

ОТКРЫТЫЙ ДОСТУП И ОТКРЫТЫЕ АРХИВЫ ИНФОРМАЦИИ

УДК 026.06:62(470–25)

М. В. Вахрушев

ГПНТБ России

Портал «Научный архив» ГПНТБ России как часть инфраструктуры научной коммуникации

Рассмотрена инфраструктура научной коммуникации на примере ГПНТБ России. Представлены функциональные особенности «Научного архива» ГПНТБ России, отмечено, какие именно научные труды индексируются на портале. Приведены статистические показатели наполняемости «Научного архива», а также пример метаданных, хранящихся в индексе. Проанализированы составляющие методики ранжирования научного труда. Приведены статистические показатели ранжирования научных трудов в срезе научных работ, например статья или учебное пособие. Подчёркнуто, что технологической основой «Научного архива» является СУБД MongoDB. Индексы научных трудов, учётные данные пользователей, статистика поиска по базе данных, метаданные научных трудов, индексирующиеся поисковыми серверами интернета, реализованы и функционируют с учётом требований протоколов концепции Открытого доступа. Отмечена функциональная особенность «Научного архива» – система депонирования, которая действует на основе передачи авторской лицензии правообладателем «Научному архиву». Сделан вывод: в информационно-технологическом обществе и рядовой научный сотрудник, и маститый учёный могут столкнуться с проблемой «нехватки» коммуникационной среды. Это свидетельствует о том, что «Научный архив» и другие порталы являются частями обширной научной среды.

Ключевые слова: научная коммуникация, научный архив, инфраструктура научной информации.

UDC 026.06:62(470–25)

Maxim Vakhrushev

Russian National Public Library for Science and Technology, Moscow, Russia

The RNPLS&T's Science Archive Portal in the scientific communication system

Scientific communication infrastructure is examined by the example of the Russian National Public Library for Science and Technology. The functional features of RNPLS&T's Science Archive are discussed. Types of scientific works to be indexed via the portal are presented.

Statistics of the Science Archive representation is given, and the metadata being stored is characterized. The author analyzes the ranking methods for academic works, and cites statistical ranking data for different types of scientific works, e. g. articles or learning aids, etc. The MongoDB DBMS makes the technological foundation for the Science Archive. Academic works indices, user account data, database retrieval data, metadata of academic works indexed by the Internet search servers are implemented and operate in compliance with the Open Access protocols. The depositing functionality is licensed to the Science Archive by a copyright owner. The author concludes that in the information and technological society a common researcher and a prominent scientists may lack the communication environment. This evidences that the Science Archive and similar portals make the essential part of the vast science environment.

Keywords: scientific communication, science archive, scientific information infrastructure.

The examples of effective implementation of the digital infrastructure of scientific communication are RINC or CyberLeninka. The RINC platform realizes the function of ranking publications according to the standards of bibliometrics and access to full text, and CyberLeninka implements the concept of open access. The development by the Russian National Public Library for Science and Technology the "Scientific Archive" was due to the additional requirements of the state and the scientific, and the "Scientific Archive" complements them since of 2017 (<https://nauchnyharchiv.rf>). It consists of two parts: the public access information system to a single digital collection of primary sources of scientific works of certified quality and the information system for the collection and processing the digital copies of monographs, scientific papers and so on. There are indexed articles, monographs, reports, textbooks, abstracts, theses, doctoral dissertations etc. The technological basis of the resource is a database based on the Mongo database, with indexes of scientific works, i.e. technical description of the work used for the purposes of the DBMS. The indexing methodology assumes reading the text layer for machine ranking of primary sources and searching for matches, borrowings, citations. The relevant information is stored in the index, the text itself of work in the index is not stored. Access to the full text of scientific work is carried out at the specified URL in accordance with current copyright laws of the Russian Federation. On the platform of the portal, the system for the deposit of scientific works is implemented, in which the user can independently post his publication. For the convenience of users with the deposit system, a video presentation was developed and posted on the YouTube portal. The document can be attributed to a predetermined class (rank) according to such characteristics of quality as originality, degree of independence and citation. For the time being the total scientific works have: the highest rank – 71 942; high rank – 143,143; average rank – 2,147,129; conditional rank – 147 222; low rank – 162,031.

С развитием информационно-коммуникационных технологий и, в частности, интернета научная коммуникация вышла на новый уровень саморегулирования научной деятельности как учёных, так и организаций. Известно, что коммуникации в сфере науки приводят к развитию экономики знаний. В этой статье рассмотрена инфраструктура научной коммуникации в отдельно взятой организации – ГПНТБ России. Реализация подобной инфраструктуры на её технологической платформе представляется обоснованной.

До настоящего времени мало кто занимался развитием всего комплекса инфраструктуры научной коммуникации. В связи с этим примечательно высказывание: «Фрагментация научной среды, изолированность научных коллективов не позволяет получить синергетический эффект от обмена знаниями в процессе научно-исследовательской деятельности» [1. С. 129].

Конечно, в России в последнее время появились примеры эффективной реализации цифровой инфраструктуры научной коммуникации – это, например, РИНЦ или КиберЛенинка. На платформе РИНЦа реализовали только функцию ранжирования публикаций и журналов по стандартам библиометрии и доступ к полному тексту с учётом авторских прав. А вот КиберЛенинка реализует концепцию открытого доступа. Но и в том, и в другом случае мы говорим о взаимодополняющих частях общей информационной инфраструктуры. Говоря языком экономики, речь идёт о разных субъектах научной деятельности, настроенных на разные целевые сегменты.

Возникновение и развитие портала, разработанного ГПНТБ России, было обусловлено дополнительными требованиями государства и научного сообщества к развитию инфраструктуры научной коммуникации. Повторюсь, «Научный архив» не является конкурентом РИНЦ и КиберЛенинке, но дополняет их.

Портал «Научный архив»

С начала 2017 г. портал «Научный архив» (<https://научныйархив.рф>) эксплуатируется на технологической площадке ГПНТБ России. Он состоит из двух взаимосвязанных частей: Информационной системы публичного доступа к единой цифровой коллекции первоисточников научных работ удостоверенного качества и Информационной системы по сбору и обработке цифровых экземпляров монографий, научных трудов и исследований. Типы трудов, которые индексируются «Научным архивом»: статья, монография, отчёт, учебное пособие, реферат, дипломная работа, кандидатская диссертация, автореферат кандидатской диссертации, докторская диссертация, автореферат докторской диссертации, сборник.

Технологическая основа ресурса – база данных на основе СУБД MongoDB с индексами научных трудов, т.е. техническим описанием работы, применяемым для целей работы СУБД. В отличие от индексов метаданных являются машиночитаемым описанием научных трудов с указанием URL источника, по которому хранятся полнотекстовые экземпляры работ, пригодных для целей библиографии.

Методология индексирования предполагает чтение текстового слоя для машинного ранжирования первоисточников и поиска совпадений, заимствований, цитирования. Соответствующая информация хранится в индексе, сам текст научного труда в индексе не хранится. Доступ к полному тексту научного труда осуществляется по указанному URL в соответствии с авторскими правами и действующим авторским законодательством РФ. На рис. 1 представлен образец заполнения полей метаданных на примере автореферата кандидатской диссертации.

The screenshot shows a web browser window with the URL <https://научный.архив.рф/documents/document?id=1023d37f-c0e2-486b-91d5-ceec13ace3c2>. The page title is "Организационно-педагогические условия развития мультикомпетентности старшекласников: технологический подход". The main content is a table of metadata, and the right sidebar contains a list of actions.

Ранг:	Высокий
Автор:	Вахрушев Максим Васильевич
Тип документа:	Автореферат кандидатской диссертации
Дата публикации:	2012
Шифр специальности ВАК:	13.00.01
ГРНТИ:	не указано Значение, определенное системой: 14
Место защиты/выпуска:	Москва
Ученый секретарь диссертационного совета:	Акунина Ю. А.
Научный руководитель или консультант:	Сладкова Ольга Борисовна

- Текст документа
- Источник
- Отчет о качестве
- Экспертные заключения (0)
- Удалить из избранного
- Найти похожее
- Служебная информация
- Статистика
- Экспорт метаданных
- Изменить метаданные
- Удалить документ

Рис. 1. Заполнение полей метаданных на примере автореферата

На платформе портала реализована Система депонирования научных трудов, в которой пользователь может самостоятельно разместить своё исследование. Для удобства работы пользователей с Системой депонирования была разработана видеопрезентация и размещена на портале *YouTube*.

Перечень научных трудов, проиндексированных в «Научном архиве», представлен в таблице.

Перечень проиндексированных трудов

Научный труд	Количество
Статья	1 874 781
Монография	2 134
Отчёт	0
Учебное пособие	24 748
Реферат	1
Дипломная работа	1
Кандидатская диссертация	327 772
Автореферат кандидатской диссертации	340 707
Докторская диссертация	46 848
Автореферат докторской диссертации	51 743
Сборник	2 031
Депонированные работы	72
Число документов в обработке [*]	629
Итого:	2 671 467

Ранжирование научных трудов

Одна из методик, на основе которых проводится индексирование работ в «Научном архиве», – методика определения ранга документа.

По таким характеристикам качества, как оригинальность, степень самостоятельности и цитируемость, документ можно отнести к заранее определённом классу (рангу). В зависимости от целей ранжирования ранг может быть определён по-разному.

Высший: научные труды высокой оригинальности, с высокой степенью уникальности, цитируемости и востребованности текста.

Высокий: научные труды высокой оригинальности, с высокой степенью уникальности текста, с любой цитируемостью, с низкой или средней востребованностью.

Средний: научные труды средней оригинальности, с высокой степенью уникальности текста, любой цитируемостью, с низкой или средней востребованностью.

Условный: работы средней или низкой оригинальности, с высокой или средней степенью уникальности текста, с любыми цитируемостью и востребованностью.

Низкий: работы низкой оригинальности, средней или низкой степени уникальности текста, с любыми цитируемостью и востребованностью.

* Документы, по которым ещё не определено ранжирование, уровень заимствования, цитирования и т.д.

На рис. 2 представлена диаграмма, иллюстрирующая соотношение рангов всех научных трудов, проиндексированных в «Научном архиве».

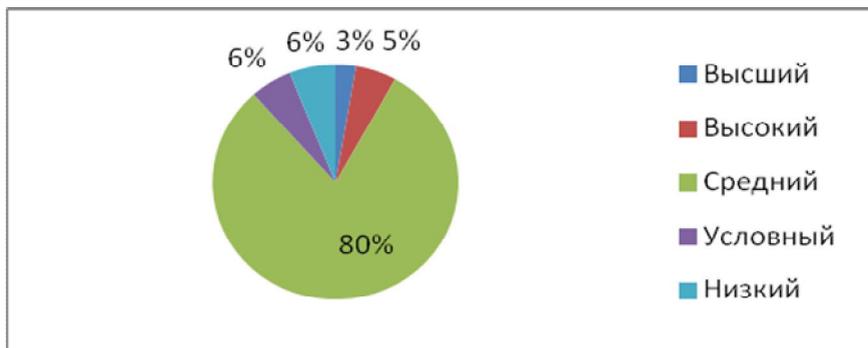


Рис. 2. Соотношение рангов научных трудов

Всего научных трудов: высший – 71 942; высокий – 143 143; средний – 2 147 129; условный – 147 222; низкий – 162 031.

Диаграмма на рис. 3 показывает соотношение рангов среди кандидатских, а на рис. 4 – среди докторских диссертаций.

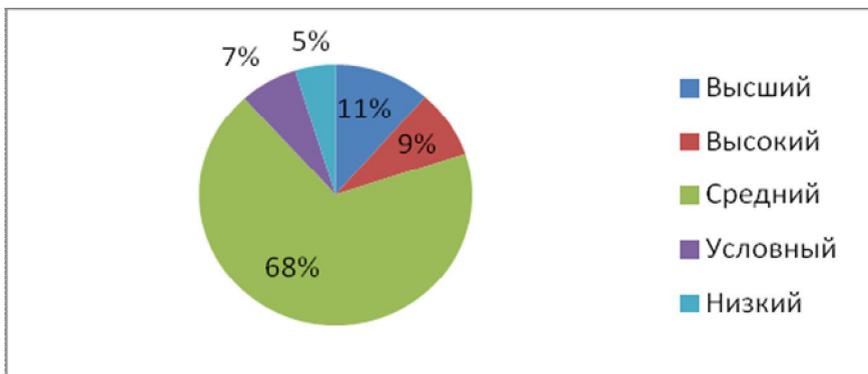


Рис. 3. Соотношение рангов кандидатских диссертаций

Кандидатские диссертации: высший – 37 388; высокий – 27 972; средний – 224 132; условный – 22 534; низкий – 15 746.

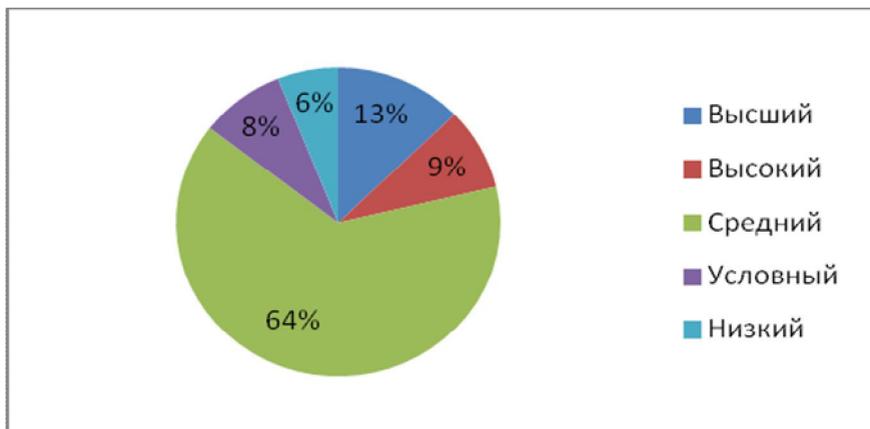


Рис. 4. Соотношение рангов докторских диссертаций

Докторские диссертации: высший – 5 976; высокий – 4 001; средний – 30 037; условный – 3 983; низкий – 2 851.

Не представлены статистические показатели по монографиям, отчётам, рефератам, дипломным работам и сборникам, потому что они не являются репрезентативными.

В современном информационно-технологическом мире и рядовой сотрудник, и маститый научный деятель могут столкнуться с проблемой «нехватки» коммуникационной среды, хотя интернет пестрит примерами технологической реализации научной коммуникации. «Научный архив» и другие порталы, реализующие научную коммуникацию, являются частями обширного научного сообщества. Их функционал всегда будет разным при кажущемся сходстве технологических решений.

СПИСОК ИСТОЧНИКОВ

1. **Савченко А. П.** Открытое информационное пространство научной коммуникации как фактор развития экономики знаний в России // Гос. и муниципал. упр. Учёные зап. СКАГС. – 2017. – № 1.

Savchenko A. P. Otkrytoe informatsionnoe prostranstvo nauchnoy kommunikatsii kak faktor razvitiya ekonomiki znaniy v Rossii // Gos. i munitsipal. upr. Uchenye zap. SKAGS. – 2017. – № 1.

2. **Сергеев С. Ю.** Роль библиотек в эффективном использовании научной информации в электронной коммуникации // Пед. образование в России. – 2013. – № 5.

Sergeev S. Yu. Rol bibliotek v effektivnom ispolzovanii nauchnoy informatsii v elektronnoy kommunikatsii // Ped. obrazovanie v Rossii. – 2013. – № 5.

Maksim Vakhrushev, *Cand. Sc. (Pedagogy)*, Senior Researcher, Russian National Public Library for Science and Technology;

vahrushev@gpntb.ru

17, 3rd Khoroshevskaya st., 123298 Moscow, Russia