

ОПЫТ СОЗДАНИЯ ПРОБЛЕМНО-ОРИЕНТИРОВАННОЙ ЭЛЕКТРОННОЙ БИБЛИОТЕКИ В ЛАБОРАТОРИИ ХИМИЧЕСКОЙ ТЕРМОДИНАМИКИ МГУ

А.Г. Богачев, Г.В. Белов, В.И. Горячева, И.Б. Куценок

Интернет-сайт кафедры или научной лаборатории в высшем учебном заведении может стать важным инструментом научной и образовательной деятельности подразделения. Для этого целесообразно помимо информации о структуре подразделения, направлений деятельности, достижений и пр. помещать на сайте учебные и научные материалы, содержание лекций, семинаров, расчетные задачи и пр. Часть этой информации должна быть доступна любому пользователю интернета, другая часть – конфиденциальная информация, предназначенная для коллектива сотрудников, аспирантов и студентов лаборатории. В настоящей публикации изложен опыт создания и ведения информационно-библиографической базы данных (ИББД) химического и химико-технологического профиля в лаборатории химической термодинамики МГУ. В настоящее время в состав системы входят полнотекстовая база учебников, научных монографий, справочников, оригинальных статей и др. источников информации по указанной тематике и смежным дисциплинам (всего около 3000 документов), а также оригинальный поисковый механизм на основе системы управления базами данных MySQL 5.1 (с дополнительными модулями), предназначенный для быстрого поиска документов различными категориями пользователей, имеющих доступ к веб-сайту лаборатории.

При создании ИББД были проанализированы имеющиеся в Интернете аналогичные базы данных (БД). Существующие электронные системы, содержащие информацию о научных публикациях, можно условно разделить на три типа: библиографические БД, электронные хранилища институтов и библиотек (Institutional repository) и БД крупных коммерческих и некоммерческих структур, оснащенные, как правило, сложными поисковыми машинами (или просто поисковые машины, настроенные на эффективную работу с научной литературой).

Первый тип систем - базы библиографических данных, как правило, не содержит сам файл электронной публикации, а лишь более или менее подробные сведения о ней. Такие БД можно, в свою очередь, разделить на два вида:

1. однопользовательские, устанавливающиеся на один компьютер. В них доступ к редактированию и чтению файлов БД имеет только один пользователь. Наиболее известными коммерческими продуктами являются Endnote и Reference Manager [1]. Из некоммерческих систем значительное распространение получили Zotero и JabRef [2];
2. сетевые программы, устанавливаемые на сервере и доступные нескольким пользователям при помощи клиентских программ или через веб-интерфейс сервера. Здесь наиболее распространены некоммерческие системы: Bibus [3], Rebase [4], а так же Aigaion, WIKINDEX.

Второй тип систем - электронные хранилища институтов и библиотек - также условно можно разделить на два вида - содержащие и не содержащие публикации в электронном виде:

- автоматизированные библиотечные информационные системы используются для управления и индексации библиотечных фондов. Среди них наиболее известна PhpMyBibli [5];
- дополнительная возможность хранения файлов электронных документов присутствует в свободно распространяемых системах-репозиториях, EPrints [6]. Из коммерческих продуктов данного вида наиболее известна Biblioscape [7].

Третий тип систем, содержащих информацию о научных публикациях, это - серверы крупных компаний, чаще всего занимающиеся электронными изданиями и имеющие в своем распоряжении серверы с базами данных журналов и книг, например, ScienceDirect (<http://www.sciencedirect.com>), Scopus (www.scopus.com), Web of Science (www.isiknowledge.com), Springer (www.springer.com) и т.д. Поисковые машины серверов таких систем позволяют искать и загружать подписанному пользователю нужные электронные издания. Такие системы обладают морфологической поддержкой поиска для английского языка, но, к сожалению, данные информационные системы обладают платной подпиской. Из систем, в которых можно получать информацию без платной подписки, можно отметить хранилище документов ArXiv.org, частично Google Books (<http://books.google.com/>), а также поисковые машины: CiteSeer (<http://citeseerx.ist.psu.edu>) и Google Scholar.

К недостаткам большинства коммерческих и некоммерческих систем поиска и хранения библиотечной информации можно отнести нередкое отсутствие отображений кириллических шрифтов и поиска русских слов. Подавляющее большинство из них не поддерживают при поиске морфологию русского языка, не производят корректировку и исправление ошибок ввода, в них фактически не заложена возможность распознавания введенных химических формул и использования специальных терминов.

Встроенная в большинство современных СУБД (Microsoft SQL Server 2008, MySQL 5.1, PostgreSQL 8.3) функция проведения полнотекстового поиска улучшает возможности получения нужной информации, но такой

механизм поиска не всегда даёт ожидаемые пользователем результаты, в частности, при запросах, содержащих химико-технологические термины.

На основании проведенного анализа с учетом специфики потребностей сотрудников лаборатории были сформулированы задачи, которые решались при создании ИББД лаборатории химической термодинамики МГУ:

- создать систему хранения электронных документов научного содержания разной структуры и объема (монографий, отдельных журнальных статей, диссертаций, трудов конференций) с их полным библиографическим описанием;
- разработать эффективную систему поиска нужного документа БД по библиографическому описанию и тексту документа с последующей его загрузкой с сервера на компьютер пользователя;
- при организации полнотекстового поиска необходимо отработать учет морфологических особенностей русского и английского языков;
- система поиска и хранения должна быть доступна через веб-сайт после введения имени и пароля пользователя;
- организовать возможность авторизованным пользователям с учетом присвоенного им статуса пополнять базу данных, удалять и загружать электронные документы.

С программной точки зрения задача создания ИББД должна удовлетворять таким критериям как масштабируемость, межплатформенная переносимость, легкость и быстрота настройки, использование некоммерческого ПО.

Созданная ИББД отвечает всем перечисленным выше требованиям. Основные модули программы написаны на языках php и JavaScript. Для хранения информации используется СУБД MySQL 5.1 с подключенным дополнительным модулем Sphinx. Кроме того, при реализации системы задействованы словари химических терминов и химических формул, а также дополнительная библиотека phpMorphy.

Помимо перечисленных выше возможностей существующая ИББД поддерживает многоязыковый интерфейс, поиск и отображение с использованием расширенной таблицы символов – русского и европейских языков, позволяет пополнять базу данных научных публикаций через веб-интерфейс, проводить коррекцию и исправление заведомых ошибок в запросах пользователей, организует поиск внутри документов форматов Acrobat (PDF), LizarTech (DJVU), MS Word (DOC), загруженных на сервер системы. Система позволяет экспортировать и импортировать записи в упоминавшиеся выше распространенные системы хранения данных EndNote и BibTex и пр., а также создавать самообновляющиеся зеркала на других серверах. Из дополнительных опций отметим возможность просмотра текста документа в веб-браузере без использования сторонних плагинов.

Для повышения эффективности работы системы ИББД соответствующим образом структурирована – разбита по темам и тегам с учетом потребностей сотрудников лаборатории. Теги и темы могут добавляться и изменяться администратором системы. Пользователи ИББД имеют различный статус - администратор, редактор, пользователь, гость, в соответствии с которым установлены различные права на работу с документами. При этом каждая запись в определенный момент времени может редактироваться только одним пользователем ИББД.

БД документов системы постоянно обновляется. В лаборатории имеется сотрудник, ответственный за пополнение ИББД новыми документами. С работой системы можно ознакомиться на сайте лаборатории химической термодинамики МГУ <http://td.chem.msu.ru/>, получив временный логин и пароль у администратора системы: techlib.tdlab@gmail.com

Работа выполнена при финансовой поддержке РФФИ, грант № 08-03-00506-а.

ЛИТЕРАТУРА:

1. Thomson Reuters EndNote www.endnote.com/
2. JABREF <http://jabref.sourceforge.net/>
3. Bibus bibliographic database, <http://bibus-biblio.sourceforge.net/>
4. Web Reference Database rebase, <http://www.refbase.net/>
5. PMb services, PhpMyBibli <http://www.sigb.net/?page=download&lang=en>
6. EPrints, <http://www.eprints.org/software/>
7. Biblioscape, <http://www.biblioscape.com/>