

УДК 0025
ББК 73

ЭВОЛЮЦИЯ ВЕБ-ОРИЕНТИРОВАННОЙ СИСТЕМЫ УПРАВЛЕНИЯ ДОКУМЕНТАЛЬНОЙ НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИЕЙ СО РАН

© Н.А. Малицкий*, Н.А. Мазов**, 2006

* Государственная публичная научно-техническая библиотека
Сибирского отделения Российской академии наук
630200, г. Новосибирск, ул. Восход, 15

** Объединенный институт геологии, геофизики и минералогии
Сибирского отделения Российской академии наук
630090, г. Новосибирск, пр. Академика В.А. Коптюга, 3

Рассматривается созданная в Сибирском отделении РАН веб-ориентированная система управления документальной научно-технической информацией, доказавшая свою эффективность и жизнеспособность за семилетний срок промышленной эксплуатации. Клиентская часть системы прошла долгий путь совершенствования и на данный момент является, по сути, информационно-поисковой системой, которая предлагает широкие функциональные возможности посредством использования многовариантного поискового интерфейса. При этом потенциал развития системы еще далеко не исчерпан, и об этом будет сказано ниже. Основная цель настоящей работы – показать, каким удалось построить столь высокоэффективный программно-технологический комплекс. Авторы надеются, что эта информация будет полезна разработчикам при создании систем подобного класса.

Ключевые слова: библиографические базы данных; программно-технологический комплекс; доступ; пользователи; информационный поиск; ведение; веб; Интернет.

Прежде дадим небольшое определение документографической информации. Под *документографической информацией* понимается информация, содержащая библиографическое описание документов и/или их рефераты. В момент создания рассматриваемой системы существовало достаточно много аналогичных информационно-поисковых систем (ИПС) как в России, так и за рубежом.

В качестве предпосылок при создании новой системы были выделены следующие потребности пользователей:

- в предоставлении оптимального (по отношению к существующим) способа поиска документографической информации;
- в комплексном решении поддержки актуальности используемой документографической информации.

Приведем пояснение, что понимается под *оптимальным способом поиска*.

Во-первых, это **оперативность** поиска. Это означает, что поиск можно провести из любого места. Для этого идеально подходят каналы связи сети Интернет.

Во-вторых, это **массовость** поиска. В сети Интернет наиболее популярны и распространены веб-ресурсы. Следовательно, необходимо было обес-

печить доступ к требуемой информации посредством веб.

В-третьих, это **удобство** и **простота** поиска. Для этого было решено использовать стандартные, наиболее распространенные инструменты. Для веб это обычный браузер, который либо уже установлен на компьютере пользователя (например, Internet Explorer для ОС MS Windows, доля которого более 90% на рынке браузеров), либо может быть легко проинсталлирован на нем. Необходимо, чтобы интерфейс системы отвечал требованиям эргономики (**удобство**), был интуитивно понятен (**простота**) и органически сочетал в себе лучшие черты существующих систем.

В-четвертых, это **функциональность** поиска. Новая система должна была аккумулировать лучшие (наиболее распространенные) идеи существующих систем и дополнить их перечнем собственных технологических решений.

Резюмируя сказанное, следует отметить, что именно эффективная реализация эргономики и функционала позволила создать более совершенную по отношению к другим системам новую поисковую систему на веб.

Теперь поясним, что имелось в виду под *комплексным решением поддержки актуальности информации*.

Во-первых, это полномасштабная **интеграция бизнес-логики** систем поиска и редактирования (актуализации). Это означает возможность проведения актуализации на веб информации, полученной в результате поиска, а также добавление новой и удаление неактуальной информации.

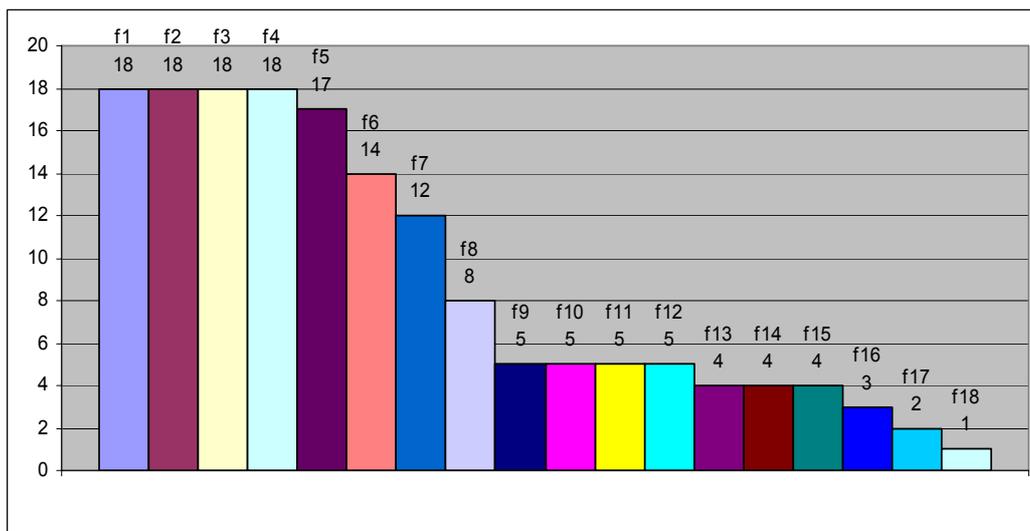
Во-вторых, это **унификация интерфейсов** систем поиска и редактирования. С учетом выбора в системе веб как среды доступа к документографической информации следует отметить, что сопутствующая система редактирования (актуализации) должна была также иметь веб-интерфейс.

Таким образом, интеграция системы поиска с системой актуализации позволила говорить о создании полнофункциональной веб-ориентированной системы управления документографической информацией. В этом состоит основная **новизна** созданной системы (ее отличие от типичных ИПС).

Для аккумуляции лучшего из существующих веб-систем был проведен их анализ в крупнейших информационных центрах и библиотеках России

(Государственная публичная научно-техническая библиотека России (ГПНТБ России), Библиотека по естественным наукам Российской академии наук (БЕН РАН), Всероссийский институт научной и технической информации (ВИНИТИ), Институт научной информации по общественным наукам РАН (ИНИОН), Центральная научная сельскохозяйственная библиотека Российской академии сельскохозяйственных наук (ЦНСХБ РАСХН), Российская национальная библиотека (РНБ), Санкт-Петербургский государственный технический университет (СПбГТУ), Объединенный институт геологии, геофизики и минералогии им. А.А. Трофимука Сибирского отделения РАН (ОИГГМ СО РАН) и за рубежом (Библиотека Конгресса США, Университет Корнели-Меллона, Университет Беркли, СОРАС, Британская библиотека, LIBRIS, LIBSYS, BIBSYS, Национальная библиотека Австралии, UNILINC).

Были выявлены и ранжированы характерные особенности их интерфейсов.



Распределение функциональных возможностей в исследуемых веб-системах:

f1 – в поисковое поле возможен ввод нескольких терминов в виде слов или фраз; *f2* – использование усечения терминов; *f3* – возможность использования логических операторов в запросе; *f4* – наличие справочного руководства и описание правил составления запроса; *f5* – представление результатов поиска в сокращенной и полной форме; *f6* – в системе имеется несколько меню формирования запроса к системе; *f7* – проведение уточняющего поиска по результатам текущего; *f8* – меню расширенного поиска в виде пар «тип поля – значение поля», с произвольным выбором типа поля; *f9* – использование словаря поисковых терминов для формирования запроса; *f10* – возможность поиска термина по всем словоформам и синонимам; *f11* – сохранение результатов поиска в открытых форматах (usmarc, gusmarc и др.); *f12* – возможность работы зарегистрированному пользователю в соответствии с правами доступа; *f13* – выбор нескольких баз данных (БД) для одновременного выполнения введенного запроса; *f14* – предварительная установка параметров выдачи (формат, порция и др.); *f15* – пересылка результатов поиска по электронной почте; *f16* – работа в сессии; *f17* – сортировка результата поиска; *f18* – сохранение и повторное использование запроса.

Результат анализа полученных данных показал, что веб-ориентированная ИПС должна отвечать следующим требованиям:

- иметь многовариантный поисковый интерфейс;

- использовать дополнительные возможности ввода в форму запроса поисковых терминов (словарь терминов, рубрикаторы и др.);
- обеспечивать гибкость в представлении результатов поиска;

- предоставлять возможность проведения уточняющего поиска;
- позволять сохранять результаты поиска;
- иметь справку о работе с системой.

На основании вышеизложенных требований была создана веб-ориентированная система управления документографической информацией, предоставляющая для пользователя следующие возможности:

В режиме поиска:

- четыре варианта поиска информации: Простой, Стандарт, Эксперт, Мульти;
- гибкое формирование запроса с выбором типа поисковых терминов для конкретной БД;
- использование поисковых терминов запроса в полной и усеченной слева форме;
- привлечение словаря поисковых терминов, тематических рубрикаторов и тезаурусов при формировании запроса;
- представление запроса пользователя в транслитерированной форме при отсутствии на клиентской машине кириллической кодовой таблицы;
- сохранение и выполнение ранее созданных поисковых запросов;
- переключение в другой режим поиска с восстановлением там ранее введенного запроса и его настроек;
- проведение уточняющего поиска в найденных данных;
- свободное перемещение по словарю поисковых терминов и рубрикаторам до момента выбора одного или нескольких терминов, включаемых в поисковый запрос;
- формулирование для поиска произвольного количества пар «тип поля – значение поля» и связывание их по правилам булевой алгебры;
- выбор параметров поиска (форматы выдачи, порция выдачи, усечение терминов) при его формировании и изменение их при просмотре его результатов;
- свободное перемещение по результату поиска: переход к любому найденному документу/странице выдачи;
- произвольное сохранение содержимого найденных записей в файл в любом требуемом формате (ASCII, ISO-2709, XML, USMARC, RUSMARC);
- быстрое отображение соотнесенного с найденным документом его реферата;
- поиск в нескольких однотипных (по поисковым терминам) базах данных (режим Мульти);
- подсвечивание в результатах поиска введенных поисковых терминов;
- проведение уточняющего поиска из результата поиска для конкретного документа по любому из его авторов.

В режиме редактирования:

- одностраничное представление интерфейса: используются фреймы и модальные диалоги;

- обеспечение функционала: создание новой записи, модификация записи, удаление записи, копирование записи;

- введение поля с подполями и повторяющимися полями, что наиболее точно соответствует структуре библиографического описания;

- сохранение введенных данных как шаблона для последующего ввода однотипных записей;

- предоставление нескольких вариантов бланков ввода библиографического описания;

- целостное представление о текущем содержимом записи в табличной форме, которое является альтернативным и удобным способом выбора содержимого полей, их повторений и подполей на редактирование;

- обеспечение многопользовательского доступа к данным;

- предоставление инструментария для создания новых бланков ввода различного содержания с соответствующими правами доступа;

- ограничение доступности к полям и их подполям для внесения изменений.

В режиме администрирования:

- ведение базы пользователей и общих настроек системы;

- группировка баз данных и определение права доступа к ним;

- прописывание параметров баз данных;

- установление параметров режима редактирования;

- настройка форматов выдачи результатов поиска.

Созданная система хорошо зарекомендовала себя в процессе промышленной эксплуатации как надежная и полнофункциональная система управления значительными объемами документографической научно-технической информации на Веб. Следует особо отметить, что эффективность системы подтверждается ее значительным сроком эксплуатации в учреждениях Сибирского отделения Российской академии наук (ГПНТБ СО РАН, информационно-библиотечный центр ОИГТМ СО РАН, информационный центр Института катализа СО РАН, библиотека Института цитологии и генетики СО РАН, г. Новосибирск, Институт вычислительной математики СО РАН, г. Красноярск, и др.), а также постоянным ростом количества информационных ресурсов, доступ к которым обеспечивает данная система (более сотни БД, содержащих десятки миллионов записей!).

Несмотря на то что система продолжает эксплуатироваться в промышленном режиме /1–3/, она требует доработок, поскольку мир информационных технологий переживает в настоящее время бурный рост.

Веб-технологии, находясь на передовых позициях, наиболее стремительно развиваются, в результате чего разработчикам предлагаются все бо-

лее совершенные средства разработки. Текущая реализация системы проектировалась и разрабатывалась с использованием технологии CGI-скриптов для серверной части и скриптовых возможностей браузеров IE и Netscape версий не выше 4 (связка JavaScript и HTML). Далее будут показаны проблемы текущей программной реализации серверной части системы и предложены возможные пути ее дальнейшего развития с учетом возможностей современных веб-технологий.

Вначале определим проблемы текущей реализации системы:

1) Независимость от серверной платформы.

На данный момент система реализована только для платформы Win32. Серверная часть системы представлена CGI-скриптами, написанными на языке C/C++. В настоящее время эта технология морально устарела и не отвечает требованиям быстрой разработки веб-решений. Перенос этих скриптов для других серверных платформ (Unix, AIX, Solaris и др.) достаточно сложен и требует их параллельного сопровождения для этих платформ.

2) Производительность и распределение серверной нагрузки. Основная проблема технологии CGI-скриптов заключается в том, что при обращении к ним веб-сервер порождает множество независимых между собой процессов операционной системы. При большом количестве обращений к веб-серверу это чревато, как минимум, резким падением производительности системы, а в худшем случае и отказом работоспособности сервера. Распределить нагрузку на несколько серверов средствами CGI-скриптов невозможно и требует от администратора веб-узла нетривиальных решений, базирующихся на использовании сетевых утилит.

3) Средства разработки. CGI-скрипты – это фактически консольные приложения, которые получают данные от веб-сервера через переменные окружения операционной системы, выполняют некоторую заложенную в них бизнес-логику и выдают в стандартный поток вывода конечный объем байтов. Отладка таких скриптов крайне затруднена. Соответственно скорость разработки достаточно невысока. Для Win32 требуется предварительная компиляция в .exe файлы и их последующее размещение на веб-сайте.

4) Готовые решения, библиотеки. HTTP – протокол без сохранения состояния сеанса связи браузера и веб-приложения. Иначе говоря, без сессии. Поэтому эту сессию приходится эмулировать. Каждый браузер при общении с веб-сервером передает ему некий уникальный идентификатор, используя который веб-сервер может сохранять состояние сессии. CGI не имеет готовых средств для этого. Все современные технологии имеют в данном случае встроенное API.

5) Поддержка технологий. CGI – это спецификация, и не более того. Причем крайне ограниченная в своих возможностях. Наиболее развитые технологии либо продвигаются такими крупными производителями, как Microsoft (.NET, ASP, ASP.NET), Sun (Java, 2EE, JSP), IBM, Macromedia (CF, Flash), либо поддерживаются сообществом разработчиков – PHP.

Итак, исходя из вышеизложенного, CGI необходимо найти замену. Для этого следует определить, какие основные цели ставятся при переходе на новую технологию и как их можно реализовать с помощью самых популярных веб-технологий. Ниже, в таблице представлены цели и технологии для развития веб-ориентированной системы.

Сравнительные характеристики современных веб-технологий

Цель	J2EE	.NET	PHP
Работа на любой серверной платформе	Java SDK и JRE для всех платформ, «написано один раз – работает везде»	Только MS Windows	Наличие интерпретаторов для любой платформы. Исходные тексты
Работа при большой веб-нагрузке	Большое количество высокопроизводительных J2EE контейнеров (серверов приложений): Resin и др. Возможность использования балансировки нагрузки, встроенных средств для многопоточной обработки запросов	Позиционируется как альтернатива J2EE	Интерпретация текстовых скриптов без предварительной (как в J2EE и .NET) компиляции
Наличие развитых средств разработки	Продукты Sun (Java Studio и др.) и сторонних производителей (JBuilder, Netbeans и др.)	MS Visual Studio.NET	В целом – текстовые редакторы
Стоимость	За серверы и средства разработки приложений, причем большинство бесплатны для некоммерческого использования	min \$800	Бесплатно
API, библиотеки	Лидер в наработках API для веб-разработок, сетевого программирования	Среда CLR, все многообразие языков .NET (VB, C#, C++, ASP.NET)	Большое количество, но несравнимо меньше с конкурентами

Таким образом, как видно из представленной таблицы, Java-J2EE превосходит .NET по кросс-платформенности, а сравнение с PHP некорректно ввиду того, что J2EE – это технология, предлагающая решение любых проблем при разработке веб-системы. PHP – это более частный случай с несравненно меньшими возможностями как по процессу разработки (нет полноценной IDE), так и по функциональным возможностям для разработчика (API).

Резюмируя сказанное, следует отметить, что технология J2EE – очевидный лидер для перевода разработки на его рельсы. Основными моментами здесь являются увеличение скорости разработки в разы и наличие законченной технологии (API, IDE), нацеленной на разработку веб-приложения.

В заключение отметим, что созданная система продолжает эффективно выполнять поставленные перед ней задачи и имеет достаточный потенциал для развития пользовательского интерфейса и серверной части системы для повышения ее произво-

дительности и расширения функциональных возможностей.

Список литературы

1. Елепов, Б. С. Сравнительный анализ Web-ориентированных систем для обработки библиографической НТИ / Б. С. Елепов, Н. А. Малицкий, Н. А. Мазов // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества : 8-я Междунар. конф. «Крым-2001» (9–17 июня 2001 г., г. Судак) : материалы конф. – М., 2001. – Т. 1. – С. 248–251.
2. Малицкий, Н. А. Расширенный инструментарий для обработки библиографической НТИ на Web / Н. А. Малицкий, Н. А. Мазов // Там же. – С. 270–272.
3. Малицкий, Н. А. Функциональные возможности современных библиографических ИПС на Web / Н. А. Малицкий, Б. С. Елепов, Н. А. Мазов // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса : 11-я Междунар. конф. «Крым-2004» (5–13 июня 2004 г., г. Судак) : тр. конф. – М., 2004. – (CD-ROM).

Материал поступил в редакцию 6.02.2006 г.

Сведения об авторах: *Малицкий Николай Александрович – аспирант ГПНТБ СО РАН, тел. (383) 219-16-36, e-mail: nmal@quik.ru;*
Мазов Николай Алексеевич – кандидат технических наук, заведующий информационно-библиотечным центром, тел. (383) 333-22-16, e-mail: mazov@uiggm.nsc.ru