

ИНФОРМАЦИОННЫЙ ПОИСК

УДК 004.42

ИНСТРУМЕНТЫ ПОДДЕРЖКИ ОНЛАЙНОВОЙ НАУЧНОЙ ПУБЛИКАЦИИ¹

© 2019 г. М. М. Горбунов-Посадов^{а,*}, Т. А. Полилова^{а,**}

^а *Институт прикладной математики им. М.В. Келдыша
125047 Москва, Миусская пл., 4, Россия*

** E-mail: gorbunov@keldysh.ru*

*** E-mail: polilova@keldysh.ru*

Поступила в редакцию 10.01.2019 г.

После доработки 10.01.2019 г.

Принята к публикации 15.01.2019 г.

Анализируется положение, сложившееся в нашей стране в области онлайн-научных публикаций: обеспечение открытого доступа к публикациям; средства надежного хранения вышедших публикаций; коррекция ошибок, обнаруженных в онлайн-публикации; живые, т.е. регулярно обновляемые публикации; мультимедиа и другие современные элементы публикаций; переход от PDF-представления публикаций к HTML.

DOI: 10.1134/S013234741903004X

1. ОТКРЫТЫЙ ДОСТУП

Сегодня агитировать авторов за публикацию их статьи в интернете, в открытом доступе, уже нет необходимости — преимущества такого размещения давно стали очевидными. Приведем здесь лишь недавний характерный пример. Уступая пожеланиям отдельных читателей, существовавшая более 10 лет в онлайн-работе [1] была издана в виде печатной монографии. Ссылка на возможность несложного приобретения печатной версии работы была размещена на почетном месте в онлайн-тексте. В результате удалось получить характерную статистику: на 10 тыс. посещений онлайн-версии приходится одна покупка печатной. Вывод ясен: у автора, желающего получить широкую читательскую аудиторию, сейчас просто нет выбора — только открытый онлайн.

Тут российский автор научной публикации может пойти по одному из трех путей. Первый, очевидный, — опубликовать статью в журнале открытого доступа.

Второй путь — опубликоваться в журнале, не обеспечивающем открытый доступ, но и не требующим от автора передачи исключительной лицензии на публикацию. В этом случае автор вправе тут же разместить публикуемую статью в открытом доступе, скажем, на сайте своей организации.

Наконец, третий путь — не связываться с журналами, а сразу разместить статью на своем сайте.

Решение уязвимое. Во-первых, с точки зрения российского администратора такая публикация не идет в зачет автору. Во-вторых, статья, опубликованная в авторитетном журнале, приобретает в глазах читателя своего рода знак качества: журнал принял статью — значит, она на уровне. Размещение статьи без такого знака качества заметно снижает ее посещаемость.

Положение с открытым доступом к научным публикациям в нашей стране относительно благополучное. Если в Европе, где, собственно, и зародилось движение открытого доступа, за пятнадцать лет с 2004 по 2018 гг. охват публикаций открытым доступом увеличился лишь с 15% до 20% [2], то у нас в стране сейчас в открытый доступ попадает до 50% опубликованных научных работ.

Дело в том, что в Европе непреодолимым препятствием на пути к открытому доступу оказался издавна сложившийся коммерческий подход к изданию научных журналов. Открытый доступ встречает мощное сопротивление коммерческих издателей, поскольку лишает их весьма внушительных прибылей. Россия же унаследовала от Советского Союза некоммерческие отношения в научной издательской среде, и поэтому переход к открытому доступу идет у нас существенно легче и энергичнее.

Наиболее заметных результатов здесь добился проект Киберленинка [3], охвативший более полутысячи, т.е. около 30% отечественных научных журналов и обеспечивший их комфортным открытым доступом. Киберленинка сегодня в

¹ Работа частично поддержана грантом РФФИ 19-01-00069 А.

своей области в определенном смысле выполнила майский (2018 года) указ Президента РФ, заняв в 2018 году во всемирном рейтинге открытых научных репозиториях почетное 5-е место [4].

Движение открытого доступа долгое время обходило стороной журналы Российской академии наук. С каждым годом положение только усугублялось. В 2017 году был исключен из открытого доступа один из ведущих журналов Академии наук — «Вестник РАН»: за скачивание статьи из «Вестника» потребовали заплатить 200 руб.

К счастью, в конце 2017 года положением дел с изданием академических журналов заинтересовалась Счетная палата. В самом деле, почему журналы, издающиеся на бюджетные деньги, не видны широкому кругу российских ученых?! Со Счетной палатой в нашей стране спорить не принято, и в начале 2018 года, после издания соответствующего распоряжения, свободный доступ к онлайн-версиям большинства академических журналов был открыт.

2. СОХРАННОСТЬ ОНЛАЙНА

Тем не менее, утверждать, что сейчас у нас в стране с ИТ-составляющей публикации онлайн-версий научных статей все обстоит благополучно, не позволяют несколько тревожных обстоятельств.

Одно из них — положение дел с обеспечением сохранности чисто онлайн-изданий. Особенно рельефно недопустимость сложившейся здесь ситуации проявилась при недавнем рассмотрении вопроса о включении таких изданий в Перечень ВАК. Вполне законный вопрос, включенный в анкету претендующего на включение в Перечень журнала, касался именно обеспечения сохранности. Требовалось ответить, какой след останется от журнала, если он и обслуживающие его веб-серверы почему-либо прекратят свое существование? Если в такой ситуации опубликованные в онлайн-журнале статьи становились недоступны, журнал, естественно, в Перечень не попадал.

Ни один чисто онлайн-журнал не смог успешно ответить на этот вопрос. А ведь еще недавно проблемой сохранности здесь успешно занимался Информрегистр [5], который получал от онлайн-издания вышедшие статьи и подобно Российской книжной палате обеспечивал их сохранность, рассылая в ведущие национальные библиотеки. Однако в 2012 году Минобрнауки прекратило финансирование этой деятельности Информрегистра, и с тех пор, несмотря на уход со своего поста принявшего это спорное решение министра, российская онлайн-наука так и не обрела жизненно необходимую надежную почву под ногами.

В то же время некоторые европейские страны не только считают своим долгом обеспечение надежного хранения онлайн-научных публикаций, но и более того — бережно сохраняют в полном объеме все содержимое своего национального домена. На этом фоне полное бездействие России выглядит особенно досадно. Ведь технологически на российских просторах организовать надежное хранение национального цифрового достояния совсем несложно. Достаточно построить треугольник со сторонами не менее 1000 км, в углах которого разместить серверы, реплицирующие сохраняемую информацию, — такой конфигурации практически не страшны никакие природные и техногенные катаклизмы.

Ради объективности заметим, что небрежение сохранением онлайн-достояния встречается и за рубежом, где крупные издательства всерьез обеспокоены неустойчивостью интернет-публикаций, ссылки на которые встречаются в их продукции. Любопытная схема обеспечения надежности онлайн-ссылок работает в проектах webcitation.org [6], perma.cc [7] и др. Ответственность за сохранность интернет-публикации там предлагается возложить не на ее автора, а на автора работы, ссылающейся на эту публикацию. В изданиях, придерживающихся этой идеологии, прежде чем разместить ссылку на чью-то статью в интернете, требуется сначала скопировать эту статью в специальный надежный депозитарий. Ссылка в таком случае содержит два адреса: непосредственный адрес публикации и (для подстраховки) адрес ее копии в депозитарии.

3. ИСПРАВЛЕНИЕ ОШИБКИ

Одно из очевидных преимуществ онлайн-представления научной работы — возможность легко исправить в этом представлении замеченные неточности, опечатки, ошибки. У человека, мало-мальски знакомого с информационными технологиями, такое исправление занимает секунды, в крайнем случае несколько минут. Ошибка исправлена — и все последующие онлайн-читатели видят корректную версию работы.

Казалось бы, какая удача: теперь допущенные в публикации ошибки можно моментально исправить, они не будут преследовать автора всю оставшуюся жизнь. Однако, увы, практически никто из издателей, размещающих полные тексты научных статей в онлайн, не разрешает авторам править обнаруженные впоследствии ошибки.

Причины и в инерции традиций печатных изданий, и в отсутствии общепринятой технологии внесения исправлений в ранее опубликованные тексты. В то же время, когда на одной чаше весов эти в общем-то незначительные причины, а на другой — жизненные потребности и автора, и чи-

тателей, естественно, всегда стремящихся иметь дело с корректными онлайн-публикациями, то, конечно же, ясно, что должно перевесить. Обнаруженная, но не исправленная в онлайн-ошибка — просто преступление перед наукой.

Масштабы проблемы хорошо видны, если вспомнить, что CrossRef, основной регистратор DOI, до 2014 года запрещал вносить изменения в научные публикации, получившие DOI. В то же время все издания, проиндексированные в Web of Science (WoS), к этому времени уже получали DOI для всех выходящих у них статей. Т.е. ни один (!) WoS-журнал не разрешал авторам исправлять в онлайн обнаруженные ошибки. CrossRef в конце концов одумался и снял свое вредоносное ограничение, однако WoS-журналы не спешат разрешать онлайн-правку.

Конечно, желательно, чтобы все вносимые в опубликованную онлайн-статью исправления тщательно протоколировались — это позволит избежать коллизий с авторскими правами на тот или иной научный результат. Конечно, такое протоколирование должно производиться в определенных общепринятых, канонизированных рамках. К сожалению, о формировании такого IT-канона остается только мечтать — как уже упоминалось, подавляющее большинство онлайн-издателей пока еще просто бездумно запрещают любые исправления.

4. ЖИВАЯ ПУБЛИКАЦИЯ

Живая публикация — размещенная в интернете в открытом доступе научная работа, которая постоянно развивается и совершенствуется ее автором. Благодаря своим очевидным преимуществам живые публикации с каждым годом приобретают все новых и новых сторонников.

Автор, отказавшийся от традиционной, статичной публикации в пользу живой, попадает в новую, существенно более комфортную и продуктивную среду. Допущенные ошибки и опечатки не носят теперь фатального характера, как отмечалось выше, не преследуют его всю оставшуюся жизнь. Круг читателей живой публикации много шире, интерес к ней со временем нередко даже усиливается: многие читатели раз за разом возвращаются к полюбившемуся тексту, не только чтобы освежить в памяти наиболее существенные моменты, но и чтобы узнать, как трансформируются взгляды автора и что нового он заметил в своей области.

Для читателя живая публикация несомненно предпочтительнее статичной. В самом деле, насколько увереннее себя чувствуешь, когда знаешь, что текст перед твоими глазами находится под неусыпным контролем автора, что в нем тщательно исправлены все неточности и ошибки, за-

меченные с момента первого размещения работы в онлайн, а также постоянно отслеживаются изменения, происходящие в данной отрасли науки.

Преимущества живой публикации перед статичной можно попытаться численно оценить на широко известном примере. В 2016 году вышли из печати последние тома Большой российской энциклопедии (БРЭ), и БРЭ была выложена в открытый доступ. Посещаемость сайта БРЭ bigenc.ru в 2018 году составила 90 тыс. обращений в месяц. Посещаемость русскоязычной Википедии ru.wikipedia.org в этот же период — 900 млн обращений в месяц, т.е. на одно обращение к статичной, быстро устаревающей БРЭ приходится 10 тыс. обращений к живой Википедии.

Не столь масштабная, но также достаточно убедительная статистика наблюдается на сайте электронной библиотеки Института прикладной математики им. М.В. Келдыша РАН. Здесь размещены несколько живых публикаций. Их посещаемость оказалась в среднем в 100 раз выше, чем посещаемость традиционных, статичных публикаций.

С точки зрения российского чиновника поддержка автором своих живых публикаций — занятие безусловно вредное, так как отвлекает его от выполнения священного плана по валовому выпуску статичных статей. Тем не менее, перечисленные выше преимущества этого жанра нередко все же перевешивают сопротивление чиновников, и ряды писателей живых публикаций множатся.

В плане технологической поддержки, к живым публикациям относятся все соображения, высказанные в предыдущем разделе в отношении необходимости не только протоколирования вносимых изменений, но и формирования канонической версии такого протокола. Есть, однако, и еще одна сторона функционирования инфраструктуры живых публикаций, нуждающаяся в инструментальной поддержке.

Как отличить живую публикацию от статичной? Просто добавить к ее представлению некий специальный значок “Публикация объявлена живой”, очевидно, недостаточно. Ведь автор мог когда-то прикрепить этот значок и благополучно забыть и о нем, и своем онлайн-тексте. Поэтому единственное достоверное свидетельство живой публикации — свежая дата ее последней редакции. Эта дата размещается на почетном месте и служит надежным ориентиром для добравшегося до нее читателя.

Хочется, однако, чтобы живую публикацию можно было заметить раньше, еще при взгляде на ссылку, ведущую к ней. Как, например, читателю, просматривающему библиографический список, отличать включенную в него живую публикацию от соседних традиционных? Для этого можно до-

бавить к библиографической ссылке конструкцию вида “Редакция от ≈27.09.2018≈”, где символы ≈ обрамляют свежую дату последней редакции.

В таком случае, разумеется, желательно, чтобы интересующая читателя свежая дата подставлялась в эту конструкцию автоматически. Ведь невозможно себе представить сверхдобросовестного автора, постоянно просматривающего библиографические списки в своих онлайн-статьях, чтобы раз за разом корректировать изменяющиеся даты последней редакции встретившихся там живых публикаций.

Выяснилось, что разработка программного инструментария для реализации автоматически обновляемой даты последней редакции вполне возможна для файлов в форматах HTML и MS Word [8], но, к сожалению, представляет собой практически неподъемную задачу для PDF — наиболее распространенного формата размещения научных публикаций.

5. МУЛЬТИМЕДИЙНЫЕ ИЛЛЮСТРАЦИИ

Мультимедиа уже давно органично и прочно вошло в мир науки. Многие выступления на конференциях, семинарах, защитах диссертаций уже просто немыслимы без мультимедийных иллюстраций. Сталкивающиеся галактики, хитроумно перемещающиеся и взаимодействующие роботы, видеозаписи важных природных и техногенных событий позволяют слушателям существенно глубже воспринимать содержание доклада.

Почему же мультимедиа до сих пор чрезвычайно редкий гость на страницах научных публикаций? Ведь и здесь потребность читателя в мультимедийных иллюстрациях ничуть не менее острая. Более того, научную публикацию помимо этого можно дополнить еще и онлайн-вычислениями, обращениями к базам данных, которые также будут с благодарностью приняты онлайн-читателем.

Одна из очевидных причин отставания мультимедийных иллюстраций — главенствующее положение и инерция печатных изданий. Если издание выходит и в печатной, и в онлайн-версиях, то издатель как правило считает, что онлайн-версия ни в чем не должна отличаться от печатной.

Соображение спорное. Как отмечалось выше, онлайн-читателей научных публикаций на порядки больше, чем любителей печатной продукции, и поэтому их интересы, очевидно, приоритетнее. Если разрешить мультимедийные отличия в онлайн-версии, отдельные любители печати действительно почувствуют себя несколько обделенными, зато основная масса читателей

получит совершенно новое качество в плане удобства восприятия.

Неуютно чувствуют себя, в частности, онлайн-читатели автореферата диссертации. Согласно сложившейся неявно рекомендуемой ВАК традиции, на сайте диссертационного совета не размещается презентация диссертанта, показанная им на защите. Т.е. онлайн-читатель автореферата сможет увидеть мультимедийные иллюстрации, только если он не поленится добраться до видеозаписи защиты и ему повезет разглядеть экран презентации в этой видеозаписи.

Строго говоря, включать мультимедиа непосредственно в текст публикации, казалось бы, нет необходимости. Автор может, ни в чем не нарушая сложившегося порядка, ограничиться включением в текст гиперссылки (URL) на мультимедийную иллюстрацию. Однако известно, что по такой гиперссылке переходят не более 20% читателей. Читатель не пойдет по ссылке, поскольку боится потерять мысль, боится встретиться с вирусами в сети, боится затруднений при возврате к исходному тексту и т.д. И напротив, если мультимедийная иллюстрация встроена непосредственно в текст публикации, более 80% читателей ею заинтересуются и инициируют соответствующий ролик.

На общем фоне блистающего мультимедийными украшениями современного интернета сектор онлайн-научных публикаций выглядит сейчас подчеркнуто скучным и сухим. Причина — не только в инерции печатных научных изданий. На пути к широкому внедрению мультимедиа в онлайн-научные публикации существуют серьезные препятствия из области информационных технологий.

Дело в том, что основная масса страниц современного интернета построена на базе файлов в формате HTML. HTML, и особенно его современную версию HTML5, мультимедийность просто пронизывает вдоль и поперек, размещать там мультимедиа — одно удовольствие. Однако для представления научной публикации HTML применяется чрезвычайно редко: основным форматом, применяемым здесь сегодня, является PDF. PDF же жестко сопротивляется внедрению элементов мультимедиа.

Рассмотрим эту проблему несколько подробнее. Авторы научных работ для их подготовки используют в основном среду TeX или MS Word. Если встроенная в текст TeX мультимедийность сохраняется при генерации PDF-файла, то все такого рода иллюстрации, включенные в файл в формате MS Word, при генерации PDF исчезают, независимо от того, как выполняется генерация: средствами MS Word или Adobe Acrobat. Автор публикации, подготовленной в MS Word, для включения в нее мультимедиа вынужден осваи-

вать весьма нетривиальный язык JavaScript для PDF (не путать с JavaScript для HTML).

Но даже когда автору удалось так или иначе внедрить мультимедиа в PDF-файл, неприятности на этом далеко не заканчиваются. Если еще несколько лет назад все обстояло вполне благополучно: браузеры, встретив файл в формате PDF, обращались для его визуализации к Adobe Reader, то теперь практически все ведущие браузеры почему-то сочли, что показать на экране PDF они способны и без посторонней помощи. Действительно, рутинные PDF-файлы такие браузеры воспроизводят без потерь, однако с демонстрацией встроенного в PDF мультимедиа ни один из них справиться пока не может.

То есть мультимедийная иллюстрация в PDF-файл научной публикации встроена, но посетитель веб-страницы публикации, просматривающий ее в браузере, эту иллюстрацию не видит. Браузеры при этом ведут себя по-разному. Google Chrome, например, делает вид, что все в полном порядке, но ничего мультимедийного на экране не показывает. Другие браузеры иногда при открытии страницы выводят малозаметную диагностику своей беспомощности, еще реже вяло отсылая посетителя к использованию для просмотра публикации Adobe Reader (который, конечно, с демонстрацией мультимедиа в PDF успешно справляется). Разумеется, сложившееся положение назвать комфортным для автора и читателя никак нельзя.

6. ОТ PDF К HTML

Разработка формата PDF — безусловно, важнейшее достижение фирмы Adobe. Этот формат безупречно работает, когда речь идет о печати документов. Однако попытки Adobe приспособить PDF для показа произвольной онлайн-информации пока в силу ряда организационных и технологических причин к успеху не привели.

Складывается впечатление, что для полноценного представления научных публикаций в онлайн-среде потребуется массово перейти к их размещению в формате HTML. В этом формате органично реализуются все необходимые манипуляции с живыми публикациями, с мультимедийными иллюстрациями, с онлайн-вычислениями, с обращени-

ями к базам данных и т.д. Кроме того, HTML имеет все необходимые средства для организации семантических связей между размещаемыми в интернете публикациями.

Для успешного осуществления перехода к HTML необходимо оснастить наиболее привычные для авторов среды TeX и MS Word новыми возможностями. PDF-файлы, по-видимому, будут продолжать генерироваться, поскольку они чрезвычайно удобны при печати. Но наряду с PDF надо будет научиться создавать в этих средах и HTML-файлы, с которыми прежде всего будет иметь дело онлайн-читатель.

Движение в направлении к HTML понемногу зарождается. Известно несколько интересных зарубежных проектов. В нашей стране на онлайн-вое HTML-представление недавно перешли [9] два-три журнала Математического института им. В.А. Стеклова РАН, размещаемых на портале MathNet.ru. По-видимому, массовый переход научных онлайн-изданий к HTML уже не за горами.

СПИСОК ЛИТЕРАТУРЫ

1. Горбунов-Посадов М.М. Интернет-активность как обязанность ученого. [б. м.]: Издательские решения, 2017. 64 с. <https://doi.org/10.20948/ridero-2017-gorbunov>
2. Ближе к людям. Европейская наука станет более открытой // Поиск. 2018. № 35. С. 5–6. <http://www.poisknews.ru/theme/international/38343/>
3. Научная электронная библиотека “КиберЛенинка”. <https://cyberleninka.ru/>
4. Transparent Ranking of Repositories (July 2018). <http://repositories.webometrics.info/en/transparent>
5. ФГУП НТЦ “Информрегистр”. <http://infoereg.ru/>
6. WebCite. On-demand archiving system for webreferences. <http://webcitation.org>
7. Websites change. Perma Links don't. <https://perma.cc/>
8. Живая публикация. Инструменты поддержки. <http://alive.keldysh.ru/>
9. Чебуков Д.Е. Об HTML версии полного текста научной статьи // Научный сервис в сети Интернет: труды XX Всероссийской научной конференции (17–22 сентября 2018 г., г. Новороссийск). М.: ИПМ им. М.В. Келдыша, 2018. С. 487–498. <https://doi.org/10.20948/abrau-2018-16>