

Интерфейс системы интеграции и хранения данных по теплофизике

Е. К. Ключкина¹, Р. И. Хайруллин²,

¹Московский государственный университет имени М. В. Ломоносова
²ОИВТ

1. Предметная область «Теплофизические свойства веществ». Специфика данных.

Теплофизика относится к дисциплинам, в которых работа с численными данными занимает центральное место. Теплофизические свойства веществ – набор естественнонаучных данных по физическим, химическим, эксплуатационным свойствам. Такие данные широко представлены в печатных пособиях и мировых БД для широчайшего круга веществ: чистых и растворов, органических и неорганических, наноструктур и материалов, характеризующихся технологией изготовления.

Проявляется 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).

Большой проблемой на пути интеграции теплофизических данных является многообразие форматов представления в отсутствие общепринятых рецептур и стандартов записи. 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 величины и т.п. Математическая форма, избавляя от интерполяции, требует повышенной тщательности в обнаружении ошибок, легко вылавливаемых 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.

2. Ограничения на предметную область.

1. Рассматриваем только чистые вещества.
2. Все физические величины делятся на три типа: функции, константы, константы описывающие вещество (например, потенциал ионизации), аргументы.
3. Все рассматриваемые физические величины - функции считаются функциями от двух аргументов — температуры и давления.
4. Множества из п.2 не пересекаются.
5. Для каждой физической величины - функции, фиксированы области определения и значения, и характер монотонности.
6. Фиксированы типы агрегатных состояний: однофазные, двухфазные, трехфазные.

3. Возможности, предоставляемые системой пользователю и администратору

1. Пользователь может просматривать терминологические словари и запрашивать численные данные, хранящиеся в системе.
2. Пользователь может выгрузить полученные численные данные в файл XLS, CSV фиксированного формата.
3. Пользователь может просматривать расчётные функции для наборов (вещество, состояние, физ. величина), хранимые в системе.
4. Пользователь может создать задание системе на расчет численных данных по заданными математическим функциям и получить результаты вычислений на электронную почту.
5. Пользователь может загрузить в систему новые численные данные, подлежащие проверке.
6. Администратор имеет возможность пополнения и редактирования терминологических перечней/словарей (физические величины, химические вещества и т.д.), для которых в дальнейшем будут введены численные данные.
7. Администратор может создать задание на проверку корректности предложенных пользователем численных данных по заданным контрольным формулам.
8. Администратор может вводить новые расчётные функции и проверочные формулы.

4. Пользовательские интерфейсы доступа к данным БД

Пользователю предоставляется доступ к словарям - спискам химических веществ, агрегатных состояний, физических свойств (свойства-функции, свойства-константы, свойства константы вещества). Формирование списков происходит с помощью запросов к соответствующим сущностям (Chemical_Substance, State, Physical_Quantity). Каждый блок — отдельный запрос (рис. 1).

На этой же странице пользователь может выбрать Вещество, Агрегатное состояние и набор Физических свойств, по которым он желает просмотреть имеющиеся в системе численные данные. После задания своих предпочтений, пользователь перенаправляется на страницу параметризации интересующих его численных данных (рис. 2). На этой странице пользователю предлагается задать диапазоны и размерности аргументов. Кроме этого ему предоставляется список источников данных, в которых есть информация об интересующих его численных данных. Пользователь может выбрать один или несколько источников, чтобы получить данные именно из них.

На основании заданных пользователем предпочтений формируется запрос к базе данных, результат которого выводится на страницу пользователя в одном из шаблонных форматов. Существование нескольких шаблонов отображения данных обусловлено тем, что набор данных может характеризоваться

- одним источником данных и одним типом неопределенности на весь набор
- одним источником данных и неопределённостями в каждой точке измерения
- источником данных в каждой строке и неопределённостями в каждой точке
- источниками и неопределённостями в каждой строке

В зависимости от характеристики найденного набора данных результат поиска выводится на страницу в соответствующем виде (рис. 3, 4, 5).

Кроме наборов данных на странице параметризации данных с рисунка 2 пользователь может выбрать величину, для которой хочет просмотреть расчётные функции. Функции, диапазоны их применимости, погрешности и источники данных представлены на странице с рисунка 6.

Использовать существующие в системе расчётные функции пользователь может для расчёта каких-либо специальных численных данных, создавая системе задания на расчёт. Результат

вычислений можно сохранить в формате XLS или получить по электронной почте. Задание создается на странице с рисунка 7.

Также пользователь может предлагать для загрузки в систему новые численные данные, описанные в XLS, CSV файле фиксированного формата. Введенные данные сохраняются во временной схеме базы данных и ожидают проверки администратором. Страница загрузки новых численных данных представлена на рисунке 8.

На основе введенных пользователем данных на уровне интерфейса формируется таблица метаданных, служащая подсказкой для пользователя, как должен быть размечен загружаемый в систему файл с данными. Страница с таблицей показана на рисунке 9.

Перейдем к возможностям администратора. Администратор имеет возможность создавать задания на проверку предложенных пользователями численных данных путем применения к ним проверочных формул. Администратор просматривает список загруженных в систему документов с кратким описанием содержащихся в них данных (рис. 10). Затем по нажатию кнопки «Проверить документ» перенаправляется на страницу, где перечислены подходящие для проверки данных в конкретном документе формулы (рис. 11). Таким образом, администратор создает задание системе проверить корректность данных, предложенных в документе, и получает отчет. Только прошедшие проверку данные попадают в основную схему базы данных.

5. Расширенные возможности.

Кроме того существует возможность интеграции системы со сторонними источниками – хранилищами теплофизических данных. Одним из примеров такого хранилища является база данных CHEBI, доступная по ссылке <https://www.ebi.ac.uk/chebi/>. Для этого данные, хранимые системой, интегрируются с соответствующими данными на стороннем ресурсе на уровне ссылок (рис. 12).

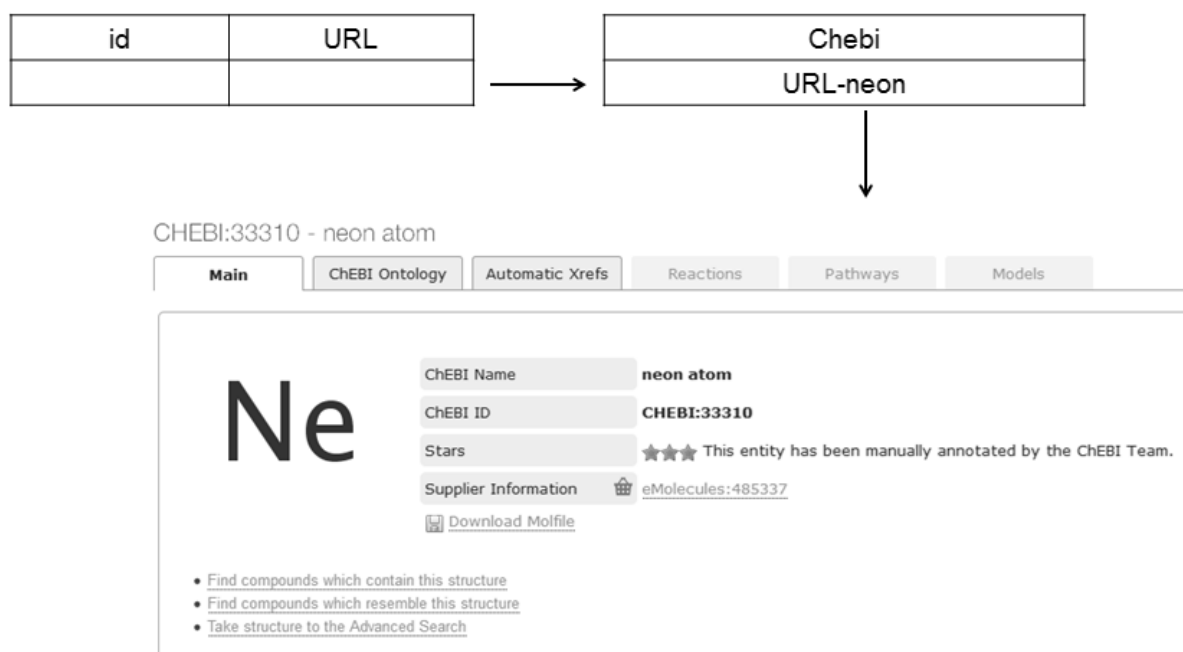


Рис. 12. Связывание со сторонним ресурсом на уровне ссылок.

Можно получать численные данные как из собственной базы данных, так и из базы интегрированного стороннего ресурса. Для этого используются веб-сервисы, предоставляемые сторонними источниками. Например, CHEBI позволяет получить данные о сущности (вещество), её онтологических предках и потомках, синонимах и пр. (рис. 13).

getCompleteEntity

Retrieves the complete entity including synonyms, database links and chemical structures, using the ChEBI identifier.

Parameter Name	Java Data Type	Value
chebid	java.lang.String	<input type="text"/>

Рис. 13. CHEBI предоставляет информацию о хранимых ею сущностях

Наконец, можно просматривать расширенную (интегрированную) модель данных. Для этого используются системы хранения моделей онтологии, которые требуют файла с онтологией .owl и копию базы данных. Одной из таких систем является VIRTUOSO.