

Информационная система SciRus: принципы построения и перспективы развития

© Сенько А. М.

Библиотека по естественным наукам РАН
Alexander.Senko@gmail.com

Аннотация

Актуальные на сегодня задачи автоматизации ввода, учёта и выдачи описаний различных документов успешно решаются в ряде систем, предназначенных для обслуживания конкретных предметных областей. Вместе с тем, схожесть решаемых задач позволяет создать универсальную систему. В данной статье описан модульный подход, использованный при создании системы SciRus — своеобразного «конструктора», на основе которого можно строить как небольшие локальные, так и распределённые системы по обработке документов.

Введение

Работа с системой SciRus (web-вариант системы «Наука России») [2, 4] и рядом других (ИСИР [1], СМИРС и др.) показала, что каждая из них слишком «завязана» на определённую предметную область или требует сложной настройки под конкретные задачи. Вместе с тем, все эти системы строятся на основе сходных бизнес-процессов и, как следствие, могут быть заменены типовой системой ввода, учёта и выдачи информации, настраиваемой под определённую предметную область и определённый круг задач [3].

В данной статье будет рассмотрена архитектура подобной типовой системы на примере SciRus с учётом её текущего состояния и перспектив развития.

Итак, система строится из одного или нескольких однотипных компонентов-ячеек, обладающих определёнными свойствами. В минимальном варианте система может состоять из одной ячейки – локального компонента.

1 Локальный компонент (ЛК)

Основная задача ЛК – ввод, редактирование, хранение данных и представление их в форме различных отчётов и справок.

Компонент состоит из нескольких слабо связанных подсистем, выполняющих определённые функции и решающих ряд задач, каждая из которых может решаться на отдельном АРМ:

- **ввод данных** (первичный ввод данных, проверка и подтверждение, редактирование ранее введённых данных);

- **поиск** (поиск по введенным данным, генерация отчётов по шаблонам);
- **администрирование** (настройка схемы данных и шаблонов, управление правами и привилегиями пользователей).

При работе ЛК использует таблицы настроек, отражающие схему данных базы, что позволяет администратору относительно легко менять и дополнять структуру данных в процессе работы системы.

В качестве интерфейса для всех компонентов системы используется web-интерфейс. Такой подход легко позволяет организовать доступ к системе как с локальных рабочих мест внутри организации, так и извне (для авторизованных пользователей).

ЛК может использоваться в качестве самостоятельной системы в случае работы с небольшими объёмами данных, либо с данными, работа с которыми ведётся, в основном, в пределах одной организации. При этом ЛК взаимодействует только с пользователями, отвечая на их запросы по протоколу HTTP; средства взаимодействия и обмена данными с другими ЛК отсутствуют.

2 Распределённый компонент (РК)

Для создания более сложных систем имеет смысл применять распределённый подход. В этом случае система строится из нескольких ЛК, объединённых в единую сеть.

При этом каждый из компонентов дополняется определённой функциональностью, позволяющей обмениваться данными с другими компонентами. В этом случае мы можем говорить уже о распределённом компоненте.

К подсистеме ввода данных добавляется функция импортирования данных, а к подсистеме поиска – возможность выгрузки (экспортирования) данных в другой РК.

Благодаря функциям загрузки и выгрузки данных система допускает различие схем (разнородность) данных в различных РК. В случае, если хранимые данные однотипны, преобразование информации (конвертация) не требуется и эти функции сводятся к тривиальной передаче информации по каналам связи – тогда можно говорить о простом обмене данными между РК.

Эти механизмы реализованы в настоящее время не полностью. Интерфейс для пакетной загрузки (импортирования) данных в SciRus практически отсутствует, хотя на системном уровне имеются средства для загрузки данных, в частности, из ранних

версий «Науки России». Система же выгрузки (экспортирования) данных реализована в достаточной степени и полностью отражена в пользовательском интерфейсе – вплоть до визуального редактора шаблонов для выгрузки данных. При помощи этого редактора можно создавать произвольные шаблоны, позволяющие производить экспорт данных и во внешние по отношению к SciRus системы. Так, например, были созданы шаблоны для загрузки данных в систему ЕНИП РАН.

РК могут быть связаны с использованием одной из двух схем:

- **интеграция в единое хранилище** – импорт данных из подчинённого компонента в вышестоящий;
- **виртуальная интеграция** – данные физически хранятся только в одном РК, передача данных между БД не происходит; поиск и получение информации выполняются с помощью распределённого запроса.

Первую схему предполагается использовать в следующих случаях:

- затруднена передача данных между РК по постоянному каналу (физически нет быстрого доступа в Интернет, повышенные требования к безопасности и т. п.);
 - данные в нижестоящей организации обновляются достаточно редко и предпочтительно иметь их экспортированную копию в вышестоящем РК;
 - объёмы данных в нижестоящей организации достаточно малы (возможно также, что в этом случае целесообразно использовать РК, а достаточно предоставить авторизованный доступ для этой организации к вышестоящему ЛК).
- Преимуществами второй схемы являются:
- снижение требований к программно-аппаратным комплексам каждого РК в отдельности (вследствие существенного снижения объёмов хранимой информации);
 - максимальная актуальность получаемых при поиске результатов (т. к. нет промежуточных синхронизаций).

Для реализации второй схемы необходимо дополнить поисковую систему каждого РК модулем поддержки распределённых запросов (предполагается использовать для этих целей технологию XMPP/Jabix). При этом также усложняется решение вопросов безопасности.

На данном этапе развития системы компоненты SciRus могут быть связаны исключительно при помощи полуавтоматической передачи данных между ними в рамках первой схемы. Но в новой версии этот обмен данными планируется максимально автоматизировать, используя в качестве транспортного протокола для обмена XMPP или широко распространённые почтовые протоколы (SMTP, POP и, возможно, IMAP).

Стоит отметить, что при организации сложных сетей с использованием вышеизложенных принципов обе схемы могут комбинироваться для создания сетей смешанной интеграции.

Заключение

Применение описанных в статье подходов позволяет, как уже показала эксплуатация SciRus в БЕН РАН, создавать не только системы для работы с библиографическими данными, но и с произвольной схемой данных.

Разрабатываемый для будущих версий SciRus генератор кода позволит поставлять конечному пользователю уже не “конструктор”, а конечную, настроенную под определенные задачи систему, что значительно облегчит процесс ее внедрения.

Литература

- 1] Бездушный А. Н., Власова С. А., Каленов Н. Е., Меденников А. М., Серебряков В. А. Подход к интеграции информационных коллекций ВИСИР РАН // Электронные библиотеки: перспективные методы и технологии, электронные коллекции. 2-я Всероссийская научная конференция. Сб. докладов. Сентябрь, 26-28, 2000. – С. 129-135
- 2] Каленов Н. Е., Васильев А. В., Власова С. А., Глушановский А. В. Автоматизированная информационная система “Наука России” // Информационно-библиотечное обеспечение науки: Проблемы интеграции информационных ресурсов: сб. статей. – М., 1995. – С. 112-115
- 3] Сенько А. М., Якшин М. М. Концепция построения типовой системы ввода, учёта и выдачи информации // Информационное обеспечение науки: новые технологии. Сборник научных трудов / под ред. Н. Е. Каленова. – М.: БЕН РАН, 2005. – С. 225-232
- 4] Якшин М. М. WEB-интерфейс системы “Наука России” // Современные технологии в информационном обеспечении науки. – М., 2003. – С. 47-52

Architecture principles and development prospects of SciRus document management system

Senko A. M.

Automated document metadata management is growing popular today. These tasks are usually carried out with a number of domain-specific systems. However, proximity of these leads to creation of a general system. This article describes modular approach used in SciRus system. It consists of several components that can be combined to build both simple local and heavily distributed document systems..