

DOI 10.32460/ion_nt-2018-0008

УДК 004.77:025.355:027.021

ББК 78.362.5+78.373

Анализ поисковых запросов в электронном каталоге АБИС «Библиобус»

Васильев А. В. (БЕН РАН, Москва)

Представлен статистический анализ выполненных поисковых запросов в служебном электронном каталоге неперIODических изданий АБИС «Библиобус». Рассмотрены стандартные ошибки, которые допускают пользователи электронных каталогов БЕН РАН при формировании запросов. Предлагаются механизмы и решения, снижающие количество ошибок при поиске библиографической информации.

Ключевые слова: электронный каталог, анализ запросов, поиск в каталоге, электронная версия, автоматизированные библиотечно-информационные системы

Here is presented a statistical analysis of search queries performed in the electronic catalogue of non-periodical publications of library-automated system «Bibliobus». The article reviews standard errors, made by electronic LNS RAS catalogues users while generating queries. It also proposes ways and solutions that reduce the number of errors when searching for bibliographic information.

Keywords: electronic catalogue, query analysis, directory search, electronic version, library-automated system

При создании моделей и поисковых алгоритмов автоматизированных библиотечно-информационных систем (АБИС), используемых в специализированных научных библиотеках, большое значение для разработчиков имеет наличие обратной связи с пользователями. Основным способом получения адекватной информации об эффективности работы поисковых механизмов АБИС, помимо прямого общения с пользователями или проведения каких-то выборочных опросов, является непосредственный анализ действий пользователей при формировании и выполнении ими запросов.

Реализованный в АБИС «Библиобус» [1, 3] механизм протоколирования операций позволят провести некоторый статистический анализ выполняемых поисковых запросов. Анализу были подвергнуты поисковые запросы к технологическому электронному каталогу БЕН РАН. Данный электронный каталог неперIODических изданий имеет статус служебного и отличается от общедоступного читательского [3] организацией пользовательского интерфейса и механизмом обработки и вывода результатов выполняемых запросов. К особенностям интерфейса данной версии ОПАК (*от англ. Open Public Access Catalog*) можно отнести его ориентацию, в первую очередь, на использование библиотекарями, что, впрочем, позволяет вполне успешно использовать его и в качестве читательского. Требования к поисковым возможностям и пользовательскому интерфейсу данной версии электронного каталога несколько отличаются от требований, обычно предъявляемых к версиям ОПАК, предназначенных для читателей [4]. Особое внимание при разработке интерфейса было уделено его удобству для библиотекарей и простоте, а также способу отображения служебной информации (данных инвентаризации, сроков обработки изданий и т. д.) (рис. 1–2).

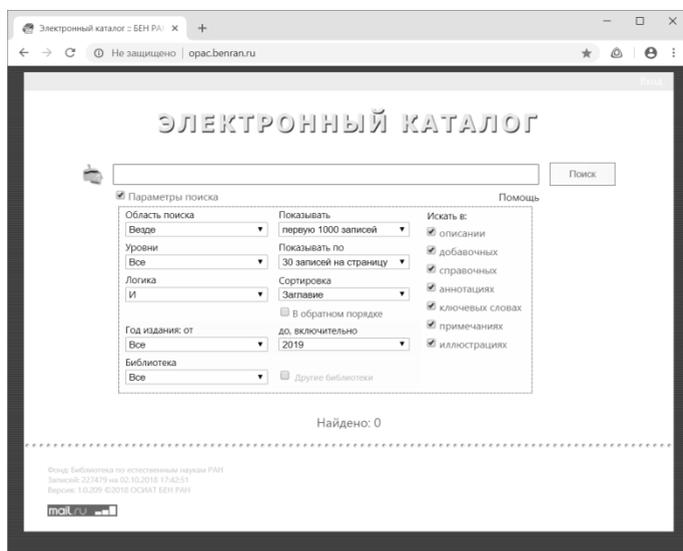


Рис. 1. Интерфейс электронного каталога неперIODических изданий

Были проанализированы запросы к данному электронному каталогу за период с конца марта 2016 г. по конец сентября 2018 г. Всего за указанный период пользователями был выполнен 78 921 запрос. Из них 17 692 (22 %) оказались безрезультатными (рис. 3).

Распределение запросов по областям поиска выглядит следующим образом:

- Поиск по всем фрагментам текста библиографического описания: 75 964 запросов, из них 13 268 использовали фильтр диапазона года издания. Удовлетворяющие запросу записи не найдены в 17 049 случаях.
- Поиск по персоне: 1 719 запросов (298 использовали фильтр года издания), из них для 293 ничего не найдено.
- Поиск по фрагментам заглавий: 1 049 запросов (394 использовали фильтр года издания), из них 316 безрезультатны.
- Поиск по коллективному автору: 21 запрос (7 использовали фильтр года издания), 19 без результата.
- Поиск по рубрикам УДК: 38 запросов (30 использовали фильтр года издания), из них 15 безрезультатны.

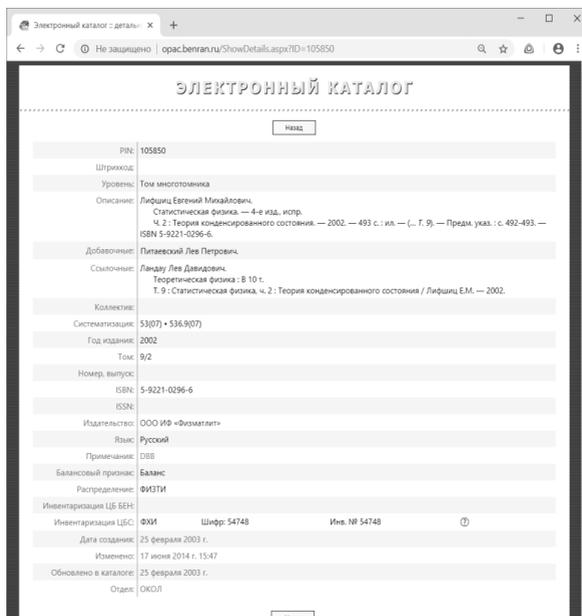


Рис. 2. Интерфейс страницы подробных сведений об издании

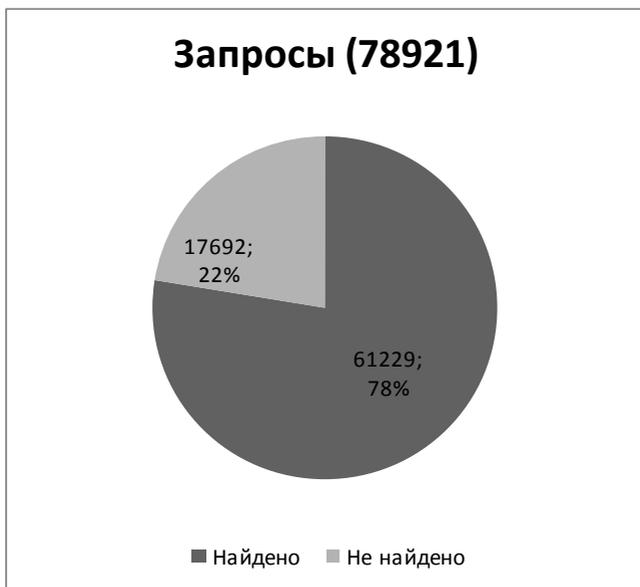


Рис. 3. Результаты выполнения поисковых запросов

На рисунке 4 представлена диаграмма распределения запросов по областям поиска. Представленные результаты подтверждают распространённое мнение о том, что подавляющее число поисковых запросов всегда выполняется в режиме по умолчанию, в нашем случае в режиме «Везде». Большая часть (85 %) запросов выполнялась без использования каких-либо уточняющих условия поиска фильтров и опций. Наиболее же востребованным и чаще всего используемым при формировании поисковых запросов является фильтр по фонду конкретной библиотеки ЦБС. Он применялся 29 358 раз, в 27 % случаев. Такое достаточно большое количество фильтраций запросов по фонду библиотеки объясняется тем, что данный ОПАК используется библиотеками ЦБС БЕН РАН для удалённого ввода в центральную систему данных локальной инвентаризации – шифров и инвентарных номеров для экземпляров изданий своего фонда. Фильтр по диапазону года издания использовался в 14 127 случаях, что составляет около 15 % от общего числа запросов. Выражения же булевой логики (И, ИЛИ, НЕ) популяр-

ностью у пользователей не пользуются и использовались всего 220 раз. Так как возможность тематического поиска в данной версии электронного каталога ограничена поиском по кодам систематизации (индексам УДК), данный режим используется крайне редко.



Рис. 4. Распределение запросов по областям поиска

Анализ неудовлетворённых запросов пользователей позволяет выделить наиболее распространённые причины получения таких результатов. Кроме очевидных попыток найти заведомо отсутствующие в базе данных издания, пользователи при создании запроса достаточно часто допускают ряд стандартных ошибок, среди которых можно выделить следующие:

- опечатки в словах запроса;
- несоответствие вводимых словоформ существующим в каталоге;
- одновременное использование символов латиницы и кириллицы в отдельных словах запроса;
- ошибки при применении поисковых фильтров.

Рассмотрим их более подробно.

Опечатки в словах запроса. Довольно часто встречаются стандартные «перебивки», типа «бибилография» или «бибдиотека», которые достаточно легко отлавливаются поисковой системой, которая информирует пользователя об отсутствии данного слова в базе данных электронного каталога (рис. 5).

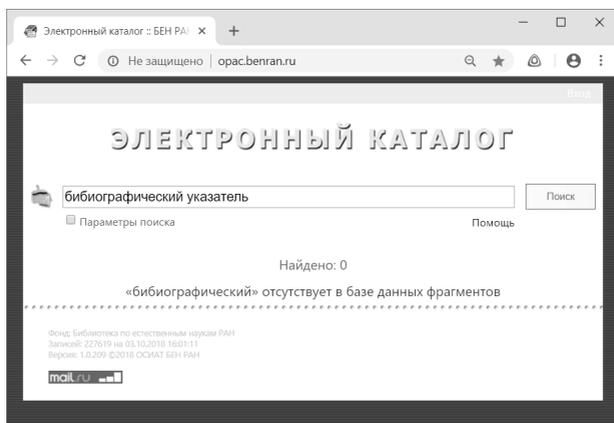


Рис. 5. Пример ошибочного запроса

Несоответствие вводимых словоформ существующим в каталоге.

Данная ошибка встречается достаточно часто и возникает, например, при попытке найти слова, которые введены в библиографическое описание издания в сокращенном виде. И, хотя в АБИС «Библиобус» работает механизм автоматического преобразования сокращений в полную форму, он применяется далеко не ко всем возможным вариантам, которые предусматривает действующий ГОСТ Р 7.0.12-2011. Так, например, в библиографическое описание может быть введено словосочетание «автомобильный завод» в виде сокращения «автомоб. з-д» и при отсутствии механизма автоматического преобразования данного выражения в полную форму оно вряд ли будет когда-либо найдено пользователем. Также запрос останется неудовлетворенным при попытке найти издание авторства «Йен Стюарт» вместо «Иен Стюарт». Причём, необходимо понимать, что в этом случае система не обязательно выдаст предупреждение о том, что данное слово («Йен») отсутствует в каталоге, так как

оно там может присутствовать в других контекстах. Тот же результат будет получен при вводе выражений типа «проектирование» вместо «проектирования» и т. д. В качестве решения данной проблемы можно предложить создание достаточно сложного аппарата лингвистического анализа словоформ и контроля орфографии, но, к сожалению, ресурсы библиотеки весьма ограничены, в отличие от коммерческих компаний типа Яндекс или, прости господи, Гугл, и возможность создания таких достаточно сложных мультязычных механизмов своими силами представляется в настоящее время весьма проблематичной. Поэтому пользователям остаётся только при составлении поисковых запросов принимать во внимание рекомендации разработчиков, а именно: учитывать возможность наличия в тексте сокращений, использовать возможности поиска по маске с левым или правым усечением и т. д. Возможность поиска с использованием левого или правого усечения, кстати, далеко не всегда понимается пользователями адекватно. Так, размещённый на странице помощи ОПАК следующий текст: «При формировании поискового запроса могут быть использованы символы левого и правого усечения – ”*”». Например, в результате выполнения поискового запроса вида “Учеб*” будут найдены все слова и слово-формы вроде “Учебник”, “Учебное”, “Учебному” и т. д.» некоторыми пользователями понимается весьма своеобразно и они, зачастую, используют запросы вида «основы проектирования*», что является абсолютно бессмысленным с точки зрения поисковой системы.

Одновременное использование символов латиницы и кириллицы в отдельных словах запроса. Данная ошибка встречается не очень часто (в нашем случае 130 раз), но также приводит к невозможности найти нужное издание. Чаще всего она возникает при вводе в строку поискового запроса текста на разных языках, например, «программирование на С#», где «С» введено на кириллице, тогда как в тексте библиографического описания данное выражение всегда вводится на латинице. В этом случае система также выдаст предупреждение об ошибке (рис. 6). Данная проблема частично решается вводом в словарь преобразования сокращений АБИС «Библиобус» выражения «С#» для взаимопреобразования его на латинице и на кириллице так, как это сделано, например, для уже введённого в словарь системы выражения «С++», которое будет найдено при написании «С» как на кириллице, так и на латинице.

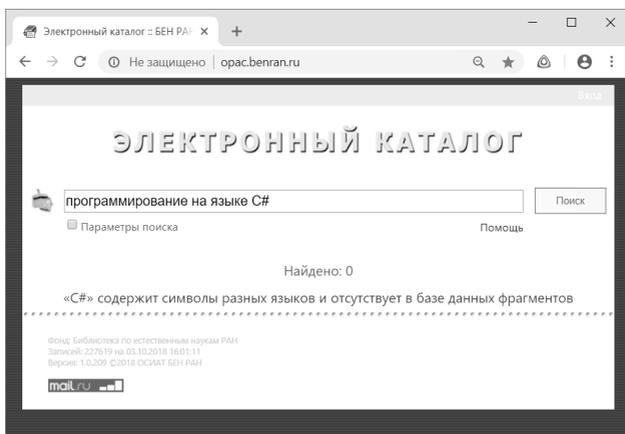


Рис. 6. Пример смещения в запросе символов разных языков

Ошибки при применении поисковых фильтров. Данный тип ошибок встречается нечасто и связан, в первую очередь, с невнимательностью пользователя при составлении поискового запроса. Так, например, ввод выражения типа «Марапулец Юрий Валентинович. Язык C++» при установленном фильтре области поиска «Персона» является заведомо бессмысленным. Пользователям можно только порекомендовать быть внимательными при формировании запросов и ознакомиться с подсказками по их составлению (переход по ссылке «Помощь» в ОПАК).

Проведенный анализ запросов позволяет выявить наиболее узкие места в обеспечении релевантности и полноты поиска библиографической информации в электронных каталогах БЕН РАН и усовершенствовать используемые поисковые механизмы.

Литература

1. Автоматизированная библиотечно-информационная система «Библиобус». Режим доступа: <http://bibliobus.benran.ru> (07.10.2018).

2. Васильев А. В. Программный комплекс «Библиобус» как ядро автоматизированной системы управления библиотечными технологическими процессами [Электронный ресурс] / А. В. Васильев, Н. Е. Каленов // Программные продукты, системы и алгоритмы. 2018. № 3.

С. 40–46. Режим доступа: <http://swsys-web.ru/the-bibliobus-software-package.html> (07.10.2018).

3. Власова С. А. Новые поисковые возможности и востребованность каталога книг и продолжающихся изданий БЕН РАН [Электронный ресурс] / С. А. Власова, Н. Е. Каленов // Информационное обеспечение науки: новые технологии: сб. науч. тр. / отв. ред. П. П. Трескова ; сост. О. А. Оганова. Екатеринбург, 2016. С. 171–178. Режим доступа: http://www.benran.ru/SEM/Sb_16/Sbornik/171.pdf (07.10.2018).

4. Скарук Г. А. Электронные каталоги библиотек в борьбе за пользователя: «старые» и новые подходы [Электронный ресурс] / Г. А. Скарук // Библиосфера. 2016. № 2. С. 7–15. Режим доступа: http://bibliosfera.spsl.nsc.ru/bibliosfera/&journal_page=archive&id=1515&article_id=32591 (07.10.2018).