

УДК 01:004.65:024

<https://doi.org/10.20913/1815-3186-2021-3-35-44>

Сравнительный анализ поискового поведения пользователей библиографических баз данных научной информации¹

К. В. Монанков



**Монанков
Кирилл Викторович,**

Национальный
исследовательский
ядерный университет
«МИФИ»,
Каширское ш., 31,
Москва, 115409,
Россия,

инженер

ORCID: [0000-0002-9267-3987](https://orcid.org/0000-0002-9267-3987)

e-mail: kmonankov@yandex.ru

Аннотация. Проведено статистическое исследование поискового поведения пользователей информационных ресурсов ИНИОН РАН при поиске научной информации за прошедшее десятилетие для двух групп пользователей: гостей и зарегистрированных пользователей. Сделано сравнение результатов исследования с результатами аналогичных исследований более ранних периодов с целью обнаружения изменений, тенденций. Сравнительный анализ показал снижение поисковой активности пользователей со статусом гостя как с точки зрения продолжительности поисковой сессии, так и с точки зрения использования уникальных поисковых операций, при этом поведение зарегистрированных пользователей практически не изменилось. Для анализа использовались следующие показатели: продолжительность поисковой сессии, количество уникальных операций за сессию, использование различных баз данных и типов интерфейсов. Такой анализ поискового поведения пользователей дает возможность оценить актуальность и эффективность информационно-поисковых ресурсов, а полученные статистические данные могут быть использованы в проектировании новых поисковых интерфейсов для различных категорий пользователей.

Ключевые слова: поисковое поведение пользователей, статистика, реферативно-библиографические базы данных, информационно-поисковые системы

Для цитирования: Монанков К. В. Сравнительный анализ поискового поведения пользователей библиографических баз данных научной информации // Библиосфера. 2021. № 3. С. 35–44. <https://doi.org/10.20913/1815-3186-2021-3-35-44>.

Статья поступила в редакцию 15.03.2021

Получена после доработки 07.06.2021

Принята для публикации 25.08.2021

© К. В. Монанков, 2021

¹ Работа выполнена при поддержке Министерства науки и высшего образования РФ (проект государственного задания № 0723-2020-0036)

Comparative Analysis of Information Retrieval Behavior of Scientific Bibliographic Databases Users

Kirill V. Monankov

Kirill Viktorovich Monankov,
National Research Nuclear University
MEPhI (Moscow Engineering
Physics Institute),
Kashirskoe shosse, 31,
Moscow, 115409, Russia,
Engineer
ORCID: 0000-0002-9267-3987
e-mail: kmonankov@yandex.ru

Abstract. Statistical study of the information retrieval behavior of information resources users, generated by Institute of Scientific Information on Social Sciences of RAN (INION RAN), in the search for scientific information over the past decade for two groups of users: guests and registered users, has been carried out. The results of the study are compared with that of the similar studies made in earlier periods to detect changes and trends. Comparative analysis showed a decrease in the search activity of users with the status of a guest, both in terms of the duration of the search session and the use of unique search operations, while the behavior of registered users has not changed. For the analysis, the following indicators were used: search session duration, number of unique operations per session, use of various databases and interfaces types. Such analysis of the information retrieval behavior is relevant for understanding effectiveness of information retrieval resources. The statistical data can be used to design new search interfaces for various categories of users.

Keywords: information retrieval behavior, statistics, bibliographic databases, information retrieval systems

Citation: Monankov K. V. Comparative Analysis of Information Retrieval Behavior of Scientific Bibliographic Databases Users. *Bibliosphere*. 2021. № 3. P. 35–44. <https://doi.org/10.20913/1815-3186-2021-3-35-44>.

Received 15.03.2021

Revised 07.06.2021

Accepted 25.08.2021

Введение

У пользователей «Яндекса» доля поисковых сессий, состоящих из одного запроса, в среднем составляет 56 %, а среднее количество запросов в сессию равно 2,4 запроса (Поиск..., 2021, с. 3). При этом поисковое поведение групп пользователей с разными постоянными поисковыми интересами несколько отличается. Так, например, представители группы «Наука» при поиске в «Яндексе» реже ограничиваются одним запросом за сессию – в 52,1 % случаев, а среднее количество запросов в сессию составляет 3,13 (Пользователи..., 2013, с. 6).

Поведение пользователей во время поиска научной информации представляет наибольший интерес, однако изучено оно недостаточно (Балуткина, Бусыгина, 2015; Голицына, Максимов, 2011; Максимов, Забегаева, 2001).

Исследования поискового поведения пользователей баз данных (БД) научной информации показали стремление отдельных групп пользователей (называемых профессионалами) к целенаправленному формированию запроса и обработке результатов поиска (Голицына, Максимов, 2011; Максимов, Забегаева, 2001). При этом характеристики большинства пользователей специализированных информационных ресурсов немногим отличаются от характеристик пользователей поисковых машин общего назначения, таких как «Яндекс»: большинство поисковых сессий короткие (1–2 поисковые операции),

а просмотр ограничивается одной страницей выдачи (Голицына, Максимов, 2011, с. 57).

Цель настоящего исследования – установить изменения и выявить тенденции поискового поведения пользователей при работе с ретроспективными библиографическими БД Института научной информации по общественным наукам Российской академии наук (ИНИОН РАН) за прошедшее десятилетие (2010–2019 гг.). В работе приведены данные, полученные в результате статистического исследования, и они сравниваются с аналогичными данными исследований других периодов: 1999–2001 гг. (Максимов, Забегаева, 2001) и 2006–2009 гг. (Голицына, Максимов, 2011).

Исходные данные и методы анализа

Для анализа поискового поведения пользователей использовались протоколы поисковых сессий в ретроспективных библиографических БД ИНИОН РАН.

Библиографические БД ИНИОН РАН по социальным и гуманитарным наукам формируются с середины 1980-х гг., данные обновляются ежегодно, и общий объем массивов составляет почти 4 млн записей (данные на 31 января 2020 г.)².

Доступ к БД ИНИОН РАН обеспечивается с помощью документальной информационно-

² Библиографические базы данных // ИНИОН РАН. URL: <http://inion.ru/ru/resources/bazy-dannykh-inion-ran/> (дата обращения: 10.03.2021).

аналитической системы xIRBIS (Максимов и др., 2020), реализующей различные четкие и нечеткие механизмы поиска. В системе предусмотрены технологии поиска:

- вербальные,
- кластерные (поиск аналогов документа или множества релевантных документов),
- ссылочные,
- с использованием таксономических фильтров,
- по обратной связи.

При формировании и реформулировке запроса могут быть использованы словари, рубрикаторы и тезаурусы, интегрированные непосредственно в систему библиографических БД³. Результаты поиска представляются в различных формах, по выборке документов могут быть построены тематические словари лексики с частотными характеристиками и временные ряды потоков публикаций.

Результат каждого запроса вносится в протокол поисковой сессии, который обеспечивает возможности возвращения к результатам ранее проведенных запросов, агрегирования результатов нескольких запросов, перехода к аналитическому представлению выборок документов. Это позволяет не только отобразить ход процесса и использовать ранее полученные результаты, но и зафиксировать оценку пользователя (Голицына, Максимов, 2011, с. 49).

Поисковая сессия представляет собой последовательность выполняемых пользователем операций (от входа в информационный массив до завершения работы), которые принято делить на три группы (Голицына, Максимов, 2011, с. 49):

- поиск и развитие поискового процесса;
- обработка результатов поиска (управление просмотром документов, использование ранее полученных результатов);

- служебные (вход/выход и выбор БД, задание параметров поиска).

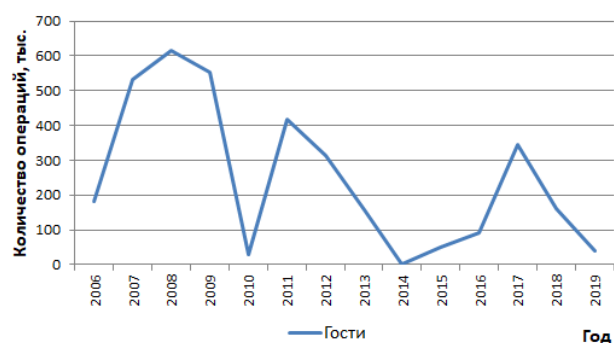
Система имеет контролируемый вход: на сайте ИНИОН РАН пользователь регистрируется самостоятельно или может воспользоваться гостевой учетной записью. В случае гостевой учетной записи система будет ограничена лишь отсутствием возможности восстанавливать и продолжать прерванную поисковую сессию, сохранять и повторно использовать запрос. Все БД и формы представления документов доступны в равной степени всем пользователям.

Для большего понимания количественных распределений вызовов операций приведем перечень протоколируемых операций (таблица).

Для определения и сравнения статистических характеристик поисковых сессий пользователей взяты пооперационные журналы поисковых сессий за три периода: 1999–2001 гг., 2006–2009 гг. и 2010–2019 гг. Рассчитанные характеристики поисковых сессий за 1999–2001 гг. взяты из работы (Максимов, Забегаева, 2001, с. 19–20), а данные за 2006–2009 гг. – из работы О. Л. Голицыной и Н. В. Максимова (Голицына, Максимов, 2011, с. 51–56) или рассчитаны заново. Статистические характеристики за 2010–2019 гг. публикуются впервые; объем соответствующего пооперационного журнала составляет 1 670 399 записей. Некоторые характеристики рассчитаны для объединенного периода – 2006–2019 гг.; общий объем соответствующего пооперационного журнала составляет 3 961 306 записей.

Количественное распределение операций по годам в зависимости от статуса пользователя представлено на рисунке 1а, б.

Статистика журналов поисковых сессий показывает преобладание пользователей со статусом гостя – 88 %.



(а)



(б)

Рис. 1. Количество операций в журнале поисковых сессий пользователей ИНИОН РАН в 2006–2019 гг.*

Fig. 1. The number of operations in the search sessions log of INION RAN's users per year for the period 2006–2019

* Данные поисковых сессий за 2014 г. отсутствуют из-за пожара в ИНИОН РАН в январе 2015 г.

³ Библиографические базы данных // ИНИОН РАН. URL: <http://inion.ru/ru/resources/bazy-dannykh-inion-ran/> (дата обращения: 10.03.2021).

Таблица. Перечень протоколируемых операций (Голицына, Максимов, 2011, с. 50)

Table. List of logged operations (Golitsina, Maksimov, 2011, p. 50)

Группа	Мнемоника	Описание
Служебные	ENTRY	Начало сессии и вход в БД
	RESET	Выбор новой БД
	CHSES	Изменение параметров сессии
	OFF	Завершение поисковой сессии
Поиск и развитие поискового процесса	FIND	Поиск по первичному запросу
	FINDT	Поиск документов, содержащих термины, отмеченные в тезаурусе
	FINDE	Поиск по отредактированному выражению после первичного поиска
	FINDX	Поиск по гипертекстовым ссылкам от элементов (терминов) отображенных документов
	ASRCH	Поиск аналогов указанного документа
	HS	Эвристический поиск по релевантным документам выдачи
	HSJ	Эвристический поиск по нескольким предложениям из протокола
	LIST	Просмотр словника релевантных документов
	SOVA	Поиск по терминам из словника релевантных документов
Обработка результатов поиска	NXT	Просмотр следующей страницы в выдаче
	PRT	Просмотр протокола поисковой сессии
	CS	Сортировка выдачи по возрастанию или убыванию значения одного из выбранных полей документа
	SDOC	Просмотр документов из Протокола
	DICT	Просмотр словаря
	JOIN	Объединение результатов поиска
	CH	Смена схемы просмотра документа
	DELQ*	Удаление строки Протокола
	SAVEQ*	Сохранение запроса (Протокола)
	SMRKCLR	Отказ от отметки (термина или документа)

* Операции только для зарегистрированных пользователей

Анализ продолжительности поисковой сессии

Анализ продолжительности поисковой сессии за 2006–2019 гг. демонстрирует, что пользователи в среднем выполняли 5,3 операции за сессию. При этом средняя продолжительность гостевой сессии составляет 4,9 операции, что существенно меньше

сессии зарегистрированного пользователя – 18,2 операции. При этом среднее количество операций за сессию в 2010–2019 гг. составляет 3,4 для гостей – 3,3, а для зарегистрированных пользователей – 19,6.

Распределение среднего количества операций за сессию по годам и в зависимости от статуса пользователя представлено на рисунке 2.

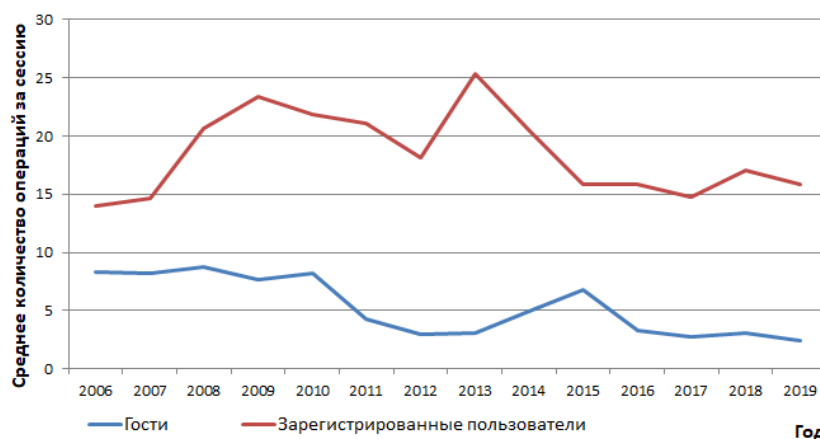


Рис. 2. Распределение среднего количества операций за сессию для двух групп пользователей в период 2006–2019 гг.*

Fig. 2. Distribution of the average number of operations per user session for two user groups for the period 2006–2019

* Значение для 2014 г. рассчитано как среднее арифметическое между значениями для 2013 и 2015 гг.

Как видно из графика, среднее количество операций за сессию для зарегистрированных пользователей лежит в интервале от 15 до 25 операций и составляет в среднем (арифметическое) 18,5 операций за весь период. Среднее количество операций за сессию за 2006–2009 гг. в среднем (арифметическое) составляет 18,2, а за 2010–2019 – 18,6. Сравнение этих периодов показывает, что продолжительность поисковой сессии для зарегистрированных пользователей практически не изменилась.

Ситуация с гостями обстоит иначе. Количество операций за сессию за 2006–2009 гг. в среднем (арифметическое) составляет 8,2, а за 2010–2019 гг. – 4,2. То есть продолжительность сессии для гостей уменьшилась почти в два раза.

На рисунках 3 и 4а, б приведено количественное распределение продолжительности поисковых сессий (в % от общего числа) за разные периоды.

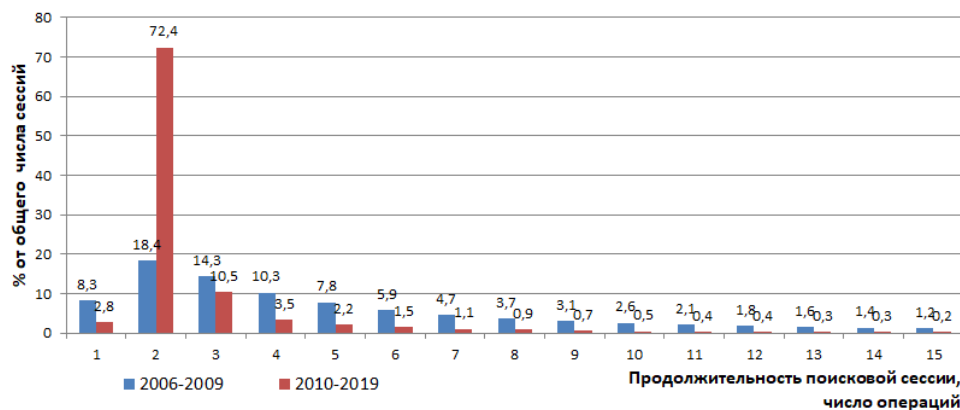
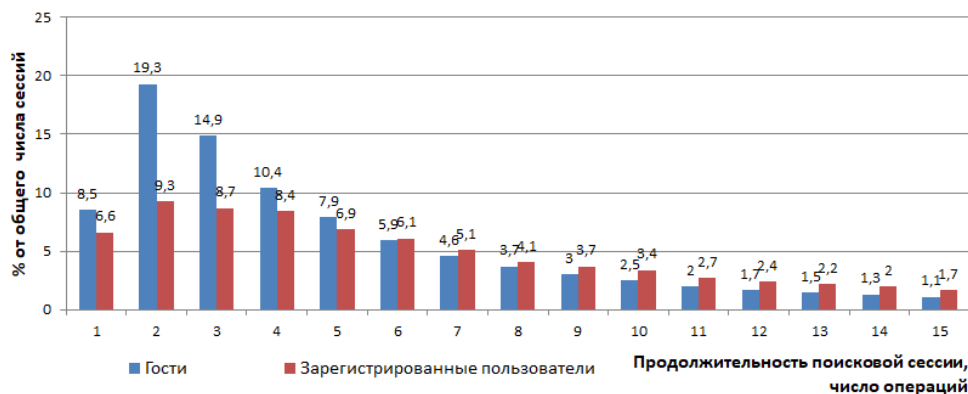
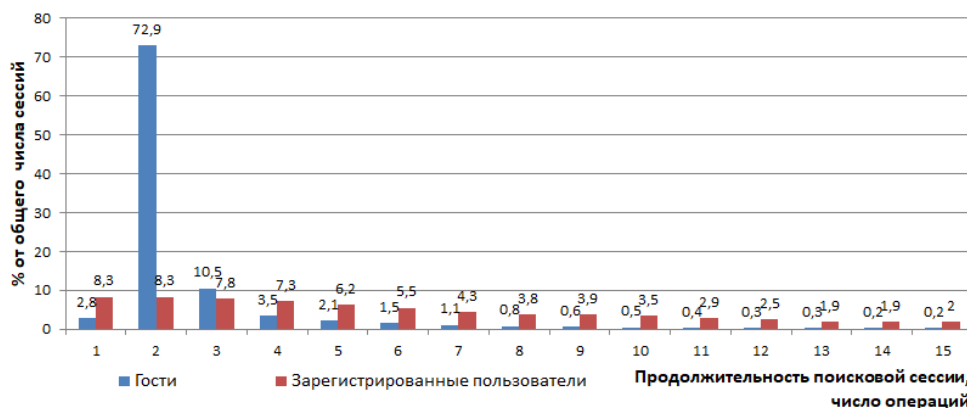


Рис. 3. Количественное распределение продолжительности поисковых сессий (в % от общего числа) всех пользователей

Fig. 3. Distribution of the search session duration (in % of the total number) of all users



(а) – 2006–2009 гг.



(б) – 2010–2019 гг.

Рис. 4. Количественное распределение продолжительности поисковых сессий

(в % от общего числа) гостей и зарегистрированных пользователей: период 2006–2009 (а), период 2010–2019 (б)

Fig. 4. Distribution of the search session duration (in % of the total number)

for guests and registered users for the period 2006–2009 (а) and for the period 2010–2019 (б)

Количественное распределение продолжительности поисковых сессий показывает разницу между гостями и зарегистрированными пользователями: гости реже задерживаются в системе и совершают меньшее количество операций. Причем в 2010–2019 гг. количество коротких сессий (в основном из двух операций) значительно возросло и превышает 70 % от общего числа сессий (рис. 4б).

Анализ используемых операций

Построены распределения количества сессий (% от общего числа сессий) в зависимости от числа используемых уникальных операций для разных групп пользователей и за разные периоды (рис. 5а, б и 6а, б).

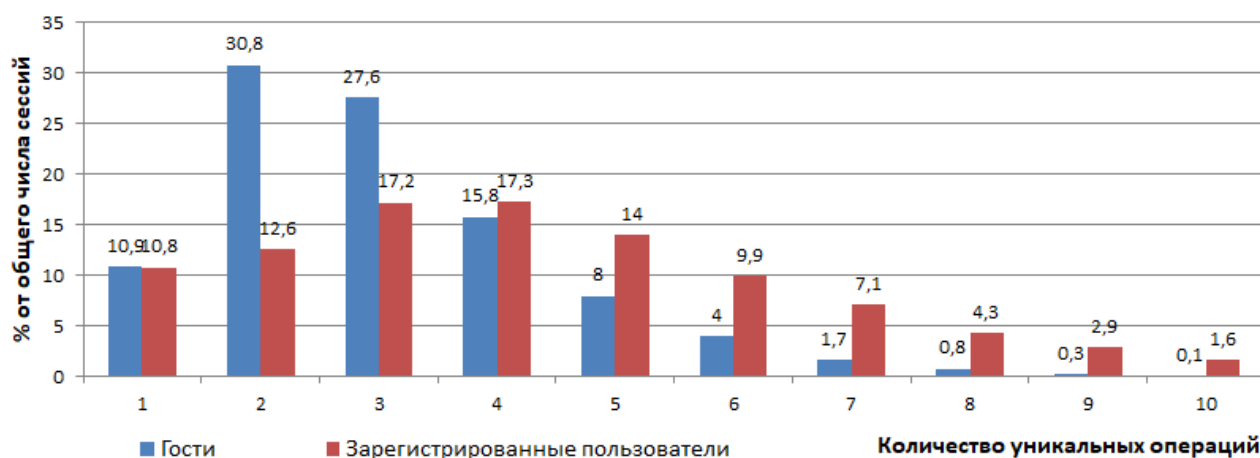
Распределение количества сессий в зависимости от числа уникальных операций говорит о том, что гости используют меньше уникальных операций системы, чем зарегистрированные

пользователи. Более того, распределение на рисунке 5б показывает, что за 2010–2019 гг. количество используемых гостями уникальных операций снизилось до двух (в более 80 % сессий).

Сравнение распределения количества сессий в зависимости от числа уникальных операций с более ранним периодом 1999–2001 гг. (рис. 6) показывает, что количество используемых уникальных операций снизилось, особенно для гостей.

Построены распределения количества вызовов операций от общего числа вызовов (рис. 7 и 8а, б).

Распределение количества вызовов показывает, что наиболее популярными операциями (помимо входа (ENTRY)) у всех пользователей являются поиск по первичному запросу (FIND), просмотр следующей страницы в выдаче (NXT) и обращение к словарю (DICT). При этом зарегистрированные пользователи применяют операцию перехода на следующую страницу



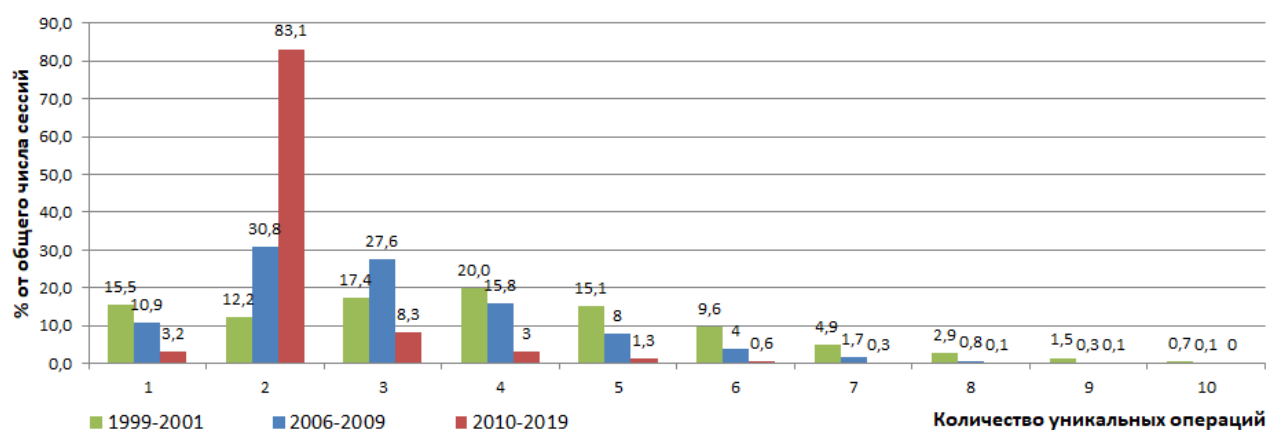
(а) – 2006–2009 гг.



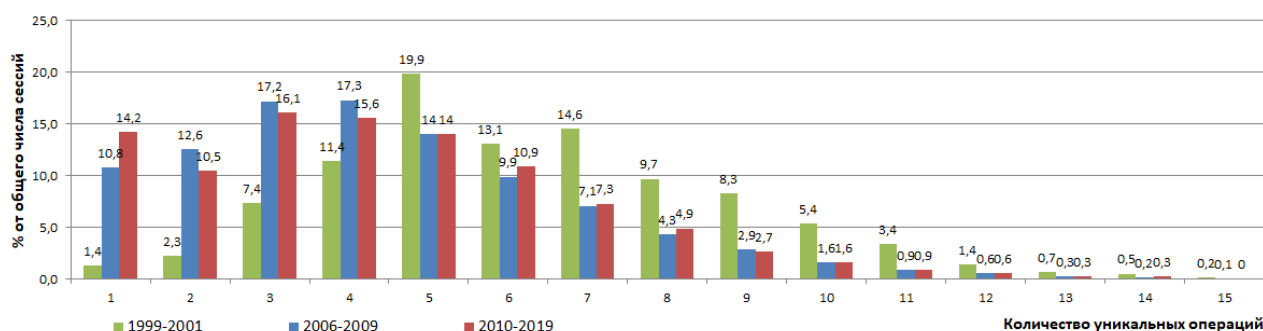
(б) – 2010–2019 гг.

Рис. 5. Распределение количества сессий (% от общего числа сессий) в зависимости от числа используемых уникальных операций: период 2006–2009 (а), период 2010–2019 (б)

Fig. 5. Distribution of the number of sessions (% of the total number of sessions) depending on the number of unique operations used for the period 2006–2009 (a) and for the period 2010–2019 (b)



(a) – гости



(b) – зарегистрированные пользователи

Рис. 6. Распределение количества сессий (% от общего числа сессий) в зависимости от числа используемых уникальных операций за разные периоды: а) для гостей, б) для зарегистрированных пользователей
 Fig. 6. Distribution of the number of sessions (% of the total number of sessions) depending on the number of unique transactions used for different periods for guests (a) and registered users (b)

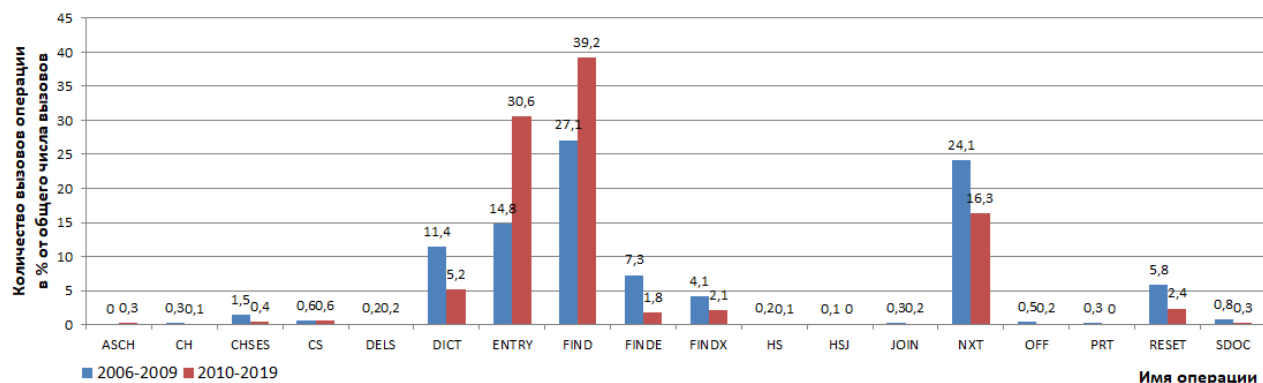


Рис. 7. Распределение количества вызовов операций от общего числа вызовов

Fig. 7. Distribution of the number of operations used from the total number

чаще, чем гости (рис. 8). Более того, за 2010–2019 гг. гости использовали операцию перехода на следующую страницу меньшее количество раз, чем входили в систему (рис. 8а), то есть после поиска гость, в отличие от зарегистрированного пользователя, не стремился просматривать результаты дальше первой страницы.

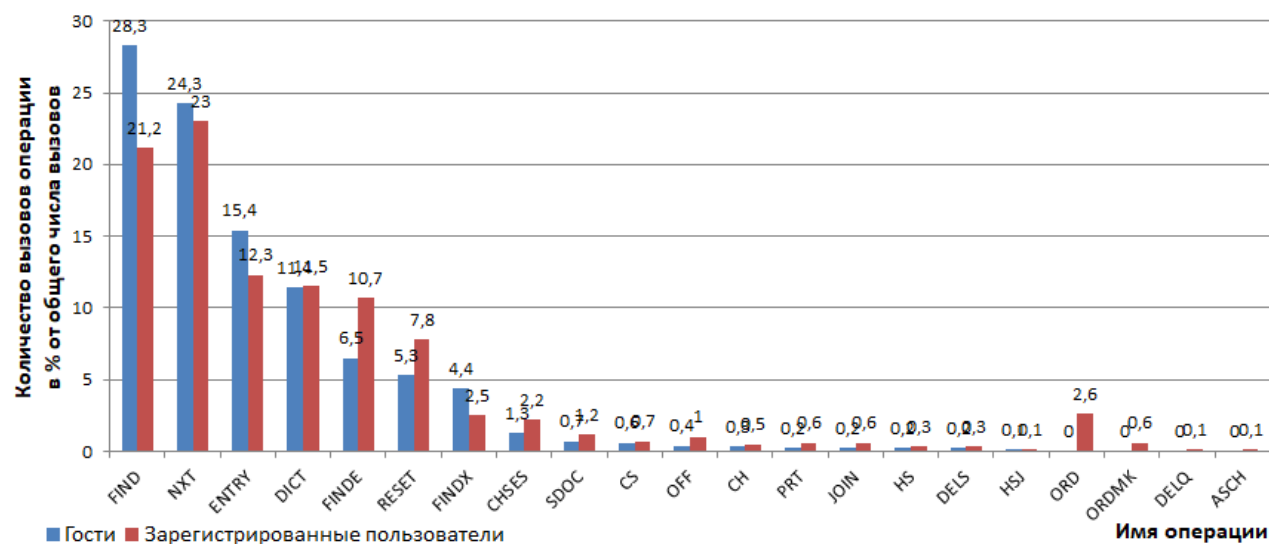
Анализ использования ресурса

Распределение количества обращений пользователей к различным БД ИНИОН РАН представлено на рисунке 9а, б.

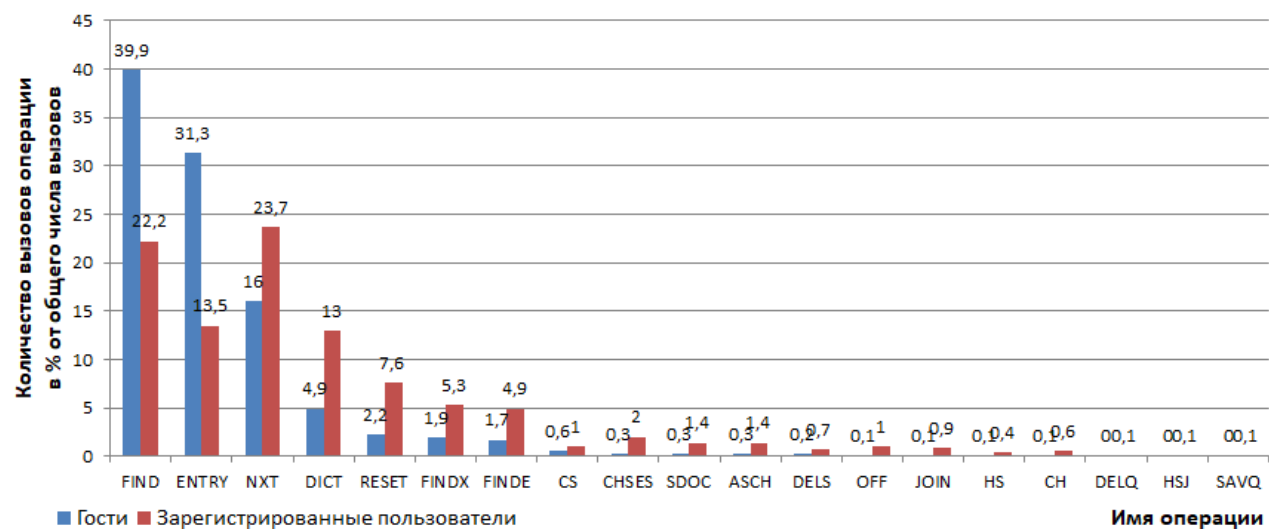
Наиболее популярной среди всех пользователей остается БД по литературоведению. Среди зарегистрированных пользователей за 2006–2009 гг. наиболее популярной была БД по языкознанию, но в 2010–2019 гг. интерес к ней снизился относительно БД по экономике и истории.

Построено распределение количества используемых БД в рамках одной сессии (рис. 10а, б).

Анализ показывает, что более 90 % пользователей в процессе поиска за одну сессию обращаются к одной БД, что говорит об устойчивости предметного интереса. При этом поиск в различных БД ведется гостями значительно реже. Эти факты



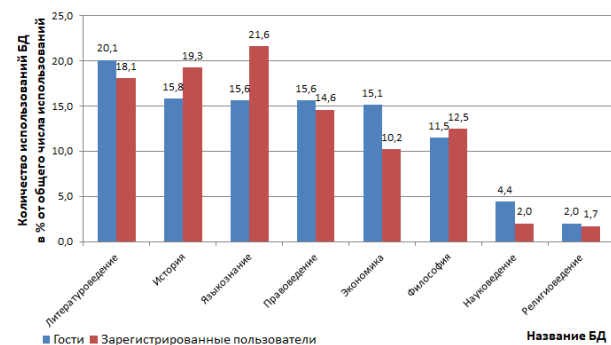
(a) – 2006–2009 гг.



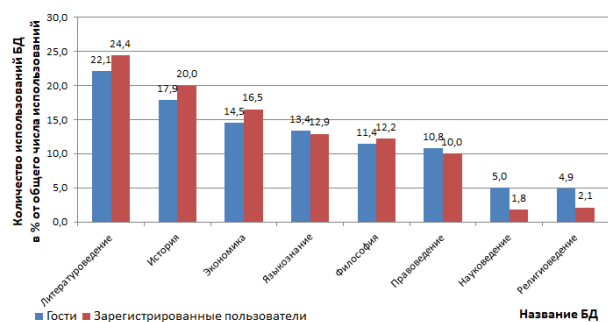
(b) – 2010–2019 гг.

Рис. 8. Распределение количества вызовов операций от общего числа вызовов: период 2006–2009 (а), период 2010–2019 (б)

Fig. 8. Distribution of the number of operations used from the total number for the period 2006–2009 (a) and for the period 2010–2019 (b)



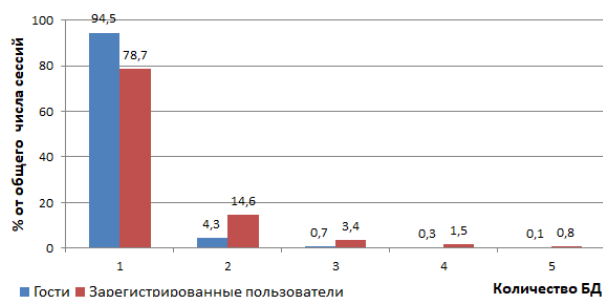
(a) – 2006–2009 гг.



(b) – 2010–2019 гг.

Рис. 9. Использование различных БД (в % от общего числа): период 2006–2009 (а), период 2010–2019 (б)

Fig. 9. Databases usage (in % of the total) for the period 2006–2009 (a) and for the period 2010–2019 (b)



(a) – 2006–2009 гг.



(b) – 2010–2019 гг.

Рис. 10. Распределение количества используемых БД: период 2006–2009 (а), период 2010–2019 (б)

Fig. 10. Distribution of the number of used databases for the period 2006–2009 (a) and for the period 2010–2019 (b)

в целом соответствуют принципам формирования ресурсов, так как БД ИНИОН РАН сформированы по тематическому принципу, и каждая из них имеет полную ретроспективу.

На рисунке 11 показано распределение объема выдач в ответ на поисковый запрос пользователя.

Чаще всего результаты поиска содержат до 300 документов. Значение этого показателя за 2010–2019 гг. близко к соответствующему показателю 2006–2009 гг.

Система xIRBIS (Максимов и др., 2020) предоставляет пользователю возможность создания поискового запроса в разных формах с использованием различных интерфейсов формирования выражения (рис. 12).

Распределение использования различных типов интерфейсов демонстрирует преобладание интерфейса «Строка запроса» над остальными

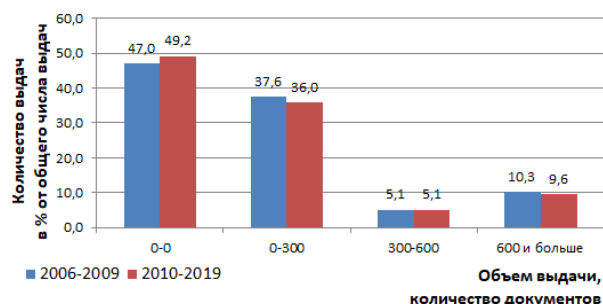
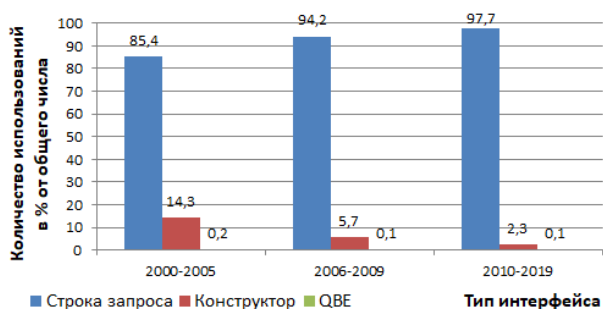
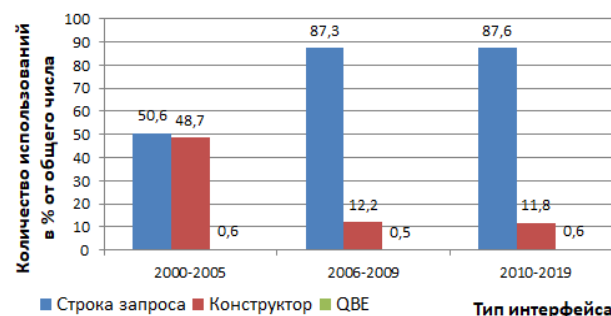


Рис. 11. Распределение объема выдач

Fig. 11. Distribution of the volume of search results



(a) – гости



(b) – зарегистрированные пользователи

Рис. 12. Использование различных типов интерфейсов: а) гости, б) зарегистрированные пользователи

Fig. 12. Using different types of interfaces for guests (a) and registered users (b)

в специализированных ресурсах научной информации является привычка «бытового поиска» в таких интернет-системах, как Google или «Яндекс». Гости, которыми часто оказываются новые пользователи, сталкиваясь с непривычными им технологиями (например, поиск по обратной связи) и интерфейсами, видимо, не готовы их осваивать и бросают поиск после первого запроса «по строке». Еще одной причиной прекращения поиска гостями может быть отсутствие в БД ИНИОН РАН полных текстов документов и ссылок на них.

Высокая поисковая активность зарегистрированных пользователей (относительно пользователей со статусом гостя), выражающаяся

высокой продолжительностью поисковой сессии и высокими показателями использования уникальных поисковых операций и интерфейсов, говорит, во-первых, о сохранении предметного интереса этой группы пользователей к БД ИНИОН РАН, а во-вторых, – об актуальности и эффективности применяемых поисковых технологий. Несмотря на то что интерфейс системы устаревает, зарегистрированные пользователи продолжают использовать и кластерные технологии поиска, и технологии развития запроса, и технологии обработки результатов, а значит, при проектировании новых поисковых интерфейсов необходимо реализовывать имеющиеся поисковые механизмы и технологии.

Список литературы / References

- Балуткина Н. А., Бусыгина Т. В. Информационно-поисковое поведение пользователей БД «Научная Сибирика» в АБИС WebIRBIS // Информационные ресурсы России. 2015. Т. 1. С. 2–5 [Balutkina NA and Busygina TV (2015) Information retrieval behavior of users of the “Scientific Sibirica” database in WebIRBIS. *Informatsionnye resursy Rossii* 1: 2–5. (In Russ.)].
- Голицына О. Л., Максимов Н. В. Статистика поискового поведения пользователей баз данных научной информации // Библиосфера. 2011. № 2. С. 47–57 [Golitsina OL and Maksimov NV (2011) Statistics of searching behaviour of database users of scientific information. *Bibliosfera* 2: 47–57. (In Russ.)].
- Максимов Н. В., Голицына О. Л., Монанков К. В., Гаврилкина А. С. Документальная информационно-аналитическая система xIRBIS (редакция 6.0): программа для ЭВМ : свидетельство о гос. регистрации № 2020661683 : дата регистрации: 29.09.2020 [Maksimov NV, Golitsyna OL, Monankov KV and Gavrilkina AS (2020) Documentary information-analytical system xIRBIS: PC software. State registration certificate no. 2020661683: date of registration 29.09.2020). (In Russ.)].
- Максимов Н. В., Забегаева Н. Н. Информационный поиск и модели поведения пользователей // Научно-техническая информация. Серия 1, Организация и методика информационной работы. 2001. № 11. С. 10–21 [Maksimov NV and Zabegaeva NN (2001) Information retrieval and user behavior models. *Nauchnaya i tekhnicheskaya informatsiya. Seriya 1. Organizatsiya i metodika informatsionnoi raboty* 11: 10–21. (In Russ.)].
- Поиск в интернете: как ищут мужчины и женщины // Исследования компании Яндекс, июнь 2011 [(2011) Internet search: what way do men and women search. *Yandex Co.* June. (In Russ.)]. URL: https://yandex.ru/company/researches/2011/ya_on_men_women_11 (дата обращения = accessed 10.03.2021).
- Пользователи Яндекса: интересы и поисковое поведение // Исследования компании Яндекс, апрель 2013 [(2013) Yandex users: interests and search behavior. *Yandex Co.* Apr. (In Russ.)]. URL: https://yandex.ru/company/researches/2013/ya_search_interests_2013 (дата обращения = accessed 10.03.2021).