

Г. А. Скарук
ГПНТБ СО РАН

Читательский поиск в электронном каталоге: итоги анализа поисковой статистики ГПНТБ СО РАН

Освещены результаты проведённого в ГПНТБ СО РАН исследования поискового поведения пользователей электронного каталога на основе информации из лог-файлов ЭК, представленной в базе данных статистики за 2016 г. Оценена частота применения при поиске информационно-поисковых языков разных типов. Первое место по активности использования занимает язык ключевых слов, второе – предметизационный ИПЯ. Охарактеризованы особенности читательского поиска в каталоге с использованием ключевых слов. Рассмотрены варианты построения поисковых предписаний на этом ИПЯ. Проанализированы индивидуальные истории поисков. Выделены типовые поисковые стратегии, способы формулировки и корректировки поисковых предписаний. Наиболее типичны: перебор простых поисковых стратегий, многократная формулировка поисковых предписаний, просмотр незначительного количества страниц результатов поиска. Сделан вывод о серьёзности проблем, возникающих у пользователей ЭК. Предложен комплекс мер по исправлению ситуации.

Ключевые слова: электронный каталог, пользователи, информационный поиск, поисковые стратегии, проблемы поиска.

Galina Skaruk

*State Public Scientific and Technological Library of the Russian Academy of Sciences
Siberian Branch, Novosibirsk, Russia*

Users' search in OPAC: Findings of search statistics analysis at RAS SB State Public Scientific and Technological Library

OPAC user search behavior study was conducted at RAS SB State Public Scientific and Technological Library based on OPAC log files information in statistical database for 2016. The frequency of information retrieval languages, the keyword language in the first place, and the information retrieval language of subject headings in the second place, was assessed. User search behavior in OPAC using keywords is characterized. Different search request structures

in this language are discussed. Individual search stories are analyzed. Typical search strategies, formulations, update methods are defined. The author concludes on the urgency of problems OPAC users face and suggests how to improve the situation.

Keywords: OPAC, users, information retrieval, search strategies, the problem of finding.

Scie-tech public library of Siberian branch of RAS has studied the search behavior of users, on the base of log files data and 2016 OPAC statistics. Pretty often there are thematic searches (subject headings, key words, LBC) and searches for the title of documents. Readers do not complicate the wording of queries or present them in any special form for computer-based search. Rarely used are advanced search strategies, and the Boolean logic. Popularity of subject headings arise because they are assigned by the Russian Book Chamber. This analysis leads us to the conclusion that the process of creating search query is difficult. Often readers sort out the wording many times, but they cannot reach the desired result. This indicates that the skills of clear wording of queries, working with search languages and logical operators, there is practically no consistent correction of the PP. Users are persistent enough: they send a lot of searches to achieve the goal and spend a lot of time on it. The main directions of improvement are: 1. Development and implementation of a system of statistical indicators to assess the quality and effectiveness of EC. In particular, information on the categories of users conducting searches, the volume of issuance, and the time spent searching will be useful. 2. Development of motivation of users to conduct high-quality search. 3. Activation of training in work with electronic search systems of different types. Inclusion of courses on training ILS users in the programs of all universities and institutes. Review user training: pay more attention to creative, intelligent search moments. According to polls, users are in demand not only technical skills, but also the ability to clearly define the topic of search, to choose terms when drafting search requirements. Training programs should be formed on the basis of analysis of information-search behavior of users. 4. Creation of high-quality search guides for users of different specialties and different skill levels, 5. Providing reverse chronological ranking of issuance in EC. 6. Introduction of mechanisms for visualizing the content of library collections. Use of a pre-established complex mechanism (knowledge map, ontology of concepts, knowledge network).

В начале эпохи электронных каталогов (ЭК) мы с восторгом говорили о неизмеримо возросших в связи с автоматизацией поисковых возможностях. Это и поиск по каждому слову библиографической записи, и форми-

рование поискового предписания (ПП) с использованием разных логических и контекстных операторов, и сочетание различных по структуре и свойствам информационно-поисковых языков (ИПЯ), и широкие возможности просмотра словарей ИПЯ и корректировки ПП, и многое другое. Сегодня ЭК стали привычной реальией библиотечной жизни, многие из названных преимуществ (хотя далеко ещё не все) предоставлены читателям. Но этого, как мы убедились, оказалось недостаточно.

В чём же дело? Нельзя не согласиться с выводом Э. Р. Сукиасяна о том, что широкие возможности ЭК в нашей стране не используются. В статье «Развитие автоматизированных библиотечных технологий в России: мысли об ожидаемом, но неосуществлённом» [1] перечислен ряд причин этой ситуации. Хочу акцентировать внимание на одной из них – мы забываем о читателе. Нельзя не учитывать: то, как будут реализованы возможности каталогов, зависит не только от их технических свойств, но и от пользователя. Как читатели работают с ЭК, применяют ли все его функциональные возможности, получают ли необходимые результаты? Информации такого рода в печати очень мало, несмотря на острую в ней необходимость. Ведь именно на этих сведениях должна основываться работа по совершенствованию ЭК и созданию системы помощи их пользователям.

Нередко мы оставляем читателя наедине с ЭК и краткой инструкцией, не интересуясь ходом поиска и возникающими у пользователя проблемами. Особенно это касается удалённого поиска в каталогах, представленных на сайтах библиотек.

ГПНТБ СО РАН провела исследование поискового поведения пользователей, ориентируясь на данные, извлечённые из лог-файлов ЭК книг ГПНТБ СО РАН и представленные в БД статистики за 2016 г. Основные методы исследования – сравнительно-статистический анализ и анализ индивидуальных историй поиска. Мы стремились выяснить, какими лингвистическими средствами и поисковыми стратегиями оперируют пользователи ЭК, достаточно ли эффективным оказывается самостоятельный читательский поиск.

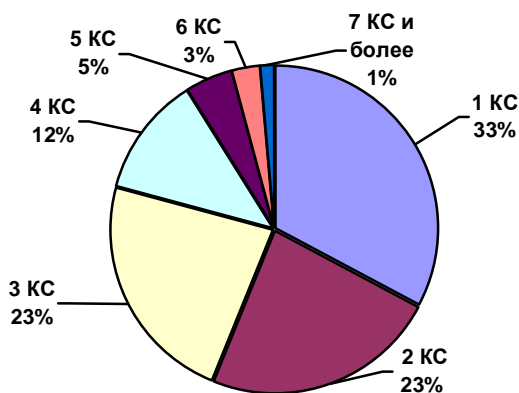
На первом этапе частота использования тех или иных ИПЯ устанавливалась по количеству префиксов, предваряющих в БД различные поисковые признаки. Результаты этого этапа отражены в табл. 1.

Частота поиска по различным поисковым признакам

Поисковые признаки	Количество поисков в САТ	Отношение к общему числу поисков в ЭК, %
По ключевым словам	2 100 104	19,3
По заглавию	70 194	0,02
По автору	3 996 909	36,8
По году издания	9 019	0,08
По характеру документа	3 221	0,03
По предметным рубрикам	2 495 641	23,0
По индексам ББК	9 017	0,08
По дате поступления	30 507	0,3

Из этих данных следует, что в ЭК чаще всего ведут тематический поиск (ИПЯ предметных рубрик, ключевых слов, ББК) и поиски по заглавиям документов (элементы языка библиографического описания). Но если над совершенствованием языка библиографического описания ведётся постоянная работа, то языкам тематического поиска в российских ЭК уделяется явно недостаточно внимания.

Как показало исследование, пользователи-читатели чаще всего применяют при поиске ключевые слова. Поэтому мы особое внимание уделяли анализу поисковых предписаний на этом языке. Прежде всего нас интересовало, насколько детально читатели формулируют запросы.



Распределение количества слов в поисковых предписаниях на языке ключевых слов

Среднее количество слов в ПП на этом ИПЯ – 2,44. Распределение количества слов в ПП в выборке из 500 поисков показано на рисунке.

Таблица 2

Лексические формы поисковых предписаний

Форма поискового предписания	Отдельные слова	Фраза на естественном языке	Набор слов, организованный по принципам предметной рубрики
Количество ПП данного типа	218 (в том числе состоящие из одного слова – 163)	139	143
Отношение к общему объёму выборки, %	43,6 (32,6)	27,8	28,6

Как видим, читатели не склонны усложнять формулировки запросов или представлять их в какой-либо специальной форме для машинного поиска. Редко используется стратегия расширенного поиска, предполагающая применение булевой логики. Формулировки в виде предметных рубрик возникают потому, что предметные рубрики, присваиваемые Российской книжной палатой, в ЭК ГПНТБ СО РАН попадают в словарь ключевых слов. При поиске пользователи внимательно относятся к контексту, в котором представлен поисковый термин, это помогает им сделать выбор, уточнить информационную потребность.

Проведённый анализ подводит нас к выводу о том, что процесс составления ПП для пользователей представляет существенную трудность. Зачастую читатели перебирают формулировки многократно, но не могут выйти на нужный результат. Это свидетельствует о том, что навыки чёткой формулировки запросов, работы с поисковыми языками и логическими операторами, последовательной корректировки ПП у них нет.

В то же время пользователи ЭК достаточно настойчивы: проводят для достижения цели большое количество поисков и затрачивают на это много времени. Проблема в том, что большинство используемых читателями стратегий не даёт ожидаемых результатов. Но пользователи пытаются исправить ситуацию.

На выборке 200 поисков были исследованы стратегии корректировки поисковых предписаний, составленных на языке ключевых слов. Применяемые пользователями стратегии корректировки поисковых предписаний представлены в табл. 3.

**Стратегии корректировки поисковых предписаний,
применяемые пользователями**

Стратегия	Количество поисков с применением данной стратегии	Отношение к общему числу корректировок, %
Перефразирование запроса без изменения объёма отражённого в нём понятия	36	18
Увеличение количества слов в ПП	34	17
Более глубокое усечение терминов	28	14
Применение формулировок, намного более широких, чем интересующая тема	26	13
Сокращение количества слов в ПП	14	7
Более узкая формулировка темы	13	6,5
Переход к поиску по словарю	12	6
Использование другого поискового признака (переход на другой ИПЯ)	11	5,5
Более широкая формулировка темы	6	3
Поиск по сочетанию лексических единиц различных ИПЯ	1	0,5
Без корректировки	38	

Как видим, пользователи ЭК применяют разнообразные стратегии, но логическая обоснованность использования тех или иных методов корректировки не всегда прослеживается. Общий подход – стремление использовать как можно больше стратегий, иногда прямо противоположных, например сокращение и увеличение количества слов в запросе.

Следует обратить внимание на стремление читателей значительно расширить формулировку предписания по сравнению с запрошенной первоначально темой. Цель понятна – включить в поле обзора все узкие темы в рамках предложенной широкой. Однако в вербальных языках эта цель таким способом не достигается. Целесообразнее перейти к индексу иерархической классификации и вести поиск с усечением. Читатели этот метод не применяют. Заметно и то, что метод подбора синонимов также не востребован. Всё это говорит о том, что в руководствах по поиску для пользователей нужно предусматривать специальные разделы, посвящённые корректировке поискового образа запроса.

Ещё одна выявленная проблема: в выборке обнаружено 13 грамматических ошибок пользователей. Решить её помогли бы составление словаря часто встречающихся ошибок и их автоматическая корректировка.

На следующем этапе исследования были рассмотрены индивидуальные истории поиска читателей. Необходимо отметить, что БД статистики ГПНТБ СО РАН не предоставляют некоторых данных, необходимых для полного анализа. К таким параметрам можно отнести категории пользователей, количество выданных в ответ на запрос БЗ, время просмотра результатов поиска. Поэтому ряд процедур анализа приходилось осуществлять вручную. Например, для определения объёма выдачи повторялся поиск, зафиксированный в лог-файлах.

В итоге мы выявили набор типовых операций, осуществляемых пользователями при поиске.

Основная тенденция – многократный перебор простых поисковых стратегий. Часто используемые приёмы – поиск в различных вариантах без просмотра выдачи, выбор наиболее результативной стратегии и повторный поиск по ней.

Достаточно часто краткие неконкретные поисковые предписания приводят к выдачам очень большого объёма. При этом установлено, что пользователь обычно делает выбор на первых трёх страницах, не просматривая дальше результаты поиска ни при чрезвычайно большой, ни при малой выдаче. Получая при поиске до нескольких тысяч БЗ, он не тратит много времени на просмотр выдачи, удовлетворяясь несколькими первыми порциями документов. Например, при поиске по части слова «ради» (тема «радио», проведено 9 поисков) выдача составила 7 271 БЗ. В итоге было просмотрено 100 первых описаний документов (5 страниц) и выбраны 4 книги. Этот результат говорит о целесообразности обратнотронологического порядка вывода результатов поискового сеанса.

Преобладающая стратегия поиска по предметным рубрикам – однократный ввод полного текста рубрики, заранее отобранной в АПУ. Другие варианты формулировок ПП не приводятся. Таким образом, можно заключить, что этим методом поиска пользуются квалифицированные пользователи – сотрудники библиотек. Та же стратегия применяется и при поиске по ББК. Результат чаще всего выводится на печать.

Следующий вывод по результатам анализа – пользователями востребован тематический поиск с глубокой ретроспективой. Необходимо создавать и дорабатывать его механизмы.

Общий итог – стратегии, используемые читателями, не обеспечивают качественный результат поиска. Очевидно, что создание системы помощи пользователям ЭК должно стать актуальной задачей. Такая система – это комплекс инструментов анализа читательских поисков, мер по совершенствованию качества поисковых образов документов в ЭК, автоматических процедур, инструктивно-методической документации, обучающих мероприятий и консультаций библиографов.

Наметим основные направления действий:

1. Разработка и внедрение системы статистических показателей, позволяющих оценить качество и эффективность ЭК. В частности, полезными будут сведения о категориях пользователей, проводящих поиск, об объёме выдачи, времени, затрачиваемом на поиск.

2. Развитие мотивации пользователей на проведение качественного поиска.

3. Активизация обучения работе с электронными поисковыми системами разных типов. Включение курсов по обучению пользователей ИПС в программы всех университетов и институтов. Пересмотр обучения пользователей: больше внимания уделять творческим, интеллектуальным моментам поиска.

По данным опросов, пользователями востребованы не только технические навыки, но и умения чётко определять тему поиска, выбирать термины при составлении поисковых предписаний. Программы обучения должны формироваться на основе анализа информационно-поискового поведения пользователей.

4. Создание качественных руководств по поиску для пользователей разных специальностей и разного уровня квалификации.

5. Обеспечение обратнхронологического ранжирования выдачи в ЭК.

6. Введение механизмов визуализации содержания фондов.

Использование заранее созданного комплексного механизма (карта знаний, онтология понятий, сеть знаний). Наглядное представление структуры и содержания библиотечных фондов – функция контролируемых ИПЯ, значение которой при удалённом самостоятельном поиске возрастает. Читателям должна быть предоставлена возможность просматривать разнообразные словари ИПЯ, структурирующие содержание библиотечных фондов в различных логиках и контекстах. Наиболее эффективный вариант – создание в ЭК комплекса лингвистических средств [2, 3].

ЭК нового поколения критикуют за то, что они игнорируют богатую схему взаимосвязей, предлагаемую контролируемым (нормализованным) предметным словарём, который может облегчить предметный просмотр. В ЭК ГПНТБ СО РАН роль такого механизма в будущем сможет выполнять авторитетный файл предметных рубрик, сопряжённых с индексами ББК.

7. Разработка системы ситуативных «всплывающих» подсказок (советов по поиску) для пользователей. Подсказки могут появляться вместе с вопросами системы по поводу удовлетворённости пользователей процессом и результатами поиска, вводиться по требованию, автоматически появляться на тех этапах поиска, где пользователи традиционно испытывают наибольшие затруднения. Они могут содержать как советы по поиску, так и ссылки на полезные ресурсы, например авторитетные файлы.

8. Внедрение механизмов помощи в корректировке ПП. Наиболее эффективным методом корректировки считается фасетная навигация. Она предполагает возможность указывать разные фасеты (параметры, желаемые признаки) искомой информации, товара, услуги. И с каждым новым параметром пользователю предлагается всё меньшее количество объектов.

Фасетный подход применяется как инструмент фильтрации (корректировки) результатов поиска, особенно проведённого с применением упрощённых стратегий (например, по ключевым словам через единое поисковое окно). В таких случаях пользователь может корректировать запрос с помощью разнообразных фасет (дата создания, язык текста, режим доступа, вид ресурса, принадлежность к определённой коллекции, место хранения, тема и т.п.). В России этот принцип реализован на базе зарубежных программных платформ, например в ЭК РНБ (система автоматизации библиотек *Primo*, фирма «*EX Libris*»).

9. Активная поддержка пользовательского поиска библиографами-консультантами. Мы на новом уровне возвращаемся к традициям консультирования у каталогов, которые в условиях автоматизации, казалось, стали не столь уж необходимы. Поскольку значительная часть поисков в ЭК проводится в дистанционном режиме, библиотеки осваивают дистанционные формы консультирования. Основной упор делается на непосредственное общение библиографов и читателей «лицом к лицу», мгновенные ответы на вопросы. Для этого используются библиотечные чаты и общение в режиме телеконференций.

Итак, мы выяснили, что автоматизация поиска не только обогащает ЭК новыми возможностями, но и привносит много трудностей. ЭК представляется читателям «чёрным ящиком», в котором неизвестно что содержится, и непонятно, как его использовать. Самостоятельный поиск – серьёзнейшая проблема, которую необходимо оперативно решать.

В заключение хотелось бы отметить, что ещё в начале XXI в. одним из специфических преимуществ ЭК (ОРАС) по сравнению с другими типами АИПС считались их ориентация на неподготовленного пользователя, учёт их проблем и организация системы помощи в поиске. Десятилетие спустя никто уже не станет отрицать, что первенство в этом вопросе перешло к поисковым машинам интернета и веб-сервисам. Значит библиотеки должны попытаться выйти за рамки традиционных моделей и активно включиться в соревнование на рынке информационных сервисов.

Каталоги нового поколения подстраивают принципы поиска в ЭК под привычные пользователям интернета механизмы и стратегии. Как и в поисковых системах интернета, в них реализованы простые поисковые окна с вводом в одно окно полного текста запроса на естественном языке, ранжи-

рование результатов поиска по релеванности, сходный с интернет-поисковиками дизайн интерфейса, использование рекомендательных сервисов, в том числе систем исправления грамматических ошибок при вводе запросов (типа «Вы имели в виду...», «Читатели, выбравшие эту книгу, выбирают также...»).

Результатами стремления приблизить библиотечную поисковую среду к привычной для пользователей социальных сетей стали применение при индексировании для ЭК фолксономии, тегирования документов; присоединение к документам комментариев и оценок пользователей.

При этом необходимо учитывать: стратегии поисковых машин ориентированы на поиск в полных текстах, что обеспечивает гораздо большую полноту выдачи. Кроме того, на разработку поисковых машин затрачиваются неизмеримо большие финансовые и кадровые ресурсы, чем на создание каталогов. Поэтому создатели ЭК должны не только подражать поисковым системам интернета, но и искать свои нишевые функции, позволяющие им занять своё место в конкурентной поисковой среде.

СПИСОК ИСТОЧНИКОВ

1. **Сукиасян Э. Р.** Развитие автоматизированных библиотечных технологий в России: мысли об ожидаемом, но неосуществленном // Науч. и техн. б-ки. – 2017. – № 4. – С. 34–43.

Sukiasyan E. R. Razvitiye avtomatizirovannykh bibliotечnykh tehnologiy v Rossii: mysl'i ob ozhidaemom, no neosushchestvlenno // *Nauch. i tehn. b-ki.* – 2017. – № 4. – S. 34–43.

2. **Скарук Г. А.** Комплексное использование лингвистических средств тематического поиска в электронном каталоге // Библиотековедение. – 2010. – № 2. – С. 34–40.

Skaruk G. A. Kompleksnoe ispolzovanie lingvisticheskikh sredstv tematicheskogo poiska v elektronnom kataloge // *Bibliotekovedenie.* – 2010. – № 2. – S. 34–40.

3. **Скарук Г. А.** Принципы комплексного использования лингвистических средств тематического поиска в электронном каталоге / Г. А. Скарук // Науч. и техн. б-ки. – 2014. – № 2. – С. 72–76.

Skaruk G. A. Printsipy kompleksnogo ispolzovaniya lingvisticheskikh sredstv tematicheskogo poiska v elektronnom kataloge / *G. A. Skaruk* // *Nauch. i tehn. b-ki.* – 2014. – № 2. – S. 72–76.

Galina Skaruk, Cand. Sc. (Pedagogy), Senior Researcher, State Public Scientific and Technological Library of the Russian Academy of Sciences Siberian Branch;

skaruk@spsl.nsc.ru

15, Voskhod st., 630200 Novosibirsk, Russia