

Н. Н. Литвинова

Российская государственная библиотека

Как пользователи находят статьи из научных журналов в электронной среде

Рассмотрены наиболее востребованные пользователями сервисы поиска научных статей – главного источника информации для современных исследователей. На основе данных глобального социологического исследования более 40 тыс. респондентов проведено сравнение сервисов по популярности среди пользователей. Выявлены наиболее часто используемые: реферативно-библиографические базы данных, научные поисковые сервисы, глобальные поисковые сервисы общего назначения, а также сайты библиотек, издателей, журналов и социальные сети. Проанализированы возможности библиотек в настройке популярных сервисов для лёгкого перехода пользователей от кратких описаний статей (библиографических записей) к их полным текстам, доступным благодаря подписке. Описаны нормативные документы National Information Standards Organization, регламентирующие обмен данными между контент-провайдерами (издателями и агрегаторами), разработчиками сервисов discovery и библиотеками для обеспечения унифицированного режима обмена и единых форматов. Освещены проблемы интеграции российских журнальных ресурсов в глобальную инфраструктуру обмена, среди которых – неполнота и противоречивость предоставляемых метаданных. Подчёркнута необходимость такой интеграции для увеличения видимости и цитируемости российских научных статей. Отмечена важность налаживания стандартизованного обмена данными между российскими контент-провайдерами, сервисами discovery и библиотеками.

Ключевые слова: поиск научных статей, реферативно-библиографические базы данных, сервисы discovery, глобальные поисковые сервисы, открытый доступ.

Nataliya Litvinova

Russian State Library, Moscow, Russia

How the users find science journal articles in the digital environment

The author discusses the most popular search services for science articles retrieval — which make the main information source for modern researchers. Based on the findings of the global sociological study of over 40,000 respondents, the services most popular with the users are compared. Among the most frequently used are: abstracted bibliographic databases, science retrieval

services, global general search services, libraries' www-sites, publishers' www-sites, journals www-sites, and social networks. The author analyzes libraries' potential for tuning popular services for the users to transit from brief article descriptions (bibliographic records) to their full texts available through subscriptions. To unify the exchange mode and formats, the requirements documents by the National Information Standards Organization (NISO) regulate data exchange between content providers (publishers and aggregators), discovery services developers and the libraries. Integration of Russian journal resources into the global exchange infrastructure is hindered by incompleteness and inconsistency of the metadata being provided. Meanwhile such integration is vital for Russian science article visibility and citation. The author also emphasizes the importance of the standardized data exchange between Russian content providers, discovery services, and the libraries.

Keywords: science articles retrieval, abstract and bibliographic databases, discovery services, global search services, open access.

In Russia the subscription of hundreds of Russian libraries is supported by state funding. Russian researchers have access to dozens of millions scientific articles on dozens of publishing platforms. It is possible with the help of global sociological surveys to obtain reliable and unbiased information on the use of various "starting points" in the search for journal content. Such studies have been carried out by Simon Inger Consulting Ltd since 2005. The last study of 2015–2016 covered 40,439 respondents. Generalized estimates showed the comparative popularity of different ways of users' movement towards the ultimate goal – the full text, namely: 1. Abstract databases. 2. Scientific Search Services. 3. Global search services of general purpose. 4. Sites of libraries. 5. Publisher sites. 6. Sites of journals. 7. Social networks. Abstracts databases remain in the first place for all four surveys, but the absolute values are steadily declining. The greatest contribution to their popularity is made by respondents who are engaged in life sciences (including medicine). This fact is well known in libraries: biologists and medical professionals are willing to use the PubMed and Medline databases. The popularity of library sites is increasing, especially in the academic sector, due to the wide introduction of discovery services. Thus, the alarming trend of the outflow of users from library sites to the search engines, which caused a wide resonance in the library community, is broken. Search services for general use lose ground, inferior to scientific search services, especially among representatives of the academic sector from high-income countries. The sites of publishers and journals show positive dynamics, which is largely due to the convenience of using them on mobile devices: publishers paid close attention to the development of mobile applications, and these sites are not so full of various options that complicate the use of mobile content. The greatest increase in popularity is expected to be demonstrated by social networks, which are rapidly developing in the last decade. However, the absolute values of estimates of their importance are not high. The most significant devia-

tions from the average estimates are generally demonstrated by library search professionals.

Наиболее востребованный источник информации для современных исследователей – статьи из научных журналов. В них содержится актуальная и оперативная информация, необходимая для научной работы. Чтобы быть в курсе новостей по интересующей учёного тематике, он должен знакомиться с большим количеством журналов, посвящённых как *своей* тематике, так и смежным с ней. Мало кто может позволить себе индивидуальную подписку на такие журналы, поэтому основным источником удовлетворения информационных потребностей исследователя становится подписка публичной библиотеки и/или библиотеки организации, в которой он работает.

Научные библиотеки всего мира тратят львиную долю бюджета на подписку на журналы, причём в основном – в форме доступа к платформам контент-провайдеров: издателей и агрегаторов.

В России в последнее десятилетие подписка сотен российских библиотек на зарубежную периодику поддерживается государственным финансированием централизованной (национальной) подписки [1]. Во многом благодаря этому российские исследователи имеют доступ к десяткам миллионов научных статей на десятках издательских платформ.

Получение объективных данных о наиболее популярных сервисах поиска журнального контента

В качестве отправных точек для поиска в огромных массивах журнальных статей пользователи могут выбрать разнообразные сервисы:

сайты библиотек, где есть электронный каталог, перечни доступных лицензионных ресурсов или сервисы *discovery*, обеспечивающие поиск по ЭК и лицензионным ресурсам через единый индекс;

реферативно-библиографические базы данных широкого тематического спектра (*Web of Science, Scopus*) или специализированные БД по интересующей пользователя тематике (*PubMed, Medline, MathSciNet* и др.);

сайты контент-провайдеров (издателей, агрегаторов, научных обществ); сайты конкретных журналов;

социальные сети, особенно научные (*Mendeley, Research Gate* и др.);

глобальные поисковые сервисы общего назначения (*Google* и подобные);

научные поисковые сервисы (*Google Scholar* и подобные).

Информация о том, как пользователи ищут и находят журнальный контент в электронной среде, чрезвычайно важна и для контент-провайдеров, и для библиотек. Она нужна издателям и агрегаторам для того, чтобы понять, какие из поисковых сервисов наиболее востребованы, и обеспечить наилуч-

шую видимость в них своего контента, а также по возможности максимальное удобство поиска и выхода на полные тексты статей.

Для библиотек также важно понимать, какими путями идут их пользователи при поиске журнальных статей, чтобы сделать эти пути удобными, предоставив информацию о том, какие журналы и статьи доступны благодаря библиотечной подписке. Если библиотека успешно выполнит эту задачу, её читатели будут активировать ссылки на полные тексты не наугад, а с уверенностью в результате поиска.

Получить надёжную и объективную информацию об использовании различных «отправных точек» при поиске журнального контента возможно с помощью глобальных социологических исследований репрезентативной выборки респондентов. Такие исследования с 2005 г. проводит компания *Simon Inger Consulting Ltd* с периодичностью раз в три года. В последнем исследовании 2015–2016 гг. участвовали 40 439 респондентов со всех континентов. С его результатами можно ознакомиться в специальном отчёте [2], опубликованном под лицензией *Creative Commons Attribution-NonCommercial 4.0*.

Респонденты оценивали важность каждого из названных выше сервисов поиска журнальных статей по десятибалльной шкале. Результаты оценки популярных сервисов поиска различными группами респондентов представлены на ряде диаграмм. Группы выделялись по следующим признакам:

область деятельности (тематика исследований);

сектор: академический (научный), корпоративный (бизнес), правительственный, неправительственный, медицина и здравоохранение;

страна, регион;

уровень доходов населения по классификации Всемирного банка: страны с высоким уровнем дохода, с доходами выше среднего уровня, с доходами ниже среднего уровня, с низким уровнем дохода.

По этим признакам оказалось возможным разделить на группы около 90% респондентов, которые ответили на соответствующие вопросы анкеты. Выборка формировалась с ориентацией на максимальный охват стран и регионов, однако наиболее представительной стала группа респондентов из Северной Америки и Европы (53%), в которую входят в основном страны с высоким уровнем дохода. Россия (около 400 респондентов) как европейская страна также входит в эту группу, хотя в последние годы её относят к странам с уровнем доходов выше среднего.

В результате исследования были выявлены пути движения пользователей к конечной цели – полным текстам статей – и проведено сравнение популярности различных сервисов; места распределились в следующем порядке:

реферативно-библиографические БД,

научные поисковые сервисы,

глобальные поисковые сервисы общего назначения,

сайты библиотек,
сайты издателей,
сайты журналов,
социальные сети.

Эти данные сравнили с полученными в ходе предыдущих опросов, что позволило судить о динамике популярности сервисов. Отметим наиболее интересные наблюдения по динамике и группам респондентов.

Реферативно-библиографические БД остаются на первом месте по результатам всех проведённых опросов, но абсолютные значения оценок неуклонно снижаются. Наибольший вклад в популярность таких БД вносят респонденты, занимающиеся науками о жизни (включая медицину). Этот факт хорошо известен в библиотеках: биологи и медики охотно пользуются базами *PubMed* и *Medline*.

Возрастает популярность библиотечных сайтов, особенно у представителей академического сектора, что объясняется широким внедрением в библиотеках сервисов *discovery*, которые обеспечили возможность поиска всех ресурсов библиотеки из одного окна. Таким образом, тревожная тенденция оттока пользователей с библиотечных сайтов в поисковые системы, вызвавшая широкий резонанс в библиотечном сообществе особенно после публикации исследования OCLC [3], переломлена.

Поисковые сервисы общего назначения теряют позиции, уступая научным поисковым сервисам, особенно среди представителей академического сектора из стран с высоким уровнем доходов.

Положительную динамику показывают сайты издателей и журналов, что в значительной степени обусловлено удобством их использования на мобильных устройствах: во-первых, издатели уделили пристальное внимание разработке мобильных приложений, во-вторых, эти сайты не так насыщены разнообразными опциями, усложняющими пользование мобильным контентом.

Наибольший прирост популярности ожидаемо демонстрируют социальные сети, бурно развивающиеся в последнее десятилетие. Однако абсолютные значения оценок их важности невысоки.

Самые существенные отклонения от средних оценок по выборке в целом демонстрируют библиотекари – профессионалы поиска.

Поддержка пользователей библиотеками в различных сервисах

Как было сказано выше, библиотекам важно знать, какие сервисы поиска журнальных статей наиболее популярны у пользователей. Чтобы сделать их максимально удобными, обеспечив надёжными ссылками на полные тексты статей, которые доступны благодаря библиотечной подписке или находятся в открытом доступе [4–6].

Наиболее просто эта задача решается на сайтах контент-провайдеров – издателей, агрегаторов и научных сообществ, если последние выступают издателями. Доступность статей обеспечивается оплаченным библиотекой правом доступа для авторизованных пользователей. Авторизация происходит либо при доступе с компьютера с зарегистрированного у контент-провайдера IP-адреса, либо путём ввода логина и пароля. Зарегистрированный IP-адрес опознаётся не только при доступе с компьютеров библиотеки, но и при работе пользователя из любого места через надёжный прокси-сервер. Краткие описания (библиографические записи) статей, как правило, снабжены индикаторами доступности, такими как зелёная пиктограмма (по принципу светофора) или текстовыми сообщениями (например, *full access*). В последнее время многие издатели стали применять также специальные фильтры, позволяющие выводить в результатах поиска только записи на доступные статьи.

Поисковые системы *discovery*, размещённые на библиотечном сайте, настраиваются аналогичным образом, чтобы пользователь сразу понимал, какие из кратких описаний статей приведут его к доступному полному тексту. Настройка заключается в том, что в глобальной базе знаний сервиса прописываются только те журналы и выпуски, которые доступны пользователю, авторизованному данной библиотекой. Переход к полным текстам обеспечивает *link resolver* – определитель ссылок, посылающий запрос на поиск статьи по её метаданным на те платформы, на которых она должна быть доступна по данным настроенной библиотекой базы знаний. Такой запрос оформляется в соответствии со стандартом *Open URL*. Таким образом, переход от краткой записи к полному тексту статьи происходит в два этапа: сначала активация ссылки на *link resolver* и переход к странице выбора доступной платформы, если их более одной, затем – переход на платформу и открытие полного текста.

Иначе обстоит дело со столь ценимыми пользователями реферативно-библиографическими БД: библиотеке придётся провести интеграцию данных о журналах, доступных её пользователям, с держателями этих БД. Конечно, можно договориться о регулярной поставке данных о доступной периодике. Но такое решение нельзя назвать удачным: регулярная и оперативная поставка данных многочисленным провайдерам БД – трудоёмкая и неудобная процедура, поскольку данные очень изменчивы. Гораздо лучше проводить изменения доступного репертуара периодики в одной точке, т.е. в базе знаний сервиса *discovery*, на который опирается корректная работа *link resolver*, и договориться с держателями реферативно-библиографических БД, чтобы они внедрили ссылки на используемый библиотекой *link resolver*. Эти ссылки будут отображаться для авторизованных пользователей библиотеки.

Такие договорённости и соответствующие процедуры широко используются библиотеками всего мира. Если договорённости достигнуты, то пере-

ход к доступным полным текстам от кратких записей в реферативно-библиографических БД будет практически таким же, как в библиотечных сервисах *discovery*.

Аналогичное решение можно использовать и для глобального научного поискового сервиса *Google Scholar*, который реализует его с помощью опции *Library Links*.

Следует подчеркнуть, что такие унифицированные решения, реализуемые множеством игроков на поле научного контента, возможны только в рамках специально выстроенной инфраструктуры обмена данными, поддерживаемой отраслевыми стандартами и/или практическими рекомендациями (*Recommended Practice*). Ведущая роль в их разработке и продвижении принадлежит Национальной организации по информационным стандартам США – *National Information Standards Organization (NISO)*. В NISO разработаны основные нормативные документы, обеспечивающие согласованный обмен данными и надёжные переходы от кратких записей на статьи к их полным текстам. Основные документы перечислены ниже.

NISO RP-19-2014. Open Discovery Initiative: Promoting Transparency in Discovery (<https://www.niso.org/publications/rp-19-2014-odi>) регламентирует процедуры, периодичность и форматы обмена данными между контент-провайдерами, провайдерами сервисов *discovery* и библиотеками.

NISO RP-9-2014. KBART: Knowledge Bases and Related Tools Recommended Practice (<https://www.niso.org/publications/rp-9-2014-kbart>) регламентирует состав и наполнение метаданных об электронных периодических изданиях и электронных книгах для наполнения баз знаний.

NISO RP-16-2013 PIE-J: The Presentation & Identification of E-Journals. Основное внимание в документе уделено представлению информации о журналах на всех уровнях: название, том, выпуск и отдельные статьи; распространённая в электронной среде практика описаний всех выпусков журнала под современным названием, без учёта переименований на протяжении его истории недопустима.

ANSI/NISO Z39.88-2004 (R2010) The Open URL Framework for Context-Sensitive Services (<https://www.niso.org/publications/z3988-2004-r2010-openurl-framework-context-sensitive-services>). Стандарт регламентирует формирование информационных пакетов для передачи от ресурса-источника к ресурсу-цели.

Российские журналы в инфраструктуре глобального обмена: проблемы интеграции

В России издаются тысячи научных журналов, статьи из которых востребованы российским научным сообществом. Поэтому очень важно сделать так, чтобы поиск статей был максимально удобным независимо от того, какую отправную точку выбирает пользователь, начиная поиск.

Как было показано выше, для достижения этой цели в мире разработана глобальная инфраструктура обмена данными между контент-провайдерами, провайдерами сервисов *discovery* и библиотеками. К сожалению, многие российские журналы и платформы, на которых они размещены, пока не вписываются в эту инфраструктуру. Поэтому их контент труднодоступен во всех сервисах-посредниках: реферативно-библиографических БД, сервисах *discovery* и глобальных научных поисковых сервисах. Как следствие, пользователи часто не доходят до полных текстов статей, привлёкших их внимание на уровне кратких описаний, поскольку к ним не обеспечен удобный переход в два-три клика. В результате уменьшаются видимость и, возможно, цитируемость российских статей, которая так важна для достижения целевых показателей деятельности учёных и организаций. Причём это относится даже к самым авторитетным российским изданиям, индексируемым в *Web of Science*, *Scopus*, *PubMed* и других глобальных базах.

Проиллюстрируем это примером. Русскоязычный журнал «Микробиология», размещённый на платформе eLibrary.ru, входит в *Scopus* и *PubMed*. Статьи журнала находятся в открытом доступе за период с 2007 по 2017 г. РГБ, пользующаяся сервисом *discovery EBSCO Discovery Service (EDS)*, прописала его в базе знаний сервиса с указанным доступным периодом и выполнила необходимые процедуры интеграции с *Google Scholar*, *Web of Science*, *Scopus*, *PubMed* и рядом других сервисов. Таким образом, описания статей из журнала «Микробиология» авторизованный пользователь РГБ может найти и в *Scopus*, и в *PubMed* со ссылками на *link resolver*, которые должны вывести его на полные тексты доступных статей. Однако фактически этот путь не так гладок, как для статей из зарубежных изданий.

Причина заключается в том, что платформа eLibrary.ru не может принять запрос определителя ссылок по стандарту *Open URL*, содержащий метаданные статьи. Поэтому определитель ссылок выводит пользователя не на конкретную статью, которая его заинтересовала, а на страницу журнала с его описанием и многочисленными опциями работы с журналом. Захочет ли пользователь, привыкший к корректной работе определителя ссылок, искать опции поиска в этом журнале или переход к оглавлениям выпусков и продолжать путь к статье?

Статистика показывает, что этого не происходит, по меньшей мере, в половине случаев: количество обращений к определителю ссылок от статей, найденных через EDS на платформе eLibrary.ru, примерно вдвое превосходит количество полных текстов, открытых на этой платформе. И это с учётом того, что часть пользователей открывают статьи, обратившись непосредственно на платформу eLibrary.ru.

Ситуация с поддержкой стандарта *Open URL* на платформах российских издателей и журналов ничуть не лучше.

Другая проблема российских научных журналов, затрудняющая интеграцию их контента в разнообразные поисковые сервисы, – неполнота и противоречивость поставляемых метаданных. Так, журнал «Микробиология» представлен в *PubMed* с транслитерированным названием *Mikrobiologiia*, а в *Scopus* и *Ulrich's Periodical Directory* он называется *Mikrobiologiya*. Описание журнала в *Ulrich's Periodical Directory* не содержит ISSN – важнейшего элемента метаданных для корректной работы ссылок на уровне журналов.

Практически все российские журналы, поставляющие свой контент в глобальные реферативно-библиографические БД, не включают в метаданные статей элемент «Тип доступа», являющийся обязательным согласно KBART. Именно этот элемент позволяет выделить в результатах поиска статьи открытого доступа и беспрепятственно переходить к их полным текстам. Если этот элемент метаданных отсутствует, тип доступа системой не определяется – и все статьи удаляются из результатов поиска при применении фильтра «Открытый доступ». Устранение этого пробела в метаданных российских статей чрезвычайно важно, поскольку статьи открытого доступа чаще выгружаются пользователями, прочитываются ими и затем цитируются. Об этом свидетельствуют недавние исследования, опирающиеся на анализ больших массивов данных [7, 8].

Не менее важным является налаживание стандартизованного обмена данными между российскими контент-провайдерами, сервисами *discovery* и библиотеками. В последние пять-семь лет ряд российских контент-провайдеров стали поставлять метаданные в *EBSCO Discovery Service*. Среди них преобладают электронно-библиотечные системы, для которых журнальный контент не является основным. Журналы в ЭБС представляются как совокупности отдельных выпусков, а статьи из них не являются отдельными объектами описания.

Крупнейший российский агрегатор журналов «Научная электронная библиотека» (*eLibrary.ru*) поставляет в базу знаний EDS только данные о примерно 1 500 журналах, распространяемых ею по подписке, а данные о доступности остальных журналов платформы (более 3 тыс.) отсутствуют. Из-за этого библиотеки не могут корректно настроить БЗ в части доступа к *eLibrary.ru* и обеспечить надёжную работу ссылок на полные тексты доступных статей.

Обобщая всё вышеизложенное, ещё раз подчеркнём важность интеграции российского журнального контента в популярные во всём мире глобальные сервисы поиска статей. Это будет способствовать повышению видимости российских научных публикаций в мировом информационном пространстве и с высокой долей вероятности – увеличению цитирования работ российских исследователей.

СПИСОК ИСТОЧНИКОВ

1. **Евстигнеева Г. А.** Национальный доступ к международным базам данных в рамках федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технического комплекса России на 2014–2020 годы» / Г. А. Евстигнеева // Науч. и техн. б-ки. – 2016. – № 5. – С. 29–43.

Evstigneeva G. A. Nacionalnyy dostup k mezhdunarodnym bazam dannyh v ramkah federalnoy tselevooy programmy «Issledovaniya i razrabotki po prioritetnym napravleniyam razvitiya nauchno-tehnicheskogo kompleksa Rossii na 2014–2020 gody» / G. A. Evstigneeva // Nauch. i tehn. b-ki. – 2016. – № 5. – S. 29–43.

2. **Gardner T., Inger S.** How Readers Discover Content in Scholarly Publications: Trends in Behaviour from 2005–2015 [Электронный ресурс]. – Режим доступа: <http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1012&context=scholcom> (дата обращения: 10.05.2018).

3. **De Rosa C., Cantrell J., Cellentani D., Hawk J., Jenkins L. and Wilson A.** Perceptions of Libraries, 2010: Context and Community, Dublin. OH: OCLC [Электронный ресурс]. – Режим доступа: <https://www.oclc.org/en/reports/2010perceptions.html> (дата обращения: 10.05.2018).

4. **Шрайберг Я. Л., Гончаров М. В., Земсков А. И., Колосов К. А.** Открытый доступ: зарубежный и отечественный опыт – состояние и перспективы / Я. Л. Шрайберг, М. В. Гончаров, А. И. Земсков, К. А. Колосов // Науч. и техн. б-ки. – 2012. – № 8. – С. 5–26.

Shrayberg Ya. L., Goncharov M. V., Zemskov A. I., Kolosov K. A. Otkrytyy dostup: zarubezhnyy i otechestvennyy opyt – sostoyanie i perspektivy / Ya. L. Shrayberg, M. V. Goncharov, A. I. Zemskov, K. A. Kolosov // Nauch. i tehn. b-ki. – 2012. – № 8. – S. 5–26.

5. **Johnson R., Fosci M.** Putting down roots: Securing the future of open-access policies. Knowledge Exchange report of Workshop held on the 10th of November 2015.

6. **Piwovar H., Priem J., Larivière V., Alperin J. P., Matthias L., Norlander B., Farley A., West J., Haustein S.** The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. // PeerJ 6:e4375 <https://doi.org/10.7717/peerj.4375>

7. **Open access availability of scientific publications** [Электронный ресурс]. – Режим доступа: http://www.science-metrix.com/sites/default/files/science-metrix/publications/science-metrix_open_access_availability_scientific_publications_report.pdf (дата обращения: 10.05.2018).

8. **Archambault E., Côté G., Struck B., Voorons M.** Research impact of paywalled versus open access papers [Электронный ресурс]. – Режим доступа: <http://www.lscience.com/oanumbr.html> (дата обращения: 10.05.2018).

Nataliya Litvinova, Cand. Sc. (Philology), Chief Librarian, Russian State Library;

nlit@rsl.ru

3/5 Vozdvizhenka st., 119019 Moscow, Russia