

ПРОБЛЕМЫ ИНФОРМАЦИОННОГО ОБЩЕСТВА

УДК 021.4(410)

А. И. Земсков

ГПНТБ России

Лондонская книжная выставка – 2016

Представлен краткий обзор материалов Лондонской книжной выставки 2016 г., приведены статистические сведения: участники, посещения и др. Особое внимание уделено таким вопросам, как учёт мнения пользователей при конструировании новой платформы, доступность информации для слабовидящих, Открытый доступ и проблемы авторского права, развитие продаж электронных книг, работа с большими массивами данных и семантические методы, перспективы развития библиотек. Рассмотрены наиболее интересные, с точки зрения автора обзора, публикации в печатных изданиях, приуроченных к Выставке. Отмечено одно из значимых событий Выставки – знакомство с новым директором Библиотеки Конгресса США.

Ключевые слова: Лондонская книжная выставка – 2016, учёт мнения пользователей, доступность информации для слабовидящих, открытый доступ, проблемы авторского права, электронные книги, большие массивы данных, семантические методы, перспективы развития библиотек, Чарльз Кларк, Карла Хэйден.

UDC 021.4(410)

Andrey Zemskov

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

London Book Fair – 2016

The author reviews the materials of the London Book Fair 2016. He addresses the issues of user feedback for building new platforms, information accessibility for visually impaired people, open access and copyright, e-book sales, processing large data arrays and semantic methods, prospects for libraries.

Keywords: The London Book Fair 2016, information user feedback, information accessibility for the visually impaired, open access, copyright, large data arrays, semantic methods, library prospects.

The exhibition atmosphere was full of optimism and confidence in the good prospects of the industry. Mr. Arnaud Nourry of the Hachette said that the most serious threat to the publishing business may be the European Council. “The extensive list of copyright exceptions... in the framework of the doctrine of “fair use” – think about the devastating consequences for European publishers. Who can stop Google from declaring itself a library and provide all the books free of charge? Their revenues based on advertising rather than pay for access?” Charles Clark Memorial Lecture “What are the publishers for?” was extremely emotional and non-standard. All power belongs to the buyer; companies like Facebook, Google, Amazon, Apple, Internet service providers, occupy a leading position. They set prices, or allow free access. They dominate the public debate and shape public opinion, in particular, that the copyright – this is an unfair monopoly. Creating a high-quality product requires a long, constant, professional employment of authors and publishers. To create great work, people should live as professionals by their work. Of course, amateurism is welcome, but it should not replace professional. “In the twentieth century the countries and societies that do not support the right to property, lost; in the twenty-first century, those who do not support the right to intellectual property, too, will lose”. Clear sounded was the theme of enhancing the role of information users in the development of new products and services. People more time spend with their smartphones and tablets, which means that they spend less time with desktop computers. The online catalog is a tool that is ideal for search and detection, but this tool is designed for use only on a desktop computer. The modern public library should move towards mobile technologies. Mr. Marmanis from the Copyright Clearance Center said “Bottle-neck is now not a weak computing power, but our lack of understanding of the processes”. He offered a good metaphor: information is similar to the oil field. Having extracted oil, you have to learn how to refine it and reprocess into a variety of valuable products.

На 45-м году своего существования Лондонская книжная выставка (12–14 апреля 2016 г., выставочный комплекс «Олимпия») достигла значительных успехов: 25 тыс. посещений за 3 дня работы; 12 тыс. посетителей из 100 стран; 1 600 участников экспозиций; более 2 млн упоминаний в социальных медиа, в том числе 12 800 лайков в *Facebook*, 38 тыс. последователей в *Twitter*, 6 тыс. членов *LinkedIn Group*, свыше 170 тыс. просмотров на канале *YouTube*.

Специальные мероприятия выставки были посвящены 400-летию со дня смерти Шекспира. Несмотря на расширение числа платных мероприятий, было предусмотрено также более 200 бесплатных семинаров.

В целом атмосфера выставки была полна оптимизма, уверенности в хороших перспективах развития отрасли. На открытии выставки её директор Джекс Томас (*Jacks Thomas*) заявила: «Определённо появились проблески надежды, продажи выросли в США, КНР, Великобритании, ЮАР, продажи и печатных и электронных книг увеличились на 4% (в 2014 г. электронные выросли на 17%, в 2013 г. – на 26%). Мы движемся в нужном направлении».

Неудивительно, что попытка повернуть вспять базовые тренды в законодательстве об авторском праве прозвучала в традиционном мемориальном докладе памяти Чарльза Кларка¹ (основные сентенции доклада: «Издатели – гаранты стабильности культуры», «Открытый доступ и *Creative Commons* – выдумки демагогов и нехороших людей», «Сплотимся против попыток изменить, смягчить законодательство об авторском праве» и т.п.). Доклад был одобрен долго не смолкающими аплодисментами.

Общее мнение многих издателей: «Не торопитесь с изменением Закона о копирайте, многое можно делать и при существующем законодательстве, используя лицензии, в том числе на глубокую обработку данных (TDM)». Выступая на Форуме издателей, руководитель «*Springer Nature*» Дерк Хаанк (*Derk Haank*) заявил: «Впустую потрачено два года. Инициатива компаний “Springer” и “Elsevier” по предоставлению учёным возможности свободно обмениваться между собой отредактированными статьями пока не поддержана всей отраслью». Он же критиковал гибридную модель Открытого доступа (ОД) за «двойное взимание платы» – с подписчиков и с авторов. Чистая модель ОД намного яснее и прозрачнее.

Представитель «*Penguin Random House*», лауреат премии *Lifetime Achievement award* Гейл Ребак (*Gail Reback*) своё выступление на конференции «*Quantum*» назвала «Падающие доходы авторов». В частности она отметила: «Только 10% писателей живут исключительно писательским трудом, половина самиздатовских авторов зарабатывают менее 500 долларов в год»; технологические продвижения имеют две стороны: всего лишь горстка авторов зарабатывает огромные деньги; очень много авторов не могут выйти на аудиторию, они «потерялись» в потоке информации. Неразумно противопоставлять печатные и электронные публикации – покупатель просто берёт книгу, неважно в каком формате.

¹ Чарльз Кларк (1933–2006) – юрист по образованию, специалист по авторскому праву (работал в *Sweet and Maxwell*, *Penguin Press*, *Hutchinson*, в Ассоциации издателей – *Publishers Association*), один из основателей *Copyright Licensing Agency* (1983) (прозвище *Lord Clark of Copyright*).

Далее приведу обзор и краткое изложение наиболее интересных, на мой взгляд, сообщений, результаты обработки презентационных докладов, а также печатных материалов, розданных участникам перед началом семинара и приуроченных к выставке периодических изданий, – приложения к журналу «Publishers Weekly» – «Ежедневник книготорговца» (*The Bookseller Daily*), «Ежедневник Лондонской выставки» (*London Show Daily*), «Цифровой прожектор» (*The Digital Spotlight*).

Учёт мнения пользователей и доступность документов

Ясно звучала тема повышения роли пользователей информации (читателей) при разработке новых товаров и услуг. В определённом смысле речь заходила даже о том, что читатель лучше, чем специалист-библиотекарь знает, что ему нужно и как это должно работать. При этом декларируемая универсальность новых устройств и разработок – это погоня не столько за обеспечением пресловутой мобильности, а скорее за пользователями, поскольку сейчас значительная их часть чаще всего работает (и развлекается), используя мобильные устройства.

На семинаре «**Учёт мнения пользователей при разработке платформы. Путешествие к открытию**» (*User-Centric Design in Platform Development: A Journey of Discovery*) был представлен совместный доклад Дженни Матиас (*Jenny Mathias*) и Викки Драммонд (*Vicky Drummond*), издательство «*Cambridge University Press*». Это издательство получило патент в 1534 г. и выпустило первую книгу в 1584 г. В 1997 г. началась поставка журналов по сети (сейчас – 360 наименований журналов, содержащих 1,6 млн статей), а с 2010 г. – книг (30 тыс. книг, 260 тыс. книжных глав). В целом издательство предоставляет доступ к 19 млн страниц на обеих платформах – журнальной и книжной. К настоящему времени накоплен позитивный опыт их работы и, как считают пользователи, назрело время их объединения.

С момента запуска в эксплуатацию первой платформы (журнальной) прошло около 20 лет, и технологии поиска, использования и обмена информацией существенно изменились:

социальные средства коммуникации (*social media*) приобрели огромное значение, клиенты обмениваются впечатлениями и делятся информацией через сети; быстро растёт применение мобильных устройств; поисковая оптимизация (*search engine optimization, SEO*) становится важным элементом технологии.

«Многие исследователи начинают научную работу с поисковых машин, и нам следует убедиться, что мы там представлены, – подчёркнуто в докладе. – Технология сделала нас нетерпеливыми – мы желаем получить мгновенный доступ и не очень любим думать».

Издательство «*Cambridge University Press*» в течение полутора лет реализует большой проект новой цифровой платформы для работы с коллекцией научных документов *Cambridge Core*. С самого начала в основу проектирования положены ощущения пользователя. Ключом к любой цифровой разработке является оценка пользователя. С учётом всех обстоятельств пришлось начинать проект с нуля, что одновременно означает большие возможности и большие сложности.

Издательство пригласило аналитическую компанию «*Nomensa*», которая распространила среди учёных, авторов, партнёров, библиотекарей около 10 тыс. анкет (в анкете 50 вопросов) и провела интервью с пользователями по всему миру. Это исследование рынка послужило базой для разработки платформы.

Пользователи информации различного уровня и научного стажа – от студентов и до профессоров – начинают научную работу с поиска в семействе Гугл (*Google, Google Scholar, Google Books*), затем (с трёхкратным снижением) следуют: библиотечные сервисы, веб-сайты ведущих научных групп, веб-сайты научных журналов, службы *Discovery*, службы индексирования и реферативные БД. Это означает, что на веб-сайт издательства будут приходить посетители от поисковых машин и библиотечных каталогов, и потому совершенно необходима качественная оптимизация поисков. Пользователю нужно дать возможность простой навигации по сайту к нужной точке и позитивного ощущения вне зависимости от типа устройства, с которым он работает, – планшет, смартфон или настольный ПК.

На следующем этапе, после завершения поиска и выхода на нужный документ, поведение различных по уровню подготовки пользователей в целом подчиняется общим правилам и закономерностям (в порядке падения приоритетов): рассмотрение ссылок, расширенный поиск, переход к другим работам того же автора, ключевые слова в данной статье, поиск альтернативных ключевых слов.

Схожие результаты получены в проектах Корнелльского университета в Итаке (США) в 2015 г. и отражены в докладе Трэси Гарднер (*Tracy Gardner*) «Каким образом читатели находят нужные документы в научных публикациях» (*How readers discover content in scholarly publications?*) (см., напр.: <https://allenpress.com/system/files/Session%20-%20-%20Tracy%20Gardner.pdf>). Хотя как правило поиск начинается не с веб-сайта издателя, однако, попав на него, около половины пользователей продолжают работу на сайте.

Невозможно удовлетворить в равной степени потребности всех участников. Результаты исследования позволили выделить четыре значимые группы: исследователи, библиотекари, авторы, социальные партнёры (группы единомышленников).

Интересный элемент исследования – проверка достоверности резуль-

татов опроса путём сравнения устных ответов на вопросы анкеты с реальным поведением, которое анализировалось по ходу изучения статистических данных сетевого трафика компанией *Webnographer*. По мере разработки регулярно тестировалась платформа с тем, чтобы конечный продукт понравился пользователям выделенных групп, отзывы которых составили информационную базу обратной связи. Предполагается, что новая платформа *Cambridge Core* появится летом 2016 г.

Семинар «*Доступность: за пределы простых истин. Как работать с насыщенным и сложным документом и улучшить ощущения пользователя*» (*Accessibility: beyond the basics how to treat rich and complex content and improve user experience*). Сложность и потенциальные возможности цифрового содержания растут, всё важнее становится проблема доступности для тех, кто не может пользоваться стандартной печатью – вследствие потери зрения, дислексии или иных заболеваний. Используя преимущества, которые раскрывает цифровой формат, издатели смогут выйти на более широкую аудиторию и улучшить ощущения для каждого.

Значительная часть семинара была отведена новым разработкам системы *ePub*. Кроме того, были изложены проблемы конверсии математических и музыкальных документов. Математические формулы считываются и распознаются на базе семантики и затем синтезируются в речь. При работе с математическими материалами нужно обращать внимание на следующее:

математика – это особый язык, нужно убедиться в правильности маркировки документа;

документ содержит исключительно математические материалы и не содержит знаков пунктуации;

необходима проверка математических материалов, в том числе правильности семантики.

Помощь слабовидящим музыкантам включает следующие направления: увеличение размера (величины) нот, модификация нотной доски, использование системы Брайля, «говорящие» ноты.

Открытый доступ и проблемы авторского права

Зачастую речь шла об оптимальном использовании технологий ОД для совершенствования процесса книгоиздания. (Типичные рекламные слоганы на стендах и в печатных изданиях выставки: «Открытый доступ быстро меняет процесс книгоиздания!», «Открытый доступ: а вы готовы к новой волне?», «Открытый доступ: следующая волна – это более совершенное обслуживание авторов»).

Очень часто обсуждение сворачивало на путь «открытого рецензирования». При этом система научного рецензирования (*peer review*) должна

рассматриваться не только как способ фильтрации, отбрасывания негодного материала, но и поддержки хорошего и очень хорошего. И с этой точки зрения характерно положительное отношение к ОД, который сможет усовершенствовать процесс книгоиздания и, конечно, использоваться в интересах слабовидящих. ОД – это не только технология доставки, это более широкий подход к сотрудничеству учёных, совместному использованию данных, расширению сотрудничества читателей. Это многослойная система, включающая сравнение деятельности авторов и производительности их труда.

В короткой вступительной речи при открытии Международного издательского конгресса в Лондоне руководитель издательства «Hachette» Арно Нурри (*Arnaud Nourry*) сказал, что ближайшей и наиболее серьёзной угрозой издательскому бизнесу может оказаться позиция Европейского Совета (говорилось о «бесмысленных атаках на копирайт»): «Перечень исключений из авторского права для библиотек, образования, а также в рамках доктрины «справедливого пользования» чрезмерен. Подумайте о разрушительных последствиях для европейских издателей, если эти предложения будут реализованы. Это то же самое, как если бы Европейская Комиссия посчитала своим приоритетом ослабить только европейскую культурную индустрию».

По мнению А. Нурри, компания «Гугл» представляет наиболее ясную опасность для индустрии книгоиздания: «Если Европейская Комиссия уступит требованиям ставленников “Гугла”, то что тогда сможет остановить их от объявления себя библиотекой и от выдачи всех книг бесплатно? Они скажут, что их доходы основаны на рекламе, а не на плате за доступ. И кто их остановит, если Европейская Комиссия выдаст благословение?» В заключение он заявил: «Разбить планы Комиссии – первейшая задача для Международной Ассоциации издателей».

Это выступление активно обсуждалось в выставочной прессе. «Ежедневник книготорговца» (*The Bookseller Daily*) опубликовал заметку «Нурри предупреждает Европейский Совет», где, в частности, сказано: «“Гугл” представляет собой большую опасность для издателей, чем “Амазон”, и позицию компании “Гугл” лишь укрепят бессмысленные атаки Европейской Комиссии на авторское право».

Этот микроскандал повторяет «Ежедневник Лондонской выставки» (*London Show Daily*): «Нурри сказал, что обширные исключения из авторского права для библиотек, образования и для справедливого пользования могут дать дорогу компании “Гугл” для того, чтобы переформатировать себя в библиотеку, открывая тем самым свои хранилища отсканированных документов для бесплатного использования и формируя свой доход на базе рекламы. Нурри вопрошает: “Почему Комиссия нацелилась на издателей?”

Похоже на то, что Комиссия выбрала своим приоритетом ослабление европейской индустрии культуры, которая является ведущей в мире. Нужно ли напоминать, что 9 из 12 крупнейших издательских компаний мира – европейские?»

В том же «*London Show Daily*» от 14 апреля 2016 г. в заметке «Библиотеки отвечают Нурри» приведён ответ ИФЛА на некоторые моменты выступления Нурри на Международном конгрессе издателей: «Очень стыдно, что Арно Нурри из «Hachette» избрал своей целью библиотеки, выступая на Международном конгрессе. Библиотеки поддерживают сбалансированную систему авторских прав, в которой каждый человек имеет доступ к информации и к творчеству, а авторы получают справедливое вознаграждение».

Интересные мысли высказывает Байрон Рассел (*Byron Russell*; компания «*Ingenta*») в заметке «Слишком много хороших вещей» (*London Show Daily*, с. 23). Открытый доступ преобразил академическое издательство в некоторых отраслях науки. Но возросшая доступность документов усложняет поиск того, что действительно нужно. Развитие ОД отчасти обусловлено экономическим кризисом последнего десятилетия и, конечно, технологическими достижениями, которые существенно упростили публикацию и архивирование.

Позиция авторов, особенно представителей точных наук, привела к тому, что сейчас примерно 43% академических документов публикуются в ОД. По оценкам компании «*Research Consulting*», в настоящее время две трети журналов предлагают опцию ОД, чаще всего в гибридном варианте, хотя это по сути последние попытки «собрать колоски» прибылей от издательства.

И всё же слишком много материалов о научных исследованиях остаются не прочтёнными и не цитируемыми, но издателей это не волнует. Поэтому задачу обнаружения источников приходится решать другим – *Google Scholar*, экспертам или поисковым машинам, таким как *TrendMD*. Поиск и индексирование усложняются обилием вариантов лицензий – не менее шести типов и с пятью версиями каждая. Наиболее распространённая лицензия – *CC-BY*. Отсутствует также перекрестное индексирование между хранилищами – *DOAJ*, *OAPEN*, институциональными репозиториями и т.п.

Интересным инновационным инструментом продвижения своих работ может оказаться система *Kudos* (название – от древнегреческого возгласа, обозначающего наивысшее достижение). Это программный сервис, помогающий авторам научных работ улучшить качество изложения для получения наивысшего воздействия работы (импакта) и её быстрого распространения. Автор описывает свою работу простым языком, сопровождает какой-либо дополнительной информацией и рассылает её по социальным сетям,

одновременно наблюдая, каков же будет эффект. Авторам демонстрируется граф, показывающий выгрузки полного текста. Компания-разработчик *Kudos* утверждает, что её сервис на 19% повышает использование информации и что каждую неделю к её услугам присоединяются более 1 тыс. авторов.)

Ещё одна позитивная идея – объединить в одном месте все распределённые хранилища, но привлекательность такой цели возрастёт и для пиратов.

Модель ОД может решить проблемы «порочного круга» в подписном книгоиздании (меньше продажи – выше цена экземпляра). (Почувствительный пример – книга «Participatory Planning for Climate Compatible Development in Maputo, Mozambique», опубликованная в 2015 г.: было продано 20 печатных экземпляров, а выгружалась книга 1 300 раз в 90 странах!) Финансирование может осуществляться по разным каналам: фонды или организация, где автор работает, либо использовать модель «*freemium*»: бесплатно доступна только сетевая версия, а форматы – электронная книга или печатная книга – платные.

Мемориальная лекция памяти Чарльза Кларка «Для чего нужны издатели?» (*Charles Clark Memorial Lecture 2016 – What are publishers for?*). В зале присутствовали ведущие специалисты в области авторского права, издательских прав, авторы и т.д. С лекцией выступил Майкл Фрэзер, профессор кафедры права Технологического университета Сиднея, бывший председатель и директор «*Australian Copyright Council*».

Лекция оказалась необычайно эмоциональной и нестандартной. Приведу лишь личные впечатления и её фрагменты (полный текст см.: http://www.pls.org.uk/media/216096/What-are-publishers-for-Charles-Clark-Lecture-2016docx.pdf?dm_i=178E,468Z3,6CXDHE,F6JYP,1).

«Моё первое наблюдение проблем копирайта относится к 1985 г., и состояло оно в следующем: чем больше люди любят какого-либо автора, тем больше они незаконно копируют его произведения, миллионы и миллионы страниц, нанося значительный ущерб и ему, и его издателю. В 1985 г. я занял в банке 105 тыс. долларов и организовал некоммерческое общество *Copyright Agency*. Мы ввели механизм упрощённой “ковровой” лицензии, использующей обобщённую статистику пользования (*blanket license*), и сейчас наше агентство распределяет в год около 2 млрд долларов».

В наш век информации вся власть принадлежит покупателю; такие компании, как *Facebook*, *Google*, *Amazon*, *Apple*, потоковые сервисы, интернет-провайдеры с их нацеленной на интересы покупателя позицией, занимают ведущее положение. Именно они – самые богатые и влиятельные – устанавливают цены или допускают бесплатное пользование (*Free ride*). Они доминируют в публичных дебатах и формируют общественное мнение,

в частности, о том, что копирайт – это несправедливая монополия, создающая барьеры творчеству и свободе выражения идей. У людей сложилось ощущение, что издатели нечестно эксплуатируют авторов, издательские контракты не предусматривают справедливых авторских отчислений. Миллионы людей поддерживают пиратство, считая его символом справедливости (сейчас в мире 43 пиратские партии); пиратство стало нормальным явлением.

Не потребитель, а автор и издатель должны решать, нужно ли оплачивать то или иное произведение.

В течение трёх веков издатели считались хранителями культуры, арбитрами качества и вкуса в писательском деле. «Сейчас они утратили доверие общества, а без него никакая охрана не поможет! Нам следует более чётко определить копирайт. Любые исключения и ограничения должны подвергаться трёхуровневой проверке, установленной Бернской конвенцией. Создание высококачественного произведения требует долгой, постоянной, профессиональной занятости авторов и издателей. Чтобы создавать отличные произведения, люди должны жить как профессионалы – за счёт своей работы. Конечно, любительство приветствуется, но оно не должно заменять профессионализм.

«В XX в. те страны и общества, которые не поддерживали право собственности, проиграли. В XXI в. те, кто не поддержит право интеллектуальной собственности, тоже проиграют».

«Должен быть учреждён специальный регистр (*copyright registry*); мы регистрируем землю, автомобили, недвижимость, патенты – точно так же должна регистрироваться интеллектуальная собственность».

В Великобритании зарегистрировано 2 160 издателей (данные 2013 г.). Доходы от публикации книг и журналов составили в 2014 г. 4,3 млрд фунтов стерлингов; 35% из них – доходы от цифровой продукции. Продажи электронных книг выросли с 258 млн фунтов в 2011 г. до 509 млн фунтов в 2013 г. Научные электронные журналы обеспечивают 79% дохода от подписки. Более четверти вузовских учебников – цифровые. В процветающих странах книготорговля по данным *International Publishers Organisation (IPO) 2014 Report* не росла или даже упала (США – 0% роста, Германия – 0%, Япония – минус 2%, Великобритания – 0%, Франция – минус 2%, Италия – минус 6%). Самиздатовские электронные книги на *Amazon's Kindle Store* составили 31% от всех продаж электронных книг.

«Я считаю неправильной идею искать баланс авторских прав между правообладателями и публикой – это ложная дихотомия. Копирайт защищает публичные интересы. Издательство – это служение обществу!»

Электронные книги

Электронные книги стали обычной компонентой процесса книготорговли и рассматривались на семинарах именно в этом качестве – не как новинка, а как «обычный бизнес», причём подчёркивалось, что развитие торговли электронными книгами стабилизировалось.

Несмотря на невесёлые выводы обзора, подготовленного *The Book Industry Study Group, Inc. (BISG)*, согласно которому только четверть читателей библиотеки пользуются электронными книгами, выдача электронных книг в американских библиотеках всё же растёт. Сложным остаётся вопрос модели ценообразования.

На семинаре «**Обозначение электронных книг и их персонализация: развитие систем управления правами на электронные документы**» (*Ebook Watermarking and Personalisation: Beyond a Better DRM*) выступил Хууб ван де Поль, руководитель компании «*BooXtream*». Он отметил: при продажах электронных книг используются различные системы поставки: а) закрытая система, работающая со средствами собственной разработки (*Amazon, Apple*) либо использующая коммерческие форматы электронных книг (например, *Mobi*), б) открытые системы независимых магазинов, напрямую продающие электронные книги по формуле D2C (наиболее употребительный формат открытых экосистем – *ePub*).

Системы защиты также могут быть замкнутыми (например, *Amazon* с собственной системой управления электронными документами *Proprietary DRM*) либо открытыми, которые используют коммерческие программы управления электронными документами *Adobe DRM, Sony DRM*, либо вообще не пользуются защитой, либо используют социальную систему (*Social DRM*), например *BooXtream* – и этот вариант лучше других!

Файлы электронных книг в формате *ePub* содержат видимую персонализацию и множество невидимых водяных знаков; однако многие упрекают DRM в недостаточной гибкости по отношению к устройствам пользователя.

В системе *BooXtream* используются многочисленные алгоритмы, кодирующие не только информацию об издателе, но и о покупателе и месте продажи (загружаются имя и электронный адрес покупателя, а также идентификатор продажи). Помеченный таким образом документ, будь он в интернете или в хранилище, связан с покупателем. Поскольку покупатель знает, что файл содержит его имя, риск нелегальной работы с ним невелик,

а так как никакая система управления правами (DRM) не используется, то покупатель может воспроизводить файл на любом своём устройстве.

Пожалуй, тема защиты электронных книг с помощью подобных систем может оказаться в ближайшее время актуальной и для библиотек, поскольку она обеспечивает нормальное и безопасное обслуживание, весьма близкое к традиционному.

В 2011 г. после внезапного краха подписного агентства электронных книг *Oyster* появились сомнения в жизнеспособности модели. Заметка о подписных электронных книгах, опубликованная в «Ежедневнике Лондонской выставки» (13 апр. 2016, с. 10–11), содержит интервью с Трипом Адлером (*Trip Adler*), руководителем крупнейшего в США поставщика электронных книг по подписке *Scribd*. «У нас хорошая база для роста: 90 млн пользователей в месяц обращаются к нашей библиотеке, в которой 80 млн документов. *Scribd* – это электронная библиотека и подписное агентство по электронным, аудиокнигам и комиксам основана в 2007 г. Подписчики могут выбирать любые книги от 1 тыс. издателей, в том числе «HarperCollins», «Simon & Schuster», «Harlequin», «Houghton Mifflin Harcourt», «Macmillan» и др. Наша система создавалась как сайт для сбора и хранения документов, иногда её называют *YouTube для документов*, поскольку любой может опубликовать на сайте своё произведение. С 2013 г. за умеренную помесечную плату пользователь может войти в библиотеку *Scribd*».

В 2009 г. против компании было возбуждено дело о нарушении авторских прав.

Семинар «Сравнение эволюции рынка электронных книг в Европе и на развивающихся рынках. Презентация доклада *Global Ebook Report 2016*» (*Comparing Ebook Market Evolution Across Europe and Emerging Markets. Presentation of the Global eBook Report 2016*).

Доклад «*Global eBook Report 2016*» (www.global-ebook.com) даёт новый взгляд на эволюцию рынка электронных книг на основе графических приближений, приводит тренды и соображения о пиратстве. За несколько последних лет традиционно консервативная отрасль промышленности, занимающая ключевые позиции в креативном секторе современного общества знаний, столкнулась с падением продаж физических книг. При этом рынок электронных книг набирал силу и рос на англоязычных рынках, особенно в США и Великобритании.

С 2013 г. рост рынка цифровых публикаций остановился. Некоторые индикаторы указывают на сдвиг в сторону нетрадиционных технологий публикаций в независимых издательствах, а также самиздата (*Self publishing*), доля которого превосходит долю издательств.

В континентальной Европе замедление продаж электронных книг началось ещё раньше, а спад продаж традиционных изданий оказался более радикальным, в том числе из-за кризиса 2008 г. Больше всех пострадали Испания и Италия.

2015 г. стал годом стабилизации, и электронные книги сумели занять определённые ниши в книготорговле. В большинстве развивающихся экономик рост рынка электронных книг либо замедлился (Китай), либо остановился совсем (Бразилия), либо упал (Россия). То, что мы наблюдаем, можно назвать «окончанием начала перехода на цифру» (*end of the digital beginning*) и переходом на новую, возможно ещё более интересную и сложную фазу, когда написание, издание и чтение трансформируются в какой-то непрерывный поток, в котором любой документ в любом формате будет доступен любому пользователю, но в условиях нестабильности в отношении как производителей продукта, так и его потребителей.

В журнале «Цифровой прожектор» («*The Digital Spotlight*»; апрель 2016, с. 1–2) Ричард Альбанезе (*Richard Albanese*) пишет: «Издательство – это цифровой бизнес. Совсем недавно мы обсуждали собственно электронные книги, а сегодня они обсуждаются уже не как продукт, а как бизнес. Ответом на нынешнее падение продаж электронных книг может оказаться развитие новых технологий чтения, просмотра (*browsing*), обнаружения (*discovery*)». И возможно, главным элементом будущей технологии продаж станет просмотр покупателем книжных полок: прикладная программа *Ncvrs* позволяет пользователю «гулять» вдоль нескончаемых рядов полок.

Пожалуй, наиболее неприятным событием 2015 г. был уход с рынка крупнейшего подписного агента электронных книг – компании «Oyster».

Таблица 1

**Продажи электронных книг за январь-ноябрь 2014 г.
и за тот же период 2015 г.**

Продажи электронных книг, млн долларов			
Сегмент рынка	2014	2015	Изменения, %
Взрослая литература	1 191,3	1 104,7	-7,9
Детская и юношеская литература	213,8	102,3	-43,7
Всего	1 405,1	1 225,0	-12,8
Доля продаж электронных книг, %			
	2014	2015	
	23,5	20,0	

Большие массивы данных и семантика

Тему использования больших массивов данных (*Big Data*) и смежную с ней тему глубокой обработки данных (*Data & Text Mining*) целесообразно осветить в одном разделе – они тесно связаны технологически и по целеустановке – заботе о повышении издательских прибылей.

На семинаре «*Как использовать в своих целях большие массивы данных*» (*Making Big Data Work for You*) выступил основатель компании «Vearsa» Гаретт Кадди (*Gareth Cuddy*). Ирландская компания «Vearsa» собирает информацию от более чем 1 тыс. книжных магазинов и 65 тыс. библиотек в 100 странах. Вначале Г. Кадди привёл интересную статистику, отражающую рост информационных массивов («Больших данных») и экологические аспекты традиционного книгоиздательства: 90% мировых данных создано за последние 2 года (информация 2013 г.), и для распечатки 1 терабайта данных нужно срубить и превратить в бумагу 50 тыс. деревьев; при двусторонней печати на это потребуется 250 млн страниц, высота их стопки составит 18 км.

Суть технологии – постоянное и оперативное отслеживание процессов продаж и возможность быстрой реакции и адаптации. *Vearsa Tracker* ежедневно отслеживает 7 млн наименований электронных книг от 35 торговцев по всему миру и информирует издателей о том, доступна ли их книга, правильно ли назначена цена, что ещё покупатели берут, за какую цену продают схожие продукты конкуренты.

В докладе приведён конкретный пример работы с низкими продажами электронных книг сектора нон-фикшн: за счёт усовершенствования и расширения метаданных, ключевых слов, рефразировки описаний, энергичного экспериментирования с ценами каждые 24 часа удалось за полмесяца повысить продажи на 56% и поднять цену примерно на 30%. Покупатели *Vearsa Tracker* увеличили свою прибыль на 7% в месяц, на 60% снизили себестоимость продукции.

В заключение Г. Кадди сказал: «Взрывной рост рынка электронных книг в последние 12 месяцев замедлился и прежний экспоненциальный темп роста вряд ли возможен. В целом портфель электронных книг достаточно хорош и рост продаж будет продолжаться, хотя и не такими темпами. Умелое манипулирование данными рынка и ценообразованием сможет обеспечить уверенный рост по разнообразным рынкам и снизить зависимость от главных рынков».

Внедрение методов семантики в процедуры обнаружения и поиска информации, в том числе в интересах совершенствования процесса книгоиздания рассматривалось на семинаре «**В чём значение семантики?**»

Добавление семантической концовки к вашему документу, начиная от авторства и до поставки» (Why semantics matter? Adding the semantic edge to your content, right from authoring to delivery). В том или ином виде базы данных используются с 1960-х гг. Технология реляционных БД неоднократно менялась и совершенствовалась, однако лишь совсем недавно слабая производительность этой технологии подтолкнула учёных к использованию альтернативных методов, в частности семантики. Семантика описывает объекты не по их внешним признакам (формат, количество страниц, год издания и т.п.), а по их сути и их связям, т.е. «о чём этот документ», «в чём его смысл».

Благодаря глубокой обработке текстов (*text mining*) и обогащению данных с помощью средств семантики (реализованных в соответствии со стандартами W3C) искусственный раздел между текстом и данными устраняется, после чего управление документами, поиск и обнаружение информации обогащаются новыми возможностями.

В компании «*Ontotext*» ясно осознают врождённую неупорядоченность окружающего нас мира, в том числе документов на естественном языке, пригодном для чтения человеком. Поэтому здесь приступили к разработке программного обеспечения, способного сортировать различные компоненты текста (*semantic tagging*) перед тем как выдавать полезные семантические метаданные. Семантические метаданные представляют окружающий мир (или документы) так, как его видит человек, а не машина, описывают документы в понятных терминах: местоположение, люди и их должности, организации, бренды и т.д.

Объединение семантической базы данных и семантического сортировочного процесса (*GraphDB + semantic tagging*) позволило создать автоматизированную систему, которая способна решать проблемы, ранее доступные лишь квалифицированным экспертам. Многие издатели сегодня подумывают об обогащении их публикаций набором метаданных для обеспечения лучшей обнаруживаемости. Это вводит нас в мир семантических технологий и системе «Умного содержания» (*Smart Content*).

Семантическая БД компании «*Ontotext*», названная *GraphDB*, организует данные посредством описания «реальных вещей» (*things in the world*) и их взаимоотношений. Такого рода описания и организация данных, подпадающих под эти описания, позволяет системе *GraphDB* в автоматизированном режиме делать динамичные выводы о новой информации. Кроме того, поскольку мир меняется, будут меняться и описания в БД и соответствующие данные. Таким образом достигаются гибкость и динамизм технологии. Семантическая БД предполагает наличие описаний существующего мира, что требует огромного опыта и колоссальных усилий.

Использование продуктов компании – *GraphDB* и *Self-Service Semantic*

Suit (S4) – формирует более гибкий, интегрированный и целенаправленный взгляд на те данные и документы, которыми располагают издатели, позволяет экономить время и деньги на управление документами и анализ текстов. С помощью структурированных данных и других сетевых семантических технологий документ разбивается на машиночитаемые фрагменты с тем, чтобы его можно было собрать заново по желанию пользователя. Для издателей важно, что с применением семантических методов совершенно новые возможности приобретает Система управления документами (*Content Management System, CMS*). Продуктами компании «*Ontotext*» пользуются *BBC, Euromoney, IET* и др.

На семинаре «*Дилемма данных*» (*The Data Dilemma*) речь шла о различных методиках фильтрации больших массивов информации с тем, чтобы повысить обнаруживаемость документа, в том числе в библиотеке. Джим Брайант (*Jim Bryant*), руководитель компании «*Trajectory*», рассказал о сотрудничестве с более чем 300 поставщиками электронных книг по всему миру.

Созданный в компании инструментарий обработки естественного языка (*Natural Language Processing*) используется для извлечения ключевых слов и другой информации, относящейся к содержанию книги. Система идентифицирует основные формы слов, части речи, имена компаний, людей, мест и т.п., описывает структуру предложения в терминах взаимозависимости фраз и слов, указывает на другие связи. В 2014 г. компания создала стандартизованный метод сравнения и сопоставления эмоций, отражённых в книге, а в 2015 г. объявила о разработке методики выдачи книготорговых рекомендаций.

В рамках проекта *Natural Language Processing* разработана система оценок чувств. Каждому слову в книге присваивается численное значение – позитивное или негативное.

Аналитики компании используют эти данные для построения эмоциональной кривой какого-либо предложения (параграфа, абзаца, главы, всей книги в целом), формируя таким образом уникальный «эмоциональный портрет» публикации. Утверждается, что подобная обработка позволяет легче обнаружить конкретный (нужный) документ или найти что-либо подобное. Инструментарий системы *Natural Language Processing* обрабатывает текст, трансформируя каждое предложение в набор данных, которые в свою очередь можно анализировать или искать.

Хараламбос Марманис (*Haralambos Marmanis*), руководящий сотрудник *Copyright Clearance Center*, отметил: данных много (в науке, технике, медицине накоплено 2 млн статей), но их «освоение» – низкое. «Узким горлышком» сейчас являются не слабая мощность компьютеров, а наше непонимание процессов. Данные – это продукт. Пока что 80–90% информации не структу-

рировано. Х. Марманис предложил хорошую метафору: «Массив информации подобен нефтяному месторождению, оцифровка и извлечение данных – это как бы нефтедобыча», после чего идут очистка, переработка и т.п. Добыв “нефть”, вы должны научиться её перерабатывать в различные ценные продукты!»

Сибил Вонг (*Sybil Wong*), представитель компании «*Sparrho*», начала своё сообщение с арифметики. Учёный читает в среднем 22 статьи в месяц, а за это же время публикуется 208 300 новых статей, т.е. осваивается немногим более 0,0001% публикаций. Созданная в компании система выдачи научных рекомендаций *Sparrho* обеспечивает персональную новостную информацию с использованием алгоритмов обучения компьютера и контекстного анализа.

Мотивации отслеживания исследователями научной литературы – попытка получить новые идеи и победить в конкурентной борьбе – не основаны на традиционном понятии релевантности и не имеют к ней отношения. Вообще определить истоки мотивации очень сложно, поэтому алгоритм поиска построен скорее на отрицании. Отбрасываются статьи, которые не соответствуют определённым параметрам, и в результате из «неподъёмных» 208 300 статей в месяц получается вполне приемлемый объём – 4 статьи (табл. 2).

Таблица 2

**Параметры фильтрации массива статей
и количество статей, остающихся после каждого этапа**

Операция, мотивы фильтрации	Число статей
Опубликовано более 30 дней тому назад	208 300
Лежат вне довольно широкого круга научных тематик	40 000
В списке ссылок нет упоминания необходимых ключевых слов	500
В статье нет упоминания необходимых ключевых слов	100
В статье нет упоминания интересных авторов или цитирований	5
В статье нет упоминания интересных журналов	4

Эта система дополняет традиционные методы поиска релевантного академического материала (такие как *Google Scholar*), а не заменяет их. Преимущества машины в том, что она: а) перерабатывает огромные массивы источников и б) может обнаружить неожиданные результаты в таких источниках, которые эксперт по какой-то причине не смотрел (например, небольшие журналы). При этом эксперт лучше распознаёт: а) альтернативные фразы и ключевые слова, б) статьи известных ему авторов, в) людей со схожими мотивациями.

Дальнейшие перспективы развития книгоиздательской и библиотечной деятельности

В дни работы выставки прошли мероприятия, посвящённые 400-летию научного книгоиздания. Обсуждались перспективы развития книгоиздания (в первую очередь печатного формата), а также академических журналов. Все сообщения были подчёркнуто оптимистичны.

На семинаре «**Война платформ: научные журналы в изменяющемся мире**» (*Platform Wars: The Scholarly Journal in a Changing World*) было отмечено, что сегодня учёные столкнулись с разнообразием платформ, одни из которых работают только с цифровыми документами, другие – только с оцифрованными печатными документами, а третьи работают на грани новых технологий. В ходе дискуссии обсуждались преимущества и недостатки существующих платформ и потенциальные угрозы действующих систем; каким образом открытость платформ и документов может повлиять на воздействие исследований на авторов и читателей.

Также была рассмотрена модернизация процесса научного рецензирования. При этом отмечено, что рецензирование постепенно превратилось в бизнес, который могут организовать сторонние компании; с учётом требований к скорости и разнообразию научных коммуникаций в качестве альтернативы проходят эксперименты с неанонимным рецензированием и с автоматизированным; переход к публикации персонального набора (коллекции) из интересующих потребителя статей при условии наличия высококачественного рецензирования и редактирования как сервиса. В качестве примера рассмотрена научная платформа открытого доступа, на базе которой выпущен журнал «*Interscript / UCL Journal of Publishing*», где обсуждаются проблемы научного издательства. Журнал выпускается аспирантами *University College London (UCL) Centre for Publishing* не в формате периодических выпусков, а как непрерывный поток статей.

Меня заинтересовала статья Тома Мерсера (*Tom Mercer, Director of Digital Products at Bibliotheca*) «Обеспечивая мобильное присутствие для библиотек» (*Creating a mobile presence for libraries*), опубликованная в «Ежедневнике Лондонской выставки» (*London Show Daily*; 13 апр. 2016, с. 22).

«*Bibliotheca*» – крупнейшая мировая компания, специализирующаяся на разработке, производстве, поставке и технической поддержке библиотечных технологических решений, созданных с целью повысить производственную эффективность и улучшить качество обслуживания читателей; RFID-технология – основное направление. Учитывая требования библиотек, компания разрабатывает и производит несколько серий RFID-оборудования: от базовых моделей до самых совершенных комплексных решений.

«Я не считаю, что электронные книги разрушат издательский бизнес и использование печатных книг. Но я вижу, что чтение и прослушивание цифровых устройств во всё большей степени составляют процесс изучения книги и наслаждения ею. Для печатных книг всегда останется место в сердце читателя – хотя бы для того чтобы знать, куда можно уйти от надоевшего экрана. Поскольку, по моему мнению, книги и чтение будут сохранять большую популярность ещё очень долго, считаю, что местные публичные библиотеки не только останутся на плаву как место, где люди могут взять и почитать физически осязаемую книгу, но постепенно превратятся в жизненный узел, сердце локального сообщества», – пишет Т. Мерсер.

Каким образом технологии выдачи электронных книг влияют на трансформацию публичных библиотек? Они начали превращать библиотеки в более мобильные, и этот тренд должен сохраняться – с тем, чтобы привлекать будущих читателей. Современные сетевые технологии в библиотеках сконструированы под использование настольных компьютеров; онлайн-каталог (ОРАС) является ключевым инструментом, обеспечивающим присутствие библиотеки в сети. Онлайн-каталог – это инструмент, идеально подходящий для поиска и обнаружения каждого библиотечного документа, но этот инструмент создан для использования только на настольном компьютере. Однако же если посмотрим вокруг, то заметим, что люди всё больше и больше времени проводят со своими смартфонами и планшетами, а значит – меньше времени перед настольными компьютерами.

В докладе «*Nielsen Total Audience report*» за 4-й квартал 2015 г. приведено несколько интересных фактов. Американцы в возрасте старше 18 лет еженедельно проводили 8 ч 24 мин со своими смартфонами, используя различные приложения, и в среднем 5 ч 9 мин – у настольных компьютеров, работая в интернете. Хотя эти цифры демонстрируют превосходство смартфонов над настольными устройствами, присмотримся к возрастной группе 18–24 года. Здесь на смартфоны в неделю «тратится» 11 ч 15 мин, а на настольные компьютеры – 4 ч 18 мин. В интернете проводится вдвое больше времени со смартфонами, чем с настольными компьютерами.

Обеспечение качественной работы мобильных устройств – ключ к успеху не только библиотек, но и всех организаций в будущем. Это неизбежное соответствие требованиям мобильности в библиотеках означает необходимость обеспечить подбор и рекомендацию печатных публикаций, возможность работать с печатными книгами «на месте», в помещениях библиотеки.

Библиотеки должны иметь электронные книги, цифровые аудиокнижки, так же как и печатные книги, и всё это должно «обнаруживаться» на новых интерфейсах, спроектированных для мобильных устройств. Пока что чаще всего библиотечные программные приложения для смартфонов представ-

ляют собой просто уменьшенные версии для настольных устройств. Сегодня нужно полностью обдумать заново, каким образом и какими средствами «мобильники» будут взаимодействовать с библиотекой.

Дело не только в том, как помочь пользователю найти документ, следует сформировать интересный и удобный режим работы, достаточно привлекательный мир, чтобы пользователь открывал библиотечные приложения так же охотно, как он обращается к *Facebook*, *Twitter*, *Instagram*, *YouTube*. Если любимое приложение пользователя начнёт функционировать в местной публичной библиотеке, за будущее библиотек в цифровую эпоху можно не беспокоиться.

«Работая в компании “*Bibliotheca*”, я имею уникальную возможность регулярно беседовать с библиотекарями о том, что они делают не только в плане внедрения цифровых технологий, но в преобразовании физического пространства библиотеки. В числе задач нашей компании – приблизить библиотеку к читателю, сделать её намного более доступной. И один из наиболее эффективных методов обеспечения светлого будущего для библиотеки – тестирование электронных платформ книговыдачи и мобильных технологий, взаимодействующих с пользователем, – как в физическом пространстве библиотеки, так и в цифровом пространстве.

Современная публичная библиотека не должна позволять себе ни пренебрегать своим каталогом печатных изданий, ни забывать о том, как меняется пользование документами, смещаясь в сторону мобильных технологий. Я с удовольствием отмечаю некоторые весьма позитивные результаты в публичных библиотеках, использующих гибридные модели, в состав которых входят и мобильные технологии. Убеждён, чем больше библиотек освоят этот подход, тем больше у них шансов оставаться жизнеспособными».

Завершая краткий обзор статьи Т. Мерсера, добавлю, что в моём обзоре работы 36-й Конференции ИАТУЛ (см.: Науч. и техн. б-ки. – 2016. – № 1. – С. 82–101) содержится информация о разработке в Университете г. Дармштадт (Германия) мобильных устройств для использования в библиотеках и музеях.

К достаточно значимым событиям в библиотечном мире можно отнести назначение нового директора Библиотеки Конгресса США. В феврале 2016 г. Президент США Обама назначил директором Библиотеки Конгресса Карлу Хэйден (*Carla Hayden*), руководителя *Enoch Pratt Free Library* в Балтиморе.

Назначения на эту должность происходят довольно редко: в истории США с 1802 г. их было всего 19, в том числе с 1900 г. – лишь 6. Предыдущий директор – Джеймс Биллингтон был назначен Президентом Рональдом Рейганом в 1987 г.

Карла Хэйден – первый библиотекарь, выдвинутый в эпоху интернета на столь высокую должность. Как замечает в редакционной статье газета «Washington Post», слушания по этому, казалось бы, безобидному предложению, могут быть нетривиальными: «Библиотека находится в глубоком кризисе, и, несомненно, её следующий руководитель столкнётся с пугающими проблемами: как сохранить и оживить символ демократии и культуры нашей страны».

Назначение К. Хэйден – первой женщины и первой афроамериканки – на должность главного библиотекаря США встречено библиотечной общественностью США с восторгом. Она – всего лишь второй библиотекарь по образованию на этом посту: первый – Лоуренс Мамфорд (*Lawrence Mummford*) – ушёл на пенсию в 1954 г.



Карла Хэйден – новый директор Библиотеки Конгресса США

Кто будет во главе национальной библиотеки, важно хотя бы потому, что библиотечная общественность США (*American Library Association*) активно выступает по вопросам законодательства, протестуя против принятия ряда проектов законов, имеющих отношение к работе библиотек. Важнейшим направлением деятельности К. Хэйден будет авторское право; в частности она предлагает вывести из Библиотеки Конгресса США такое подразделение, как *Copyright Office* (Служба авторского права) и превратить его в независимое агентство. Это предложение поддержано издателями, но библиотечное сообщество, а также компании *Google*, *Amazon* и другие против.

Наш журнал желает успеха Карле Хэйден!

Andrey Zemskov, Cand. Sc. (Physics and Mathematics), Associate Professor, leading researcher, Russian National Public Library for Science and Technology;

andzem@gpntb.ru

17, 3rd Khoroshevskaya st., 123436 Moscow, Russia