

# ИНФОРМАЦИОННО-ПОИСКОВЫЕ ЯЗЫКИ И СИСТЕМЫ

УДК 025.036

DOI: 10.33186/1027-3689-2021-7-87-96

**Ю. В. Смирнов**

*ГПНТБ России, Москва, Российская Федерация*

## **Тематический поиск в современных библиотечных информационно-поисковых системах**

**Аннотация:** Статья посвящена тематическому поиску в современных поисковых системах. Рассмотрено понятие «тематический поиск»; дано его определение (в ГОСТах оно отсутствует); перечислены виды – предметный поиск, поиск по ключевым словам и систематический; приведена их краткая характеристика. В библиотеках активно используются все три вида тематического поиска, а в поисковых системах интернета только один – по ключевым словам. Отмечено, что каталогизация обычно не используется в универсальных поисковых системах, таких как Google или Yandex, однако возможность тематического поиска по ключевым словам заложена в html-код интернет-страницы и систему тегирования. Исследовано использование систематического поиска в интернете. Отмечено, что этот вид поиска активно используется при обращении к ресурсам, связанным с библиотечной или научной деятельностью (ЭК библиотек, реферативные БД). Высказано предположение: со временем популярная в интернете система тематического поиска – тегирование – будет развиваться в сторону классификационного подхода и систематического поиска.

Работа подготовлена в рамках государственного задания ГПНТБ России – 730000Ф99.1 БВ09АА00006.

**Ключевые слова:** тематический поиск, систематический поиск, поиск по ключевым словам, предметный поиск, тегирование, АБИС, поисковые системы.

# INFORMATION RETRIEVAL LANGUAGES AND SYSTEMS

UDC 025.036

DOI: 10.33186/1027-3689-2021-7-87-96

**Yury V. Smirnov**

*Russian National Public Library for Science and Technology,  
Moscow, Russian Federation*

## **Subject search in modern library information retrieval systems**

**Abstract:** The author reviews in brief the library catalogs as the origins of subject search in the modern search systems. He examines the concept of “subject search”, offers its definition (lacking in the state standards – GOSTs), specifies the types (subject search, search by keywords, systematic search) and characterizes them. The libraries use intensively all three types of subject search while the Internet-based search engines offer just one type, i.e. search by keywords. The universal search systems, like Google and Yandex, do not entail cataloguing, however the possibility of subject search by keywords is designed within html-code and tagging system. The author investigates into the application of systematic search on the Internet. He concludes that this search type emphasizes is intensively applied to the resources related to libraries or science and research (library e-catalogs, abstract databases, etc.). He suggests that in time tagging, the Internet popular subject search system, will tend to classifying and systematic search.

The paper is prepared within the framework of the State Order to RNPLS&T № 73000F99.1 BV09AA00006.

**Keywords:** subject search, systematic search, search by keywords, subject-matter search, tagging, ALIS, search systems.

Один из способов получения новой информации или ответов на возникшие вопросы – это поиск. Сегодня он всё чаще осуществляется с помощью поисковых систем интернета, прочно вошедших в нашу жизнь. Ещё недавно за информацией нужно было идти в библиотеки, так как именно в них создавались и развивались различные поисковые средства, упрощающие процесс поиска. Наиболее популярное из них – каталог. Он и сегодня – важнейший элемент справочно-поискового аппарата.

Древнейшими считаются каталог на глиняных табличках в Ниппуре (Шумер, XX в. до н. э.) и выгравированный на стене храмовой библиотеки «Перечень ящиков, содержащих книги в больших свитках из кожи» (Эдфу, Египет, XVIII в. до н. э.) [1. С. 47].

Изначально каталогом называли список имеющихся книг. С развитием библиотек появились карточные каталоги, которые используются и сегодня. По наиболее значимым для поиска признакам и принципам построения выделяют три основных вида карточных каталогов: алфавитный, предметный и систематический.

С появлением компьютеров все каталоги трансформировались в единый электронный каталог (ЭК) библиотеки, так как создавать его менее затратно: на каждый документ формируется только одна запись БД ЭК, которая может использоваться для различных видов поиска, в том числе и тематического.

К сожалению, в ГОСТах определение тематического поиска отсутствует, однако есть определение термина *тематическая картотека* – «библиографическая картотека, отражающая документы по определённой теме, организованная в систематическом или предметном порядке» [2. С. 32]. Основываясь на нём, можно заключить, что тематический поиск – поиск документов, использующий предметную рубрику или классификационную систему в качестве поискового образа документа.

Не следует забывать и про язык ключевых слов, который отличается как от языка предметных рубрик, так и от различных классификаций, но позволяет осуществлять поиск по теме, заданной таким словом. Поиск по ключевым словам стал возможен с появлением ЭК – в карточных каталогах ключевые слова обычно не использовали, так как для тематического поиска было достаточно предметных и систематических каталогов.

Таким образом, можно сформулировать определение: *тематический поиск* – поиск документов, при котором в качестве поискового образа документа используется предметная рубрика, ключевое слово или индекс классификационной системы.

В ЭК библиотек активно используются три вида тематического поиска: предметный, по ключевым словам, систематический – по классификационным индексам.

Для организации систематического поиска в российских библиотеках обычно используются:

Библиотечно-библиографическая классификация – ББК (разработка РГБ);

Универсальная десятичная классификация – УДК Консорциума УДК на английском языке (перевод на русский язык – ВИНИТИ РАН);

Государственный рубрикатор научно-технической информации, созданный в СССР в 1978–1979 гг. информационными центрами (ВИНИТИ, ГПНТБ СССР и др.) для единой тематической систематизации научно-технической информации в организациях, осуществляющих научно-информационную деятельность.

Классификации имеют преимущества перед языком предметных рубрик и языком ключевых слов. Перечислим основные:

наглядность классификационной схемы и смысловых связей между понятиями благодаря представлению рубрик в виде иерархического дерева;

управление полнотой выдачи при тематическом поиске – возможность сужать или расширять тематические границы поиска, перемещаясь по иерархическому дереву;

использование в качестве лексических единиц кодов, а не слов естественного языка.

Классификации – довольно сложные системы со своими правилами и грамматикой. Их словарь состоит из огромного количества индексов. Например, словарь УДК на русском языке содержит 68 194 индекса, а ГРНТИ – 8 140 [3. С. 269].

При обучении библиотечных специалистов индексированию и поиску с помощью индексов уделяется много времени.

Отметим недостатки классификационных языков:

жёсткая структура, обусловленная фиксированным порядком классификационных делений;

обновление и модернизация классификационной схемы затруднительны, поскольку введение нового понятия может нарушить структуру классификации;

необходимость поддерживать баланс ёмкости классификации, так как большое количество индексов может привести к дублированию, а малое – к информационному шуму;

- сложная автоматизация поиска;
- длинный классификационный код: для отражения всех аспектов темы документа необходимо использовать несколько кодов из разных таблиц.

В отличие от систематического, предметный поиск и поиск по ключевым словам намного легче осваиваются пользователями: они основаны на естественном языке, а не на кодах, как в классификации. Внедрение этих видов поиска не вызывает сложностей и у разработчиков.

Библиотеки активно используют все три вида тематического поиска, а поисковые системы интернета – только поиск по ключевым словам. Терминология в библиотеках и поисковых системах интернета немного различается: под предметным поиском в интернете обычно подразумеваются разные виды тематических каталогов, а под поиском по ключевым словам – поиск не только по ключевым словам, но и по тегам.

Каталоги как поисковое средство не получили в интернете широкого распространения, хотя попытки каталогизации сайтов предпринимались. Например, в каталоге Яндекса редакторами вручную собраны аннотированные ссылки на интернет-сайты [4].

Ключевые слова в интернете обычно используются в html-коде страницы и указываются её разработчиками (например: «<meta name="keywords" content="Ключевые слова документа">»).

На большинстве современных сайтов для тематического поиска применяется система тегирования. Слово «тег» многозначно [5], в данном контексте оно понимается как ключевое слово для категоризации. Теги являются одним из видов ключевых слов, отличающихся оформлением (например: со знаком «#» (решётка) и без пробелов – «#инстаграманет»), а иногда – и авторской орфографией (написание с ошибками, часто намеренными, например: «#модаа»).

Систематический поиск с использованием какой-либо классификации не применяется в поисковых системах интернета по нескольким

причинам: трудность в освоении пользователями; необходимость интеллектуального индексирования ресурсов, что влечёт за собой необходимость найма квалифицированных работников; сложность реализации поисковых алгоритмов для раскрытия всех возможностей систематического поиска, например расширение и сужение поискового запроса.

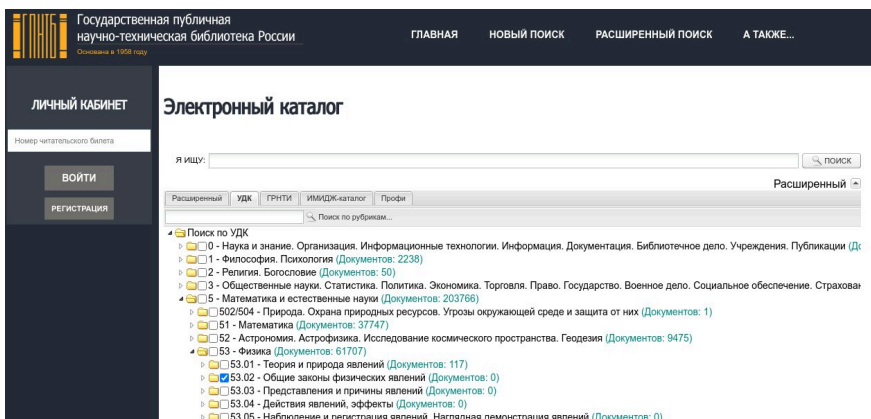
Отметим, что сужение и расширение запроса не всегда возможны в используемых библиотеками АБИС. Очень часто ожидается, что пользователь формулирует новый поисковый запрос с более широкой или узкой темой. Таким образом, систематический поиск реализуется как совпадение указанных в поисковом запросе индексов и индексов документа. Это снижает возможности систематического поиска и приводит к непониманию его сути, которая заключается не в банальном совпадении индексов документа и поискового запроса, а в возможности сужения и расширения тематики поиска.

Не следует думать, что систематический поиск не используется в интернете вообще, – например, на ресурсах, связанных с библиотечной или научной деятельностью, он довольно востребован.

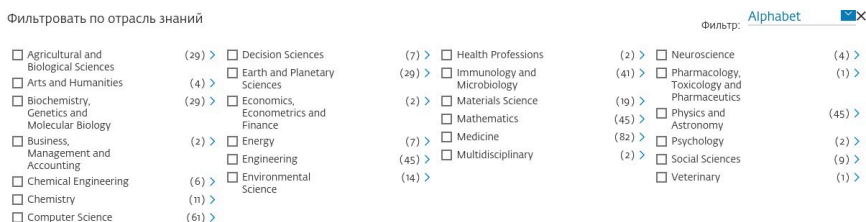
Традиционные библиотечные классификации используются при организации поиска в ЭК библиотек, представленных в интернете. Например, в ЭК ГПНТБ России [6], реализованном на базе Системы автоматизации библиотек ИРБИС64, предлагается систематический поиск по УДК и ГРНТИ (рис. 1).

В международных реферативных БД (*Web of Science, Scopus*) обычно используются классификации собственной разработки, коды и структура которых не отображаются для пользователей – только наименования рубрик (рис. 2). Сами классификаторы также нигде не опубликованы.

Коды и структура классификаторов *WoS* и *Scopus* были выявлены и описаны ВИНТИ РАН при работе над проектом по построению онтологии информационных ресурсов в виде сети библиографических классификаций на основе ГРНТИ [7]. Например, классификатор *Scopus* [8] является двухуровневым, содержащим 336 рубрик, а *WoS* [9] – алфавитным двухуровневым иерархическим классификатором с разной глубиной иерархии отдельных рубрик, которых 265.



**Рис. 1. Иерархическое дерево УДК в ЭК ГПНТБ России**



**Рис. 2. Отрасли знаний в Scopus**

РИНЦ, в отличие от зарубежных реферативных БД, не скрывает используемую для систематического поиска классификацию: она доступна на сайте *eLibrary* [10] и активно используется для систематизации не только научных публикаций, но и научных журналов (рис. 3). Также в РИНЦ реализована возможность поиска научных журналов по иерархическому дереву тематического рубрикатора.

На сайте *elibrary.ru* этот рубрикатор везде называется «Рубрикатор ГРНТИ», однако ВИНТИ РАН в вышеупомянутой работе по построению онтологии информационных ресурсов в виде сети библиографических классификаций на основе ГРНТИ выделил этот рубрикатор как отдельную классификацию, без какой-либо связи с ГРНТИ [7]. Причина этого не ясна, поскольку структуры рубрикаторов в РИНЦ и ГРНТИ практически идентичны (в последнем присутствуют ссылки, ключевые

слова и разделы). Они различаются количеством рубрик: 8 337 – в ГРНТИ [11] и 7 879 – в РИНЦ [12], что можно объяснить использованием в РИНЦ неактуализированной версии ГРНТИ либо исключением рубрик, которыми ни один журнал не индексируется. Ссылки, ключевые слова и разделы были исключены из рубрикатора РИНЦ как излишние.

Код	Название рубрики	Журналов
00.00.00	Общественные науки в целом	828
02.00.00	Философия	1772
03.00.00	История. Исторические науки	3613
04.00.00	Социология	2683
05.00.00	Демография	569
06.00.00	Экономика. Экономические науки	5586
10.00.00	Государство и право. Юридические науки	2871
11.00.00	Политика. Политические науки	2646
12.00.00	Науковедение	388
13.00.00	Культура. Культурология	2625
14.00.00	Народное образование. Педагогика	3626
15.00.00	Психология	2572
16.00.00	Языкознание	2186
17.00.00	Литература. Литературоведение. Устное народное творчество	1952
18.00.00	Искусство. Искусствоведение	1428
19.00.00	Массовая коммуникация. Журналистика. Средства массовой информации	713
20.00.00	Информатика	1159
21.00.00	Религия. Атеизм	912

Рис. 3. Тематический рубрикатор РИНЦ

Из приведённых примеров видно, что систематический поиск применяется на специализированных ресурсах, связанных с библиотечной и научной деятельностью (ЭК библиотек, реферативные БД). Каталогизация обычно не используется универсальными поисковыми системами, такими как *Google* или *Yandex*, однако возможности тематического поиска заложены в поиске по ключевым словам в html-код интернет-страницы и систему тегирования.

Тегирование приобрело популярность среди пользователей и разработчиков поисковых систем интернета потому, что оно позволяет быстро формулировать поисковый запрос по интересующей теме. На данный момент это основная система тематического поиска на универсальных ресурсах.



По нашему мнению, необходимость тематической категоризации интернет-ресурсов приведёт систему тегирования к следующему шагу – построению классификационных систем тегов и систематическому поиску.

## СПИСОК ИСТОЧНИКОВ

1. **Справочно-поисковый** аппарат библиотеки : учеб. / под ред. И. С. Пилко. – Санкт-Петербург : Профессия, 2015. – 288 с.
2. **ГОСТ 7.76-96.** Комплектование фонда документов. Библиографирование. Каталогизация. Термины и определения : межгос. стандарт : введён впервые : дата введения 1998-01-01 / разработан Рос. гос. б-кой [и др.]. – Москва : Изд-во стандартов, 1997. – III, 53 с. – (Система стандартов по информ., библи. и изд. делу). – 53 с.
3. **Гиляревский Р. С.** Рубрикатор как инструмент информационной навигации / Р. С. Гиляревский, А. В. Шапкин, В. Н. Белоозеров. – Санкт-Петербург : Профессия, 2008. – 351 с.
4. **Яндекс** Каталог. – Режим доступа: <https://yasa.yandex.ru/> (дата обращения: 17.01.2021).
5. **Тег** // ru.wikipedia.org = Википедия. – Режим доступа: <https://ru.wikipedia.org/wiki/Тег> (дата обращения: 17.01.2021).
6. **Электронный** каталог ГПНТБ России // gpntb.ru = ГПНТБ России. – Режим доступа: <http://cat.gpntb.ru/?id=ЕС> (дата обращения: 17.01.2021).
7. **Антошкова О. А.** Построение онтологии информационных ресурсов в виде сети библиографических классификаций / О. А. Антошкова, В. Н. Белоозеров, Е. Ю. Дмитриева и др. // Перспектив. направления науч. исслед. и крит. технологии в классификацион. системах : материалы науч.-практ. конф., 25–27 окт. 2017 г., Москва. – Москва, 2017. – Режим доступа: <http://www.udcc.ru/MATERIALS/2017/CONFERENCE2017.pdf>. – С. 20–25 (дата обращения: 17.01.2021).
8. **Классификатор** Scopus // viniti.ru = ВИНИТИ РАН. – Режим доступа: <http://scs.viniti.ru/MapService/treeInfo.aspx?tree=Scopus> (дата обращения: 17.01.2021).
9. **Классификатор** Web of Science // viniti.ru = ВИНИТИ РАН. – Режим доступа: <http://scs.viniti.ru/MapService/treeInfo.aspx?tree=WoS> (дата обращения: 17.01.2021).
10. **Тематический** рубрикатор // eLibrary.ru. – Режим доступа: <https://www.elibrary.ru/rubrics.asp> (дата обращения: 17.01.2021).
11. **Государственный** Рубрикатор НТИ России (ГРНТИ) // viniti.ru = ВИНИТИ РАН. – Режим доступа: <http://scs.viniti.ru/MapService/treeInfo.aspx?tree=RGNTI> (дата обращения: 17.01.2021).
12. **Классификатор** Российского индекса научного цитирования // viniti.ru = ВИНИТИ РАН. – Режим доступа: <http://scs.viniti.ru/MapService/treeInfo.aspx?tree=RINC> (дата обращения: 17.01.2021).

## REFERENCES

1. **Spravочно-poiskovyy** apparat biblioteki : ucheb. / pod red. I. S. Pilko. – Sankt-Peterburg : Professiya, 2015. – 288 s.
2. **GOST 7.76-96**. Komplektovanie fonda dokumentov. Bibliografirovanie. Katalogizatsiya. Terminy i opredeleniya : mezhgos. standart : vveden v pervye : data vvedeniya 1998-01-01 / razrabotan Ros. gos. b-koy [i dr.]. – Moskva : Izd-vo standartov, 1997. – III, 53 s. – (Sistema standartov po inform., bibl. i izd. delu). – 53 s.
3. **Gilyarevskiy R. S.** Rubrikator kak instrument informatsionnoy navigatsii / R. S. Gilyarevskiy, A. V. Shapkin, V. N. Beloozerov. – Sankt-Peterburg : Professiya, 2008. – 351 s.
4. **Yandex** Katalog. – URL: <https://yaca.yandex.ru/>.
5. **Teg** // ru.wikipedia.org = Википедия. – URL: <https://ru.wikipedia.org/wiki/Teg>.
6. **Elektronnyy** katalog GPNTB Rossii // gpntb.ru = GPNTB Rossii. – URL: <http://cat.gpntb.ru/?id=EC>.
7. **Antoshkova O. A.** Postroenie ontologii informatsionnykh resursov v vide seti bibliograficheskikh klassifikatsiy / O. A. Antoshkova, V. N. Beloozerov, E. Yu. Dmitrieva i dr. // Perspektiv. napravleniya nauch. issled. i krit. tehnologii v klassifikatsion. sistemah : materialy nauch.-prakt. konf., 25-27 okt. 2017 g., Moskva. – Moskva, 2017. – URL: <http://www.udcc.ru/MATERIALS/2017/CONFERENCE2017.pdf>. – С. 20–25.
8. **Klassifikator** Scopus // viniti.ru = VINITI RAN. – URL: <http://scs.viniti.ru/MapService/treeInfo.aspx?tree=Scopus>.
9. **Klassifikator** Web of Science // viniti.ru = VINITI RAN. – URL: <http://scs.viniti.ru/MapService/treeInfo.aspx?tree=WoS>.
10. **Tematicheskyy** rubrikator // eLibrary.ru. – URL: <https://www.elibrary.ru/rubrics.asp>.
11. **Gosudarstvennyy** Rubrikator NTI Rossii (GRNTI) // viniti.ru = VINITI RAN. – URL: <http://scs.viniti.ru/MapService/treeInfo.aspx?tree=RGNTI>.
12. **Klassifikator** Rossiyskogo indeksa nauchnogo tsitirovaniya // viniti.ru = VINITI RAN. – URL: <http://scs.viniti.ru/MapService/treeInfo.aspx?tree=RINC>.

### Информация об авторе / Information about the author

**Смирнов Юрий Викторович** – канд. техн. наук, научный сотрудник ГПНТБ России, Москва, Российская Федерация  
[yu.smirnoff@gmail.com](mailto:yu.smirnoff@gmail.com)

**Yury V. Smirnov** – Cand. Sc. (Engineering), Researcher, Russian National Public Library for Science and Technology, Moscow, Russian Federation  
[yu.smirnoff@gmail.com](mailto:yu.smirnoff@gmail.com)