



**Павел Юрьевич  
Лушников,**

заместитель директора по  
цифровизации — руководитель  
проектного офиса НЭБ  
Российской государственной  
библиотеки (Москва)

## Оцифрованное достояние

П. Ю. ЛУШНИКОВ

В конце 2018 г. Национальная электронная библиотека (НЭБ) представляла собой информационную систему для использования в библиотеках, в которой было около 4,5 млн объектов. По сути, она являлась большим хранилищем цифровых документов, разобраться в котором без профессиональной помощи было непросто. С тех пор НЭБ сильно изменилась, обрела другой дизайн и обзавелась новыми функциями и приложениями. Что поменялось в главном архиве цифровых документов за несколько лет, и что ждет портал в ближайшем будущем?

Прежде чем мы поговорим о развитии НЭБ, я бы хотел начать с глобальных изменений в предпочтениях наших читателей, потенциальных пользователей библиотеки. Ведь разговор о разработке продукта цифровизации невозможен без точного понимания аудитории — тех, для кого мы его делаем. В нашей работе мы руководствуемся ответами на два вопроса: что мы делаем и кому это нужно. Это простые вопросы, но иногда за множеством планов и операционной деятельностью они забываются.

Новое поколение, сохранившее прекрасную привычку читать, не всегда готово использовать прежние модели потребления печатной литературы. Для подтверждения этих слов не надо далеко ходить, проводить какие-либо исследования и опросы, хотя и они у нас есть. Зайдите в час пик в вагон метро или автобус, зал ожидания в аэропорту или любую парковую зону в мегаполисе и посмотрите, что делают люди. Читают, смотрят фильмы, играют, готовят доклады, просто стримят. И все это при помощи мобильного телефона.

Нет смысла разбираться, хорошо это или плохо — это факт. Поэтому в 2019 г. была переформулирована миссия НЭБ: обеспечить максимально быстрый, удобный и законный доступ к авторитетным ис-

### Книжные памятники

Возможность познакомиться с историческим, культурным и художественным наследием России



#### Популярное



**Не только читать, но и слушать!**  
Получите аудиокниги по самым популярным произведениям.



**Больше, чем спорт**  
Получите электронные версии самых популярных спортивных книг.



**Царская рыба**  
Получите электронные версии популярных книг о царской России.



**Гражданская авиация**  
Получите электронные версии популярных книг о гражданской авиации.



**100 лет СССР**  
Получите электронные версии популярных книг о советском Союзе.

#### Истории писателей



**Государевы забавы**  
История Пушкина о царских забавах и народных играх. Узнайте, как развивались забавы, игры, шутки и развлечения в царской России.



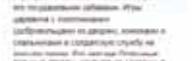
**Слова тебе, безысход...**  
История Толстого о жизни и творчестве.



**«Вздох, крестьян, а...»**  
История Достоевского о жизни и творчестве.



**Врач, написавший 300...**  
История Горького о жизни и творчестве.



**Минус тридцать слов**  
История Пастернака о жизни и творчестве.



**Немеренные имена**  
История Солженицына о жизни и творчестве.



**Я весь ушел в стихи**  
История Ахматовой о жизни и творчестве.

#### Коллекции



**Литературный дуэт XX века**  
Книжки о жизни и творчестве писателей XX века.



**Изучаем Крым по книгам**  
Книжки о истории и культуре Крыма.



**Камчатские экспедиции**  
Книжки о экспедициях на Камчатку.

#### Изображения



#### Профессионалам



**Научно-издательская деятельность РНБ**  
Получите электронные версии научных изданий.



**Профессионалам Библиотечного дела**  
Получите электронные версии профессиональных книг.



**Оцифровано в НЭБ**  
Получите электронные версии оцифрованных документов.

точникам знаний. В соответствии с ней возникла и новая продуктовая стратегия портала, отражающая главные направления его развития.

Согласно этому плану, НЭБ из единого целого трансформируется в экосистему связанных между собой решений и продуктов, предназначенных для разных потребителей. Фактически некоторые направления развития мы выделяем в самостоятельные проекты. Поэтому наряду с порталом rusneb.ru читатели получили сайт «Книжные памятники» (<https://kp.rusneb.ru/>), приложение для смартфонов «Свет! Главные книги бесплатно» (<https://svetapp.rusneb.ru/>), а в ближайшем будущем начнет действовать раздел «Наука». Для специалистов были созданы отдельный реестр книжных памятников (<https://knpam.rusneb.ru/>) и ресурс с профессиональной литературой по библиотечному делу ([https://rusneb.ru/collections/887\\_prof/](https://rusneb.ru/collections/887_prof/)).

Кроме того, перед нами стоят стратегические задачи по обеспечению технической и юридической доступности документов. Совершенствуются механизмы оцифровки, распознавания, поиска и предоставления материалов, например, широкое распространение получают аудиоинструменты для чтения. Проводится большая работа по формированию новых схем взаимодействия с издателями и иными правообладателями. Расскажу обо всем немного подробнее.

### Обновления на портале НЭБ

При всей нашей любви к кириллическим доменам надо признать, что их эффективность на сегодня с точки зрения SEO-продвижения ниже, чем доменных имен с латинским написанием. Это первое видимое изменение, которое произошло с НЭБ: она переехала на адрес с новым именем — rusneb.ru. Может быть, не столь элегантно, как нэб.рф, но сам факт переезда позволил увеличить трафик на портале в несколько раз. В цифровой среде мы должны всегда делать поправку на выбор средства информирования пользователей о своих ресурсах, и поисковая оптимизация — один из важнейших инструментов. Впрочем, кириллический домен по-прежнему действует, и по нему сайт можно найти.

Следующими большими обновлениями портала стали редизайн и изменение поиска. Новый, бо-

лее современный и привлекательный интерфейс главной страницы, центральное место которой заняли подборки, — результат глубокой исследовательской работы. Расширенный поиск позволяет с большей эффективностью найти нужный документ, а редакционные подборки фокусируют внимание на интересных и трендовых материалах. Ведь, согласитесь, среди множества объектов НЭБ не всегда просто найти то, что нужно, их количественные показатели вовсе не означают, что они качественные. Снова проведя исследование, мы получили интересную картину обращений к ним. Большую часть источников открывали один раз за все время их нахождения на сайте. А другие, напротив, постоянно используются. Для нас это сигнал: значит, на портале размещены как материалы, которых больше нет нигде в российском сегменте Интернета на условиях бесплатного легитимного использования, так и редкие, ценные, узкоспециализированные издания, которые необходимы исключительно профильным специалистам.

Как объединить две такие совершенно разные аудитории? Вероятно, самый правильный ответ — разработать востребованные продукты внутри портала. Таким решением стали редакционные подборки. Рекомендации — хороший маркетинговый инструмент, который позволяет создавать тематические коллекции, сегментированные по интересам аудитории и наиболее полно отражающие тренды в конкретный момент. А для региональных библиотек это способ заявить о себе и рассказать о своей миссии читателям.

За последний год мы запустили десятки спецпроектов, в том числе совместных — с партнерами, которые выступают прежде всего в качестве экспертов. Например, благодаря коллегам из Континентальной хоккейной лиги (КХЛ) в НЭБ появился материал об истории становления клубного хоккея в нашей стране (<https://khl.rusneb.ru/>). Уверен, он будет интересен всем поклонникам этого вида спорта — здесь содержатся архивные черно-белые фотографии спортсменов, газетные вырезки советского периода, книги и диафильмы о хоккее.

Совместно с Федерацией конного спорта России мы подготовили и представили исторический материал, посвященный выездке и воспоминаниям олимпийских чемпионов. В качестве эксперта выступил С. Вяльшин, судья международной категории по

конному спорту. Он рассказал, можно ли сейчас изучать основы конной выездки по книгам вековой давности, как поменялся спорт за прошедшие несколько лет. О том, актуальны ли эти книги сейчас, вы можете прочитать в нашей публикации «По лошади виден и ездок» (<https://horses.rusneb.ru/>).

Также сделали подборку по истории уголовно-исполнительной системы страны совместно с Федеральной службой исполнения наказаний России (<https://fsinhistory.rusneb.ru/>). В нее вошли оцифрованные архивы, которые раньше не были доступны широкому кругу читателей — книги и фотографии о жизни заключенных до революции 1917 г. и в советский период. И, конечно, составили рекомендации, посвященные знаменательным датам нашей страны — Дню Победы (<https://pobeda.rusneb.ru/>), Дням воинской славы России (<https://dnislavy.rusneb.ru/>).

Мы слышали отзывы многих профессиональных специалистов, что слишком все упростили и переориентировались на «ленивую» аудиторию, но выросшие показатели посещаемости и книговыдачи говорят, что такой подход принес успех. Для узких специалистов мы постараемся разработать еще более удобные инструменты в рамках экосистемы НЭБ.

Важной частью развития остается направление для профессионалов — ресурс, который адресован специалистам: библиотечным и информационным работникам, библиографам, преподавателям, а также рекомендован студентам вузов и колледжей культуры и искусств, библиофилам ([https://rusneb.ru/collections/887\\_prof/](https://rusneb.ru/collections/887_prof/)). В нем мы размещаем материалы по вопросам библиотечного дела. Примечательно и то, что такие подборки есть не только у Российской государственной библиотеки (РГБ), на портале каждая библиотека-участница НЭБ может составлять такие рекомендации самостоятельно. И коллеги из Российской национальной библиотеки (РНБ) активно этим пользуются.

Стоит отметить разделы, адресованные специалистам в других областях. Например, подборка, посвященная детской книге ([https://rusneb.ru/collections/512\\_russkaya\\_detskaya\\_kniga/](https://rusneb.ru/collections/512_russkaya_detskaya_kniga/)). В ней несколько тысяч наименований, и мы видим ее целевой аудиторией прежде всего учителей и исследователей. Впрочем, любознательным родителям она тоже будет полезна.



## Книжные памятники

В конце 2019 г. стартовал большой проект «Книжные памятники», состоящий из нескольких частей. В первую очередь это раздел НЭБ — Книжные памятники (<https://kp.rusneb.ru/>). Его суть в том, что мы собрали в одном месте объекты, являющиеся достоянием письменной и печатной культуры страны. Десятки тысяч оцифрованных в высоком разрешении печатных изданий и рукописей, которые не только обладают высокой материальной и духовной ценностью, но и представляют особое историческое, научное и культурное наследие России. Книжные памятники — это не только старинные книги или рукописи, но и гравюры, плакаты, нотные издания, географические карты. Многие из них — настоящие произведения искусства. Изучая их, можно глубже понять реалии времени, в которое они были созданы, или по-новому взглянуть на привычные исторические события. Важное отличие ресурса от многих аналогичных проектов в том, что мы

ориентировали его не только на профессионалов, хотя уверены, что и им будет интересно, но и на широкую аудиторию — всех тех, кто ценит историю и хочет знать больше.

Второй составляющей проекта является реестр книжных памятников (<https://knpam.rusneb.ru/>), в который фондодержатели могут вносить сведения о своих объектах.

Третья составляющая — работа по оцифровке книжных памятников, которая финансируется в рамках национального проекта «Культура» (федеральный проект «Цифровая культура»). Сейчас на страницах реестра можно увидеть более 8 тыс. документов из фондов РГБ и РНБ. К концу 2020 г. количество наименований удвоится, среди них будут представлены также материалы из коллекций других библиотек страны. До 2024 г. должно быть оцифровано не менее 48 тыс. документов. Участие в проекте доступно для всех библиотек — держателей памятников.

## Приложение «Свет»

К началу нового учебного года в App Store и Google Play появилось новое мобильное приложение «Свет! Главные книги бесплатно» (<https://svetapp.rusneb.ru/>). Его основная задача — увлечь молодое поколение чтением. Оно модное и интерактивное, в нем удобная и интересная мотивационная программа, которая, как мы надеемся, поможет выработать привычку читать. При разработке продукта мы уделили большое внимание отбору книг и приобретению лицензионных прав: в библиотеке есть и охраняемые авторским правом произведения, любимые всеми, например, «Три мушкетера» Александра Дюма и «Волшеб-

СВЕТ! ГЛАВНЫЕ КНИГИ БЕСПЛАТНО



ник Изумрудного города» Александра Волкова, труды Бориса Пастернака и других авторов. Стоит отдельно подчеркнуть, что это не самая большая библиотека в смартфоне, и мы не стремимся сделать ее такой. Главная ценность приложения — абсолютно бесплатный доступ к золотому фонду российской и мировой литературы.

---

Юрий Сапрыкин, редакционный партнер приложения «Свет», руководитель образовательного проекта «Полка»: «Я часто вижу, как школьников и молодых людей пытаются убедить, что “читать модно”. По-моему, для людей, которые выросли со смартфоном в руках, нужно прежде всего сделать так, чтобы читать было удобно — не менее удобно, чем листать сториз в Инстаграме. С тем, что модно или нет, они как-нибудь разберутся сами».

---

### Свежая пресса

Это еще один продукт из экосистемы НЭБ (<https://press.rusneb.ru/>), а также мобильное приложение «НЭБ.пресса»), где мы собрали свежую федеральную периодику, доступ к которой может получить каждый гражданин Российской Федерации бесплатно, нужно только авторизоваться через портал «Госуслуги». В разделе есть как деловая пресса, так и научно-популярная, например, журнал National Geographic, а также периодика, доступная по подписке.

ПРИЛОЖЕНИЕ «НЭБ. ПРЕССА» ПРЕДОСТАВЛЯЕТ ДОСТУП К СОВРЕМЕННОЙ ПЕРИОДИКЕ



Деловой Петербург



Известия



РБК Газета



Российская газета



Пока мы громко не говорим об этом разделе, но со временем рассчитываем наполнить его большим количеством изданий, особенно региональными медиа, доступ к которым затруднен за пределами территорий их распространения.

### НЭБ наука

Сложный раздел, который требует привлечения большого количества экспертов, и в целом его непросто развивать, так как он предназначен для весьма искушенной и требовательной аудитории. Однако мы уверены, что научному сообществу такой ресурс с доступом к научным материалам на русском языке и специальными инструментами будет весьма полезен. С одной стороны, мы делаем все, чтобы запустить его как можно скорее, с другой — не можем себе позволить сделать его недостаточно качественным.

### Технологические решения

Это главные проекты, на развитии и продвижении которых сконцентрированы наши усилия. Но мы работаем и над технологическим совершенствованием процессов, без которого достичь цели — «сделать чтение удобным» — невозможно.

В современной цифровой реальности просто хранить документы недостаточно. Электронной библиотеке необходимо предоставлять доступ к материалу в удобном и понятном виде. Например, при помощи электронной базы с возможностью поиска, сортировки материалов, анализа необходимых частей объекта.

Один из примеров — в 2019 г. КХЛ запустила беспрецедентный проект по сбору и оцифровке протоколов матчей клубов высшего эшелона, начиная с первого чемпионата СССР, состоявшегося в 1946 году. Итогом его реализации должна стать достоверная электронная база клубной статистики игроков и команд, участвовавших в чемпионатах и кубках СССР и России по хоккею.

Но как оцифровать газеты, чтобы они не превратились в архив джипегов, которые удобно смотреть, но невозможно полноценно анализировать? Проще говоря, как перейти от классической модели оцифровки, результатом которой становится файл в формате изображения, к удобным инструментам поиска необходимой и достоверной информации? Даже простейший поиск по слову «хоккей» в материалах 1940-х гг. скорее всего будет посвящен хоккею с мячом, а не с шайбой. Перед нами стояла задача — не просто оцифровать, а превратить сканированный документ в связный текст.

Если вы когда-нибудь работали с системами распознавания машинного текста, то, вероятно, знаете, что занятие это непростое. Много ошибок, сложности с переносами и форматированием. Встречающиеся в реальности тексты, особенно в нашей практике, далеки от идеала. Грязные изображения, проблемы качества сканирования, следы старения. Из-за того, что документ ксерокопировали или микрофильмировали, или со временем он стал настолько ветхим, что подвергся реставрации, часто при его распознавании возникают дополнительные разрывы и слияния символов. Любой из перечисленных дефектов может заставлять OCR-системы ошибаться, побуждая их находить несуществующие знаки

СОВМЕСТНЫЙ ПРОЕКТ НЭБ С КОНТИНЕНТАЛЬНОЙ  
ХОККЕЙНОЙ ЛИГОЙ



или пропускать целые строки различных для глаза слов. А все потому, что некоторые из систем полагают, что непрерывная область изображения должна быть одиночным символом. Учитывая это, несложно представить, что шанс получить нечитаемый текст в конечном результате весьма велик, и кажется, что проще набрать его заново, нежели работать с полученным материалом.

Исходные данные нашей задачи состояли в том, чтобы создать программу, которая как раз может выполнять весь описанный пул работ, доставляя минимум неудобств редактору.

В итоге нам удалось разработать технологию, которая распознает газетную верстку почти так, как если бы ее заново произвел человек. Это гарантирует отсутствие большого количества ошибок от неправильно идущих или соединенных строк. Документ проходит три этапа обработки. Сначала при помощи визуально-семантического анализа производится деление страницы на блоки — статьи; статьи — на составные части: заголовки, автор, фото, подпись к фото, текст и т. п. Затем каждый из блоков проходит распознавание с учетом конкретной шрифтовой и лингвистической модели. А после нейронная сеть вычленяет ключевые смысловые понятия: упоминания персон, географические названия, наименования организаций. Для спортивных изданий полезно выделить виды спорта, соревнований и прочие ключевые понятия, даты проведения спортивных событий. Это позволит в перспективе создать принципиально новые инструменты для доступа к этим данным.

Еще одно направление работы — конвертация традиционных текстовых материалов в новые форматы, например в аудио, мы реализуем ее совместно с коллегами из Yandex. Уже сегодня некоторые документы в НЭБ можно не только прочитать, но и прослушать, начиная с любого места. Безусловно, озвучка техническая, и иногда в ней могут возникать ошибки, но это первый шаг в данном направлении и скоро ее качество станет лучше! Аудиоматериалов пока будет немного, в пределах 4–5 тыс. документов, чтобы оценить реакцию читателей, востребованность технологии и произвести необходимые доработки.

У нас большие и амбициозные планы, и я абсолютно уверен, что мы справимся с поставленными задачами! ■