

О развитии электронной библиотеки “Научное наследие России” как составляющей Единого цифрового пространства научных знаний

Н.Е. Каленов, К.П. Погорелко, А.Н. Сотников

*Межведомственный суперкомпьютерный центр (МСЦ) РАН – филиал ФГУ ФНЦ НИИСИ РАН,
Москва, Россия*

Поступила в редколлегию 30.08.2023

Аннотация—В работе рассматривается новая версия электронной библиотеки “Научное наследие России” (ЭБ ННР), разработанная с учетом перспектив ее интеграции в Единое цифровое пространство научных знаний (ЕЦПНЗ). Задача создания ЭБ ННР как общедоступного ресурса была сформулирована в рамках целевой программы президиума РАН в 2007 году. Разработка структуры библиотеки и ее программной оболочки осуществлялась специалистами ряда академических организаций во главе с Межведомственным суперкомпьютерным центром (МСЦ РАН). Библиотека строится по принципу связанных данных и отражает информацию об ученых, их публикациях (включая оцифрованные полные тексты), связанные с учеными архивные материалы и музейные предметы. Начиная с 2010 г., ЭБ ННР представлена в открытом доступе в интернет по адресу <http://e-heritage.ru>. С появлением в 2019 году концепции и основных принципов создания Единого цифрового пространства научных знаний (ЕЦПНЗ) как системы связанных данных, объединяющей разнородные научные ресурсы, возникла задача реорганизации ЭБ ННР (как ресурса, содержащего важнейшие элементы ЕЦПНЗ) в соответствии с подходами, заложенными в ЕЦПНЗ. К 2020 г. были полностью переработаны и введены в эксплуатацию технология формирования ЭБ ННР, ее внутренняя и внешняя программные оболочки, поисковые алгоритмы. В статье описывается структура ЭБ ННР, включающая два WEB-ориентированных блока – административный и пользовательский. Административный блок, доступный авторизованным пользователям, обеспечивает ввод, редактирование, контроль качества создаваемых ресурсов; пользовательский блок, представленный в открытом доступе, предоставляет широкие возможности поиска ресурсов с использованием булевой логики; он обеспечивает наглядную визуализацию и развитую навигацию по связанным ресурсам. В статье приводятся примеры обработки поисковых запросов и навигации по ресурсам ЭБ ННР.

КЛЮЧЕВЫЕ СЛОВА: пространство научных знаний, информационные ресурсы, научное наследие, цифровые библиотеки, сетевые технологии, связанные данные

DOI: 10.53921/18195822_2022_22_3_155

1. ВВЕДЕНИЕ

Электронная библиотека “Научное наследие России” (далее ЭБ ННР) начала разрабатываться в рамках целевой программы Президиума РАН в 2007 году. Целью проекта было формирование развернутой информации об ученых, внесших существенный вклад в развитие российской науки, и предоставление ее широкому кругу отечественных и зарубежных пользователей. В качестве информационной базы ЭБ ННР выступали биографические сведения об ученых, их библиография, оцифрованные основные публикации, связанные с ними архивные и музейные материалы.

Разработкой принципов организации и программной оболочки ЭБ ННР занимался коллектив специалистов ряда академических организаций во главе с МСЦ РАН [1-4]. В формировании контента ЭБ ННР участвовало более 20 организаций РАН, включая центральные библиотеки, Архив РАН, ряд музеев. В 2010 году ЭБ ННР была введена в промышленную эксплуатацию, и представлена в открытом доступе в интернет по адресу <http://e-heritage.ru/>.

ЭБ ННР успешно функционировала как элемент целевой научной программы РАН. После 2014 года, в связи с реорганизацией Академии наук, целевое финансирование было прекращено, и развитие библиотеки поддерживалось грантами РФФИ. К 2019 году в библиотеке была представлена информация о более чем 5500 ученых, начиная с 18 века и кончая первой третью 20 века; более 26 000 оцифрованных публикаций, изданных в этот период; связанные с учеными архивные документы и музейные предметы (в том числе, 3D-модели). ЭБ ННР является достаточно востребованным ресурсом со стороны как российских, так и зарубежных пользователей [5, 6]. В 2019 году было зарегистрировано около 140 000 обращений к полным текстам публикаций, размещенным в библиотеке.

С появлением в 2019 году концепции и основных принципов создания Единого цифрового пространства научных знаний (ЕЦПНЗ) как системы связанных данных, объединяющей разнородные научные ресурсы [7-9], возникла задача разработки автоматизированной системы формирования и поддержки ЕЦПНЗ, к решению которой приступили специалисты МСЦ РАН. К середине 2020 г. была разработана первая очередь автоматизированной системы поддержки пространства научных знаний (АС ППНЗ), и ЭБ ННР была поэтапно, без перерывов в доступе, переведена под ее управление [10–12]. Новая версия ЭБ ННР размещена на сервере МСЦ РАН по адресу <http://heritage1.jssc.ru>. Для удобства пользователей, постоянно работающих с ЭБ ННР, обеспечена автоматическая переадресация со старого адреса (<http://e-heritage.ru/>).

Ниже приводится описание реализации первой очереди АС ППНЗ на примере современной версии ЭБ ННР.

АС ППНЗ оперирует с разнородными объектами и их связями и реализована в виде комплекса двух WEB-ориентированных блоков – пользовательского и административного.

Административный блок обеспечивает ввод, технологический контроль и редактирование информации, включая возможность формирования отдельных коллекций разнородных ресурсов, связанных некоторыми признаками; доступ к функциям блока осуществляется авторизованными пользователями, каждый из которых обладает теми или иными правами, определяемыми пользователем со статусом “администратор”. Пользовательский блок обеспечивает многоаспектный поиск информации и навигацию по связанным ресурсам; доступ к нему имеет любой пользователь.

2. ИНФОРМАЦИОННАЯ БАЗА СИСТЕМЫ

Информационная база АС ППНЗ может включать различные виды объектов и их попарных связей, которые могут быть добавлены по мере развития системы. В настоящее время она оперирует с видами и атрибутами объектов ЭБ ННР, определенными на предыдущих стадиях развития библиотеки. Среди таких объектов:

- персоны;
- публикации;
- архивные материалы;
- музейные предметы;
- изображения.

Персоны представлены следующими атрибутами, каждый из которых является поисковым элементом:

- фамилия;
- имя;
- отчество;
- дата рождения;
- место рождения;
- дата смерти;
- область научных интересов (выбирается из списка научных направлений)
- биография (поиск осуществляется по полному тексту).

Атрибуты публикаций включают следующие поисковые элементы:

- название (поиск по сочетанию слов);
- год издания;
- издательство;
- язык публикации (выбирается из списка);
- тематика (рубрика ГРНТИ);
- оглавление (поиск по тексту);
- тип публикации (выбирается из списка).

Связи между персонами и публикациями могут принимать значения “автор”, “редактор” и “о нем”. Соответственно, осуществляется поиск по заданной связи.

Архивные материалы отражены в действующей версии системы в виде связей с персонами (указываются ссылки от персон на соответствующие им материалы, отраженные в Архиве РАН).

Музейные предметы имеют следующие поисковые атрибуты:

- название;
- описание (поиск по тексту);
- год поступления в музей.

Связи между музейными предметами и персонами могут принимать значения “автор” и “автор сбора” (для естественнонаучных коллекций).

Изображения в действующей версии представлены оцифрованными фотографиями или портретами персон и связаны с ними соответствующими ссылками.

Метаданные публикаций включают ссылки на их нераспознанные оцифрованные тексты.

3. ПОЛЬЗОВАТЕЛЬСКИЙ БЛОК ЭБ ННР

На рис. 1 представлена начальная страница пользовательского блока ЭБ ННР. На ней динамически отражаются текущие количественные характеристики контента ЭБ ННР и поддерживаются в актуальном состоянии разделы “События” и “Памятные даты”. Первые служат “рекламой” научных мероприятий, связанных с контентом ЭБ ННР, вторые отражают, в первую очередь, юбилейные даты, связанные с объектами, отраженными в ЭБ ННР, и содержат ссылки на эти объекты.

Вкладка “Поиск” позволяет выбрать вид объекта, который требуется найти (персоны, публикации, музейные предметы, коллекции), и затем сформулировать запрос на поиск. Запрос может включать практически неограниченное число условий, связанных булевыми операторами. Условия отбора элементов зависят от вида атрибута. Например, для персон фамилия, имя и отчество допускают задание условий “содержит”, “начинается с”, “равно” (по умолчанию – “содержит”). Года рождения и смерти допускают задание условий “равно”, “больше”, “меньше”. Область деятельности предполагает выбор из списка основных разделов науки. Фрагмент

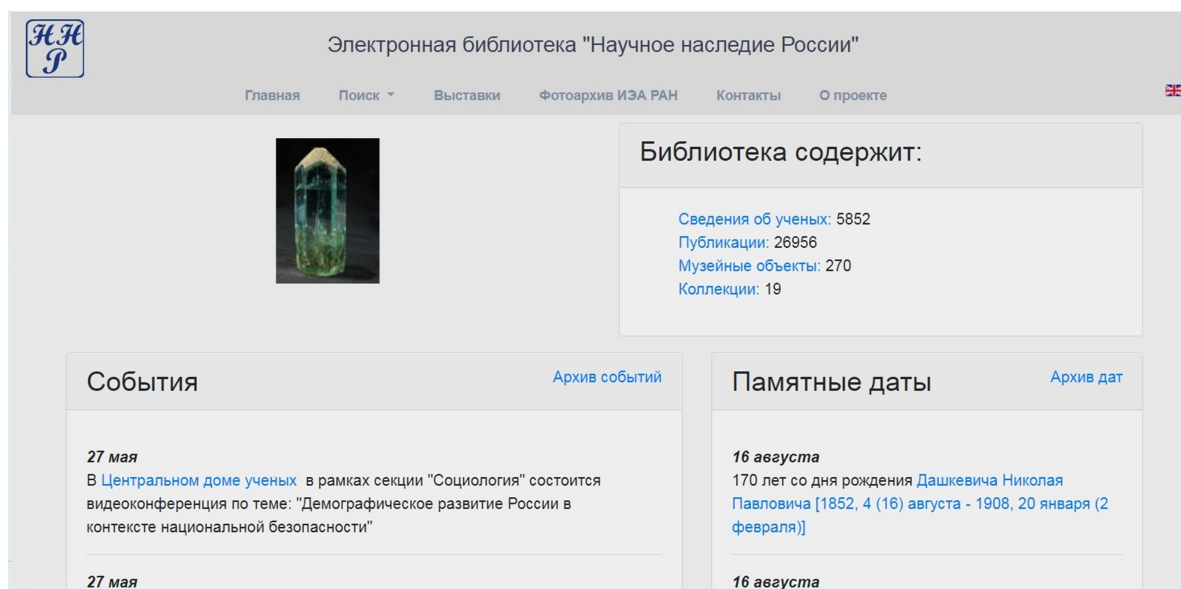


Рис. 1. Начальная страница сайта ЭБ ННР.

биографии подразумевает полнотекстовый поиск по представленной в ЭБ ННР биографии. Результирующий список персон может быть отсортирован по фамилии, году рождения или году смерти и включать заданное количество элементов, выводимых на страницу.

На рис. 2 представлен запрос на поиск персон, научная деятельность которых связана с науками о Земле или палеонтологией, год рождения – меньше 1900, а в биографии присутствует слово “Волга”. Результат поиска, отсортированный по году рождения персон, включает сведения о 46 ученых.

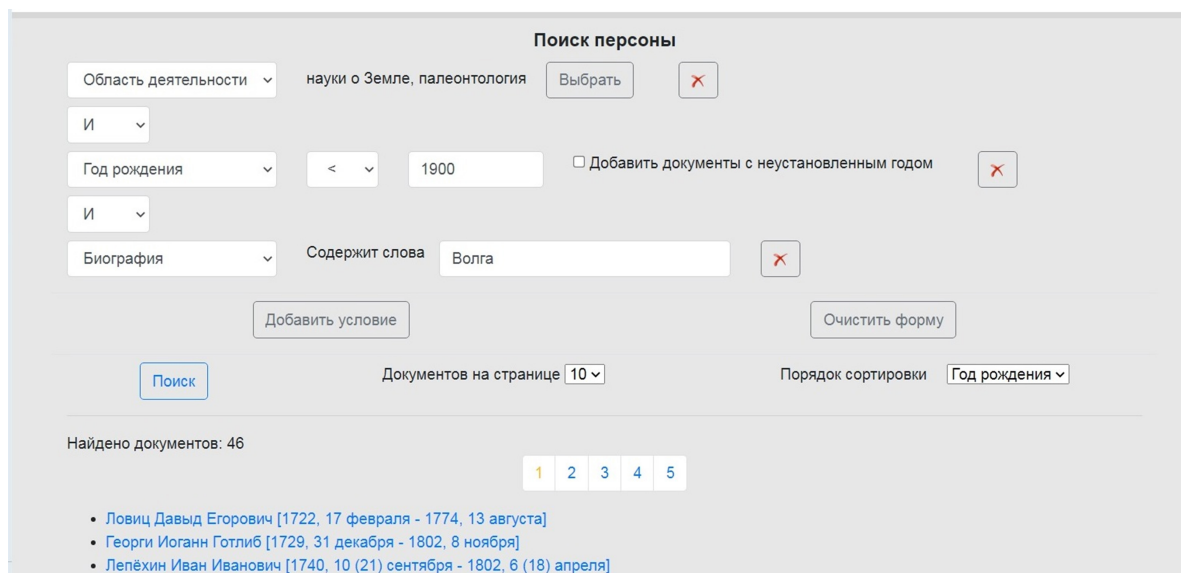


Рис. 2. Пример запроса на поиск персон.

Каждый элемент результирующего списка является активной ссылкой, открывающей страницу, относящуюся к данному ученому. На этой странице отражаются портреты ученых, информация о годе и месте рождения и смерти, раскрывающиеся элементы “Биографическая

справка”, “Архивная информация” и “Публикации”. Пример такой страницы, относящейся к И.М. Губкину, приведен на рис. 3.

Рис. 3. Пример страницы с информацией о персоне.

Если раскрыть “биографическую справку”, легко увидеть текст, относящийся к Волге:

“Большую теоретическую и практическую ценность имели работы Губкина о создании нефтяной базы между Волгой и Уралом”.

Ссылка под названием “Архивная информация” приводит на страницу Архива РАН, относящуюся к И.М. Губкину (рис. 4).

Рис. 4. Пример архивной информации.

Раскрыв раздел “публикации” (рис. 3), получим список оцифрованных материалов И.М. Губкина, отраженных в ЭБ ННР:

- Губкин, И.М. Вторая нефтяная база на востоке СССР и Урало-Эмбенский нефтеносный район. - М. ; Л. : ОНТИ, Гл. ред. горно-топл. лит., 1936. 37, [2] с., 3 л. черт., карт. : ил.
- Иван Михайлович Губкин (1871–1939) / Сектор сети спец. б-к АН СССР, Всесоюз. кн. палата ; редкол.: М.П. Востокова, О.В. Исакова (отв. ред.), В.М. Корнилов, Н.Т. Толкачев. - М. : изд-во Всесоюз. кн. палаты, 1941. - 39, [1] с., 1 л. портр.
- Губкин, И.М. К вопросу о генезисе нефтяных месторождений Северного Кавказа // Международный геологический конгресс. XVII сессия. СССР, 1937 : Труды. - М., 1940. - Т. 4.- С. 9-22.
- Губкин, И.М. Майкопский нефтеносный район. Нефтяно-Ширванская нефтеносная площадь. - СПб., 1912. - VI, 169 с. - (Тр. Геол. ком. Нов. сер. ; вып. 78).
- Губкин, И.М. Мировые запасы нефти // Международный геологический конгресс. СССР, 1937. Труды 17 сессии. 1937 г. - Т. 1. - М. : ГОНТИ, 1939. - С. 167-179 : ил.

Каждый элемент этого списка является активной ссылкой, раскрывающей страницу данной публикации, содержащую развернутую информацию и ссылку на оцифрованный полный текст (рис. 5).

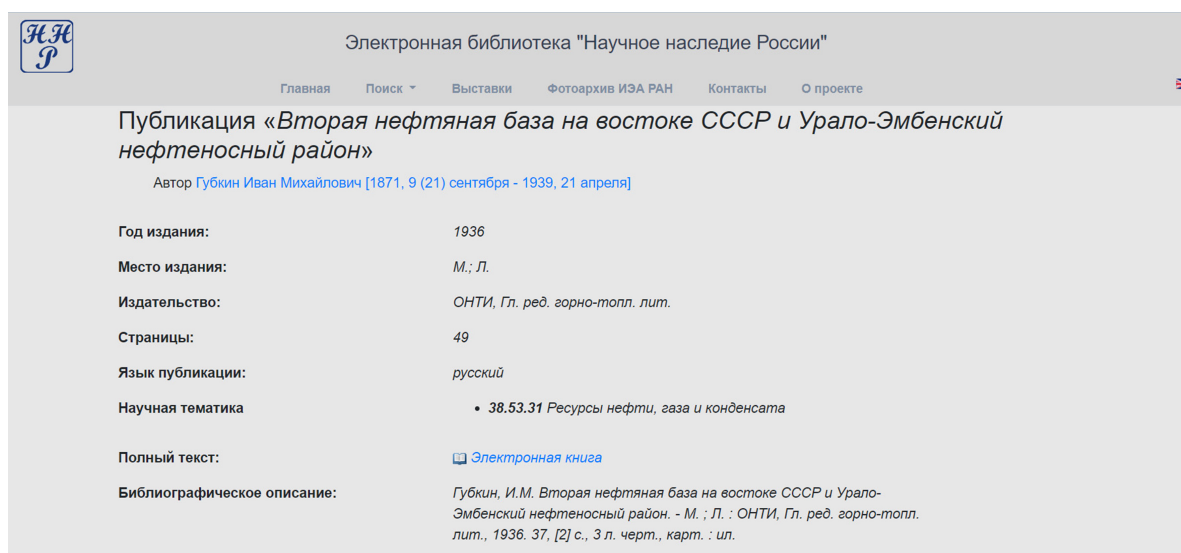


Рис. 5. Пример страницы информации о публикации.

Текст публикаций в ЭБ ННР не распознается, но реализована навигация по элементам оглавления. Переход по ссылке “Электронная книга” позволяет читать издание, увеличивая, поворачивая, при необходимости, страницу переходить на нужный раздел оглавления или требуемый номер страницы (рис. 6).

Поиск публикаций, так же, как и поиск персон, может включать различное число условий, связанных булевыми операторами. Поиск возможен по словосочетаниям из библиографического описания, году издания (с возможными условиями “равно”, “больше”, “меньше”), фамилии автора или редактора, виду публикации (выбирается из списка, включающего 15 вариантов), рубрикам ГРНТИ (выбираются с помощью реализованной навигации по полному рубрикатору), словосочетаниям из оглавлений.

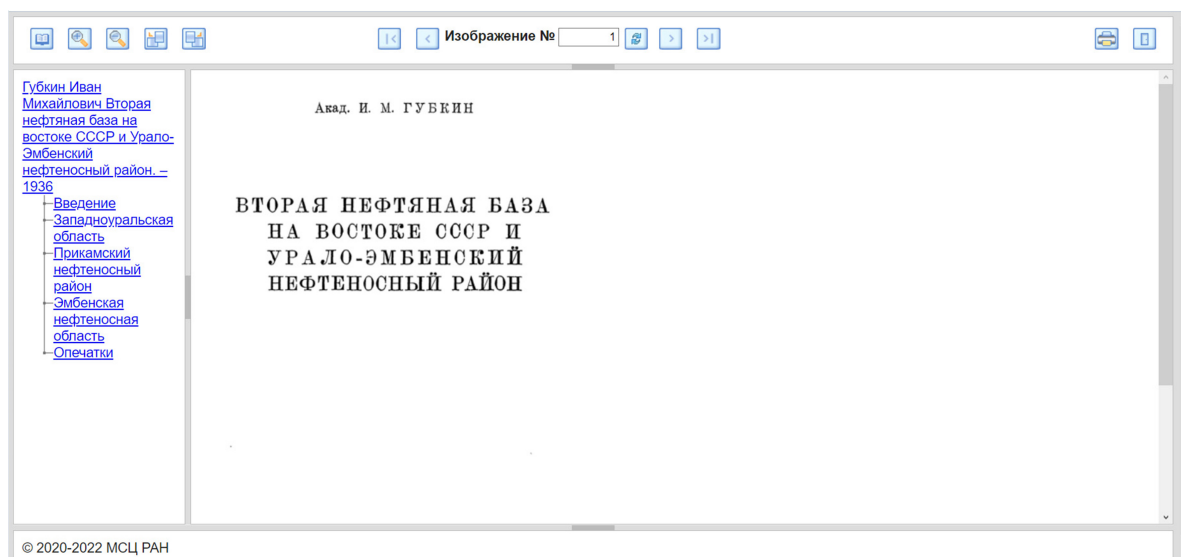


Рис. 6. Пример первой страницы просмотра текста публикации.

Результатом поиска является список публикаций, по каждому элементу которого можно перейти на страницу, аналогичную представленной на рис. 5, и далее - на просмотр полного текста.

Через интерфейс пользовательского блока доступны все элементы ЭБ ННР, прошедшие полный цикл обработки и контроля качества. Эти элементы получают статус “опубликовано”, который присваивается выпускающим редактором.

4. АДМИНИСТРАТИВНЫЙ БЛОК ЭБ ННР

Для работы с элементами, находящимися на стадии технологической обработки, а также для формирования блоков “События” и “Памятные даты” в ЭБ ННР, как уже было указано выше, существует административный блок. Вход в него осуществляется по логину и паролю, присваиваемым администратором. Начальная страница административного блока представлена на рис. 7. Пользователю сообщается о его правах (в примере на рис. 7 пользователь обладает всеми возможными правами работы с ЭБ ННР) и предлагается выбрать один из режимов работы – редактирование (с выбором видов редактируемых объектов), работа с письмами пользователей, формирование блока новостей, расширенный поиск данных.

Режим редактирования позволяет вводить новые записи и редактировать существующие и предоставляет оператору возможности выбора объекта по широкому кругу характеристик, в том числе связывающих объекты разных видов, в частности, персоны с музейными предметами и публикациями. На рис. 8 и 9 приведены поисковые формы блоков редактирования персон и публикаций. В этих формах значительная роль отведена отбору объектов по связям с другими объектами, а также по элементам, связанным с технологией, что позволяет проводить динамический анализ контента и оперативно контролировать ход наполнения ЭБ ННР.

После обработки запроса система представляет список активных ссылок на найденные объекты, перейдя по которым редактор может корректировать необходимые данные. После сохранения изменений объекты, имеющие статус “опубликовано”, отражаются в пользовательском блоке в скорректированном виде.

Одной из возможностей административного блока является включение в контент внешних ссылок, относящихся к тому или иному объекту. Внешняя ссылка характеризуется наименованием и URL. Аппарат внешних ссылок позволяет связывать объект ЕЦПНЗ (а в частном

Администрирование ННР

Пользователи Редактирование Письма Новости Поиск Пользователь *** Выйти

Система администрирования данных Электронной библиотеки "Научное наследие России"

Здравствуйте, Пользователь *****

Ваши права:

- Осуществлять поиск
- Вводить новые записи и редактировать свои записи
- Редактировать все записи
- Публиковать записи
- Управлять пользователями
- Редактировать новости

Изменить пароль

Рис. 7. Начальная страница административного блока.

Администрирование ННР

Пользователи Редактирование Письма Новости Поиск Пользователь *** Выйти

Администрирование персон

Начало фамилии

Год рождения

Фрагменты комментария

Год смерти

Есть биография

Количество фотографий

Слова из биографии

Внешних ссылок

Наличие открытых ошибок

Количество публикаций

Фрагменты ошибок в

Музейных объектов

Ввод завершен

Опубликовано

Автор записи

Сортировать по

Даты

Найти Очистить форму Новая персона

Рис. 8. Поисковая форма блока редактирования персон.

Администрирование ННР

Пользователи Редактирование Письма Новости Поиск Пользователь *** Выйти

Администрирование публикаций

Номер МСЦ

Количество авторов

Фрагменты названия

Количество муз. объектов

Фрагменты рус. названия

Количество коллекций

Фрагменты биб.описания

Год издания

Фрагменты комментария

Наличие открытых ошибок

Фрагменты доп. информации

Фрагменты ошибок в

Статус

Статус книги

Автор записи

Сортировать по

Даты

Найти Очистить форму Новая публикация

Рис. 9. Поисковая форма блока редактирования публикаций.

случае объект ЭБ ННР) с объектами, представленными в других сетевых научных ресурсах. К таким ресурсам относятся сайты, посвященные конкретным ученым, электронные библиотеки, базы данных и т.п. На момент подготовки данной публикации в текущей версии ЭБ ННР у 757 ученых имеются ссылки на связанные с ними внешние (по отношению к ЭБ ННР) ресурсы. (см. результат выполнения запроса, представленный на рис. 10). У 149 из них – более одной ссылки.

Администрирование персон

Начало фамилии		Год рождения	--- ▾	
Фрагменты комментария		Год смерти	--- ▾	
Есть биография	не важно ▾	Количество фотографий	--- ▾	
Слова из биографии		Внешних ссылок	> ▾	0
Наличие открытых ошибок	не важно ▾	Количество публикаций	--- ▾	
Фрагменты ошибок в	открытых ▾	Музейных объектов	--- ▾	
Ввод завершен	не важно ▾	Опубликовано	да ▾	
Автор записи	--- ▾	Сортировать по	Фамилия ▾	

Даты

Найдено 757

№	ФИО	Дата рождения	Дата смерти	Ошибки	Кто ввел	Статус
1	... Авакян Артавазд Аршакович	1907, 8 (21) июля	1966, 16 ноября	Нет	arran	☒
2	... Агаджанян Николай Александрович	1928, 28 января	2014, 27 декабря	Нет	haryb	☒

Рис. 10. Запрос на подсчет количества персон, имеющих ссылки на внешние ресурсы.

В настоящее время в рамках совместных работ МСЦ РАН и Росатома в ЭБ ННР и ЭБ “История Росатома” (<http://elib.biblioatom.ru>) введены взаимные ссылки на информацию о персонах, участвовавших в “Атомном проекте” СССР. Например, со страницы ЭБ ННР, отражающей информацию о А.Ф. Иоффе, (рис. 11), по первой из внешних ссылок можно перейти на страницу ЭБ “История Росатома”, содержащую полный текст автобиографии А.Ф. Иоффе (рис. 12), а по второй – на страницу, посвященную А.Ф.Иоффе (рис. 13).

Как видно из структуры поисковых форм, административный блок позволяет не только обеспечивать технологию формирования контента ЭБ ННР, но и проводить многоаспектный анализ данных, отраженных в библиотеке.

5. ЗАКЛЮЧЕНИЕ

Отработка новых решений в области развития АС ППНЗ проводится на сайте <http://cdssk.jssc.ru>. После их тестирования и доработки они переносятся на сайт ЭБ ННР. Ближайшими планами развития АС ППНЗ является реализация возможности добавления новых видов объектов и связей различного типа между ними с расширенными возможностями поиска и навигации по разнородным ресурсам и связям.

Работа в области развития ЭБ ННР как части единого цифрового пространства научных знаний проводится в МСЦ РАН в рамках государственного задания по теме FNEF-2022-0014.

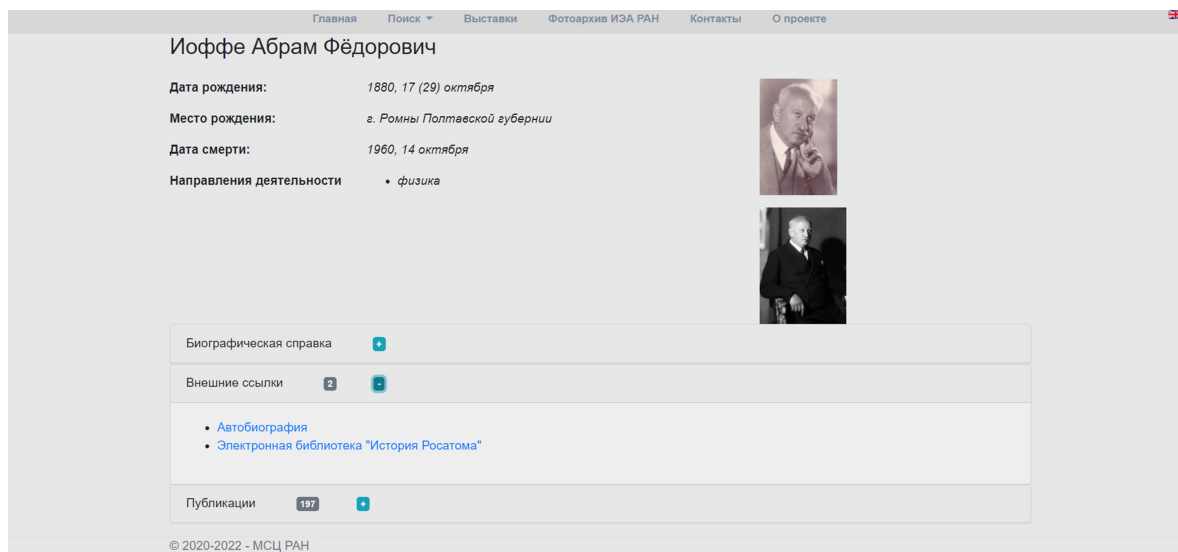


Рис. 11. Информация о А.Ф.Иоффе.

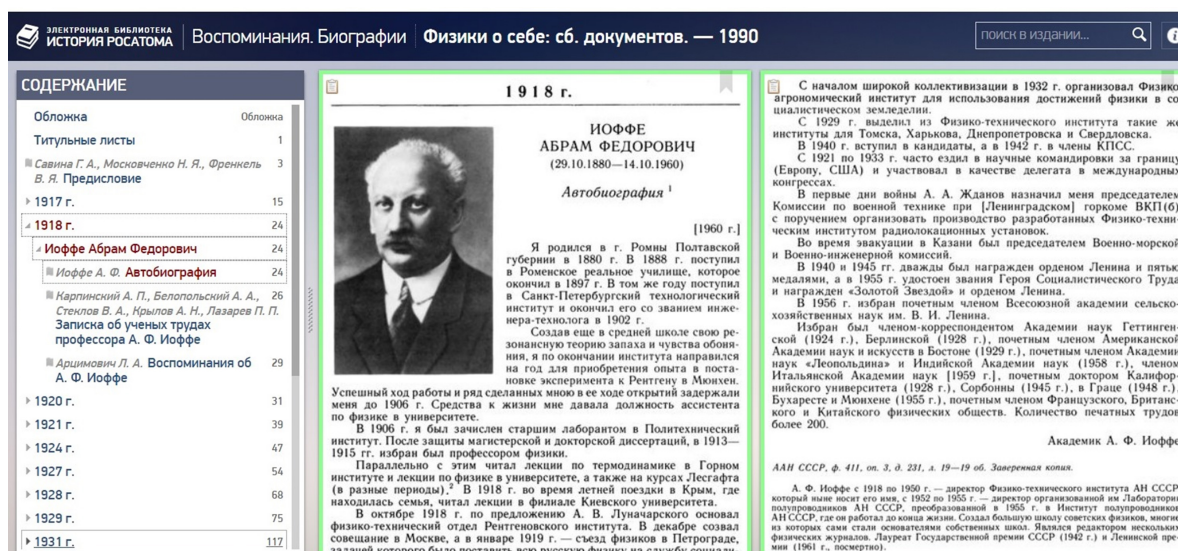


Рис. 12. Текст автобиографии А.Ф. Иоффе в ЭБ “История Росатома”.

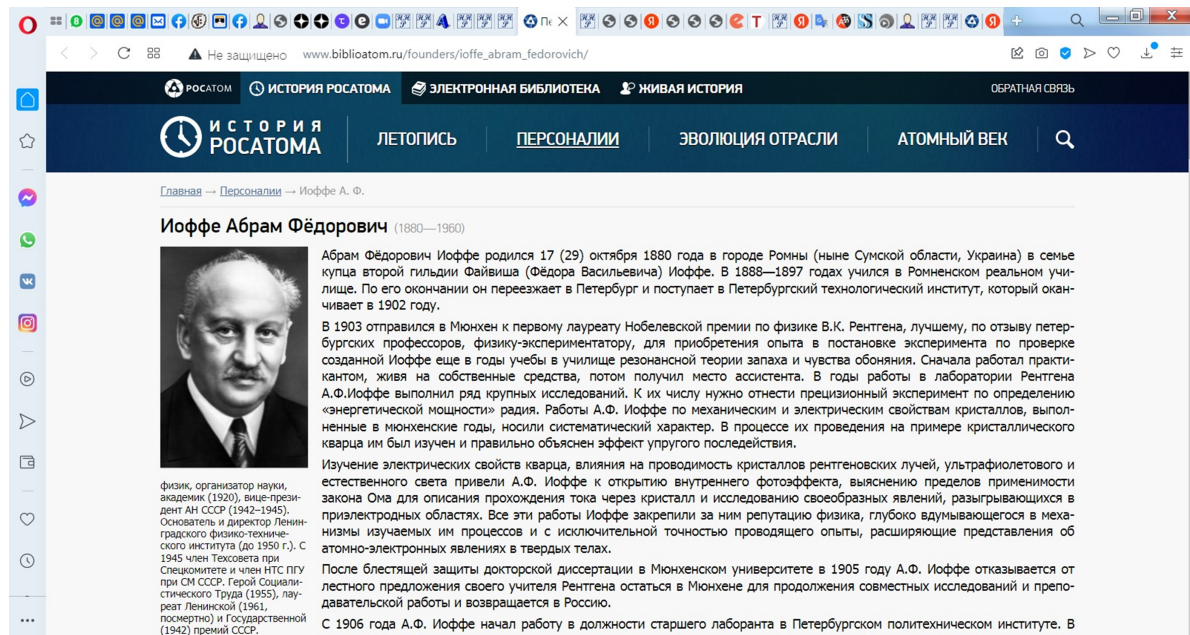


Рис. 13. Страница сайта «История Росатома», посвященная А.Ф. Иоффе.

СПИСОК ЛИТЕРАТУРЫ

1. Каленов Н.Е., Савин Г.И., Сотников А.Н. Электронная библиотека «Научное наследие России» // Информационные ресурсы России, 2009. № 2. С. 19–20.
2. Каленов Н.Е., Савин Г.И., Серебряков В.А., Сотников А.Н. Принципы построения и формирования электронной библиотеки «Научное наследие России» // Программные продукты и системы, 2012. Т. 4. № 100. С. 30–40.
3. Забровская И.Е., Кириллов С.А., Шабанов Б.М. Структура и фонды электронной библиотеки «Научное наследие России» // Информационное обеспечение науки: новые технологии : сб. науч. тр. М. : БЕН РАН, 2013. С. 20–31.
4. Якшин М.М. Платформа SciRus – основа технологического комплекса электронной библиотеки «Научное наследие России» // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: XVI Всероссийская научная конференция RCDL-2014 Дубна, 2014. С. 362–368.
5. Погорелко К.П. Анализ востребованности электронной библиотеки «Научное наследие России» // Информационное обеспечение науки: новые технологии: Сборник научных трудов - М.: БЕН РАН, 2015. - С. 191–199.
6. Погорелко К.П. Динамика использования электронной библиотеки «Научное наследие России» // Информационное обеспечение науки: новые технологии: Сборник научных трудов.- М.: БЕН РАН, 2017. - С. 192–200.
7. Антопольский А.Б. и др. О едином цифровом пространстве научных знаний // Вестник Российской академии наук, 2019. - Т. 89, - № 7. - С. 728–735. DOI 10.31857/S0869-5873897728-735
8. Савин Г.И. Единое цифровое пространство научных знаний: цели и задачи // Информационные ресурсы России, 2020. - № 5. - С. 3–5. DOI: 10.51218/0204-3653-2020-5-3-5
9. Каленов Н.Е., Серебряков В.А. Об онтологии Единого цифрового пространства научных знаний // Информационные ресурсы России, 2020. - № 5. - С. 10–12. DOI: 10.51218/0204-3653-2020-5-10-12
10. Кириллов С.А. Технологическая платформа формирования цифровых ресурсов электронной библиотеки «Научное наследие России» // Информационные ресурсы России, 2020. № 5. С. 25–27. DOI: 10.51218/0204-3653-2020-5-25-27

11. Погорелко К.П. Новая версия программного обеспечения электронной библиотеки “Научное наследие России” // Информационные ресурсы России, 2020. № 5. С. 27–29. DOI: 10.46920/0204-3653_2021_02180_9
12. Погорелко К.П. Развитие программного обеспечения электронной библиотеки “Научное наследие России” // Единое цифровое пространство научных знаний: проблемы и решения : сборник научных трудов – Москва; Берлин : ДиректМедиа Пабблишинг, 2021. С. 199–207. DOI: 10.51218/978-5-4499-1905-2-2021-199-207

On the evolution of the digital library “Scientific Heritage of Russia” as a component of the Common Digital Space of Scientific Knowledge

Nikolay Kalenov (nekalenov@mail.ru), Konstantin Pogorelko (konstpog@yandex.ru),
Alexander Sotnikov (asotnikov@jscc.ru)

The paper considers a new version of the digital library “Scientific Heritage of Russia” (DL SHR), developed taking into account the prospects of its integration into the Common Space of Scientific Knowledge (CDSSK). The task of creating the DL SHR as a public resource was formulated within the framework of the target program of the Russian Academy of Sciences Presidium in 2007. The development of the library structure and its software shell were carried out by specialists of a number of academic organizations headed by the Joined Supercomputer Center (JSCC). The library is built on the principle of linked data and reflects information about scientists, their publications (including digitized full texts), archival materials related to scientists and museum items. Since 2010, the DL SHR has been publicly available on the Internet at <http://e-heritage.ru>. In 2019, the concept and basic principles of CDSSK creating as a system of related data that unites diverse scientific resources were published. The DL SHR contains the most important elements of the CDSSK, therefore, the task arose of reorganizing the Library in accordance with the approaches laid down in the CDSSK. By 2020, the technology for the formation of the DL SHR, its internal and external software shells, and search algorithms were completely redesigned and put into operation. The article describes the structure of the DL SHR, which includes two WEB-oriented blocks – administrative and user. The administrative block, accessible to authorized users, provides input, editing, quality control of created resources; the user block, presented in the public domain, provides ample opportunities for resource search using Boolean logic; it provides visualization and advanced navigation of related resources. Examples of processing search queries and navigation through the resources of the DL SHR are given.

KEYWORDS: scientific knowledge space, information resources, scientific heritage, electronic libraries, network technologies, linked data