

Развитие платформы SciRus

*М. М. Якшин
(БЕН РАН)*

Платформа «SciRus» разрабатывается в БЕН РАН с 2004 года как обобщение ранее разработанной системы «Наука России». Платформа изначально позиционировалась как инструмент для создания настраиваемых учетных систем, позволяющих хранить сильно структурированную информацию о наборе сущностей и связях между ними. В базовые задачи типовой системы входит предоставление многопользовательского доступа операторам для редактирования, поиска и выдачи информации [1, 2].

За годы использования на базе «SciRus» было реализовано несколько десятков информационных систем — от ряда внутренних учетных систем БЕН РАН (базы данных «Публикации сотрудников», «Доклады на конференциях», «Библиотеки ЦБС БЕН РАН», «Редкие книги» и т. д. [1, 2]), до более заметных внешних проектов, таких как:

1. Технологическая подсистема ЭБ «Научное наследие России» [5];
2. Информационная система «Волоконная оптика» [4];
3. Информационная система «История геологии и горного дела».

Предпосылки создания новой версии платформы

В процессе эксплуатации систем на основе предыдущих версий платформы был выявлен ряд принципиальных недостатков, устранение которых в рамках развития текущей ветки разработки было нецелесообразно. В качестве примеров можно привести следующее:

- В каждый момент времени в СУБД хранилось лишь текущее состояние схемы, полей и связей сущностей, представляющих хранимую информацию. Хотя некоторые элементы транзакционности и отслеживания истории версий/изменений были введены (в частности, записывалось авторство и временная метка создания записи и последнего изменения),

целенаправленного хранения транзакции как самостоятельной сущности в СУБД не было.

- Сам механизм работы с записями по своей реализации был тесно связан с пользовательским интерфейсом, что осложняло доступ к БД формата SciRus из других приложений, порождало ряд возможных несовместимостей и в лучшем случае приводило к дублированию кода. В ряде проектов на базе SciRus были предприняты попытки изоляции части кода абстракции доступа к БД в виде REST-подобных веб-сервисов, но единого API не было, каждый такой веб-сервис требовалось реализовывать вручную (согласовывая с его потребителями), а морально устаревшая структура URL делала фактически невозможной реализацию REST API в существующем пространстве имен.
- Существующий пользовательский интерфейс давал возможность редактировать несколько сущностей одновременно в пределах сессии, но при этом никак не демонстрировал ход этой сессии (используемую стековую машину), что зачастую приводило к недопониманию со стороны редакторов и, как следствие, неправильной последовательности действий, приводящей к неумышленной порче данных.
- Существующий интерфейс поиска не был рассчитан на постраничную выдачу записей, что приводило к выдаче десятков тысяч записей в случае неудачных запросов. Такие запросы выполнялись десятки секунд, что вызывало справедливое недовольство пользователей, а в ряде случаев приводило к прекращению работы системы из-за превышения времени ожидания.
- Существующая система не была изначально спроектирована с учетом предоставления неограниченного публичного доступа на чтение пользователям сети Интернет:
 - имевшаяся структура URL записей и ссылок на поисковые запросы далека от рекомендуемых с точки зрения поисковой оптимизации;
 - поисковые запросы, выдающие большое количество записей, могут приводить к сильной нагрузке на сервер/контейнер.

- Техническая реализация SciRus основывалась на сравнительно старых технологиях (HTML 4, CSS 1, PHP 4) и была привязана к одному серверу СУБД — MySQL (и совместимым с ним системам типа Persona, MariaDB и т. п.). С одной стороны, это обеспечивало очень широкую поддержку на большинстве серверов и браузеров (вплоть до текстовых браузеров) с минимальными системными требованиями, с другой — не использовало современные возможности проектирования интерфейсов и затрудняло использование систем на основе платформы SciRus с мобильных устройств.

Отображение сущностей, отношений и транзакций в схему данных СУБД

Для хранения полной информации о транзакциях и предоставления возможности выбирать произвольное состояние записи в прошедшие моменты времени по идентификатору, потребовалось существенно изменить формат отображения сущностей и отношений в схему данных, хранящуюся в СУБД. Кроме того, решено было отказаться от использования специфичных для СУБД типов данных (таких, как «set» или «enum» в MySQL), что потенциально сделало возможным использование любых SQL-совместимых СУБД.

Формат хранения строк с данными и транзакций

Каждая сущность хранится в виде двух отдельных таблиц:

1. Таблица с именем сущности хранит актуальные на текущий момент строки с данными. При этом в поля SQL отображаются лишь те поля строки, по которым будет нужна индексация (и может производиться поиск). Все остальные поля хранятся в виде JSON-документа (или его бинарных аналогов — например, в случае PostgreSQL может быть использован BSON) в отдельном поле. В случае необходимости чтения можно прочитать запись целиком — читается только JSON-поле. В случае необходимости проведения поиска используются индексированные образы.

2. Таблица с именем сущности с суффиксом «_h» хранит все транзакции, связанные с этой сущностью. Формат таблицы фиксирован и существенно проще, чем динамический формат основной таблицы:
 1. hid — первичный ключ таблицы
 2. _id — ссылка на основную запись
 3. _data — структурированное содержимое транзакции в формате JSON-документа
 4. ts — временная метка в формате UNIX timestamp
 5. user_id — ссылка на пользователя, инициировавшего эту транзакцию

Формат хранения метаданных

В предыдущих версиях SciRus формат метаданных (описаний типов полей сущностей и связей между ними), во многом, по историческим причинам, был сильно децентрализован и хранился во множестве различных таблиц в структурированном (с точки зрения SQL) виде. С одной стороны, это позволяло (в теории) масштабировать его, с другой (для реальных задач) — куда полезнее оказалась возможность иметь всю метаинформацию сразу в оперативной памяти, а не подгружать ее по частям с диска (особенно, учитывая объемы оперативной памяти современных систем). В связи с этим было решено свести всё, что имеет отношение к метаинформации о структуре БД, в единую таблицу «_desc» с одной строкой и двумя колонками: полем, хранящим собственно всю информацию в виде JSON-документа, и полем, хранящим версию формата JSON-документа (что позволит в будущем легко расширять формат, сохраняя совместимость).

Программные аспекты реализации

Новая версия платформы «SciRus» была реализована на языке Ruby с использованием инструментария Ruby on Rails. Для реализации клиентского взаимодействия было использовано средство быстрого прототипирования интерфейсов Twitter Bootstrap с готовыми наборами стилей. В интерфейсе

широко используются современные возможности JavaScript и динамической подгрузки контента с помощью AJAX.

Весь код, обращающийся с БД, реализован в виде отдельной библиотеки, что позволяет использовать его не только в проектах, связанных с реализацией пользовательского веб-интерфейса для редактирования БД.

Литература

1. Сенько А. М., Якшин М. М. Концепция построения типовой системы ввода, учета и выдачи информации // Информационное обеспечение науки. Новые технологии. — М.: БЕН РАН, 2005. — С. 225–232.
2. Якшин М.М. WEB-интерфейс системы «Наука России» // Современные технологии в информационном обеспечении науки. — М., 2003. — С. 47–52.
3. Сенько А. М. Информационная система SciRus: принципы построения и перспективы // Электронные библиотеки: Перспективные Методы и Технологии, Электронные коллекции 2006 (сб. ст.). — Суздаль, 2006. — URL: http://rcdl.ru/doc/2006/paper_58_v1.pdf
4. Л. Исхакова, Н. Каленов, С. Васильев, Т. Полякова, Д. Чешков, М. Якшин. Разработка интерактивной информационной системы «Волоконная оптика» // Информационные ресурсы России. — 2011. — №5. — С. 19–25.
5. Якшин М. М. Платформа SciRus — основа технологического комплекса электронной библиотеки «Научное наследие России» // Труды 16-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL-2014, Дубна, Россия, 13–16 октября 2014 г. — С. 362–368.