

INFORMATION SYSTEMS CREATION USING ONTOLOGY

Yurii I. Molorodov¹, Kirill E. Vishnev²

¹Institute of Computational Technologies SB RAS, Novosibirsk, Russia

²Novosibirsk State University, Novosibirsk, Russia

Abstract

The approach based on ontology of subject domain to development of architecture of information systems for work with data of thermophysical properties of the materials which are stored in the distributed information sources is described. For processing of basic data of a physical experiment it is offered to use Savitzky-Golay filter with the subsequent approximation of the received dependences classical Chebyshev's orthogonal polynomials.

Keywords: ontology, thermophysics, physical properties, materials, approximation

ПОСТРОЕНИЕ ИНФОРМАЦИОННЫХ СИСТЕМ НА ОСНОВЕ ОНТОЛОГИИ

Молородов Ю.И.⁽¹⁾, Вишнев К.Е.⁽²⁾

¹ Институт вычислительных технологий СО РАН, Новосибирск

² Новосибирский государственный университет, Новосибирск

Описан подход, основанный на онтологии предметной области, к разработке архитектуры информационных систем для работы с данными теплофизических свойств материалов, хранящихся в распределенных источниках информации. Для обработки исходных данных физического эксперимента предложено использовать фильтр Савицкого-Голея с последующей аппроксимацией полученных зависимостей классическими ортогональными полиномами Чебышева.

Ключевые слова: онтология, теплофизика, физические свойства, материалы, аппроксимация.

Введение. В теплофизике, как и в любой другой науке, накапливаются огромные объемы информации, которые необходимо хранить, систематизировать, а также предоставлять содержательный доступ. Всё это вызывает потребность в создании и проектировании крупных информационных систем, которые помогут эффективно решать задачи связанные с использованием такой информации на практике.

Теплофизика - наука о свойствах вещества и материалов. В ней на первое место выходит работа с непосредственно количественными данными, которые описывают состояния веществ. Такие данные могут быть представлены в различных формах: табличная, графическая визуализация таблиц и хранимые процедуры.

Создаваемые в настоящее время информационные системы позволяют обеспечить человеку содержательный доступ к интересующей его информации, но и привлечь к обработке эмпирического материала современные методы аппроксимации и интерполяции. Их использование позволит получить необходимые знания.

Такой доступ сможет объединить различные источники данных, и дать возможность пользователю искать информацию в нескольких местах сразу. Существуют так называемые каталоги ресурсов. Такой каталог представляет собой простую базу данных, которая агрегирует сведения о различных источниках информации по интересующей предметной области, и даёт возможность централизованного поиска в этих источниках.

Именно по этой причине создание информационных систем, обеспечивающих доступ исследователей к достоверным экспериментальным данным по теплофизическим свойствам различных классов веществ и материалов является в настоящее время актуальной задачей [1-2]. Это обеспечивает, также и возможность прогнозирования свойств неисследованных материалов.

Построение онтологии предметной области. Основной задачей рассматриваемой научной среды является получение, интеграция и предоставление данных и знаний в интересах фундаментальных научных исследований в предметной области. Концептуальной основой систематизации знаний и информации предметной области являются онтологии. Как правило, они исполняют роль модели предметной области [3]. Онтология является ядром, базовым компонентом информационной модели портала. Она не только описывает систему знаний портала, но и задает формальные структуры для представления его контента. Онтология содержит понятия моделируемой области, связывающие их отношения, атрибуты понятий и отношений, ограничения на значения атрибутов, а также аксиомы, определяющие семантику понятий и отношений. Формализм, используемый в технологии построения порталов научных знаний обеспечивает описание понятий проблемной и предметной областей портала и разнообразных семантических связей между ними, а также выстраивание понятий в иерархию «общее-частное и поддержку наследования свойств по этой иерархии.

Чтобы портал знаний мог предоставлять пользователям описанные выше возможности, он должен не только иметь гибкие средства представления разнородной информации и содер-

жательного доступа к ней, но и обеспечивать оперативное управление своим информационным наполнением (контентом). Этим целям служит информационная модель портала знаний «Интеллектуальный научный интернет ресурс» (ИНИР) [2], которая объединяет модели его предметной и проблемной областей, а также описывает типы представляемой в его контенте информации. В качестве модели предметной области обычно выступает ее онтология. Онтология является ядром, базовым компонентом информационной модели портала. Она не только описывает систему знаний портала, но и задает формальные структуры для представления его контента. Онтология содержит понятия моделируемой области, связывающие их отношения, атрибуты понятий и отношений, ограничения на значения атрибутов, а также аксиомы, определяющие семантику понятий и отношений.

При построении любого портала научных знаний его онтология строится на основе двух базовых технологий: онтологии научного знания и онтологии научной деятельности.

Онтология научного знания, по своей сути, является метаонтологией. Она содержит метапонятия, задающие структуры для описания предметной области (области знаний) портала, такие как Раздел науки, Предмет исследования, Объект исследования, Метод исследования, Научный результат, позволяющие выделить в данной науке значимые разделы и подразделы, задать типизацию предметов, объектов и методов исследования, описать результаты научной деятельности.

Онтология научной деятельности является онтологией верхнего уровня и включает базовые понятия, относящиеся к организации научно-исследовательской деятельности, такие как Научный результат, Объект исследования, Персона, Публикация, используемые для описания результатов научной деятельности, мероприятий, научных программ и проектов, различного типа публикаций. В эту онтологию также включено понятие «Информационный ресурс», которое служит для описания информационных ресурсов, представленных в сети интернет.

Свойства каждого понятия описываются с помощью атрибутов и ограничений, наложенных на область их значений. Понятия базовых онтологий связаны между собой ассоциативными отношениями, выбор которых осуществлялся не только исходя из полноты представления проблемной и предметной областей портала, но и из удобства навигации по его информационному пространству и поиска информации.

Контент портала образует множество разнотипных информационных объектов (ИО) – экземпляров классов онтологии, связанных между собой отношениями, заданными в онтологии.

Функциональные возможности портала ИНИР (рис. 1) обеспечиваются с помощью взаимозаменяемых компонент (сервисов), которые обеспечивают обработку хранящейся информации и предоставляют непосредственный доступ к ней.

Понятия онтологии научной дисциплины, изучающей теплофизические свойства веществ, являются реализациями метапонятий онтологии научного знания и организованы в несколько иерархий «общее-частное», каждая из которых соответствует одному из метапонятий, представленных в этой онтологии. Эти иерархии могут быть связаны посредством ассоциативных отношений, часть которых наследуется из базовых онтологий, а часть отражает специфику конкретной предметной области.

Помимо задачи систематизации данных, должна быть решена задача визуализации этих данных в удобном для пользователя виде и обеспечена возможность ознакомления пользователя с работой методов обработки информации, принятых в данной области знаний (ОЗ). Пользователь портала не просто получит информацию об интересующем его методе или ссылку на реализацию метода. Он может тут же, на ресурсе, посмотреть примеры использования метода, запустить его, проанализировать его работу с разными входными данными.

Сервис-ориентированная составляющая. Важной возможностью ресурса является обеспечение доступа к реализующим методы web-сервисам, которые пользователи разработчики могут встраивать в создаваемые ими программные продукты и которые снабжены семантическими описаниями.

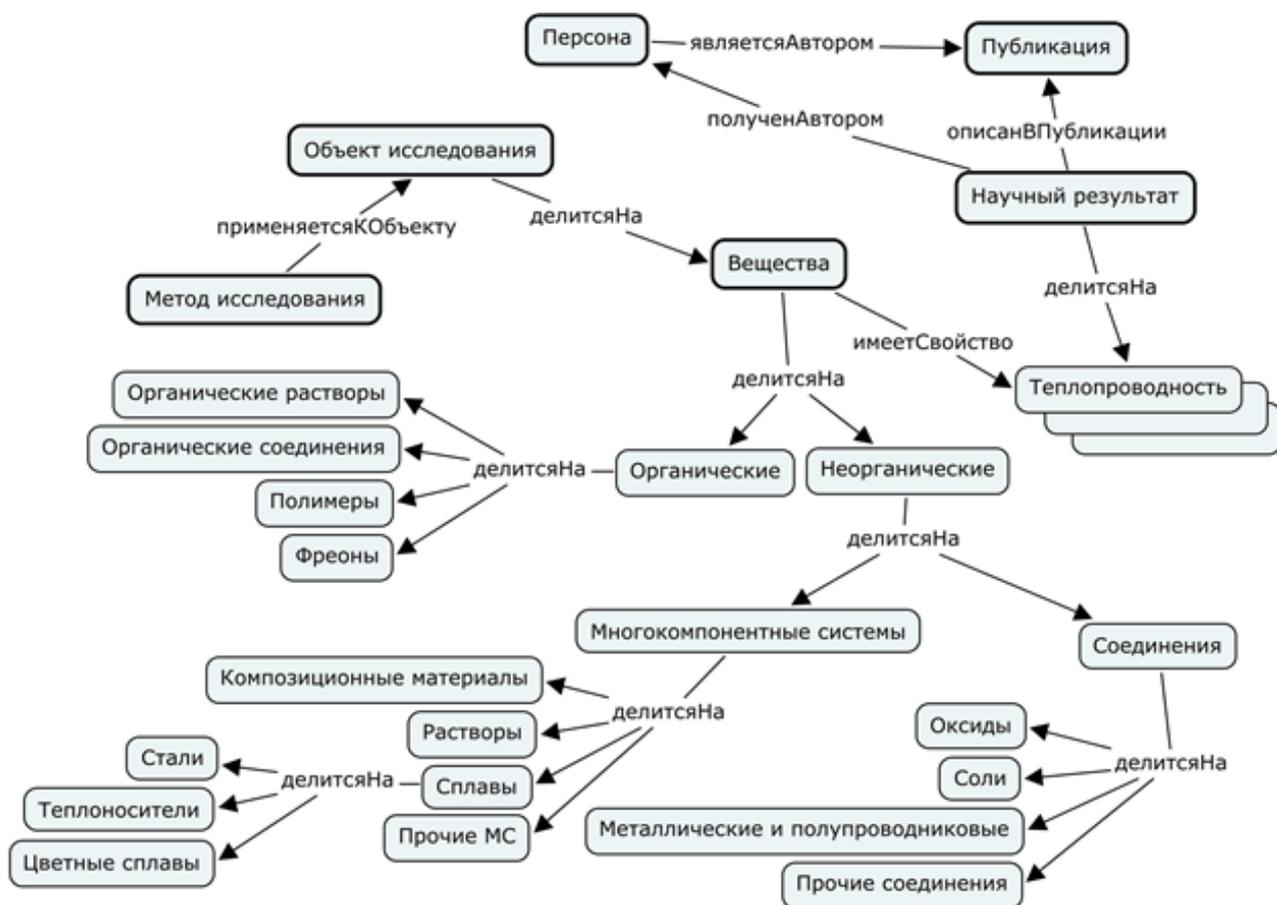


Рис. 1. Онтология портала «Теплофизические свойства химических веществ».

Основной парадигмой используемой в рассматриваемой архитектуре является сервис-ориентированный подход. Он основан на использовании распределённых, слабо связанных заменяемых компонентов, называемых сервисами, оснащённых стандартизированными интерфейсами для взаимодействия по стандартизированным протоколам.

Используя этот подход, не только разработчики портала, но и сторонние разработчики, могут создавать различные сервисы для работы с данными.

Такой подход обеспечивает несколько преимуществ:

- сервисы могут быть написаны на совершенно разных языках программирования, реализовываться на разных программно-аппаратных платформах и быть взаимозаменяемыми.
- сервисы могут размещаться на удалённых серверах, независимо от места положения ИНИР.

Поскольку взаимодействие между сервисами происходит по стандартным протоколам обмена информацией, таким как, HTTP, SOAP, WSDL, сервисы могут объединяться друг с другом для решения более сложных задач.

Несмотря на перечисленные выше преимущества, на данный момент, портал «Теплофизические свойства химических веществ» не предоставляет удобного доступа, непосредственно к экспериментальным данным о веществах, так как он хранит только формальное описание сущностей данной ОЗ. Именно этот недостаток решает сервис, разработанный в рамках данной работы.

ИНИР должна обеспечить доступ к реализующим методы web-сервисам, которые пользователи разработчики могут встраивать в создаваемые ими программные продукты и которые снабжены семантическими описаниями.

Архитектура системы доступа к экспериментальным данным. Ранее мы говорили, что онтология ИНИР содержит только метаописания информационных объектов ОЗ. Поэтому

нужно решить проблему построения системы, которая предоставит доступ к конкретным экспериментальным данным через web-интерфейс ИНИР (рис. 2).

Данная система состоит из двух независимых сервисов:

1. *Сервис загрузки данных.* В первую очередь этот сервис обрабатывает входящие `rest` запросы от онтологии ИНИР. Эти запросы содержат метаданные об информационных объектах, которые выбрал пользователь в процессе работы с онтологией. Данные объекты, как правило, относятся к категории “Научный результат”, и описывают некоторую часть таблицы со свойствами вещества и данные об удаленном хранилище (БД) где хранится эта таблица.

Загрузчик имеет собственную БД, которая содержит адреса сторонних источников и информацию. Это даёт возможность построить конкретный запрос к ним, для доступа к экспериментальным данным. Здесь могут использоваться как шаблоны SQL запросов, так и иные описания формата запроса на нужный ресурс. На основе этого и описания и полученных от ИНИР параметров, строится и выполняется запрос к указанному источнику, а полученные данные передаются на сервис визуализации данных.

2. *Сервис визуализации данных:* Представляет собой web-сервис, который помимо простой табличной и графической визуализации, позволяет пользователю, в интерактивном режиме, применить нужный метод обработки к выбранным данным и просмотреть результат. Сервис написан на фреймворке `jango` языка Python, а графическая визуализация производится с помощью `javascript` библиотеки `Plotly`, что позволяет получать динамический масштабируемый график.

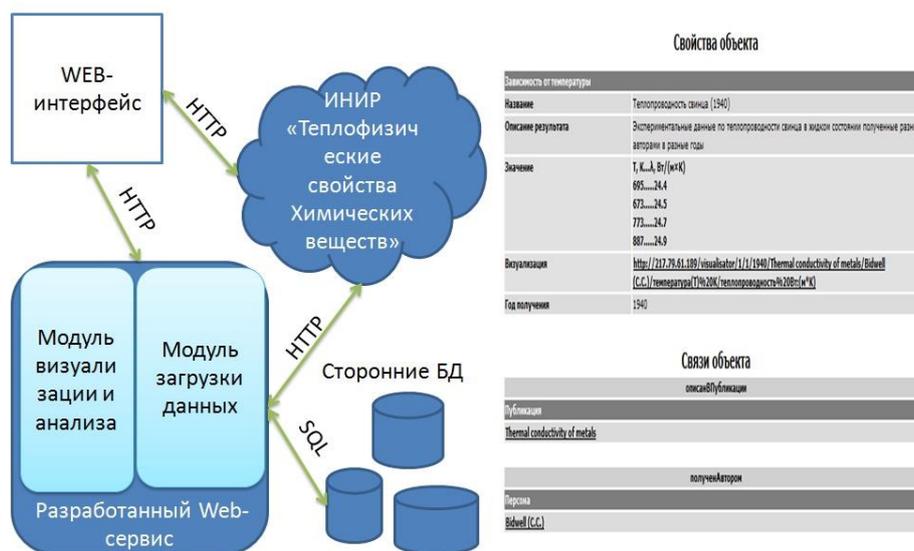


Рис. 2. Архитектура взаимодействия информационного ресурса.

Заключение. Описан способ разработки информационных систем, в основе которого положен симбиоз интеллектуального портала знаний, построенного на основе онтологического подхода. Ресурс становится «единой точкой входа», которая позволит пользователю, анализировать информацию из различных источников. Причём, пользователь получает не просто ссылки на источники с данными, но и сами данные.

Становится возможным процесс, автоматического пополнения онтологии портала ИНИР, новыми информационными объектами, через интерфейс загрузчика, а также подключение новых хранилищ с данными.

Данные из загрузчика может использовать не только визуализатор, но и другие сторонние сервисы, входящие в состав портала ИНИР.

Приложения написаны на языке Python и использовались вместе с графической библиотекой `Matplotlib` [5], что предоставляет широкие возможности для построения различных графических изображений. Пакет `SciPy` содержит все необходимые алгоритмы для обработки данных [6].

*Работа выполнена при финансовой поддержке гранта ведущих научных школ
НШ-7214.2016.9.*

ЛИТЕРАТУРА

- [1] Загорулько Г.Б., Молородов Ю.И., Федотов А.М. Систематизация знаний по теплофизическим свойствам веществ // Вестник Новосибирского государственного университета. Серия: Информационные технологии. - 2014. - Т.12. - № 3. - С.48-56.
- [2] Дударев, В.А. Интегрированная информационная система по свойствам неорганических веществ и материалов // Труды XVII Международной конференции DAMDID/RCDL '2015, «Аналитика и управление данными в областях с интенсивным использованием данных» Обнинск, 13-16 октября 2015. С 41-48
- [3] Загорулько Ю. А. Технология разработки интеллектуальных научных интернет-ресурсов, ориентированная на экспертов предметной области // Инфраструктура научных информационных ресурсов и систем. Сборник избранных научных статей. Труды Четвертого Всероссийского симпозиума (С.-Пб., 6–8 октября 2014 г.). Под ред. Е.В. Кудашева, В.А. Серебрякова. М.: ВЦ РАН, 2014 Т.1. С.69-86.
- [4] Savitzky A., Golay M. J. Smoothing and differentiation of data by simplified least squares procedures // Anal. Chem. 1964. Vol. 36. P. 1627–1639.
- [5] **<http://matplotlib.org>**
- [6] **<http://www.scipy.org>**