

Качественное улучшение поиска информации в электронных каталогах: проблемы, пути, перспективы

Горшкова Галина Анатольевна, начальник отдела компьютерной обработки документов и создания каталогов, Научная библиотека Петрозаводского государственного университета

В докладе рассматриваются вопросы разработки в АИБС «Фолиант» справочного аппарата к электронным каталогам, который позволяет выполнить большинство запросов, сформированных пользователями, привыкшими искать в Интернете. В силу специфики создания электронных каталогов пользователи при поиске информации по ним нередко вынуждены обращаться к дежурным библиографам. Как показывает практика, количество обращений к электронным каталогам с компьютеров, находящихся вне библиотеки, постоянно возрастает. В этом случае пользователи не могут рассчитывать на помощь библиографа.

Поиск информации в различных поисковых системах постоянно совершенствуется. Электронные каталоги как поисковая система продукт довольно специфичный. Закладывая определенные правила в созданные библиографические записи, сотрудники библиотек тем самым предлагают, а, можно сказать, и навязывают пользователям определенные правила поиска информации по электронным каталогам, отличающиеся в значительной степени от поиска в поисковых системах Интернета. При этом предполагается, что, если пользователи сами не разберутся при поиске информации в электронном каталоге, то в библиотеке помогут им в этом и научат. Для этого проводятся библиотечно-библиографические занятия, в залах с каталогами дежурит библиограф. Все это очень хорошо и правильно.

Однако если учитывать практику работы библиотек Карелии, когда электронные каталоги представлены только через Интернет и доступны всем, а для поиска информации совсем необязательно посещать библиотеку, то уже сегодня получается очень интересная картина. Анализируя количество обращений к электронным каталогам, мы можем констатировать, что их количество извне постоянно возрастает.

Если пользователь пришел в любую из библиотек, то у него есть возможность поискать информацию не только в электронном каталоге этой библиотеки, но и в электронных каталогах других библиотек. Пользователи часто производят поиск информации на своих рабочих местах, как, например, в Петрозаводском государственном университете. Да и дома сегодня компьютеры с выходом в Интернет уже не является редкостью.

Поэтому в условиях, когда пользователи не всегда могут рассчитывать на помощь библиографа, появилась настоятельная необходимость модернизировать поисковую систему электронных каталогов так, чтобы улучшить качество поиска информации в них.

Создаваемый в АИБС «Фолиант» справочный аппарат в значительной степени расширяет поисковые возможности электронных каталогов. Этот справочный аппарат реализуется с использованием Российского коммуникативного формата представления авторитетных / нормативных записей. Реализация такого справочного аппарата в "доморощенном" формате, как показывает практика, привела бы к тому, что с его использованием возникли бы затруднения. Конвертирование данных из одной АИБС в другую, из одной версии программы в другую привело бы к потерям информации и потребовало значительных временных затрат на последующее редактирование данных.

Отмечу, что правила создания справочного аппарата не являются темой рассмотрения данной статьи. Чтобы объяснить суть создаваемого справочного аппарата, не слишком вдаваясь в лингвистические сложности, и показать его возможности в реализации поиска информации по электронным каталогам обратимся к конкретному примеру.

Допустим, что при поиске по электронному каталогу пользователь формирует свой запрос следующим образом: «История России 90-х годов XX в.». Однозначно можно сказать, что при запросе, сформированном таким образом, если что-то и найдется, то это будет очень незначительная часть из имеющейся в электронном каталоге информации. В большинстве электронных каталогов в результатах поиска, вероятнее всего, будет нулевой результат, так как вносится информация сотрудниками библиотек совсем не так, как представлено в запросе.

Что касается электронного каталога Научной библиотеки Петрозаводского государственного университета, то этот запрос будет выполнен, а в результатах поиска мы можем гарантировать максимально полный результат. Такой же точно результат будет получен, если пользователь запросит: «Российская история 90-х годов XX в.», «Русская история 90-х годов XX в.», «История России 90-х гг. 20 в.», «История России в 90-е гг. 20 в.», «История Россия 90-е 20 в.», «История Россия 20 в. 90-е» и другие варианты. Причем в результатах поиска при любом из этих запросов будет найдено одно и то же количество библиографических записей.

За счет чего же достигается такой результат? Происходит это потому, что в одной из баз данных справочного аппарата к электронному каталогу присутствует конкретная справочная запись. Мы называем ее справочной записью, несмотря на то, что внешне она похожа чем-то на авторитетную запись, и даже создана в том же формате. Однако на нынешнем этапе работы об авторитетной записи у библиотек уже сложилось определенное мнение, поэтому пока мы не рискуем называть нашу запись авторитетной, хотя по своей сути она таковой является. Привожу пример справочной записи:

- 250 \$a История \$y Россия \$z 90-е 20 в.
- 450 \$a История России 90-х гг. 20 в.
- 450 \$a Российская история 90-х гг. 20 в.
- 450 \$a Русская история 90-х гг. 20 в.
- 450 \$a История России в 90-е гг. 20 в.

450 \$a Российская история в 90-е гг. 20 в.
450 \$a Русская история в 90-е гг. 20 в.
450 \$a История-Россия-90-е гг. 20 в.
450 \$a История России 90-х гг. XX в.
450 \$a Российская история 90-х гг. XX в.
450 \$a Русская история 90-х гг. XX в.
450 \$a История России в 90-е гг. XX в.
450 \$a Российская история в 90-е гг. XX в.
450 \$a Русская история в 90-е гг. XX в.
450 \$a Россия \$x История \$z 90-е гг. 20 в.
450 \$a История-90-е гг. 20 в.-Россия
450 \$a История-90-е гг. 20в.-Россия
450 \$a История \$y Россия \$z 20 в. 90-е
450 \$a Российская история 90-х годов 20 в.
450 \$a Российская история 90-х годов XX века
450 \$a Российская история 90-х годов 20 века
450 \$a Русская история 90-х годов XX века
450 \$a Русская история 90-х годов 20 века
450 \$a История России 90-х годов XX века
450 \$a История России 90-х годов 20 века
450 \$a Русская история 90-х годов 20 в.
450 \$a История России в 90-е годы 20 в.
450 \$a Российская история в 90-е годы 20 в.
450 \$a Русская история в 90-е годы 20 в.
450 \$a История-Россия-90-е годы 20 в.
450 \$a История России 90-х годов XX в.
450 \$a Российская история 90-х годов XX в.
450 \$a Русская история 90-х годов XX в.
450 \$a История России в 90-е годы XX в.
450 \$a История \$y Россия \$z 20 в. 90-е гг.
450 \$a Российская история в 90-е годы XX в.
450 \$a Русская история в 90-е годы XX в.
450 \$a История-90-е годы 20 в.-Россия
450 \$a История-90-е годы 20в.-Россия
450 \$a История России 90-х 20 в.
450 \$a Российская история 90-х 20 в.
450 \$a Русская история 90-х 20 в.
450 \$a История России в 90-е 20 в.
450 \$a Российская история в 90-е 20 в.
450 \$a Русская история в 90-е 20 в.
450 \$a История-Россия-90-е 20 в.
450 \$a История России 90-х XX в.
450 \$a Российская история 90-х XX в.
450 \$a Русская история 90-х XX в.
450 \$a История России в 90-е XX в.

- 450 \$a Российская история в 90-е XX в.
- 450 \$a Русская история в 90-е XX в.
- 550 \$a Исторические науки
- 550 \$a История
- 550 \$a История \$y Россия
- 550 \$a История \$y Россия \$z 20 в.
- 686 \$a 63.3(2)6 \$2 rubbk
- 940 \$a 63.3(2)7

Если внимательно посмотреть на эту запись, то она состоит из нескольких частей. В поле 250 находится предметная рубрика – основной вариант, который всегда переносится в создаваемую библиографическую запись. Хотя теперь абсолютно все равно, какой из вариантов будет приведен в библиографической записи. Тем более, что для удобства поиска предметные рубрики объединены с неконтролируемыми предметными терминами (ключевыми словами). От понятия «неконтролируемые предметные термины» мы постепенно уходим. Предполагается наладить со временем полный контроль даже за приводимыми в библиографических записях ключевыми словами.

В полях 550 находятся все вышестоящие предметные рубрики, которые при создании библиографической записи также автоматически переносятся в нее вслед за основной предметной рубрикой. Они помогают систематизаторам при раскрытии содержания документов, а также помогают организовать иерархический поиск информации без использования индексов УДК и ББК, которые чужды пользователю.

А вот 450-е поля содержат многочисленные варианты предметной рубрики, внесенной в поле 250. Многие из представленных вариантов, мы обнаружили в электронных каталогах библиотек при поиске информации, а особенно при заимствовании. Заимствуя библиографические записи из каталогов различных библиотек, мы систематически анализируем, каким образом вносятся те или иные сведения каталогизаторами и библиографами. Это дает определенный положительный результат, помогает нам находить варианты для справочного аппарата.

Кроме того, создавая справочную запись, мы следим за отражением в ней единственного и множественного числа основного термина, а также прямой и инверсированной форм. Эти варианты в электронных каталогах библиотек точно встречаются.

И еще при обслуживании читателей по электронным каталогам дежурный библиограф следит, какие варианты запросов формируются, особенно, если в результатах поиска на мониторе появился «0». Это для того, чтобы взять эти запросы на заметку.

Также в АИБС «Фолиант» ведутся лог-файлы, в которых автоматически фиксируются все запросы к электронным каталогам, а затем анализируются. Можно узнать, какие запросы являются наиболее частыми. Особенно интересен анализ запросов с нулевым результатом поиска. Такой анализ мы проводили уже неоднократно, тем более что результаты выдаются в виде таблиц. К сожа-

лению, пока ничего не можем поделаться с запросами, при формировании которых допущены грамматические ошибки. А вот правильно сформированные запросы, по которым ничего не найдено, обязательно отражаем в справочном аппарате как варианты к основным предметным рубрикам. В следующий раз при таких же точно запросах поиск будет результативным.

Что касается разработки справочного аппарата к электронному каталогу, то можно отметить следующее. Сегодня этой работой заняты только сотрудники Научной библиотеки Петрозаводского государственного университета, и только тогда, когда остается время после выполнения основной работы. Работа по созданию справочного аппарата ведется третий год, но даже за это время удалось кое-что сделать. Например, в базе данных «Предметные рубрики» на 1 апреля 2009 г. присутствует 43805 справочных записей, содержащих хотя бы одно поле 450 (вариант, синоним). Эти записи содержат 328707 вариантов и синонимов. Число справочных записей, содержащих полный набор вышестоящих предметных рубрик составляет 19584.

Также работа ведется в базах данных «Имя лица», «Организации» и т. п. Определенные базы данных справочных записей закрепляются за определенными полями, предназначенными для создания библиографических записей.

Наполнение баз данных справочными записями дело трудоемкое. Мы прекрасно понимаем, что сразу объять необъятное невозможно. Радует и вдохновляет тот факт, что любая созданная сегодня справочная запись сразу же начинает участвовать в поиске информации по электронному каталогу.

К созданию справочного аппарата стараемся подойти рационально. Так, например, создаем справочные записи на какую-то науку или область знаний, всегда начиная с вышестоящих терминов. Затем происходит постепенное создание нижестоящих терминов и понятий, то есть дальнейшая детализация науки или области знаний. При этом создавать справочную запись низшего уровня можно путем копирования с уже созданной записи вышестоящего уровня, что существенно экономит время.

Можно создавать справочные записи блоками, что также в значительной степени ускоряет работу. Создаются, например, самые полные справочные записи на блок «Уголовное право». Далее, используя этот опорный блок записей, можно путем копирования создавать записи на государственное право, лесное право и т. д. При этом исправления во вновь создаваемых справочных записях минимальны, потому что эти науки похожи по своей структуре.

Таких блоков можно обнаружить очень много во всех областях знания.

Если для создания записей опорного блока обязательно нужен подготовленный сотрудник, то для дальнейшей работы по созданию справочных записей путем копирования возможно привлечение дополнительной рабочей силы, например, сотрудников других отделов библиотеки или студентов.

Как же происходит поиск в электронных каталогах с учетом создаваемого справочного аппарата? После формирования запроса происходит обращение к базам данных справочного аппарата.

Далее возможны два сценария развития поиска информации в электронных каталогах. Первый заключается в том, что при обращении к базам данных спра-

вочного аппарата не обнаружена справочная запись, содержащая поля с вариантами и синонимами. Если это так, то происходит отказ от использования справочного аппарата, и запрос выполняется по электронному каталогу в обычном порядке.

Если справочная запись, содержащая варианты и синонимы, обнаружена, то автоматически формируется новый запрос, включающий в себя основной и дополнительные варианты. Теперь уже этот сводный запрос отправляется для поиска информации по электронному каталогу. И в этом случае можно вести речь о наиболее полных результатах поиска, которые способны удовлетворить пользователей.

Теперь остановимся на поиске информации с учетом иерархии, имеющейся в науке или области знаний, к которой относится тематика запроса. Такие запросы актуальны, когда необходимо найти новинки по определенной теме или области знаний.

Мы давно отказались от того, чтобы наши пользователи производили поиск по классификационным индексам. Они совершенно непонятны массовому пользователю. Чтобы сохранить иерархию в библиографической записи, в ней приводятся вышестоящие предметные рубрики и термины.

Если иерархические уровни разработаны в справочном аппарате, то с их перенесением в библиографические записи проблем не возникает. Но, если справочных записей нет, то при внесении вышестоящих предметных рубрик и терминов в библиографические записи все зависит от квалификации систематизаторов и предметизаторов. Именно поэтому в библиографических записях далеко не полностью приведены вышестоящие термины иерархической цепочки.

Отсутствуют полные иерархические цепочки также в записях, которые мы заимствуем из других библиотек. И если библиографические записи на книги мы дорабатываем «под себя», то записи на статьи из сборников и журналов загружаем в электронный каталог без всякого редактирования. Исправляются только грамматические ошибки и опечатки по мере их обнаружения.

Решить проблему поиска с учетом иерархических уровней можно, скрыто используя индексы УДК и ББК при тематическом поиске. Справочные записи содержат индексы УДК и ББК, то есть каждая основная предметная рубрика или термин классифицированы. Классификационные индексы объединяются при поиске в электронном каталоге с предметными рубриками и ключевыми словами. Организован их совместный поиск. При поиске по вышестоящим терминам находятся все нижестоящие, входящие в него

Подводя итоги, следует отметить следующее:

1. Создание справочного аппарата к электронным каталогам, хотя и является процессом трудоемким, но как показывает пусть пока и небольшой опыт нашей работы, полностью себя оправдывает. Уже сегодня реально улучшился поиск в электронном каталоге. Он постепенно упрощается для пользователя.
2. С помощью справочного аппарата упростилась работа по созданию библиографических записей. Нет никакой необходимости при заимствовании библиографических записей редактировать их под правила ве-

дения своего электронного каталога. Ведь то редактирование электронных каталогов, которое проводят сотрудники библиотек, «чтобы было единообразно», требует очень много времени и дополнительные штаты. При нашем подходе снижаются трудозатраты при работе с электронным каталогом. А главное, по-разному внесенные сведения никак не отражаются на качестве поиска информации в электронном каталоге. Наличие справочной записи нейтрализует негативные моменты, имеющиеся в библиографических записях.

3. Можно до бесконечности редактировать библиографические записи в электронных каталогах, пытаясь довести их до идеального уровня. Можно даже договориться и вносить сведения абсолютно одинаково во всех библиотеках России. Только зачем? Поисковая система электронных каталогов так и останется специфической, построенной по определенным правилам, которые придется постоянно разъяснять пользователям.