

РАЗРАБОТКА ИНТЕРАКТИВНОГО WEB-ПРИЛОЖЕНИЯ ДЛЯ РАБОТЫ С ДАННЫМИ РЕНТГЕНОВСКОГО МАЛОУГЛОВОГО РАССЕЯНИЯ

Кучко Артём Владимирович, Смирнов Александр Витальевич

Аннотация

В работе обсуждаются требования, предъявляемые к интерактивному web-приложению, предназначенному для удаленного хранения, структурирования, визуализации и анализа данных эксперимента по малоугловому рентгеновскому рассеянию и полученной из него структурной информации. Рассмотрен опыт реализации такого приложения.

Ключевые слова: хранилище данных, параметрический поиск кривых, рентгеновское малоугловое рассеяние, интерактивное представление данных.

1. ВВЕДЕНИЕ

Малоугловое рентгеновское рассеяние (МУРР) — эффективный инструмент изучения структуры систем различной природы в диапазоне размеров 1 нм... 100 нм [1–3]. Разработка новых материалов и увеличение парка экспериментальных установок непрерывно расширяют круг объектов, в исследовании которых применяется МУРР. Благодаря росту вычислительных мощностей произошло заметное увеличение эффективности методов обработки и моделирования МУРР: появилась возможность решать за короткий срок достаточно объёмные вычислительные задачи [4, 5], были разработаны универсальные программные пакеты обработки экспериментальных данных [6–8]. Указанные факторы ведут к интенсивному накоплению экспериментального материала и результатов его анализа, а это, в свою очередь, ставит проблему организации легкодоступного хранения, упорядочения и визуализации как исходных данных натурального и модельного эксперимента по МУРР, так и полученной из этих данных структурной информации. Имеющиеся в Интернете ресурсы — сайты научных журналов или такие, как ресурс сети biosys [9], конечно, решают задачу представления результатов МУРР эксперимента. Однако такие ресурсы имеют исключительно иллюстративный характер и не предназначены для хранения данных и интерактивной работы с ними. Отметим, что существенный недостаток пакетов обработки экспериментальных данных по МУРР заключается в их зависимости от платформы, что накладывает жесткие требования на пользовательское ПО. Поэтому в последнее время активно разрабатываются web-сервисы для обработки данных МУРР [10–13]. Однако эти сервисы предназначены либо для узкоспециализированной обработки одной кривой в определенном формате, либо для расчета кривой рассеяния от определенной модельной системы и в принципе не позволяют хранить, визуализировать и проводить сравнительный анализ данных нескольких экспериментов. Выхо-

дом может служить создание интерактивного web-приложения, позволяющего хранить в разных форматах и структурировать данные об имеющихся кривых МУРР, условиях измерения кривых или параметрах их моделирования, а также создавать различные варианты визуализации данных и выполнять математические процедуры, используемые при обработке данных малоуглового рентгеновского эксперимента. В настоящей работе обсуждаются требования, предъявляемые к такому приложению, и рассмотрен пример реализации такого приложения (“SAXSDB” — Small Angle X-Ray Scattering Data Base) с помощью активно развивающихся web-технологий создания клиентских интерфейсов и серверных сценариев: EXTJS5, DbSimple, JGraph, php [14–17].

2. РАБОТА С ВХОДНЫМИ ДАННЫМИ

Основные результаты эксперимента по МУРР имеют матричный вид. Поэтому обсуждаемое web-приложение должно быть предназначено для работы с двумерными массивами числовых данных, и основная его задача состоит в агрегировании таких массивов, приведении их к единому формату хранения и выполнении последующих вычислений с ними. Данные прямого эксперимента, полученные разными исследователями и на разных установках, исходно могут быть представлены в нескольких разных форматах. Так, в качестве абсциссы кривой рассеяния может выступать угол рассеяния, модуль вектора рассеяния и пропорциональные ему величины, в качестве ординаты — количество отсчетов ФЭУ, усредненный сигнал с пикселей ПЗС-матрицы, абсолютная интенсивность или ее неопределенность. Данные в прямом пространстве могут быть представлены, например, функцией распределения парных расстояний, радиальным распределением электронной плотности внутри одного рассеивателя или распределениями объемных долей рассеивателей по характеристикам их формы и размеров. Поэтому приложение должно обеспечивать широкие возможности по работе с исходными данными разного формата. Предлагаемый способ решения этой задачи состоит в том, чтобы независимо задавать (конструировать при загрузке) формат данных, параметры парсинга данных, формат графического представления, текстовое описание содержания импортируемых колонок и заголовков осей при визуализации, а также значение дополнительного поля сортировки «тип». Панель настроек загрузки данных, реализующая такое решение в SAXSDB, представлена на рис. 1.

Тип данных (экспериментальная кривая рассеяния, модельная кривая рассеяния, функция распределения рассеивателей по радиусам инерции и т. д.) указывает пользователь, выбирая пункт из выпадающего списка при загрузке. Формат данных представлен в пользовательском интерфейсе так же, как выпадающий список, каждый элемент которого включает формат парсинга, текстовое описание и соответствующий формат визуализации. Такой подход позволяет настроить систему для упорядочения и визуализации практически любых матричных данных. Формат парсинга данных может быть реализован в трех вариантах:

- выбор формата из списка имеющихся форматов;
- выбор формата непосредственно из импортируемого файла;
- задание формата пользователем при загрузке.

Все три варианта включены в SAXSDB, причем в третьем варианте формат автоматически добавляется в список доступных форматов панели настроек. При сравнительном анализе данных представляют интерес не только основные данные (кривая рассеяния,

Settings

Parsing Mode Filter Type **Format** User

Format -- All fields are required:

Parse Format:

Description:

Plot Format:

Abscissa Title:

Ordinates Title:

Format List

...	Parse Format	Plot(X, Y)	Description
1	THETA,I,dTHETA,dI	THETA,I,THETA,dI	Scatering angle (mrad) ,Intensity (arb.un.)...
2	THETA,I	THETA,I	Scatering angle (mrad) , Intensity (arb. un.)
3	q,I	q,I	Magnitude of the scattering vector(1/nm), ...
4	q,I,sp_term	q,I	Simulated intens was gotten by scatterin...
5	Rg, PDF,dPDF	Rg, PDF,Rg,dPDF	Radius of gyration, probality distribution fu...
6	THETA,I	THETA,I	Scatering angle mrad ,Intensity
7	Q,I,dQ,dI	Q,I,Q,dI	Magnitude of the scattering vector(1/nm) ,...
8	q,I,sp_term	q,I	Simulated intens was gotten by scatterin...

Рис. 1. Лист конструирования формата вводимых числовых данных и формата графического представления

кривая распределения рассеивателей по размерам и т. д.), но и характеристики объекта исследования, условия проведения эксперимента, параметры установки. Для организации работы с такой дополнительной информацией, последнюю следует описать как набор параметров, ассоциированных с данной матрицей. Для реализации этого требования в SAXSDB пользователю предлагается задать количество параметров, описание каждого из параметров и его значение. Приложение SAXSDB может принимать на вход файлы, содержащие матричные данные в prn, csv и им родственных форматах с различными вариантами разделителя колонок. В базе данных для каждой матрицы создается отдельная таблица, где каждому числу соответствует одна ячейка таблицы. Такой подход позволяет посредством интерфейса библиотеки DbSimple взаимодействовать с БД, не используя сложных запросов. Для большей надежности хранения информации импортируемые данные дублируются на сервере в csv/prn файлах.

3. ОРГАНИЗАЦИЯ ДОСТУПА. МЕТОДЫ СТРУКТУРИРОВАНИЯ И ВИЗУАЛИЗАЦИИ ДАННЫХ

Задачи, возникающие при работе с данными в хранилище, в первую очередь — поиск и группировка данных по различным критериям. Поэтому существенным элементом приложения являются инструменты для создания и применения поисковых фильтров. Кроме того, для приложения, работающего с удаленным хранилищем, важнейшим качеством является быстрота поиска. Для создания необходимых инструментов, было

предложено использовать стандартные механизмы библиотеки EXTJS. В SAXSDB это позволило реализовать поиск данных по любому из полей, группировку числовых данных по полю, выбираемому пользователем при импорте данных, сортировку их по разным критериям (дате изменения или добавления, по названию, по числу строк и т. д.), поиск кривых по ассоциированным с ними параметрам. Применение указанной библиотеки позволило также реализовать «живой поиск», то есть поиск в реальном времени, включающий в себя опции поиска по регулярному выражению [18], по подстроке и регистронезависимости. Приложение, предназначенное для анализа данных МУРР, должно обеспечивать обзор данных как в матричном, так и в графическом виде. Интенсивность МУРР в области исследуемых углов рассеяния может уменьшаться на значительное число порядков (семь и более), причем на некоторых участках она может падать степенным образом. В этом случае наиболее наглядным является представление кривой рассеяния не в линейном, а в полулогарифмическом или двойном логарифмическом масштабе. Следовательно, приложение должно предоставлять возможность соответствующего перемасштабирования графиков. Для сравнительного анализа нескольких интенсивностей должна быть также предусмотрена возможность их одновременного изображения на одном графическом поле. В SAXSDB не только учтены указанные требования, но также есть возможность изменить параметры графического представления (наличие или отсутствие сетки, тип линии и т. д.). Предполагается, что с хранилищем могут работать разные пользователи независимо друг от друга. Поэтому необходимыми требованиями к организации доступа являются:

- возможность экспорта матричных данных и выбранного графического представления;
- возможность независимого изменения пользователями основных параметров графического представления данных и экспорта/импорта данных;
- сохранение и использование настроек при повторном обращении пользователя к тем же данным.

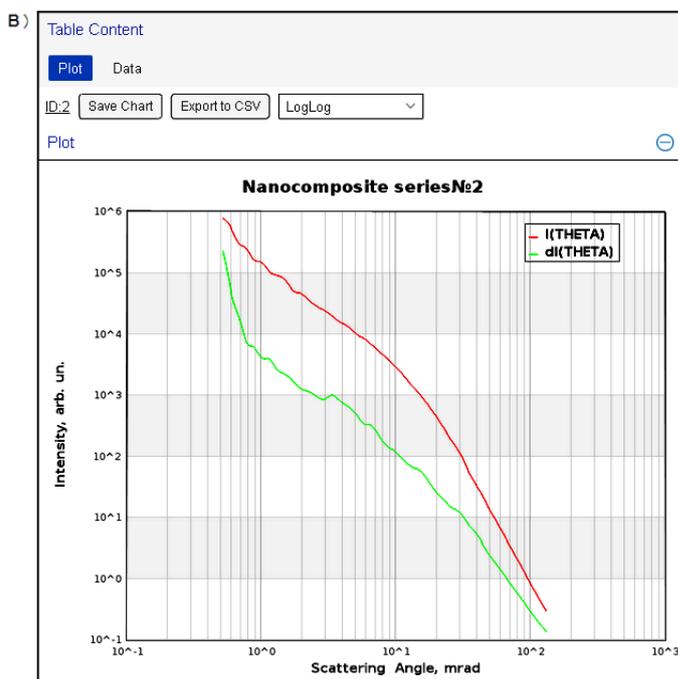
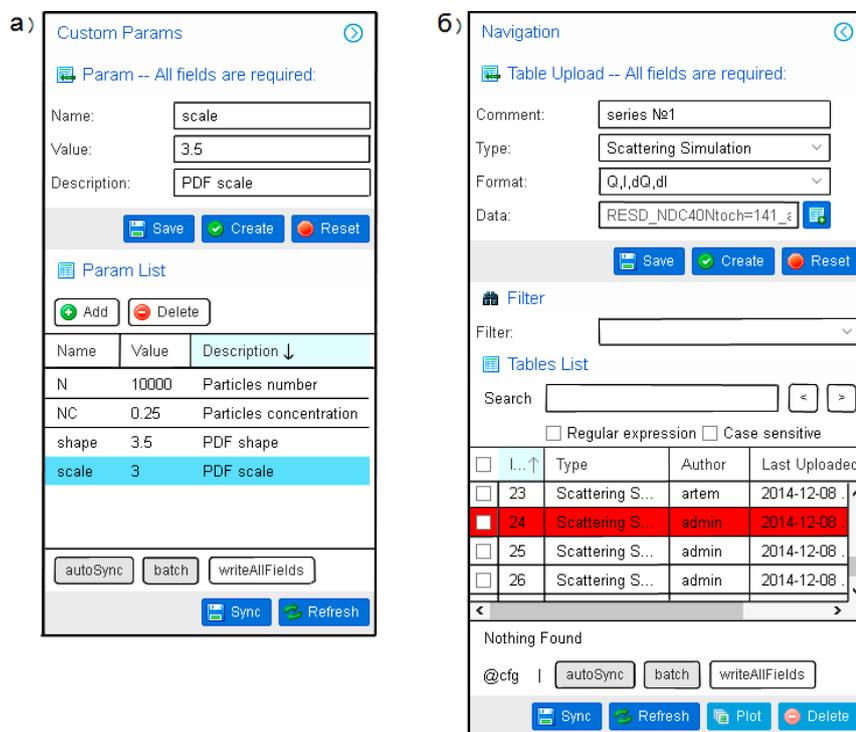
Эти требования реализованы в SAXSDB, благодаря тому что информация о всех пользовательских настройках визуализации, а также настройках парсинга сохраняется в базе данных. Кроме того, матричные данные и выбранное графическое представление по запросу пользователя можно экспортировать на его компьютер в csv и png форматах, соответственно.

4. ИНТЕРФЕЙС

С использованием библиотеки EXTJS в SAXSDB удалось реализовать адаптивный интерфейс, настраиваемый каждым пользователем под свои нужды и сохраняемый приложением для последующих обращений того же пользователя. Предложенный вариант интерфейса состоит из трех основных разделов: «навигация», «содержание таблицы» и «параметры» (см. рис. 2).

Раздел «навигация» отвечает за импорт новых данных и поиск данных среди уже имеющихся. Раздел «содержание таблицы» позволяет выбрать числовое или графическое представление для содержания выбранной матрицы (матриц), определить параметры представления и экспортировать числовые и графические данные на пользовательский компьютер. Последний раздел показывает ассоциированные с выбранной кривой параметры съёмки или объекта исследования (в случае реального эксперимента), а

также параметры модели (в случае компьютерного эксперимента). Кроме того, с помощью инструментов этого раздела осуществляется редактирование перечня параметров, их описаний и значений.



(а)

Рис. 2. Интерфейс web-приложения SAXSDB: а) раздел «навигация»; б) раздел «параметры»; в) раздел «содержание таблицы»

5. ТЕХНИЧЕСКАЯ РЕАЛИЗАЦИЯ

Кривые, измеренные на современном оборудовании, могут содержать несколько тысяч точек. Поэтому основной акцент при разработке приложения должен быть сделан на удобстве и скорости использования функций представления данных: сортировке, группировке, построении графиков, перемасштабировании и т. д. В SAXSDB операции, требующие обращения к серверу, реализованы с использованием технологии Ajax [19], заключающейся в «фоновом» обмене данными браузера с веб-сервером, и при обновлении данных веб-страница не перезагружается полностью. Отметим, что те вычислительные операции, которые уже реализованы в SAXSDB, осуществляются без использования ресурсов компьютера-клиента. Для хранения данных в SAXSDB используется СУБД MySQL 5. Работа с СУБД реализуется с помощью библиотеки DbSimple. При создании web-приложения использовано серверное ПО: PHP Version 5.3.27-1 dotdeb.0, mysqlnd 5.0.8-dev — 20102224, Apache/2.2.22. Обмен данными между сервером и клиентом осуществляется в формате JSON. Интерфейс реализован с помощью элементов, предоставляемых сайтом <http://www.famfamfam.com/lab/icons/>. В основе системы SAXSDB лежит паттерн программирования RPC [20]. Возможно переключение на паттерн REST [21]. Код написан в соответствии со стандартом кодирования, рекомендуемым разработчиками Zend Technologies Ltd. [22]. Благодаря этому система поддерживается всеми современными браузерами: Ie6++, FF2++, Opera 6++, Chrome, Safari 2++, единственное требование — в браузере должна быть включена поддержка языка Javascript. Файл конфигурации серверного ПО размещен на сайте <http://demo.saxslab.org>. На данный момент создано две версии web-приложения SAXSDB: 1.01 и 2.11. В версии 1.01 для рендеринга графиков используется стандартный механизм EXTJS4. В версии 2.1.1 для визуализации и сохранения используются библиотеки EXTJS5 и JGraph. Сохранение графиков в формате png реализовано с помощью web-сервиса svg.sencha.io. Созданные версии приложения при необходимости позволяют обеспечить конфиденциальность данных с помощью стандартного механизма «digest authentication». Для повышения уровня защищенности данных возможно также ограничение доступа по IP.

6. ЗАКЛЮЧЕНИЕ

База данных web-приложения хранится на удаленной рабочей станции, что естественным образом решает проблему резервного копирования данных. В оптимальном варианте архитектура системы должна быть спроектирована так, чтобы сделать нагрузку на стороне компьютера-клиента минимальной, в частности, за счет того, что все математические операции будут выполняться на стороне сервера. Таким образом, в случае расширения возможностей системы при переходе на следующие версии, включения сторонних разработок или расширения перечня математических операций не увеличивается нагрузка на вычислительный ресурс компьютера-клиента. Авторы рассматривают web-приложение как инструмент для увеличения эффективности работы отдельных исследователей и совместной работы в области МУРП удаленных друг от друга научных групп. Разработанный вариант приложения SAXSDB успешно опробован авторами для хранения и представления данных о рассеянии углеродными наноконструкциями [23], другими наноразмерными углеродными системами [24] и порошками оксидов, получаемыми при гидротермальном синтезе [25]. Работа с СУБД в SAXSDB реализуется с помощью библиотеки DbSimple, предоставляющей единый интерфейс

для работы с различными базами данных. Поэтому базу данных можно использовать практически любую без изменения исходного кода системы. Используемый подход к определению формата данных позволяет настроить систему, подобную SAXSDB, для упорядочения практически любых данных. Это дает возможность эксплуатировать приложение в других предметных областях. Использование возможностей EXTJS5-framework Sencha позволило создать лаконичный интерфейс, реализующий максимум возможностей при минимуме пользовательских элементов управления. Кроме того, SAXSDB может служить примером успешного использования библиотек EXTJS и JGraph для работы с данными научного эксперимента. Web-приложение SAXSDB распространяется по общедоступной лицензии GNU-GPL3 [26]. Демоверсия приложения размещена в свободном доступе по адресу <http://demo.saxslab.org>.

Список литературы

1. Guinier A., Fournet G. Small-angle Scattering of X-rays. New-York: Wiley, 1955.
2. Glatter O., Kratky O. Small Angle X-ray Scattering. London: Academic Press, 1982.
3. Свергун Д.И., Фейгин Л.А. Рентгеновское и малоугловое рассеяние. М.: Наука, 1986.
4. Захаров Д.Д. Разработка интерактивной программной среды для решения некорректных интегральных уравнений. Ее применение к обработке результатов малоуглового рентгеновского эксперимента // Изв. вузов. Приборостроение, 2011. Т. 54, № 9. С. 82–83.
5. Zakharov D.D., Smirnov A.V., Fedorov B.A. Collimation problem in SAXS for anisotropic objects: statement and solving // J. of Appl. Cryst, 2014. V. 47. P. 1552–1561.
6. Kohlbrecher, J, Bressler, I. Software package SASfit for fitting small-angle scattering curves (2013). [Электронный ресурс]: <https://kur.web.psi.ch/sans1/SANSSoft/sasfit.html> (дата обращения: 25.02.2015).
7. Ilavsky J., Jemian P.R. Irena: tool suite for modeling and analysis of small-angle scattering // J. Appl. Cryst, 2009. V. 42. P. 347–353. Irena package for analysis of small-angle scattering data. [Электронный ресурс]: <http://usaxs.xray.aps.anl.gov/staff/ilavsky/irena.html> (дата обращения: 25.02.2015).
8. Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V., Svergun, D.I. New developments in the ATSAS program package for small-angle scattering data analysis // J. Appl. Cryst, 2012. V 45. P. 342–350. Data analysis software ATSAS 2.5.2. [Электронный ресурс]: <http://www.embl-hamburg.de/biosaxs/software.html> (дата обращения: 25.02.2015).
9. [Электронный ресурс]: <http://www.bioisis.net> (дата обращения: 25.02.2015).
10. Sokolova A. V., Volkov V. V., Svergun D. I. Prototype of a database for rapid protein classification based on solution scattering data // J. Appl. Cryst, 2003. № 36. P. 865–868.
11. Hansen S. Update for BayesApp: a web site for analysis of small-angle scattering data // J. Appl. Cryst, 2014. № 47. P. 1469–1471.
12. Liu H., Hexemer A., Zwart P. H. The Small Angle Scattering ToolBox (SASTBX): an open-source software for biomolecular small-angle scattering // J. Appl. Cryst, 2012. № 45. P. 587–593.
13. Schneidman-Duhovny D., Hammel M., Sali A. FoXS: a web server for rapid computation and fitting of SAXS profiles // Nucleic Acids Research, 2010, Vol. 38. P. W540–W544.
14. Официальный сайт Sencha, Inc. [Электронный ресурс]: <http://sencha.com> (дата обращения: 25.02.2015).
15. Котеров Д. Официальный сайт библиотеки DbSimple. [Электронный ресурс]: <http://dklab.ru/lib/DbSimple> (дата обращения: 25.02.2015).
16. Официальный сайт библиотеки JGraph. [Электронный ресурс]: <http://jgraph.net/> (дата обращения: 25.02.2015).
17. Официальный сайт PHP. [Электронный ресурс]: <http://php.net> (дата обращения: 25.02.2015).
18. Fitzgerald M. Introducing Regular Expressions. Cambridge: O'REILLY, 2012.
19. Дари К., Бринзаре Б., Черchez-Тоза Ф., Бусика М. AJAX и PHP: Разработка динамических веб-приложений. СПб.: Символ-Плюс, 2006.

20. Bloomer J. Power Programming with RPC. Sebastopol, CA : O'REILLY, 1991.
21. Richardson L., Ruby S. RESTful Web Services. Sebastopol, CA: O'Reilly, 2007.
22. Официальный сайт Zend, Ltd. [Электронный ресурс]: <http://framework.zend.com/manual/1.12/en/coding-standard.coding-style.html> (дата обращения: 25.02.2015).
23. Уханова Е.А., Смирнов А.В., Федоров Б.А. Расчет функции распределения сферических частиц по размерам по данным рентгеновского малоуглового рассеяния // Научно-Технический Вестник Информационных технологий, механики и оптики, 2009. № 6 (88), С. 66–75.
24. Смирнов А.В., Федоров Б.А., Петров Н.А., Соколов .В., Томкович М.В., Гусаров В.В. Особенности строения углеродных наночастиц, полученных хлорированием нанопорошка β -SiC. // Доклады Российской академии наук. Физическая химия. 2214. Вып. 458, №. 4. С. 430–435.
25. Смирнов А.В., Федоров Б.А., Томкович М.В., Альмяшева О.В., Гусаров В.В. Особенности строения наночастиц переменного состава со структурой типа флюорита, сформированных в гидротермальных условиях на основе систем ZrO₂-Y₂O₃ и ZrO₂-Gd₂O₃ // Журнал общей химии, 2014. Т. 84, № 5. С. 710–716.
26. Официальный сайт лицензии GPL. [Электронный ресурс]: <http://www.gnu.org/copyleft/gpl.html> (дата обращения: 25.02.2015).

DEVELOPMENT OF THE INTERACTIVE WEB-APPLICATION FOR SMALL-ANGLE X-RAY SCATTERING DATA MANIPULATION

Kuchko A. V., Smirnov A. V.

Abstract

The present paper discusses requirements for the interactive web-application for remote storage data of simulated and natural getting by SAXS experiment and for manipulation with structure information obtained from this data. This application also provides ability to sort, render and analyze. Experiment of the development and implementation was considered.

Keywords: *data warehouse, parametric search of curves small-angle X-ray scattering, interactive data view.*



Наши авторы, 2015.
Our authors, 2015.

Кучко Артём Владимирович,
программист ЗАО «Рамакс Интернейшнл»,
artemkav@gmail.com

Смирнов Александр Витальевич,
кандидат физико-математических наук,
доцент кафедры физики Университета
ИТМО,
smirnav@phd.ifmo.ru