

А. И. Бродовский

ГПНТБ России

Новый продукт семейства ИРБИС – ИРБИС64+: система для создания электронной библиотеки

Представлен очередной новый продукт семейства ИРБИС, предназначенный для создания электронной библиотеки. Отмечено, как разработчики трактуют электронную библиотеку: это программное обеспечение для системы, в которой реализуется концепция полноценных баз данных – фактографической и полнотекстовой. Раскрыты принципы работы и возможности ИРБИС64+; перечислены все модули, входящие в систему.

Ключевые слова: Система автоматизации библиотек ИРБИС, ИРБИС64+, электронная библиотека.

UDC 026.06

Alexander Brodovsky

Russian National Public Library for Science and Technology, Moscow, Russia

IRBIS64+ – a new product of IRBIS family: The system for building e-library

A new product of IRBIS family is introduced, designed for building e-libraries. The designers define electronic library as a software for the system where the concept of full functional – factographic and fulltext – databases can be realized. IRBIS64+ operational principles and possibilities are revealed\$ system modules are listed.

Keywords: IRBIS Library Automation System, IRBIS64+, electronic library.

We propose a new software product IRBIS64+ that is an instrument for organizing the electronic library. Functional requirements to the electronic library consist of the following elements: full-text search on the ground of automatic partitioning text onto the pages followed by word by word indexing; presentation of results in order of relevance of documents, and inside a single document – in descending order of relevance of found pages; provision to the end user a possibility to browse the full texts and marking of found words of the request; access to the full text; accounting statistical system. Search result is a list of biblio-

graphical records, in order of decreasing relevance. The end user interface consists of two levels. On the first level you get a list of bibliographic records and a page with the first fragment of the full text. It also displays the document ranking on the basis of end-user evaluations. The second level provides you access to the full text. Depending on the user rights one has two options: download the full text and / or view it page by page. You can scroll text; move from one page to another; make personal bookmarks in the text, and look previously made bookmarks. In the case of downloading the right user gets the opportunity to work with the accounting system, view the history of applications to the full text (both the electronic one and printed); get the list of personal bookmarks and use them to apply to the relevant pages of previously read texts, This functionality can be used in university libraries in the formation of the database of final qualifying works of students that now is an actual problem, ie, students can provide an opportunity to include their diploma work – the complete text and description – in the corresponding database of the electronic library. It is also planned to implement the possibility of assessment the full text originality (to process the text by the Antiplagiat system).

В 2016 г. мы предлагаем новый продукт семейства ИРБИС – **ИРБИС64+** – он представляет собой систему автоматизации, предназначенную для создания и ведения электронной библиотеки, а не просто электронного каталога.

Наши пользователи (и не только они) под электронной библиотекой понимают самые разные вещи. Некоторые называют электронной библиотекой ЭК, где к библиографическим описаниям привязаны ссылки на полные тексты, которые не участвуют в поиске и выступают в роли ресурса для просмотра. Такую «электронную библиотеку» можно создавать с помощью нашего действующего продукта – обычного ИРБИС64. Другие пользователи электронной библиотекой называют то, что можно создавать с помощью другого нашего продукта – ИРБИС64 ПБД, т.е. полнотекстовые базы данных, в которых возможен поиск по словам.

Разрабатывая ИРБИС64+, электронную библиотеку мы трактовали более широко. Но при этом мы не претендуем на то, чтобы давать строгое определение собственно *электронной библиотеке*. Это задача для специалистов, пишущих монографии и учебники. Но в любом случае важно отметить: с точки зрения читателя, основное назначение электронной библиотеки – это поиск нужной *информации*, в то время как назначение электронного каталога – поиск нужной *книги* (документа).

Если говорить о практической стороне дела, то перед нами стояла задача создать программное обеспечение для системы, в которой реализуется объединённая концепция полноценной БД – и библиографической и полнотекстовой. Иными словами, говоря на языке пользователей ИРБИС, – создать систему, которая полностью включает возможности ИРБИС64 и ИРБИС64 ПБД и сверх того – имеет дополнительный функционал, связанный со спецификой электронной библиотеки, в нашем, разумеется, понимании этой проблемы.

В общем функционал, обусловленный требованиями именно электронной библиотеки, с нашей точки зрения, состоит из следующих элементов:

- полнотекстовый поиск на основе автоматического разбиения текстов на страницы и их пословной индексации;

- представление результатов поиска в порядке убывания релевантности документов, а в рамках одного документа – в порядке убывания релевантности найденных страниц;

- предоставление конечному пользователю возможности постраничного просмотра полных текстов с обеспечением навигации по релевантным страницам и маркировки найденных слов запроса;

- развитая система прав доступа конечных пользователей к полным текстам;

- система учёта обращений пользователей к полным текстам.

Таким образом, основой нового продукта ИРБИС64+ являются БД, представляющие собой совокупность связанных библиографических данных и полных текстов изданий.

Библиографические данные – на основе коммуникативного формата RUSMARC. В качестве полных текстов используются распознанные PDF-файлы, которые подвергаются автоматическому разбиению на страницы и индексированию по словам.

Поиск в системе ведётся по словам полного текста и/или любым элементам библиографического описания. Результат поиска в общем случае представляется в виде списка документов (т.е. библиографических описаний изданий), как было сказано, в порядке убывания их релевантности. При этом с каждым документом связан список релевантных страниц полного текста (также в порядке убывания).

Релевантность определяется на основе оригинального критерия, который учитывает количество и контекстную близость слов запроса, найденных в полном тексте. При сравнении слов используется механизм морфологии русского языка. Найденные термины на страницах полного текста маркируются цветом.

Поисковый интерфейс, адресованный конечному пользователю, состоит из двух уровней. На первом пользователю предлагается список библиографических описаний документов, найденных в результате поиска. При этом, если выполнялся полнотекстовый поиск, в каждое описание документа включается фрагмент первой релевантной страницы полного текста. Здесь же выводится рейтинг документа, сформированный на основе оценок конечных пользователей. В случае представительного результата поиска формируются фасеты. Под фасетами понимается распределение найденных документов по значениям определённых элементов библиографического описания (в частности: автор, год издания, тематика), т.е. выделяются авторы, годы издания, тематика, которые максимально представлены в результатах поиска.

На втором уровне осуществляется работа с полным текстом. В зависимости от прав доступа пользователю могут быть предоставлены две возможности: скачать полный текст и/или просматривать его постранично.

В случае постраничного просмотра каждая страница полного текста представляется в виде только графического слоя, т.е. в виде «картинки», что, с одной стороны, защищает текст от копирования, с другой – исключает наличие каких-либо плагинов для работы с PDF-текстами. В этом случае, работая с полным текстом, пользователь имеет возможности:

- листья текст последовательно вперёд и назад;
- переходить от одной релевантной страницы к другой (т.е. листать найденные страницы полного текста);
- делать и, разумеется, удалять личные закладки в тексте, и соответственно листать текст по ранее сделанным закладкам;
- менять масштаб изображения страницы;
- скачивать (т.е. копировать) конкретные страницы, конечно, при наличии соответствующих прав;
- выставлять личную оценку тексту.

При наличии права скачивания пользователь получает возможность работать с полным PDF-текстом. Если в браузере клиента установлен плагин *Acrobat Reader*, то скачанный полный PDF-текст открывается непосредственно в браузере и при этом отображается список найденных слов запроса с возможностью позиционировать страницы, на которых они найдены.

Теперь – собственно о правах доступа. В зависимости от прав доступа, связанных с конкретным документом, пользователю может быть:

- запрещено обращение к полному тексту;
- разрешён постраничный просмотр полного текста;
- дополнительно к постраничному просмотру разрешено скачивание полного текста.

Разрешение на просмотр и скачивание полного текста может иметь количественные ограничения, которые задаются или в процентах от общего объёма документа, или в виде непосредственного количества страниц.

При определении прав доступа к полным текстам могут учитываться следующие данные:

- временной интервал разрешённого доступа (с точностью до дня);
- идентификационные признаки пользователя (такие как номер читательского билета, категория читателя, факультет, семестр, специальность студента и т.п.), которые могут задаваться в виде маски;

- IP-адрес клиента (в виде маски);

- доменное имя клиента.

Собственно значение права доступа вырабатывается в момент обращения конкретного читателя к конкретному полному тексту на основе того, что в ИРБИСе называется *формат*. Этот формат, как и все прочие, находится в распоряжении пользователей системы, а это значит, что, изменяя формат, пользователь может уточнять или как угодно менять алгоритм формирования права доступа к полным текстам.

Теперь – о системе учёта обращений к полным текстам. При обращении конечного пользователя (т.е. читателя) к полному тексту фиксируется электронная книговыдача. Единичной выдачей считается обращение к любым страницам полного текста в течение одного сеанса. Сеанс определяется как временной интервал (по умолчанию – одни сутки). На основании этих данных можно получать статистические формы, связанные с обращением и использованием полных текстов. Необходимо отметить, что это не отменяет технологии традиционной бумажной книговыдачи.

Вход в электронную библиотеку осуществляется на основе авторизации читателя (читатель указывает свои логин и пароль). При этом ему предлагаются элементы личного кабинета, а именно, он имеет возможность просматривать:

- историю своих обращений к полным текстам, а также список источников, связанный с традиционной (бумажной) книговыдачей (что и когда брал, что находится сейчас на руках);

- список личных закладок и с их помощью обращаться к соответствующим страницам ранее прочитанных текстов;

- корзину своих заказов на «бумажную» книговыдачу.

И наконец, конечному пользователю может быть предоставлена возможность загрузить в электронную библиотеку собственный полный текст и его краткое описание. На этой возможности следует остановиться подробнее.

Этот функционал может использоваться в вузовских библиотеках при формировании БД выпускных квалификационных работ студентов, что сейчас является актуальной задачей, т.е. студенты смогут самостоятельно включать свои дипломные работы – полный текст и его описание – в соответствующую БД электронной библиотеки. В дальнейшем планируется также реализовать возможность оценки включаемого таким образом полного текста на оригинальность (т.е. пропускать этот текст через систему антиплагиата).

В заключение перечислим входящие в состав системы ИРБИС64+ модули: Сервер баз данных; АРМ «Администратор», АРМ «Каталогизатор», АРМ «Книговыдача», АРМ «Комплектатор», Web-шлюз для доступа конечного пользователя.

*Alexander Brodovsky, Head of ALIS and Special Databases Design and Support Department, Russian National Public Library for Science and Technology;
alio@gpntb.ru
17, 3rd Khoroshevskaya st., 123436 Moscow, Russia*