

ОТКРЫТЫЙ ДОСТУП. ОТКРЫТЫЕ АРХИВЫ ИНФОРМАЦИИ

УДК 022.42+930.25

DOI: 10.33186/1027-3689-2020-12-45-58

Я. Л. Шрайберг, М. В. Гончаров, К. А. Колосов

ГПНТБ России, Москва, Россия

Московский государственный лингвистический университет, Москва, Россия

О разработке концепции Открытого архива информации ГПНТБ России

Аннотация: Рассмотрены принципы и технологии, используемые при создании Открытого архива ГПНТБ России. Особое внимание уделено актуальности поставленной задачи в условиях цифровизации. В качестве основной цели выступает создание единого Открытого архива информации (ЕОАИ) ГПНТБ России, объединяющего все электронные полнотекстовые ресурсы, создаваемые или собираемые библиотекой на правовой основе. Приведены базовые принципы ЕОАИ ГПНТБ России как целостной информационной системы. Раскрыты функциональные возможности и структура таблиц используемых протоколов для разработки специального программного и технологического обеспечения для ЕОАИ ГПНТБ России. Приведена и обоснована перспективная функциональная схема архива. Рассмотрены возможные схемы программно-технической реализации сервера приложений и метаданных ЕОАИ ГПНТБ России, а также вопросы взаимодействия со смежными и родственными крупными информационными системами, в том числе Национальной электронной библиотекой. Отмечено, что реализация системы будет важным этапом на пути к созданию типовой модели электронной библиотеки, функционирующей как участник взаимодействия в системе открытых архивов. Это упростит технологию обмена данными в информационных системах библиотек, научно-информационных подразделениях научных организаций, университетов, общеобразовательных учреждений и обеспечит поддержку поиска информации через единую точку доступа на веб-порталах. Статья подготовлена в рамках Государственного задания № 075-01300-20-00 «Разработка и совершенствование системы Открытого архива интегрированных информационно-библиотечных ресурсов ГПНТБ России как современной системы управления знанием в цифровой среде: на пути к Открытой науке» на 2020–2022 гг.

Ключевые слова: открытый доступ, открытые архивы, библиотечные автоматизированные системы.

OPEN ACCESS. OPEN INFORMATION ARCHIVES

UDC 022.42+930.25

DOI: 10.33186/1027-3689-2020-12-45-58

Yakov L. Shrayberg, Mikhail V. Goncharov and Kirill A. Kolosov

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

Moscow State Linguistic University, Moscow, Russia

On the concept of RNPLS&T's Open Archive

Abstract: The authors examine the principles and technologies of RNPLS&T's Open Archive development. Special attention is given to the urgency of the task under the circumstances of digitalization. The main goal is seen as building the Single Open Archive of Information (SOAI) of the Russian National Public Library for Science and Technology that is to accumulate every digital full-text resources designed or acquired legally by the Library. The key principles of SOAI as an integrated information system are specified. The functionalities and protocol table structure for RNPLS&T's SOAI soft- and hardware are discussed. The projected functional structure of the archive is described and explained. Possible software and hardware solutions for SOAI application and metadata server are discussed along with the interaction with large related information systems like the National Electronic Library (NEL). The system implementation will become the next important stage in building the typical model of e-library as a player in the open archive system. This is to facilitate the technology of data exchange in library information systems, units of science information within research organizations, universities, general educational institutions, and to support information retrieval via the single access point at web-portals. The article is prepared within the framework of the State Order № 075-01300-20-00 "Design and development of RNPLS&T's Open Archive of Integrated Information Library Resources as a modern system of knowledge management within the digital environment: On the way to Open Science" for the years 2020–2022.

Keywords: Open access, open archives, automated library systems.

Постановка задачи

В ГПНТБ России с начала 2000-х гг. создаётся собственная электронная библиотека и ряд отраслевых электронных полнотекстовых коллекций (в первую очередь авиационно-космического и химического профилей). А в 2005 г. началась совместная работа с РГБ и РНБ по формированию и развитию ресурсов Национальной электронной библиотеки (НЭБ), продолжающаяся до сих пор. Кроме того, в пул электронных полнотекстовых ресурсов нашей библиотеки включаются оцифрованные копии редкой книги; полные тексты сборников трудов Крымского форума и конференции «ЛИБКОН», других научно-образовательных мероприятий, организованных ГПНТБ России; коллекции трудов сотрудников библиотеки; авторефераты и диссертации защитившихся сотрудников или представителей других организаций, проходивших через базовую кафедру при МГИК и передавших свои труды с правом публичного доступа к ним; другие полнотекстовые материалы, переданные в электронную библиотеку ГПНТБ России на правовой основе.

Ставится задача: создать единый Открытый архив информации (ЕОАИ) ГПНТБ России, объединяющий все электронные полнотекстовые ресурсы, формируемые или собираемые библиотекой на правовой основе с целью расширения пользовательского сервиса и создания базы для научных исследований и обеспечения научно-образовательной деятельности. Этот архив является главным ответом ГПНТБ России на современные вызовы цифровизации.

Принципы формирования и использования ЕОАИ

ЕОАИ разрабатывается на принципах построения информационных систем (ИС), включая базовые принципы для систем автоматизации библиотек [1]. Перечислим основные из них.

Открытость. Ресурсы ЕОАИ создаются и используются исключительно как ресурсы открытого доступа, предоставляемые всем зарегистрированным пользователям ГПНТБ России беспрепятственно на бесплатной основе.

Легитимность. Размещение и использование ресурсов осуществляется на легитимной основе в рамках существующих правил отечественного законодательства об авторском праве и интеллектуальной собственности и соответствующих норм международного права.

Интероперабельность. Способность программно-технического обеспечения (ПТО) и интерфейса системы беспрепятственно взаимодействовать с другими ПТО в доступе и, при необходимости, – в выгрузке.

Мобильность. Обеспечение работы с любыми мобильными коммуникациями, в том числе с использованием мобильных приложений.

Кроссплатформенность. Способность программного обеспечения (ПО) работать с несколькими операционными системами или аппаратными платформами.

Коммуникабельность. Возможность обмена информацией с другими системами открытых архивов или системами автоматизации библиотек на физическом или логическом уровне.

Адаптивность. Лёгкая адаптация к любым возникшим функциональным и технологическим изменениям.

Единое ПО и интегрируемость с другими ПО внутри системы и снаружи. Обеспечение совместной работы ПТО и аппаратного комплекса других генераторов или владельцев информации в рамках одной системы связи.

Источники формирования контента системы ЕОАИ

ГПНТБ России имеет большой опыт по созданию, обработке, хранению и предоставлению читателям электронных ресурсов [2, 3]. Расширение количественного состава источников контента, создание новых пользовательских сервисов для повышения качества доступа к ресурсам ЕОАИ являются нашими приоритетными задачами [4].

Источниками для формирования контента системы станут:

электронная библиотека редкой книги, содержащая оцифрованные издания из фондов ГПНТБ России;

тематические полнотекстовые коллекции, созданные и создаваемые библиотекой;

авторефераты и полные тексты диссертаций защитившихся через базовую кафедру при МГИК или МГЛУ или других авторов по договорам безвозмездной передачи;

полные тексты статей журнала «Научные и технические библиотеки», ежегодного межведомственного сборника трудов ГПНТБ России, трудов Крымского форума, конференции «ЛИБКОН», конференции «БиблиоПитер», школ НАББ, других научно-образовательных мероприятий ГПНТБ России;

монографии, учебники, учебные и методические пособия, учебные программы и другие издания, которые подготовили сотрудники ГПНТБ России в рамках служебных заданий и в инициативном порядке (по согласованию) с аффилиацией Московского государственного института культуры и Московского государственного лингвистического университета;

монографии, учебники, исследовательские отчёты, обзоры и др., авторы которых – отечественные или зарубежные специалисты, передающие эти произведения в систему ЕОАИ на условиях дарения или по свободной лицензии;

материалы отчётов, обзоров и исследований, выполненных в ГПНТБ России, за исключением конфиденциальных или закрытых тем;

полнотекстовые издания, переданные другими владельцами (генераторами), например, полные тексты в ЭКБСОН, «Научный архив» и др. (по согласованию с правообладателями);

видеоархив и комплекс видеоматериалов, создаваемых в ходе научно-образовательных мероприятий ГПНТБ России;

специальные коллекции оцифрованных материалов, создаваемые в ГПНТБ России планово или по инициативе сотрудников;

другие полнотекстовые материалы, переданные на законных основаниях с правом их обнародования.

Принципы взаимодействия системы ЕОАИ с другими информационными системами

Принцип совместимости, подразумевающий построение открытой ИС, ориентированной на максимальное использование стандартов программного, технического и иного обеспечения, является одним из базовых принципов создания единого Открытого архива информации. Основные используемые протоколы и форматы представлены на рис. 1.

Протоколы обмена данными	Форматы метаданных	Форматы представления данных
OAI-PMH SRU/SRW Z39.50 SOAP	MARC (MARC21, RUSMARC) DUBLIN CORE (DC) MODS	XML ISO 2709

Рис. 1. Протоколы и форматы, используемые в ЕОАИ

Открытые архивы – ИС, содержащие массивы научных публикаций и информационных объектов, доступных в интернете, для которых поддерживается репозиторий с набором описывающих их метаданных. Обычно они работают по протоколу *Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)* [5].

В мире, по данным системы *OpenDOAR* [6], на 8 окт. 2020 г. насчитывается 5 558 сетевых архивов открытого доступа (репозиториев). Система собирает и предоставляет информацию исключительно о сайтах, полностью реализующих концепцию открытого доступа к полным текстам документов, нужных вузовским учёным. По состоянию на 8 окт. 2020 г. в базе данных *OpenDOAR* зарегистрирован 151 репозиторий русскоязычных документов.

Использование протокола *OAI-PMH* в открытых архивах даёт определённые преимущества для распространения и продвижения информации [7]. Так, кроме возможности обмениваться метаданными

в рамках инициативы открытых архивов, порталы, поддерживающие протокол *OAI-PMH*, имеют преимущества при индексировании поисковыми машинами интернета, такими как *Google* и *Yandex*.

Протокол *OAI-PMH* удобен и при взаимодействии с интеграторами, такими, как *EBSCO Discovery*, *Summon*, *ExLibris Primo*. Кроме того, он используется в базовом варианте системы передачи в НЭБ полных текстов и метаданных оцифрованных документов, подготовленных библиотеками – участниками проекта. Различие между форматами хранения препятствует обмену данными с системами автоматизации библиотек (САБ). Его преодоление могло бы позволить быстрее и эффективнее формировать контент цифрового архива [8].

Функциональная схема ЕОАИ

Общая функциональная схема ЕОАИ приведена на рис. 2. Для выполнения информационных ресурсов предусмотрены следующие варианты [9]: загрузка электронных ресурсов, подготовленных в ГПНТБ России; загрузка ранее подготовленных и размещённых коллекций полнотекстовых документов; сбор метаданных от *OAI*-источников в автоматическом режиме; самостоятельная загрузка документов авторами; загрузка ресурсов, полученных от авторов в полуавтоматическом режиме.

Ресурсы, созданные в библиотеке, загружаются при участии редакторов подготовки электронных ресурсов и библиографов.

Для загрузки авторских документов используется форма на веб-сервере. Данные, полученные от автора, обрабатываются с использованием скриптов, после чего полные тексты документов передаются на сервер хранения полных текстов. Полученные метаданные обрабатываются конвертором, а затем передаются в модуль загрузчика библиографических данных. На этом этапе их контролирует и, в случае необходимости, редактирует каталогизатор. Подготовленные таким образом метаданные загружаются на сервер хранения метаданных.

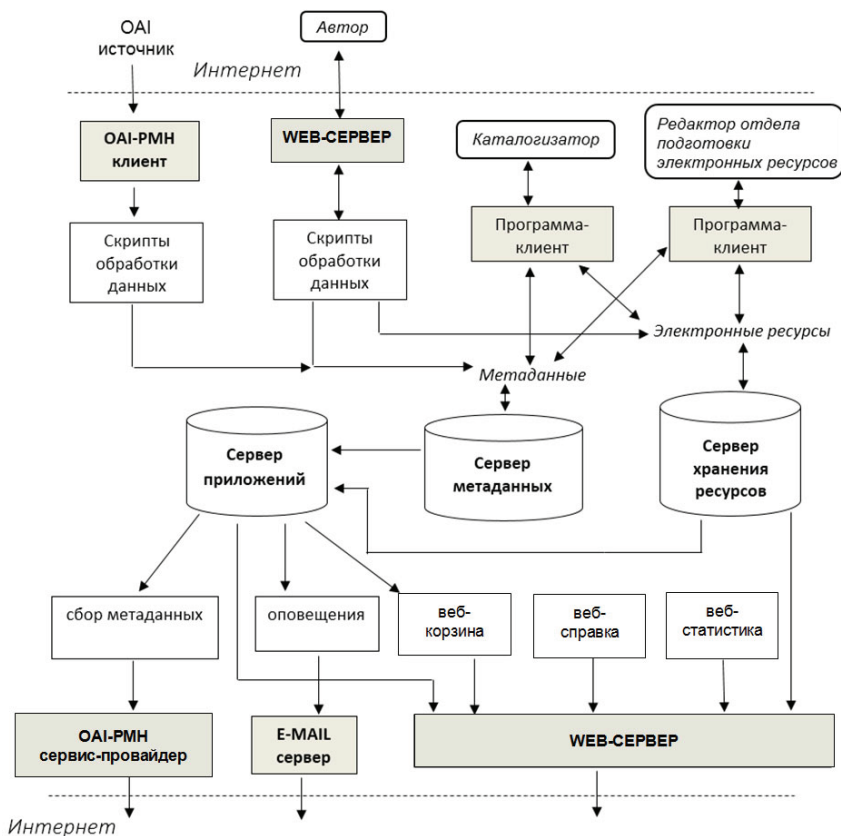


Рис. 2. Общая функциональная схема ЕОАИ ГПНТБ России

Метаданные, полученные в автоматическом режиме, также обрабатываются с помощью конвертора библиографических записей (БЗ), а затем поступают в модуль загрузчика БЗ. На этом этапе они проверяются и могут быть отредактированы каталогизатором, который использует для этого специальный веб-интерфейс.

Обслуживание запросов пользователей и взаимодействие с сервис-провайдерами открытых архивов осуществляет сервер приложений. Он получает метаданные от сервера метаданных и использует их

для выполнения следующих функциональных задач: передачи метаданных открытого архива ГПНТБ России сервис-провайдеру *OAI-PMH*; обслуживания поисковых запросов пользователей, поступающих через веб-сервер; оповещения пользователей о новых поступлениях с учётом персонального профиля (по электронной почте, через *RSS*-ленту новостей или личный кабинет читателя).

Программно-техническая реализация сервера приложений и сервера метаданных

Планируется, что основным ПО при реализации ЕОАИ будет САБ ИРБИС64, широко распространённая в библиотеках России, включая значительную часть вузовских библиотек. Новый функционал ИРБИС64+, обусловленный требованиями электронной библиотеки, поддерживает ряд специальных функций [10], а именно:

полнотекстовый релевантный поиск на основе автоматического разбиения текстов на страницы и их пословной индексации;

представление результатов поиска в порядке убывания релевантности документов, а в рамках одного документа – в порядке убывания релевантности найденных страниц;

постраничный просмотр полных текстов с обеспечением маркировки найденных слов запроса и навигации по релевантным страницам, оглавлению и личным закладкам.

Для передачи данных во внешние ИС ИРБИС64+ имеет следующие возможности: взаимодействие по протоколу *Z39.50*, поддержка протокола *SRU/SRW*, работа в режиме веб-сервиса. Для передачи БЗ и полных текстов документов в НЭБ используется провайдер данных, взаимодействующий с технологической платформой НЭБ. Это позволяет поддерживать обмен данными с использованием протокола *OAI-PMH*. Однако существующее программное решение имеет ряд ограничений, касающихся форматов передачи данных, технологических особенностей установки ПО и др.

Чтобы обеспечить стандартный обмен данными между участниками взаимодействия по протоколу *OAI-PMH* и ИРБИС64+, был разработан специальный провайдер данных. Он даёт возможность экспортировать метаданные из САБ ИРБИС64+ во внешние репозитории (открытые архивы), например, *DSpace*, что позволяет рассматривать САБ ИРБИС64+ как открытый архив.

Следует отметить, что функции поддержки полнотекстового поиска в САБ ИРБИС реализуются на уровне сервера (сервер ИРБИС или *RHP*-модуль) и не привязаны к варианту поставки системы.

Внедрение разрабатываемого варианта системы ЕОАИ позволит в автоматическом режиме передавать метаданные ресурсов, хранящихся в системе, участникам взаимодействия открытых архивов, что обеспечит их видимость на порталах, использующих протокол *OAI-PMH* для сбора метаданных. При этом сохранится гибкость представления электронного ресурса пользователю, поскольку вариант его отображения будет определяться на уровне администрирования системы. Так, для текстовых материалов возможны варианты: передача *PDF*-файла, использование веб-версии программы постраничного показа документа, отсылка к просмотру БЗ. Мультимедийные материалы могут размещаться в файловом или облачном хранилище. В системе ЕОАИ будут храниться их метаданные.

Реализация системы ЕОАИ будет важным этапом на пути к созданию типовой модели электронной библиотеки, функционирующей как участник взаимодействия в системе открытых архивов, что упростит технологию обмена данными в информационных системах библиотек, научно-информационных подразделений научных организаций, университетов, общеобразовательных учреждений и обеспечит поддержку поиска информации через единую точку доступа на веб-порталах.

СПИСОК ИСТОЧНИКОВ

1. **Шрайберг Я. Л.** Принципы построения автоматизированных библиотечно-информационных систем и сетей : дис. в виде науч. докл. ... д-ра техн. наук. : 05.25.05, 05.25.03 / Я. Л. Шрайберг. – Москва, 1999. – 39 с.
2. **Гончаров М. В., Колосов К. А.** Электронная библиотека ГПНТБ России: динамика пополнения, технологии, ресурсы / М. В. Гончаров, К. А. Колосов // Науч. и техн. б-ки. – 2018. – № 12. – С. 34–41.
3. **Гончаров М. В., Михайленко И. И.** Интеграция информационных ресурсов ГПНТБ России в рамках Системы открытого архива / М. В. Гончаров, И. И. Михайленко // Там же. – 2018. – № 4. – С. 5–13.
4. **Гончаров М. В., Колосов К. А.** Разработка системы открытого архива ГПНТБ России / М. В. Гончаров, К. А. Колосов // Там же. – 2018. – № 12. – С. 42–48.
5. **Земсков А. И.** Открытый доступ: роль библиотек / А. И. Земсков // Там же. – 2016. – № 6. – С. 41–61.
6. **The Directory of Open Access Repositories** : сайт. – URL: <http://www.openoar.org/index.htm> (дата обращения: 08.10.2020). – URL: свободный.
7. **Вахрушев М. В.** Научная библиотека вуза в роли открытого архива / М. В. Вахрушев // Науч. и техн. б-ки. – 2018. – № 4. – С. 14–22.
8. **Ковязина Е. В.** Цифровой архив: обмен данными с системой автоматизации библиотеки // Информ. технологии, компьютер. системы и издат. продукция для б-к. – 2019. – С. 15–19.
9. **Открытый доступ: история, современное состояние и путь к открытой науке** : моногр. / Вахрушев М. В., Гончаров М. В., Засурский И. И. [и др.] ; под общ. и науч. ред. д-ра техн. наук, проф., чл.-корр. Рос. акад. образования Я. Л. Шрайберга. – Санкт-Петербург [и др.] : Лань, 2020. – 165, [1] с. : ил. – ISBN 978-5-8114-5034-3.
10. **Колосов К. А.** Развитие решений и новые возможности системы автоматизации библиотек ИРБИС64: электронная библиотека «ИРБИС64+» и её использование как открытого архива, взаимодействие с ЕПГУ, поддержка технологии веб-сервисов / К. А. Колосов, А. И. Бродовский // Электронный век культуры : сб. тез. докл. XVI Ежегод. междунар. науч.-практ. конф. (Пермь, 18–22 сент.). – Москва, 2017. – С. 39–45.
11. **Alipour-Hafezi M. et al.** A model for creating a virtual library in Iranian universities of medical sciences // The Electronic Library. – 2019. – URL: <https://doi.org/10.1108/EL-01-2019-0017>.

12. **Ahammad N.** Open source digital library on open educational resources // The Electronic Library. – 2019. – URL: <https://doi.org/10.1108/EL-11-2018-0225>.

REFERENCES

1. **Shrayberg Ya. L.** Printsipy postroeniya avtomatizirovannykh bibliotечно-информационных систем i setey : dis. v vide nauchn. dokl. ... d-ra tehn. nauk. : 05.25.05, 05.25.03 / Ya. L. Shrayberg. – Moskva, 1999. – 39 s.

2. **Goncharov M. V.** Elektronnaya biblioteka GPNTB Rossii: dinamika popolneniya, tehnologii, resursy / M. V. Goncharov, K. A. Kolosov // Nauch. i tehn. b-ki. – 2018. – № 12. – S. 34–41.

3. **Goncharov M. V.** Integratsiya informatsionnykh resursov GPNTB Rossii v ramkah Sistemy otkrytogo arhiva / M. V. Goncharov, I. I. Mihaylenko // Nauch. i tehn. b-ki. – 2018. – № 4. – S. 5–13.

4. **Goncharov M. V.** Razrabotka sistemy otkrytogo arhiva GPNTB Rossii / M. V. Goncharov, K. A. Kolosov // Nauch. i tehn. b-ki. – 2018. – № 12. – S. 42–48.

5. **Zemskov A. I.** Otkrytyy dostup: rol bibliotek / A. I. Zemskov // Nauch. i tehn. b-ki. – 2016. – № 6. – S. 41–61.

6. **The Directory of Open Access Repositories** : сайт. – URL: <http://www.openoar.org/index.htm>.

7. **Vahrushev M. V.** Nauchnaya biblioteka vuza v roli otkrytogo arhiva / M. V. Vahrushev // Nauch. i tehn. b-ki. – 2018. – № 4. – S. 14–22.

8. **Kovyazina E. V.** Tsifrovoy arhiv: obmen dannymi s sistemoy avtomatizatsii biblioteki / E. V. Kovyazina // Informatsionnye tehnologii, kompyuternye sistemy i izdatelskaya produktsiya dlya bibliotek. – 2019. – S. 15–19.

9. **Otkrytyy** dostup: istoriya, sovremennoe sostoyanie i put k otkrytoy nauke: monografiya / Vahrushev M. V., Goncharov M. V., Zasurskiy I. I. [i dr.] ; pod obshch. i nauch. red. d-ra tehn. nauk, professora, chlena-korrespondenta Rossiyskoy akademii obrazovaniya Ya. L. Shrayberga. – Sankt-Peterburg [i dr.] : Lan, 2020. – 165, [1] s. : il. – ISBN 978-5-8114-5034-3.

10. **Kolosov K. A.** Razvitie resheniy i novye vozmozhnosti sistemy avtomatizatsii bibliotek IRBIS64: elektronnaya biblioteka "IRBIS64+" i ee ispolzovanie kak otkrytogo arhiva, vzaimodeystvie s EPGU, podderzhka tehnologii veb-servisov / K. A. Kolosov, A. I. Brodovskiy // Elektronnyy vek kultury : sb. tezisev dokladov XVI Ezhegodnoy mezhdunar. nauch.-prakt. konf. (Perm, 18–22 sent.). – Moskva, 2017. – S. 39–45.

11. **Alipour-Hafezi M. et al.** A model for creating a virtual library in Iranian universities of medical sciences // The Electronic Library. – 2019. – URL: <https://doi.org/10.1108/EL-01-2019-0017>.

12. **Ahammad N.** Open source digital library on open educational resources // The Electronic Library. – 2019. – URL: <https://doi.org/10.1108/EL-11-2018-0225>.

Информация об авторах / Information about the authors

Шрайберг Яков Леонидович – доктор техн. наук, профессор, член-корреспондент Российской академии образования, научный руководитель ГПНТБ России, главный редактор журнала «Научные и технические библиотеки», заведующий кафедрой электронных библиотек и наукометрических исследований Московского государственного лингвистического университета, заслуженный работник культуры РФ, Москва, Россия
gpntb@gpntb.ru

Гончаров Михаил Владимирович – канд. техн. наук, доцент, ведущий научный сотрудник, руководитель группы перспективных исследований и аналитического прогнозирования ГПНТБ России, доцент Московского государственного лингвистического университета, Москва, Россия
goncharov@gpntb.ru

Yakov L. Shrayberg – Dr. Sc. (Engineering), Professor; Corresponding Member of Russian Academy of Education; Academic Director, Russian National Public Library for Science and Technology; Editor-in-Chief, “Scientific and Technical Libraries” Journal; Head, Department for Electronic Libraries and Scientometric Studies, Moscow State Linguistic University, Moscow, Russia
gpntb@gpntb.ru

Mikhail V. Goncharov – Cand. Sc. (Engineering), Associate Professor, Leading Researcher, Head, Group for Perspective Research and Analytic Forecasting, Russian National Public Library for Science and Technology; Associate Professor, Moscow State Linguistic University, Moscow, Russia
goncharov@gpntb.ru

Колосов Кирилл Анатольевич – канд. техн. наук, ведущий научный сотрудник ГПНТБ России, доцент Московского государственного лингвистического университета, Москва, Россия
kolosov@gpntb.ru

Kirill A. Kolosov – Cand. Sc. (Engineering), Leading Researcher, Russian National Public Library for Science and Technology, Associate Professor, Moscow State Linguistic University, Moscow, Russia
kolosov@gpntb.ru