

ИНФОРМАЦИОННО-ПОИСКОВЫЕ ЯЗЫКИ

УДК 025.4.03:[025.355:004.738.5]

<https://doi.org/10.33186/1027-3689-2022-10-52-65>

Лингвистические средства информационного поиска в электронных каталогах крупнейших библиотек России

Е. М. Зайцева¹, Ю. В. Смирнов²

¹, ²ГПНТБ России, Москва, Российская Федерация

¹Московский государственный лингвистический университет,
Москва, Российская Федерация

¹katja@gpntb.ru, <https://orcid.org/0000-0001-7554-3032>

²yury@gpntb.ru, <https://orcid.org/0000-0003-0363-7229>

Аннотация. Статья представляет собой результат проведённого исследования состояния лингвистических средств информационного поиска, используемых в электронных каталогах (ЭК) крупнейших библиотек России. В список таких библиотек были включены те, объём фонда которых превышает 4 млн единиц хранения. Проанализированы представленные в поисковом веб-интерфейсе ЭК библиотек виды поиска, выявлены следующие его разновидности: однострочный (простой), стандартный (базовый), расширенный и профессиональный. Наиболее популярные виды – однострочный и расширенный, они применяются во всех указанных библиотеках. Рассмотрены дополнительные поисковые возможности: системы уточнения, ранжирования и сортировки результатов поиска, по крайней мере одна из которых в обязательном порядке присутствует в любом из изученных ЭК, а также системы поисковых словарей и рекомендаций дополнительных документов и особенности использования этих систем. Проанализированы применяемые информационно-поисковые языки (ИПЯ) и сделан вывод: ЭК всех рассмотренных библиотек предоставляют возможности поиска по различным наборам элементов библиографического описания, самые распространённые из которых – автор, заглавие, год издания, а также тематического поиска с использованием вербальных языков (все рассмотренные библиотеки) и классификационных языков (большинство рассмотренных библиотек). Сформулированы общие черты, присущие всем ЭК (или большинству) крупнейших библиотек России, в плане применения видов поиска, дополнительных поисковых возможностей и лингвистических средств информационного поиска. Определены рекомендации по развитию и эффективному использованию рассмотренных средств при совершенствовании ЭК.

Публикация подготовлена в рамках Государственного задания ГПНТБ России на 2022 г. по выполнению работы № 720000Ф.99.1.Б385АА03000 по теме № 1021062311369-1-1.2.1;5.8.2.

Ключевые слова: лингвистические средства, информационный поиск, информационно-поисковые языки, электронные каталоги, библиотеки России

Для цитирования: Зайцева Е. М., Смирнов Ю. В. Лингвистические средства информационного поиска в электронных каталогах крупнейших библиотек России / Е. М. Зайцева, Ю. В. Смирнов // Научные и технические библиотеки. 2022. № 10. С. 52–65. <https://doi.org/10.33186/1027-3689-2022-10-52-65>

INFORMATION RETRIEVAL LANGUAGES

UDC 025.4.03:[025.355:004.738.5]

<https://doi.org/10.33186/1027-3689-2022-10-52-65>

Information retrieval linguistic instruments in e-catalogs of Russian large libraries

Ekaterina M. Zaitseva¹, Yury V. Smirnov²

^{1,2}*Russian National Public Library for Science and Technology,
Moscow, Russian Federation*

¹*Moscow State Linguistic University, Moscow, Russian Federation*

¹katja@gpntb.ru, <https://orcid.org/0000-0001-7554-3032>

²yury@gpntb.ru, <https://orcid.org/0000-0003-0363-7229>

Abstract. The article is based on the findings of the study of linguistic instruments of information search used in e-catalogs of Russian largest libraries. For the study, the libraries with the collections of over 4 million entries were selected. The authors analyze search types applicable to e-catalog search interface.

They specify the following search types applied in the above libraries: single-line (simple) search, standard (basic), advanced and professional search. The advanced search options are discussed, namely refinement systems, ranking and sorting of search results; at least one of these options is used in the e-catalogs under study. The authors also analyze the systems of search dictionaries, guidelines and use characteristics. The popular information retrieval languages are discussed with the conclusion made that the e-catalogs under study offer search options for various sets of bibliographic elements, among them author, title, year of publication, and subject search based on verbal languages (every library under study) and classification languages (majority of the libraries). The authors formulate the common features inherent in the libraries' e-catalogs and information retrieval linguistic instruments. They also offer recommendations on how to develop and efficiently use these instruments for upgrading library e-catalogs.

The article is prepared within the framework of the Government Order No. 1021062311369-1-1.2.1;5.8.2 to RNPLS&T for 2022, Project № 720000F.99.1.BZ85AA03000.

Keywords: linguistic instruments, information search, information retrieval languages, e-catalogs, Russian libraries

Cite: Zaitseva E. M., Smirnov Yu. V. Information retrieval linguistic instruments in e-catalogs of Russian large libraries / E. M. Zaitseva, Yu. V. Smirnov // Scientific and technical libraries. 2022. No. 10. P. 52–65. <https://doi.org/10.33186/1027-3689-2022-10-52-65>

Одна из основных функций библиотек – обеспечение поиска и предоставления информации для читателей, для чего применяются различные лингвистические средства, в частности ИПЯ.

Каждая библиотека формирует свой собственный набор лингвистических средств, который определяется сложившейся практикой индексирования, поисковыми потребностями пользователей и возможностями используемой системы автоматизации. На основе этого набора формируются поисковые опции ЭК библиотеки.

Цель работы: проанализировать использование лингвистических средств информационного поиска в ЭК крупнейших библиотек России.

Объектами исследования выступили ЭК, представленные на сайтах крупнейших библиотек России. В список таких библиотек включены те, объём фонда которых превышает 4 млн единиц хранения [1]:

Российская государственная библиотека (ФГБУ «РГБ») (Москва);
Российская национальная библиотека (РНБ) (Санкт-Петербург);
Библиотека Российской академии наук (БАН) (Санкт-Петербург);
Библиотека по естественным наукам Российской академии наук (БЕН РАН) (Москва);

Институт научной информации по общественным наукам Российской академии наук (ИНИОН РАН) (Москва);

Государственная публичная научно-техническая библиотека Сибирского отделения Российской академии наук (ГПНТБ СО РАН) (Новосибирск);

Научная библиотека Московского государственного университета им. М. В. Ломоносова (НБ МГУ) (Москва);

Государственная публичная научно-техническая библиотека России (ГПНТБ России) (Москва);

Научная библиотека Санкт-Петербургского государственного университета (НБ СПбГУ) (Санкт-Петербург);

Государственная публичная историческая библиотека России (ГПИБ России) (Москва);

Научная библиотека Казанского (Приволжского) федерального университета (НБ КФУ) (Казань);

Донская государственная публичная библиотека (ГБУК «ДГПБ») (Ростов-на-Дону);

Всероссийская государственная библиотека иностранной литературы им. М. И. Рудомино (ВГБИЛ) (Москва);

Самарская областная универсальная научная библиотека (ГБУК «СОУНБ») (Самара);

Кировская государственная универсальная областная научная библиотека им. А. И. Герцена (КГУОНБ) (Киров);

Дальневосточная государственная научная библиотека (ДВГНБ) (Хабаровск).

Настоящий список можно считать представительным для проведения анализа, поскольку в нём присутствуют библиотеки разных уровней и направлений, базирующиеся на разных системах автоматизации (различные версии систем автоматизации библиотек – ИРБИС64, Alerph, KoHa, OPAC-Global и др.): две общероссийские национальные библиотеки, ещё пять библиотек федерального уровня, а также другие

крупные научные и публичные библиотеки. Интересно отметить, что из 16 крупнейших библиотек России только семь находятся в Москве. Остальные расположены в других крупных городах: в Санкт-Петербурге (3), Новосибирске (1), Казани (1), Ростове-на-Дону (1), Самаре (1), Кирове (1), Хабаровске (1).

Каждая библиотека из этого списка имеет актуальный ЭК, представленный на официальном сайте. У ДВГНБ на сайте отражены два ЭК: ведущийся с 1995 г. по 2013 г. и ведущийся с 2014 г. При проведении сравнительного анализа рассматривался последний.

Для начала рассмотрим виды поиска, имеющиеся в поисковом интерфейсе ЭК крупнейших библиотек:

однострочный (простой) – поисковый запрос задаётся в одной строке без указания поля базы данных (БД), при этом поиск может осуществляться не только по всем полям БД, но и по имеющимся полным текстам документов (как, например, в ЭК ГПНТБ России);

стандартный (базовый) – поиск ведётся только по одному выбранному элементу;

расширенный – поиск осуществляется по одному или нескольким полям БД, часто с использованием булевых операторов («И», «ИЛИ», «НЕ»);

профессиональный – поиск можно вести практически по всем полям БД с использованием булевых операторов и различных поисковых выражений, характерных для поисковой системы.

Результаты анализа видов поиска представлены в табл. 1.

Таблица 1

Виды поиска в ЭК крупнейших библиотек России

Библиотека	Однострочный (простой) поиск	Стандартный (базовый) поиск	Расширенный поиск	Профессиональный поиск
ФГБУ «РГБ»	+	–	+	+
РНБ	+	–	+	–
БАН	+	+	+	+
БЕН РАН	+	+	+	–

Библиотека	Однострочный (простой) поиск	Стандартный (базовый) поиск	Расширенный поиск	Профессиональный поиск
ИНИОН РАН	+	–	+	+
ГПНТБ СО РАН	+	+	+	+
НБ МГУ	+	+	+	–
ГПНТБ России	+	–	+	+
НБ СПбГУ	+	+	+	+
ГПИБ России	+	–	+	–
НБ КФУ	+	+	+	–
ГБУК «ДГПБ»	+	+	+	–
ВГБИЛ	+	+	+	–
ГБУК «СОУНБ»	+	–	+	–
КГУОНБ	+	+	+	+
ДВГНБ	+	+	+	+
<i>Итого</i>	16	10	16	8

Однострочный и расширенный виды поиска наиболее популярны и присутствуют в ЭК всех рассмотренных библиотек. Единственную оговорку необходимо сделать относительно ЭК ИНИОН РАН: при неоднократном обращении к сайту ИНИОН РАН не удалось увидеть однострочный поиск в действии, работают только простой и профессиональный виды поиска в конкретных БД ЭК, при этом поиск, названный простым, по своей сути является расширенным. Стандартный и профессиональный виды поиска оказались менее распространёнными.

Стандартный поиск представлен в десяти ЭК, а профессиональный – в восьми ЭК.

Следует отметить, что довольно часто вышеуказанные виды поиска объединяются с целью упростить поиск для пользователя. Например, для однострочного поиска предоставляется возможность дополнительно выбрать из небольшого списка поле БД, наиболее популярное при составлении поискового запроса: автор, заглавие, предметная рубрика, ключевое слово. Таким образом, однострочный поиск комбинируется со стандартным поиском, что можно наблюдать в ВГБИЛ, НБ КФУ, КГУОНБ, ДВГНБ на первой поисковой странице и в БЕН РАН, ГПНТБ СО РАН, НБ МГУ, ГБУК «ДГПБ» – после проведения первичного поиска с использованием единой строки.

Обеспечение простого однострочного поиска в сочетании с многоуровневым поисковым интерфейсом отмечается в качестве одного из ведущих требований к онлайн-каталогу в документе «Руководство по обеспечению тематического доступа в национальной библиографии», разработанном ИФЛА: «Обеспечьте интерфейс с возможностью работы с различной степенью детализации. Рекомендуется предложить в дополнение к интерфейсу с простым окном поиска один или несколько других уровней, предлагающих опции расширенного поиска, включающие более точный тематический поиск или возможность уточнения запросов с помощью булевой логики» [2]. К этому положению, очевидно, нужно добавить требование дифференцированности видов поиска, выполнение которого делает поисковый интерфейс более наглядным и понятным в отличие от комбинирования видов поиска.

С видами поиска тесно связаны такие дополнительные поисковые возможности, как системы уточнения (фильтрации), ранжирования и сортировки результатов поиска, по крайней мере одна из которых в обязательном порядке присутствует в любом из изученных ЭК, а также системы поисковых словарей и рекомендаций дополнительных документов, использование которых в ЭК отражено в табл. 2.

**Дополнительные поисковые возможности
в ЭК крупнейших библиотек России**

Библиотека	Уточнение (фильтрация) результатов поиска	Ранжирование результатов поиска	Сортировка результатов поиска	Использование поисковых словарей	Рекомендация дополни- тельных документов
ФГБУ «РГБ»	+	+	+	-	-
РНБ	+	+	+	-	-
БАН	+	+	-	+	+
БЕН РАН	+	+	+	-	-
ИНИОН РАН	-	-	+	+	-
ГПНТБ СО РАН	+	+	-	+	-
НБ МГУ	-	-	+	+	-
ГПНТБ России	+	-	+	+	-
НБ СПбГУ	+	-	+	+	-
ГПИБ России	+	-	-	+	-
НБ КФУ	+	+	+	+	-
ГБУК «ДГПБ»	+	-	+	+	-
ВГБИЛ	+	+	+	+	-
ГБУК «СОУНБ»	+	-	+	-	-
КГУОНБ	+	-	-	+	-
ДВГНБ	+	-	-	+	-
<i>Итого</i>	14	7	11	12	1

Уточнение (фильтрация) обычно производится после представления пользователю результатов поиска и помогает ему сузить количество полученных результатов. Такая система является самой распространённой и используется в 14 ЭК. Ранжирование результатов поиска по релевантности применяется в семи ЭК, а сортировка результатов поиска – в 11 ЭК.

Системы фильтрации (использования поисковых фасетов) и ранжирования результатов поиска особо значимы и входят в число необходимых характеристик библиотечных ЭК [3]. Фильтрация может осуществляться по набору различных элементов библиографического описания и тематических элементов библиографической записи. Ранжирование результатов поиска по релевантности по своей сути является семантическим или семантико-статистическим ранжированием, базируется на сложном анализе результатов поиска и поэтому пока не нашло широкого применения в ЭК. Теоретически ранжирование результатов поиска можно проводить по другим критериям, таким как, например, популярность документа [Там же], но самым эффективным, безусловно, считается семантическое ранжирование. Сортировка результатов поиска обычно проводится по автору, заглавию, году издания.

Поисковые словари применяются для предоставления подсказок пользователю при вводе запрашиваемых данных или для выбора конкретного термина из поискового словаря без ввода поискового запроса с клавиатуры. Системы поисковых словарей представлены в большинстве библиотек – в 12 ЭК. Система рекомендации дополнительных документов («найти похожие») реализуется как поиск с ранжированием по запросу, составленному из ключевых слов выбранной записи ЭК, и отмечена только в БАН.

В каждой библиотеке используется свой набор ИПЯ: предметные рубрики (ПР), ключевые слова (КС), Библиотечно-библиографическая классификация (ББК), Универсальная десятичная классификация (УДК), Государственный рубрикатор научно-технической информации (ГРНТИ), Рубрикатор Высшей аттестационной комиссии (Рубрикатор ВАК). В библиотеках могут применяться не только вышеперечисленные ИПЯ, но и ИПЯ собственной генерации, например, Рубрикатор ИНИОН РАН, индексы НБ МГУ, индексы ГПИБ России. Кроме тематических ИПЯ, во всех рассмотренных ЭК для поиска, безусловно, используется язык

библиографического описания (ЯБО), при этом поиск по автору, заглавию и году издания присутствует во всех библиотеках. Применение ИПЯ в ЭК крупнейших библиотек России отражено в табл. 3.

Таблица 3

ИПЯ в ЭК крупнейших библиотек России

Библиотечка	ЯБО автор	ЯБО заглавие	ЯБО год издания	ПР	КС	ББК	УДК	ГРНТИ	Рубрикатор ВАК	Другая классификация
ФГБУ «РГБ»	+	+	+	-	+	+	-	-	+	
РНБ	+	+	+	+	+	+	-	-	+	
БАН	+	+	+	+	+	+	-	+	+	
БЕН РАН	+	+	+	+	+	-	-	-	-	
ИНИОН РАН	+	+	+	+	+	-	-	-	-	Рубрикатор ИНИОН РАН
ГПНТБ СО РАН	+	+	+	+	+	+	-	+	+	
НБ МГУ	+	+	+	+	+	-	-	-	-	Индексы НБ МГУ
ГПНТБ России	+	+	+	+	+	-	+	+	+	
НБ СПбГУ	+	+	+	+	+	+	-	-	-	
ГПИБ России	+	+	+	+	+	-	-	-	-	Индексы ГПИБ
НБ КФУ	+	+	+	+	+	-	-	-	-	
ГБУК «ДГПБ»	+	+	+	+	+	-	-	-	-	
ВГБИЛ	+	+	+	+	+	-	-	-	-	
ГБУК «СОУНБ»	+	+	+	-	+	+	-	-	-	
КГУОНБ	+	+	+	+	+	+	+	-	-	

Библиотека	ЯБО автор	ЯБО заглавие	ЯБО год издания	ПР	КС	ББК	УДК	ГРНТИ	Рубрикатор ВАК	Другая классификация
ДВГНБ	+	+	+	+	+	+	+	-	-	
<i>Итого</i>	16	16	16	14	12 + 4	8	3	3	5	3

Проанализированы применяемые ИПЯ, и отмечено, что ЭК всех рассмотренных библиотек предоставляют возможности поиска по различным наборам элементов библиографического описания, а также тематического поиска с использованием вербальных языков (все рассмотренные библиотеки) и классификационных языков (большинство рассмотренных библиотек).

В результате анализа данных можно сделать следующие более детальные выводы:

поиск по различным наборам библиографических элементов, самые распространённые из которых – автор, заглавие, год издания, используется в ЭК всех крупнейших библиотек. Следует, однако, отметить, что в ФГБУ «РГБ» и в БЕН РАН год издания фигурирует не в основном поисковом наборе, а в блоке фильтрации (в табл. 3 такое использование элемента «год издания» отражено пометой «+*»), что может быть не всегда удобно пользователю, который ожидает увидеть самые востребованные элементы библиографического описания в одном поисковом ряду;

тематический поиск с применением вербальных языков представлен во всех рассмотренных библиотеках, а с использованием классификационных языков – в большинстве рассмотренных библиотек (12 ЭК);

ПР очень популярны и применяются в 14 ЭК;

КС используются в ЭК всех библиотек, однако в 12 ЭК – непосредственно как поисковое поле, а в 4 ЭК (помета «+*» в табл. 3) данный поиск присутствует, но не выделен в отдельный вид поиска;

как правило, в ЭК сочетается поиск по ПР и КС;

наиболее популярная классификация – ББК (8 ЭК);

менее распространённые, но активно используемые классификации – Рубрикатор ВАК (5 ЭК), УДК (3 ЭК) и ГРНТИ (3 ЭК);

ряд библиотек пользуются классификациями собственной разработки – ИНИОН РАН, НБ МГУ и ГПИБ России;

классификационный поиск не представлен в 4 ЭК;

в большинстве библиотек, имеющих классификационный поиск, применяется несколько классификационных ИПЯ: два языка (4 ЭК) или три языка (3 ЭК);

только один классификационный ИПЯ используется в 5 ЭК: ББК (2 ЭК), рубрикатор собственной разработки (3 ЭК);

в виде иерархического дерева классификации представлены довольно редко – в ЭК только четырёх библиотек: ББК (ФГБУ «РГБ»), УДК (ГПНТБ России), ГРНТИ (ГПНТБ России, ГПНТБ СО РАН), индексы НБ МГУ (НБ МГУ).

Несмотря на различие лингвистических средств информационного поиска, используемых в ЭК крупнейших библиотек России, можно выделить общие черты их поисковых интерфейсов:

активное использование двух видов поиска: однострочного и расширенного;

использование нескольких (двух-четырёх) систем, предоставляющих дополнительные поисковые возможности;

наряду с поиском по различным библиографическим элементам активно используется тематический поиск, представленный различными ИПЯ;

в ЭК всех библиотек для тематического поиска предлагаются вербальные языки: либо предметные рубрики, либо ключевые слова, либо, что встречается чаще, оба языка;

в ЭК большинства библиотек (11 ЭК) для тематического поиска используются различные классификационные ИПЯ, при этом обычно применяется несколько языков такого типа.

Наряду с выявлением общих черт ЭК крупнейших библиотек России проведённый сопоставительный анализ позволил наметить определённые пути совершенствования представленных в них поисковых средств:

более последовательное соблюдение дифференцированной многоуровневости поиска;

отражение основных поисковых опций (названий основных видов поиска) ЭК на первой поисковой странице;

сочетание в ЭК библиографического, тематического и полнотекстового поиска;

обязательное отражение элементов поискового запроса в найденных записях;

обязательное использование систем уточнения (фильтрации), ранжирования и сортировки результатов поиска, а также системы поисковых словарей;

более активное внедрение системы рекомендаций дополнительных документов;

внедрение средств классификационного поиска во всех ЭК;

более активное использование иерархического представления классификаций в ЭК;

использование систем соответствий классификационных ИПЯ и в перспективе – классификационных и вербальных ИПЯ.

Предлагаемые направления развития поисковых возможностей ЭК должны обеспечить предоставление пользователям наиболее комфортного и понятного поискового интерфейса и эффективных поисковых инструментов.

Список источников

1. **Список** крупнейших библиотек России. Текст : непосредственный // Wikipedia.org : [сайт]. URL: https://ru.wikipedia.org/wiki/Список_крупнейших_библиотек_России (дата обращения: 19.04.2022).

2. **Руководство** по обеспечению тематического доступа в национальной библиографии / Международная федерация библиотечных ассоциаций и учреждений, Российская национальная библиотека ; [пер. с англ.]. Санкт-Петербург, 2017 // International Federation of Library Associations and Institutions : официальный сайт. URL: <https://www.ifla.org/wp-content/uploads/2019/05/assets/hq/publications/series/45-ru.pdf> (дата обращения: 19.04.2022).

3. **Breeding M.** Next-Generation Library Catalogs. Chapter 1: Introduction // Library Technology Reports. 2007. Vol. 43. № 4. P. 5–14. URL: <https://librarytechnology.org/document/18344> (дата обращения: 19.04.2022).

References

1. **Spisok** krupnei`shikh bibliotek Rossii. Tekst : neposredstvenny`i` // Wikipedia.org : [sai`t]. URL: https://ru.wikipedia.org/wiki/Список_крупнейших_библиотек_России (data obrashcheniia: 19.04.2022).
2. **Rukovodstvo** po obespecheniiu tematiceskogo dostupa v natsional`noi` bibliografii / Mezhdunarodnaia federatciia bibliotchny`kh assotiatcii` i uchrezhdenii`, Rossijskaia natsional`naia biblioteka ; [per. s angl.]. Sankt-Peterburg, 2017 // International Federation of Library Associations and Institutions : ofitcial`ny`i` sai`t. URL: <https://www.ifla.org/wp-content/uploads/2019/05/assets/hq/publications/series/45-ru.pdf> (data obrashcheniia: 19.04.2022).
3. **Breeding M.** Next-Generation Library Catalogs. Chapter 1: Introduction // Library Technology Reports. 2007. Vol. 43. № 4. P. 5–14. URL: <https://librarytechnology.org/document/18344> (data obrashcheniia: 19.04.2022).

Информация об авторах / Information about the authors

Зайцева Екатерина Михайловна – канд. филол. наук, ведущий научный сотрудник, руководитель группы информационно-лингвистического обеспечения ГПНТБ России, доцент кафедры электронных библиотек и наукометрических исследований Московского государственного лингвистического университета, Москва, Российская Федерация
katja@gpntb.ru

Смирнов Юрий Викторович – канд. техн. наук, старший научный сотрудник группы информационно-лингвистического обеспечения ГПНТБ России, Москва, Российская Федерация
yury@gpntb.ru

Ekaterina M. Zaisteva – Cand. Sc. (Philology), Leading Researcher, Head, Information and Linguistic Support Group, Russian National Public Library for Science and Technology; Associate Professor, E-Libraries and Scientometric Studies Chair, Moscow State Linguistic University, Moscow, Russian Federation
katja@gpntb.ru

Yury V. Smirnov – Cand. Sc. (Engineering), Senior Researcher, Information and Linguistic Support Group, Russian National Public Library for Science and Technology, Moscow, Russian Federation
yury@gpntb.ru