знаменитый справочник Р. Халтгрена с соавторами (R. Hultgren, P. D. Desai, D. T. Hamkins, M. Gleiser, K. K. Kelley. Selected values of the thermodynamic properties of binary alloys. ASM. Metal Park. Ohio. 1973), не говоря уже о более поздних изданиях. Недоступны специализированные журналы, например, Bull. of Phase Diagrams, CALPHAD и т. п. Именно поэтому создание сетевых баз данных, позволяющих собрать, систематизировать научную информацию и сделать ее доступной широкому кругу исследователей, является актуальной задачей. До последнего времени создаваемые разными коллективами базы данных были доступны лишь ограниченному кругу пользователей. Это касается и поддерживаемого в Институте металлургии УрО РАН банка данных АСТРА.OWN [13]. Однако на современном этапе именно с помощью разработки и создания сетевой базы данных на основе www-технологии возможно сделать собранную информацию доступной максимально широкому кругу пользователей и, таким образом, существенно повысить эффективность научно-исследовательской работы.

Конечной целью предлагаемой работы является создание фактографической сетевой базы данных по широкому кругу физико-химических свойств материалов.

Для достижения указанной цели нам нужно решить следующие задачи:

1. разработать действующую модель фактографической базы данных по физико-химическим свойствам неорганических материалов;
2. организовать экспертный отбор опытных данных (как полученных в результате проведения эксперимента, так и взятых из литературных источников) для наполнения базы данных;
3. разместить в базе данных информацию о диаграммах состояний и физико-химических свойств материалов;
4. реализовать новые возможности информационного обеспечения исследователей, решающих фундаментальные задачи в области расчетов и прогнозирования свойств, чтобы создавать новые технологии и материалы с заданными свойствами.

Онтологический подход

Теплофизика – одна из дисциплин, в которых центральное место занимает работа с численными данными. При работе с ними приходится учитывать, что в публикациях и БД используются 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 при передаче данных, легко контролируема в отношении пропусков, ошибок в знаке или порядке величины и т. п. Математическая форма, избавляя от интерполяции, требует повышенной тщательности в обнаружение ошибок, легко вылавливаемых в табличной форме. Доминирующей формой в экспериментальных работах и справочниках является именно табличная форма.

Чтобы разрабатываемый портал знаний мог предоставлять пользователям описанные выше возможности, он должен не только иметь гибкие средства представления разнородной информации и содержательного доступа к ней, но и обеспечивать оперативное управление своим информационным наполнением. Всем этим целям служит информационная модель портала знаний, описывающая предметную область.

Перспективным подходом при проектировании информационной системы является моделирование предметной области с использованием онтологии [6].

Классическое определение онтологии было дано Грубером [7] и модифицировано Борстом [8]: *Онтология* – это формальная спецификация разделяемой концептуализации». Под *концептуализацией* понимается строгое описание понятий предметной области и семантических связей (отношений) между ними, в том числе средствами естественного языка. Концептуализация должна быть *разделяемой*, то есть фиксировать общие знания, которые признаны некоторой группой, а не частные знания конкретного индивидуума. Под формальной спецификацией понимается представление знаний (концептов) в формально определённом формате, например, в терминах некоторого языка, понимаемого машиной.

Таким образом, онтология позволяет описать предметную область на формальном языке, что позволяет оперировать понятиями предметной области в рамках информационной системы [9].

Именно такой подход мы используем для работы с данными о теплофизических свойствах веществ и материалов [10; 11]. Разрабатываемая онтология должна достаточно полно описывать данную предметную область. Она должна служить основой для информационных систем, работающих с теплофизическими свойствами веществ и материалов.

Одной из наиболее популярных и адекватных формальных систем для описания онтологий является OWL (Web Ontology Language), основанная на логиках описаний (Description Logics, DL). Вокруг OWL сформировано сообщество, разработаны программные интерфейсы (API), машины рассуждения и вывода (reasoners), прикладные программы.

Знания о предметной области можно разделить на два вида: первые более стабильны и постоянны, а вторые более подвержены модификациям. В соответствии с этим делением знания, записываемые на языке OWL, подразделяются на набор терминологических аксиом (*TBox, terminological box*) и набор утверждений об индивидах (*ABox, assertion box*). Понятия «класс» и «индивид» являются базовыми в OWL. Совокупность аксиом о классах и утверждений об индивидах вместе составляют базу знаний.

Объектами исследования в теплофизике являются материалы, в том числе вещества. Материал – объект, свойства которого могут в заметной степени определяться сферой производства: изготовитель, марка, технология, условия хранения и прочее. Вещество – объект, свойства которого определяются его природой: стехиометрической формулой, составом, фазой и др. На рис. 1 показаны основные понятия (концепты), которые мы выделяем среди данных о свойстве вещества.

Тогда основная запись о теплофизических данных выглядит следующим образом: «Данное вещество в данном состоянии имеет данное свойство с данным значением, полученным из данного источника».



Рис. 1. Базовые концепты в теплофизике

Существует несколько способов идентификации вещества:

- 1) номер CAS (Chemical Abstracts Service);
- 2) название;
- 3) химическая формула;
- 4) состав (для смесей).

Вещество может находиться в различных состояниях.

1. Однофазные состояния.

Фазы:

- твёрдые (отличаются ориентационным и трансляционным порядком);
- жидкости (отличаются ориентационным порядком);
- газ;
- плазма.

2. Двухфазные состояния.

Межфазные границы (линии равновесия):

- «жидкость-газ» – кривая насыщения (кипения);
- «твёрдое-газ» – кривая сублимации;
- «твёрдое-жидкость» – кривая плавления;
- «твёрдое-твёрдое»;
- «жидкость-жидкость».

3. Трёхфазные состояния.

Особые точки (точки равновесия):

- тройные точки;
- критическая точка.

Значения свойств веществ.

1. Константа.
2. Функция одной переменной:
 - аналитическая;
 - табличная (одномерная).
3. Функция двух переменных:
 - аналитическая;
 - табличная (двумерная).

Замечание. Чаще всего свойство зависит от температуры и измеряется в ходе эксперимента при различной температуре. В результате получается одномерная таблица. Если свойство зависит ещё и от давления, и измеряется при различных давлениях, то получается двумерная таблица.

Источником данных в теплофизике могут служить:

- 1) базы данных;
- 2) статьи;
- 3) данные физического эксперимента.

Архитектура информационной системы

При реализации информационной системы (ИС) используется многоуровневая архитектура клиент-сервер. По принципу разделения ответственностей бизнес-логика выносится в отдельный модуль. Графический веб-интерфейс (Web UI) взаимодействует с основным приложением (сервером) через программный интерфейс (API) по протоколу HTTP. В качестве формата запросов и ответом выбран формат JSON ². Другие приложения также могут взаимодействовать с сервером через данный программный интерфейс.

Все числовые данные, описывающие свойства материалов хранятся в нереляционной базе данных MongoDB ³. При математической обработке числовых данных могут понадобиться библиотеки алгоритмов анализа. Для этого нам необходима интеграция с пакетом прикладных математических программ, например, с пакетом Scilab [12].

Функциональные возможности ИС

Важной компонентой информационной системы является графический интерфейс пользователя. С его помощью становится возможным отображение классификация материалов (рис. 3).

На рис. 4 представлена фазовая диаграмма воды для возможных ее состояний. Она строится автоматически, исходя из имеющихся данных.

На рис. 5 представлен результат обработки данных эксперимента по исследованию зависимости плотности Никеля от температуры. Исследуются две фазы вещества (твердая и жидкая), включая фазовый переход между ними. Мы не можем обработать всю совокупность данных (из-за наличия фазового перехода), поэтому мы обрабатываем участки с гладкими кривыми. На каждом участке вычисляется и отображается аппроксимация данных. В качестве аппроксимирующих функций могут использоваться степенные полиномы, сплайн-функции или кривые Безье.

² JSON – *JavaScript Object Notation* – текстовый формат обмена данными, основанный на JavaScript <http://www.json.org/>.

³ MongoDB – документо-ориентированная система управления базами данных с открытым исходным кодом, <https://www.mongodb.org/>.

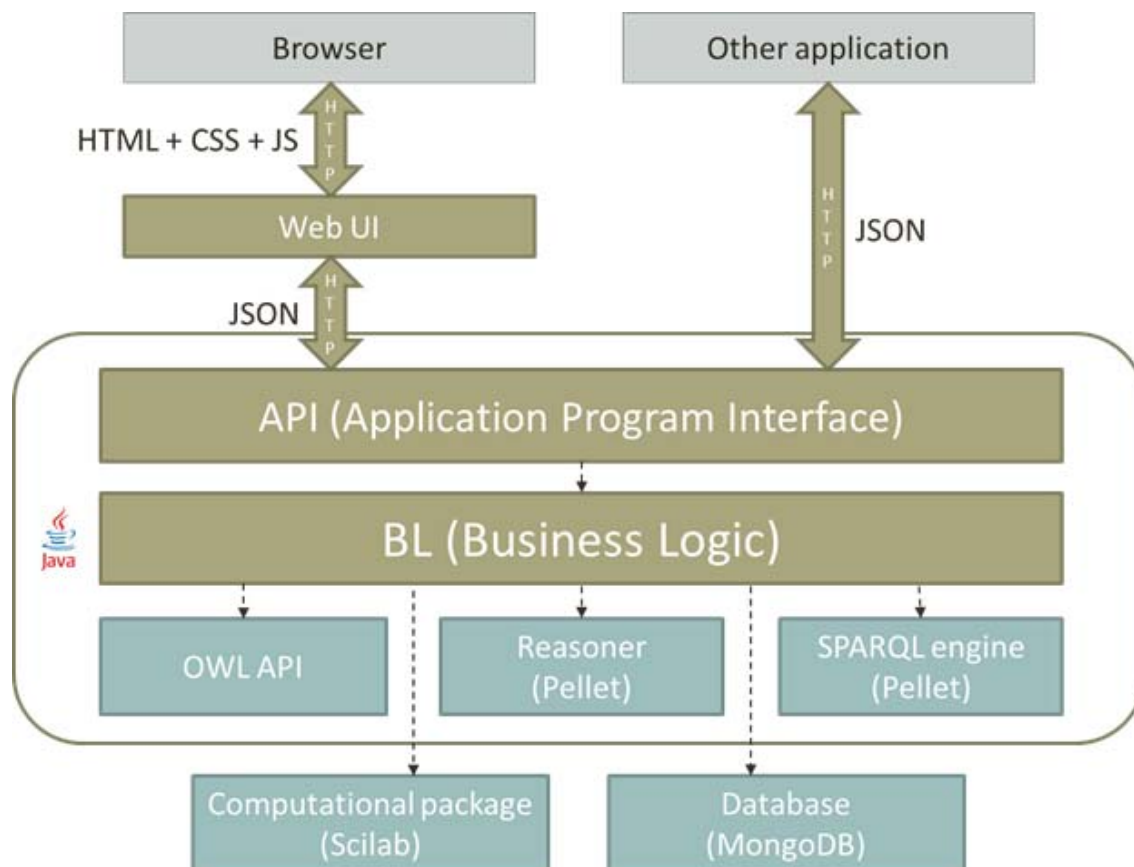


Рис. 2. Архитектура ИС

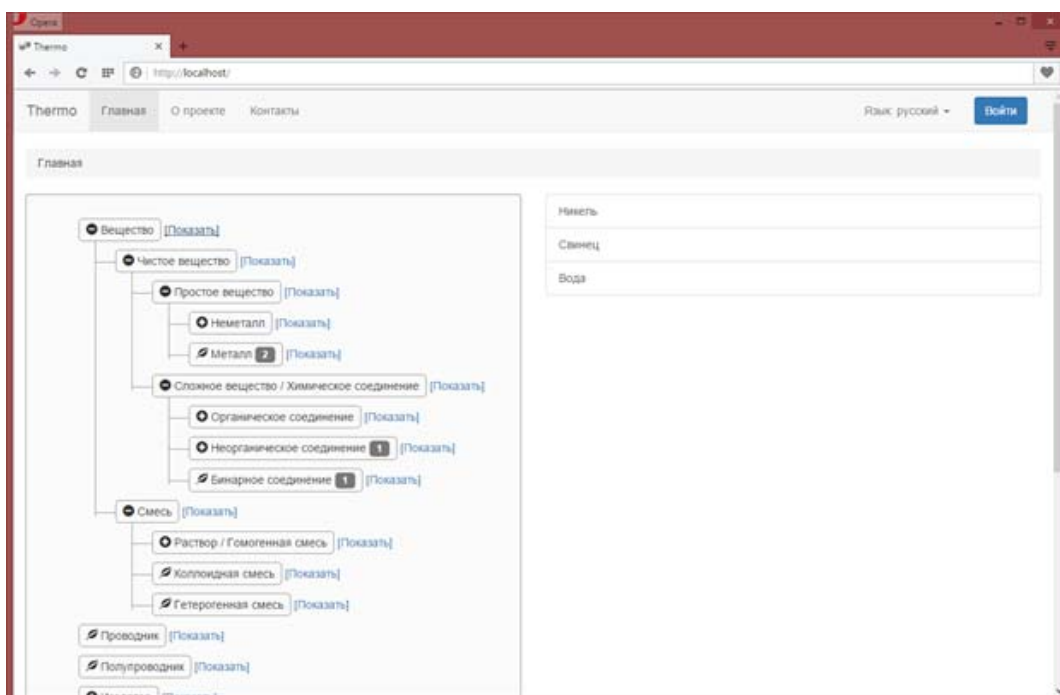


Рис. 3. Классификация материалов

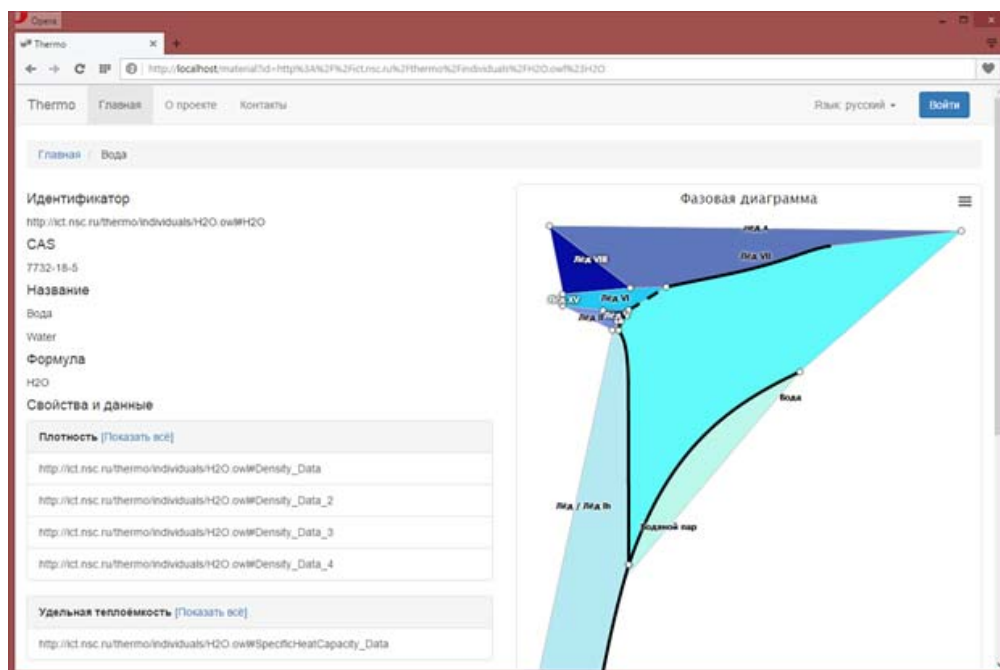


Рис. 4. Фазовая диаграмма воды

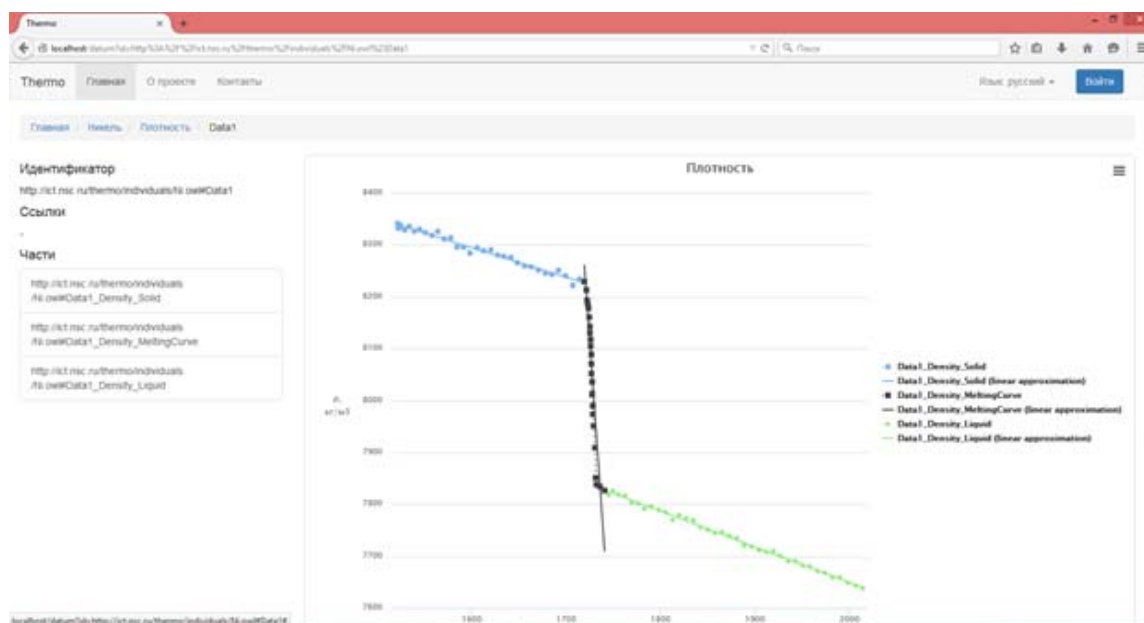


Рис. 5. Динамика изменения плотности Ni от температуры

Заключение

Разрабатываемая информационная система обеспечивает систематизацию и интеграцию знаний, первичных данных экспериментов и данных исследований, представленных в научной литературе, относящихся к изучению свойств материалов в единое информационное пространство, и содержательный доступ к ним.

Реализована возможность импорта данные из файлов, их просмотр и редактирование. Результат обработки может быть представлен как в виде таблиц, так и в виде графиков, построенных с использованием алгоритмов аппроксимации.

Список литературы

1. База данных по свойствам материалов университета Пердью. <http://www.purdue.edu/>
2. NIST Materials Database, <http://www.mst.gov/chemistry-portai.cfm>.
3. NMS Materials Database, http://mits.nims.go.jp/db_top_eng.htm.
4. *Belov G. V., Iorish V. S., Yungman V. S.* IVTANTHERMO for Windows – database on thermodynamic properties and related software. CALPHAD. 1999. Vol. 23 (2). P. 173–180.
5. *Дударев В. А., Киселева Н. Н.* Интегрированная информационная система по свойствам неорганических веществ и материалов. // Тр. XVII Междунар. конф. DAMDID/RCDL'2015 “Аналитика и управление данными в областях с интенсивным использованием данных”. Обнинск, 2015. С. 41–48.
6. *Erkimbaev A. O., Zhizhchenko A. B., Zitserman V. Y., Kobzev G. A., Son E. E., Sotnikov A. N.* Integration of Databases on Substance Properties // Automatic Documentation and Mathematical Linguistics. 2012. Vol. 46. No. 4. P. 170–176.
7. *Gruber T. R.* Toward principles for the design of ontologies used for knowledge sharing // Human-Computer Studies. 1993. No. 43. P. 907–928.
8. *Borst W. N.* Construction of engineering ontologies for knowledge sharing and reuse // СТІТ Ph. D-thesis series No. 97–14. Universiteit Twente. Enschede, 1997. 243 p.
9. *Загорюлько Ю. А., Боровикова О. И.* Информационная модель портала научных знаний // Информационные технологии. 2009. № 12. С. 2–7.
10. *Загорюлько Г. Б., Молородов Ю. И., Федотов А. М.* Систематизация знаний по теплофизическим свойствам веществ // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2014. Т. 12, № 3. С. 48–56.
11. *Барахнин В. Б., Молородов Ю. И., Станкус С. В., Федотов А. М.* Информационные технологии для задач теплофизических свойств веществ // Информатика и системы управления. Автоматизированные системы и комплексы. 2013. № 4 (38). С. 149–157.
12. Open source software for numerical computation. URL: <http://www.scilab.org>
13. *Моисеев Г. К., Ватолин Н. А., Маршук Л. А., Ильиных Н. И.* Температурные зависимости приведенной энергии Гиббса некоторых неорганических веществ. Екатеринбург, 1997. 231 с.

Материал поступил в редколлегию 18.09.2015

A. M. Fedotov, Yu. I. Molorodov, A. M. Zelenchuk

*Institute of Computational Technologies SB RAS
6 Acad. Lavrentjev Ave., Novosibirsk, 630090, Russian Federation*

*Novosibirsk State University
2 Pirogov Str., Novosibirsk, 630090, Russian Federation*

fedotov@sbras.ru, yumo@ict.sbras.ru, andrey.zelenchuk@gmail.com

THE CONCEPTION AND ARCHITECTURE OF THE INTERNET PORTAL FOR THE STUDY OF THERMOPHYSICAL PROPERTIES OF MATERIALS

The article is devoted to the creation of an information system (IS), providing intelligent analysis and integration of knowledge and information resources for the study of thermal properties of materials, including organic and inorganic materials in a wide temperature range. EC provides a systematization and integration of available information resources in a single information space. Using

ontology to describe the domain of information systems allows semantically structured its content and organizes it to navigate and find meaningful information.

The information system includes a system of import data from files in the database (DB), which can be locally or on another network resource. To display the data in the form of interactive graphs, using smoothing algorithms, interpolation approximation and the resources are used, based on the free software.

Keywords: Information system, thermophysical properties of materials, information resources, ontology, meaningful search, ontology-driven navigation.

References

1. <http://www.purdue.edu/>
2. NIST Materials Database, <http://www.mst.gov/chemistry-portai.cfm>
3. NMS Materials Database, http://mits.nims.go.jp/db_top_eng.htm.
4. Belov G.V., Iorish V.S., Yungman V.S., IVTANTHERMO for Windows – database on thermodynamic properties and related software. CALPHAD. 23(2) p. 173-180, 1999.
5. Dudarev V.A., Kiselyova N.N. Integrated Information System on the properties of inorganic substances and materials. // Tr. XVII International Conference DAMDID / RCDL'2015 "Analytics and data management in areas with intensive use of data", Obninsk, Russia, October 13-16, 2015. On 41-48. . (in Russian).
6. Erkimbaev A.O., Zhizhchenko A.B., Zitserman V.Y., Kobzev G.A., Son E.E., and Sotnikov A.N. Integration of Databases on Substance Properties: // Automatic Documentation and Mathematical Linguistics. 2012. Vol. 46. No. 4. pp. 170–176.
7. Gruber T.R. Toward principles for the design of ontologies used for knowledge sharing // Human-Computer Studies, No. 43, 1993. pp. 907-928.
8. Borst W.N. Construction of engineering ontologies for knowledge sharing and reuse // CTIT Ph. D-thesis series No. 97-14. Universiteit Twente. Enschede. 1997. 243 pp.
9. Zagorulko Y.A., Borovikova O.I. Information Model Portal of scientific knowledge // Information technologies. – 2009. – № 12. P. 2-7. (in Russian).
10. Zagorulko G.B., Molorodov Y.I., Fedotov A.M. systematization of knowledge on thermo-physical properties of substances // Vestnik NSU. Series: Information technology. – 2014 – Volume 12, Issue № 3. – P. 48-56. – ISSN 1818-7900. (in Russian).
11. Barakhnin V.B., Molorodov Y.I., Stankus S.V., Fedotov A.M. Information Technology for the problems of thermal properties of substances. // Informatics and control systems. Automated systems and complexes. 2013, № 4 (38) .- P. 149-157. (in Russian).
12. Open source software for numerical computation. [Electr. resource]. – URL: <http://www.scilab.org>
13. Moiseev GK, Vatolin NA Marchuk LA, Ilyin NI The temperature dependencies of reduced Gibbs energy of some inorganic substances // Yekaterinburg: Institute of Metallurgy UB RAS, 1997. – 231 p. (in Russian).