

Новые подходы к каталогизации заимствованием в ИРБИС-корпорации
New Approaches to Cataloguing by Borrowing in IRBIS Corporation
Нові підходи до каталогізації за допомогою запозичення в ІРБІС-корпорації

К. Е. Соколинский

*Международная ассоциация пользователей и разработчиков
электронных библиотек и новых информационных технологий,
Учебно-информационный центр Северо-Западного
заочного технического университета, Санкт-Петербург, Россия*

Kirill Sokolinsky

*International Association of Users and Developers
of Electronic Libraries and New Information Technologies
North-Western Extra-mural Technical University.
Educational Information Center, St. Petersburg, Russia*

К. Є. Соколинський

*Навчально-інформаційний центр Північно-західного
заочного технічного університету, Санкт-Петербург, Росія*

В докладе рассматривается новый программный интерфейс для работы с ИРБИС-корпорацией и применённые в нём методы повышения эффективности заимствования записей. Обосновывается необходимость использования специальной технологий кэширования, нормализации формы запросов и ранжирования результатов поиска по качественным признакам.

The paper covers a new program interface for work with IRBIS corporation and methods of increasing efficiency of record borrowing that are used in it. The necessity of using a special caching technology, normalizing request form and ranging search result according to qualitative attributes is grounded.

У доповіді розглянуто новий програмний інтерфейс для роботи з ІРБІС-корпорацією і застосованими в ньому методами підвищення ефективності запозичення записів. Обґрунтовано необхідність використання спеціальної технології кешування, нормалізації форми запитів і ранжирування результатів пошуку за якісними ознаками.

Использование в каталогизации готовых библиографических записей можно считать неотъемлемой частью современной библиотечной технологии. Поэтому большинство российских АБИС включают средства заимствования. Как правило, предусматривается заимствование по протоколу Z39-50¹, но в некоторых случаях поддерживаются и специализированные технологии². Распространение получили также интеграционные порталы³, позволяющие выгружать записи без установки дополнительных программ. Таким образом, разными способами обеспечивается доступ к значительному массиву записей, который может использоваться как в текущей каталогизации, так и при ретроконверсии.

Тем не менее, значительная часть библиотек либо вообще не применяет каталогизацию заимствованием, либо применяет, но без существенного выигрыша от этого[3]. Нередко трудоёмкость получения и редактирования готовых записей превышает трудоёмкость их самостоятельного создания. Косвенным свидетельством неэффективности заимствования допустимо считать сохранение прежних штатов каталогизаторов[1, с. 57] и предпочтение, отдаваемое многими библиотеками ресурсоёмкой технологии имиджкаталогов перед ретроконверсией.

Можно указать на совокупность как организационных, так и технических причин недостаточного применения заимствования. К первой группе причин относится разрозненность библиографических ресурсов, высокие барьеры для вступления в корпоративные объединения и несогласованность стандартов. Ко второй – недостаточная эффективность инструментальных средств заим-

¹ Например, в АБИС ИРБИС, АБИС Руслан, АБИС Академия+.

² Режимы заимствования из OPAC Global, WEB ИРБИС и д.р.

³ Например, порталы: АРБИКОН, Корпоративной сети московских библиотек, RUSLANet, Сигла и д.р.

ствования. И если существующую организационную структуру приходится рассматривать как данность, то технические препятствия устранимы и должны быть устранены.

Сегодня в российских технологиях распределённого поиска допустимо выделить следующие проблемы:

1. Низкая скорость выполнения запросов. При последовательном поиске она определяется количеством серверов, при параллельном – скоростью самого медленного сервера. В зависимости от принципа работы поисковой системы время ожидания может оказаться весьма значительным⁴.

2. Трудоемкость ввода поискового запроса. В большинстве случаев поиск по оптимальному идентификатору записи – ISBN возможен только при условии, что он введён с разделителями(тире)⁵. Поэтому цифровое значение, которое может быть легко считано с обложки книги сканером штрихкодов, оказывается для поиска непригодным. При поиске по автору зачастую некорректно интерпретируются инициалы. Они или вообще игнорируются или их требуется вводить с обязательным соблюдением пунктуации (с пробелом и после запятой)⁶.

3. Сложность выбора оптимальной записи в результатах поиска. Как правило, библиографические записи при поиске отображаются в соответствии с порядком сканирования серверов-источников. Нередко при этом выводится несколько десятков вариантов одной записи и выбор наилучшей оказывается проблематичным.

4. Недостатки реализации клиентов. При использовании стандартных windows-клиентов в проксируемой сети возникает необходимость дополнительной настройки прокси-сервера и установки специального программного обеспечения. Автономная работа клиентов делает так же невозможным эффективное кэширование и фильтрацию неработоспособных серверов. С другой стороны, при использовании web-приложений обычно не удаётся обеспечить отображение результатов по мере выполнения запроса, поэтому пользователю приходится ожидать сканирования всех источников.

Таким образом, можно сделать заключение о необходимости создания программного средства, позволяющего устранить все описанные выше недостатки. И в настоящее время для АБИС ИРБИС ведётся его разработка. Планируется, что оно сделает более эффективным использование имеющихся режимов.

ИРБИС располагает сегодня самым широким набором режимов заимствования. Непосредственно из АРМа Каталогизатор могут быть получены записи с серверов Z39-50, из OPAC Global и ИРБИС-корпорации. При этом поддерживаются все распространённые схемы данных: RUSMARC, USMARC, UNIMARC. Однако наиболее перспективным является заимствование в ИРБИС-корпорации. Это связано с распространённостью АБИС ИРБИС, применением в корпорации единого формата и отсутствием искажений вносимых конвертерами. Поэтому создаваемый в настоящее время новый программный интерфейс ориентирован, в первую очередь, на интеграцию библиографических ресурсов пользователей ИРБИС. В то же время заимствование возможно не только через АРМ Каталогизатор, но и любой браузер. По своей форме интерфейс является AJAX-ориентированным web-приложением. Архитектура решения даёт возможность соединить простоту использования web-технологий с эргономичностью windows-интерфейса.

Главный акцент при разработке интерфейса делается на скорости получения данных и как ключевые факторы быстродействия рассматриваются эффективное кэширование, параллелизация процессов, блокировка сбойных серверов.

Сходство принципов комплектования многих видов библиотек (например, публичных) позволяет прогнозировать и сходство библиографических потребностей. Поэтому постоянный опрос всех источников нецелесообразен, и оправдано многократное предоставление пользователям единойжды собранных данных – кэширование. При определении времени жизни кэша очевидна необходимость учёта специфики библиографической деятельности. Например, если нужна запись

⁴ Даже в эффективном программном обеспечении АРБИКОН среднее время выполнения широковеб-запроса составляет 37.7 секунд.[2]

⁵ В Z39-50 клиенте Lib Navigator при отсутствии разделителей поиск не даёт результатов. На портале RUSLANet поиск по ISBN без разделителей позволяет выявить меньшее количество записей, чем с разделителями.

⁶ Например, инициалы не учитываются на портале RUSLANet и в клиенте Lib Navigator.

может быть позаимствована в одной из пяти библиотек, библиографу, скорее всего, окажется достаточно найденных записей и повторное сканирование всех источников ради альтернативных вариантов не потребуется. Более редким должно быть обновление кэша на издания советского периода, так как появление описаний на них в каталогах менее вероятно, чем на литературу последних лет выпуска. Всё это уместно назвать принципами «библиографического кэширования». Ожидается, что они будут уточнены и дополнены на основе фактических данных статистики заимствования.

Поскольку новый интерфейс предполагает сортировку и быстрое извлечение больших объёмов данных, вместо традиционного файлового кэширования используется кэширование в локальную реляционную базу данных. Оно осуществляется в фоновом режиме и практически не отражается на скорости выполнения запроса.

Вторым средством повышения скорости является распараллеливание. Запросы множеству серверов выполняются и обрабатываются одновременно самостоятельными процессами. За счёт этого нивелируется разница во времени отклика источников и записи извлекаются максимально быстро. Кроме того, более эффективно используются вычислительные ресурсы, так как данные одного запроса обрабатываются в момент ожидания выполнения другого и при этом может быть задействован весь аппаратный потенциал (процессоры, ядра процессоров, память).

Третьим средством повышения эффективности выступает мониторинг работоспособности серверов. Хорошо известно, что запросы к неработоспособным серверам крайне длительны. Они продолжаются до максимального времени, отведённого на запрос (таймаута). Поэтому целесообразно минимизировать их количество. Сбойный сервер на сутки блокируется, и новые обращения к нему, в случае повторения неудачных запросов, происходят через экспоненциально увеличивающиеся промежутки времени.

Таким образом, скорость выполнения запросов оптимизируется за счёт целого ряда как давно апробированных, так и оригинальных технологий. Но это не рассматривается как единственный фактор эффективности. Не меньшее внимание уделяется пользовательскому интерфейсу. К его ключевым особенностям относятся упрощённый механизм ввода запросов, динамический вывод результата, оптимизированный принцип сортировки записей и поддержка клавиатурных команд.

В отличие от большинства WEB интерфейсов, где результаты демонстрируются после завершения выполнения запроса, в описываемом решении их вывод обеспечивается при получении новых данных. Кроме того, сам процесс поиска начинает выполняться автоматически при указании каталогизатором минимальных идентификационных сведений. По мере ввода запроса возникают и динамические «подсказки» в стиле Google, которые предлагают применить данные, использованные в предшествующих запросах. Всё это позволяет свести к минимуму время ввода. Той же цели служат упрощённые правила формулировки запроса.

ISBN может вводиться как с разделителями(дефис), так и без них. Благодаря этому создаются условия для получения записей на современные издания путём простого считывания сканером книжного штрихкода. Для книг без ISBN обеспечивается возможность поиска по библиографическим данным. Данные об авторе могут вводиться как по библиографическому ГОСТУ(с запятой после фамилии и пробелами между инициалами), так и в предельно упрощённой форме – без каких-либо знаков препинания, даже без точек после инициалов(например, иванов и и). Из заглавия может вводиться лишь несколько первых слов. Год задаётся путём указания двух последних цифр, так как библиографические записи на книги столетней давности мало востребованы.

Как показывает статистика, фамилия и инициалы автора, нескольких символов заглавия (в общей сложности 10–12 знаков), позволяют идентифицировать нужную запись в 99.5% случаев. Поэтому в основу концепции интерфейса был положен принцип однозначной идентификации.

Вместе с тем предлагаются и оптимальные условия для выбора нужных записей в результатах поиска. Записи, полученные из разных источников, выводятся не просто в порядке сканирования серверов, но ранжируются по формальным признакам качества. В роли таких признаков выступают:

1. Авторитетность источника;
2. Количество ранее выполненных заимствований;
3. Количество заполненных полей и подполей записи;
4. Объём записи.

Значимость каждого из этих критериев может быть самостоятельно определена пользователем с помощью настроек. Индивидуальные предпочтения сохраняются и учитываются в следующих сеансах заимствования.

И хотя ранжирование не претендует на решение проблемы оптимального выбора, но позволяет структурировать хаотичный набор записей, при этом оставляя последнее слово за человеком. Для визуальной оценки может использоваться краткая библиографическая запись, полный набор полей записи, название библиотеки-источника и автоматически сгенерированный индекс качества. В то же время проектируется и технология консолидации записей, которая предполагает объединение всех версий записи в одну. Возможно, она позволит полностью снять вопрос о выборе оптимальной записи.

Поскольку очевидно, что отрыв каталогизатора от клавиатуры для использования мыши требует времени и снижает координацию действий, для выполнения технических операций предлагается употреблять сочетания клавиш. Переход между полями может осуществляться сочетаниями CTRL и клавишами управления курсором, выбор записи – применением CTRL и цифровых клавиш. Но применение мыши, ставшей самым привычным инструментом для многих каталогизаторов, также остаётся допустимым.

Таким образом, в интерфейсе ИРБИС-корпорации предусматриваются все средства для повышения эффективности ввода запроса, технической реализации поиска и выбора оптимальных записей. По сравнению с другими программными решениями, его преимуществом являются функции:

1. Специализированного кэширования широковещательных запросов;
2. Динамического вывода результатов поиска и их ранжирования по формальным критериям качества;
3. Алгоритмической нормализации данных пользовательского запроса;
4. Автодополнения элементов поискового запроса.

Поэтому можно рассчитывать, что интерфейс ИРБИС-корпорации изменит отношение библиографов к каталогизации заимствованием и позволит доказать её высокую эффективность.

Литература

1. Саринкова, Т.Г. Внедрение новых компьютерных технологий в работу отдела комплектования и научной обработки фонда [Текст] / Т.Г. Саринкова // Современные тенденции развития библиотечно-информационных технологий вузовских библиотек: материалы конф.(Россия, Хабаровск, 21-23 октября 2003 г.). – Хабаровск: Изд-во ХГТУ, 2003. – С.72-79.
2. Усманов, Р.Т. Оценка эффективности работы распределенной сети Z39.50 серверов АРБИКОН [Электронный ресурс] / Р.Т. Усманов, А.А. Кузнецов // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: материалы конф. (Судак, Авт. Респ. Крым, Украина, 4–12 июня 2005 г.). – Электрон. текстовые дан. – М.:ГПНТБ, 2005 . – Режим доступа: <http://www.gpntb.ru/win/inter-events/crimea2005/disk/217.pdf>.– Загл. с экрана.
3. Фронкин, А.В. Современная библиотека в контексте корпоративной деятельности [Электронный ресурс] // Труды научно-практической конференции посвященной 145-летию Тверской областной библиотеки им. А.М. Горького(Тверь, Россия, 20 мая 2005). – Электрон. текстовые дан. – Режим доступа: <http://corbis.library.tver.ru/cbs-ot08.htm>.– Загл. с экрана.