

Об HTML версии полного текста научной статьи

Д.Е. Чебуков

Математический институт им. В.А. Стеклова Российской академии наук

Аннотация. Традиционной формой представления полного текста научной статьи в электронной форме является PDF файл, который является полной копией печатной версии. Несколько лет назад у западных издателей появился устойчивый тренд на представление полных текстов научных публикаций в формате HTML. Версия научной статьи в формате HTML содержит ту же информацию, что и PDF файл, но при этом уже не является полной копией ее печатной версии и может предоставлять читателям расширенный функционал, недоступный в PDF: браузер иллюстраций, возможность экспорта иллюстраций и таблиц в PowerPoint, ссылки со списка литературы на библиографические базы данных. В данной статье приводятся примеры и обсуждаются преимущества и недостатки HTML версий научных публикаций.

Ключевые слова: форматы представления научной статьи, сайт издателя научных журналов, расширенный функционал HTML версии научной статьи.

About HTML version of the full-text article

D.E. Chebukov

Steklov Mathematical Institute of Russian Academy of Sciences

Abstract. PDF is a traditional format of the full-text scientific article in electronic form, and it is a full copy of the print article. Recently a new stable trend to prepare full-text articles in HTML format has appeared among western publishers. HTML and PDF versions of the full-text articles contain the same scientific information, but HTML version is not an exact copy of the print version and can contain extra functionality for readers unavailable in the PDF version: figures browser, tables and pictures export to PowerPoint, hyperlinks from references to bibliographic resources. In this article we discuss examples of the HTML full-text articles, their advantages and disadvantages.

Keywords: formats of the full-text scientific article, scientific journal publisher website, extra functionality of the HTML full-text article.

Основной задачей научной публикации является изложение результатов научного исследования, для подтверждения которых широко используются иллюстрации, графики, таблицы, представляющие экспериментальные данные,

иногда довольно объемистые. Читатель, как правило, стремится не только поверить на слово выводам автора, но и убедиться в их верности, анализируя приведенные экспериментальные данные. Объем традиционной публикации в печатном издании обычно регулируется правилами, принятыми в конкретном журнале, что ограничивает автора при приведении экспериментальных данных, и он отдает предпочтение обсуждению полученных результатов, но не раскрывает детальной картины, каким образом эти результаты получены. Для изложения подробностей могут использоваться дополнительные материалы, которые, однако, не входят в саму публикацию и оказываются оторванными от основного текста статьи, для их получения от читателя потребуются дополнительные усилия.

Структура научной статьи не меняется столетиями. С началом цифрового века стандартом для публикации научной статьи в Internet стал PDF файл, который является полной копией печатной версии и наследует все ее преимущества и недостатки. Преимущества PDF неоспоримы: его легко создать из разных источников, любые программы для редактирования и верстки текстов имеют экспорт в PDF; формат публикации удобен для чтения на всех устройствах, а также на печати, а PDF файл выглядит одинаково и именно так, как был задуман автором или редактором; его легко распространять, в том числе offline. Однако, несмотря на то, что функционал PDF файлов совершенствовался, он остается привязанным к печатной версии, он сделан для того, чтобы его распечатали на принтере. Любые интерактивные вставки, иллюстрации высокого разрешения, анимация в PDF файле если и возможны, то непропорционально увеличивают его размер и ограничивают его совместимость с программами чтения. Полный просмотр возможен только в Adobe Acrobat, в то время как многие современные браузеры (Google Chrome, FireFox) имеют собственный функционал просмотра PDF, который не поддерживает мультимедийные дополнения, аналогичная ситуация наблюдается и с программами просмотра PDF сторонних производителей.

Возможности добавления мультимедиа-материалов в PDF файл подробно рассмотрены в работе [1].

Подавляющее большинство традиционных новостных и аналитических изданий, имеющих печатную версию, а также интернет версии газет и журналов, публиковали и продолжают публиковать свои online материалы в формате HTML. В HTML публикациях присутствуют анимированные иллюстрации, видео и аудио блоки, важным компонентом является контекстная реклама. Издатели могут дополнительно распространять PDF версии отдельных статей или всего номера газеты-журнала, как правило, распространяя их по подписке, но это является дополнением, а не основной версией публикации.

Издатели (западных) научных журналов осознали, что им тоже нужна большая гибкость в отображении публикуемого материала. В 2011-2013 гг. у большинства западных издателей научных журналов появился устойчивый тренд на представление полных текстов научных статей не только в формате

PDF, но и в HTML. При этом обе версии несут одну и ту же научную информацию, но HTML версия призвана предоставлять дополнительные возможности читателям. Проекты по созданию HTML версий носят кричащие названия, они рекламируются среди подписчиков. Приведем несколько примеров рекламных текстов:

- Elsevier, платформа ScienceDirect: "*The Article of the Future* initiative aims to revolutionize the traditional linear format of the academic paper to make it more dynamic and user-friendly" [2];
- IOP Publishing, платформа IOPScience: "*The Article evolution* project designed to improve the online delivery of research articles, enabling readers to interact with research in new ways" [3];
- American Physical Society, платформа Physics Review Journals: "A high-fidelity, robust full-text HTML display of content has been a long-term goal for our journals" [4];
- Springer, платформа SpringerLink: "All content can be viewed in PDF and a significant portion of our newer content is also available in full-text HTML" [5].

Основной функционал PDF и HTML версий полных текстов научных статей

Рассмотрим основные требования, предъявляемые к полным текстам современных научных публикаций, и возможность их реализации в форматах PDF и HTML (табл. 1). Ячейки, подкрашенные зеленым цветом, указывают на параметры, которые легче реализуются или имеют преимущество перед другим форматом. Красным указан функционал, который сложнее реализуется или не возможен. Белым цветом указаны равнозначные возможности реализации функционала.

Функционал	PDF	HTML
Создание файла: набор и верстка	Все системы набора и верстки: LaTeX, InDesign, MS Word.	Необходимы специальные утилиты для создания XML/HTML файла.
Распространение online	Можно распространять линк на ресурс (URL).	Можно распространять линк на ресурс (URL).
Распространение offline и хранение	Можно распространять (пересылать по эл. почте, на флешке). Можно хранить. Доступ необходим только в момент скачивания файла.	Можно распространять линк на ресурс (URL), а не сам файл. Хранить невозможно. Необходим постоянный доступ к ресурсу.
Сходство с "оригинальной" печатной версией	На всех устройствах выглядит одинаково и является копией печатной версии.	Вид может варьироваться в разных браузерах, при разном разрешении экрана или размере окна.
Наличие номеров страниц	да	нет

Функционал	PDF	HTML
<p>Обновление, живая публикация [6]. Индексация поисковыми системами.</p>	<p>При обновлении текста создается новый PDF файл (новая версия). URL может быть изменен. Ранее скаченная и сохраненная offline версия не обновится. При сохранении старого URL возможны проблемы с доступом - PDF файлы кешируются браузерами. Поисковые системы их сканируют с задержкой.</p>	<p>URL обновленной версии сохраняется прежним, читатель всегда получает обновленную версию. Быстрая индексация поисковыми системами.</p>
<p>Навигация, гиперссылки</p>	<p>Возможны гиперссылки внутри PDF и внешние гиперссылки из текста.</p>	<p>Возможны внутренние и внешние гиперссылки, интерактивные всплывающие меню, всплывающие подсказки, "share links" (ссылки на реферативные и библиографические ресурсы)</p>
<p>Удобство чтения с экрана монитора</p>	<p>Программа просмотра PDF предоставляет необходимый функционал для масштабирования и листания</p>	<p>Интернет браузер предоставляет необходимый функционал для масштабирования и листания</p>
<p>Удобство чтения с экрана мобильного телефона</p>	<p>Чтение возможно, но не удобно, если при создании PDF не была учтена возможность просмотра на маленьких экранах.</p>	<p>Современные информационные системы адаптированы для просмотра web-страниц на экранах мобильных устройств.</p>
<p>Дополнительные материалы. Вставки мультимедиа материалов, анимации и видео</p>	<p>Возможно, но сильно утяжеляет PDF файл, требуется дополнительное программное обеспечение [1]. Совместим не со всеми программами просмотра PDF.</p>	<p>Возможно добавление мультимедиа файлов, видео, анимации, которые проигрываются браузером. Возможны доп. материалы в виде отдельных файлов или линков на внешние ресурсы.</p>

Функционал	PDF	HTML
Отображение математических формул	Формулы создаются с учетом требований системы набора и верстки (LaTeX, InDesign, MS Word) и их расширений.	Формулы оформляются в виде картинок и/или в LaTeX или MathML, отображение в браузере осуществляется при помощи библиотеки MathJax [7]
Обновляемая статистика просмотров (article metrics)	нет	да
Браузер иллюстраций и таблиц	нет	да
Расширенный функционал списка литературы	Только гиперлинки на DOI или одну реферативную базу данных	Просмотр ссылки из текста статьи при наведении мыши, гиперлинки на различные базы данных, экспорт списка литературы
Функция "Переслать" (Share links) для элементов и всей публикации	нет	да
Обновляемый список цитирования	нет	да
Возможности экспорта	нет	Экспорт иллюстраций и таблиц в PowerPoint. Экспорт метаданных и списка литературы в форматы BIBTEX, RIS и т.п.
Обсуждение публикации	нет	Возможна вставка блока обсуждения публикации

Таблица 1. Основные требования, предъявляемые к современным научным публикациям, и их реализация в форматах PDF и HTML

Важно отметить, что обе версии статьи (PDF и HTML) содержат одну и ту же научную информацию и имеют идентичную научную ценность. При

цитировании не важно, какая версия была прочитана. Отличия носят технический характер, каждая версия имеет свои преимущества и недостатки.

PDF версия имеет следующие основные преимущества перед HTML:

1. Широко распространенные системы набора и верстки.
2. Хранение и распространение offline.
3. Полное сходство с "оригинальной" печатной версией при просмотре на любых устройствах и при печати.

HTML версия имеет следующие основные преимущества перед PDF:

1. Возможность добавлять (встраивать в текст статьи) мультимедиа материалы - видео, аудио, анимированные рисунки, анимированные графики.
2. Расширенный функционал гиперссылок и списка литературы.
3. Возможность экспорта элементов (иллюстраций и таблиц) публикации в PowerPoint.
4. Функционал "share links" (ссылки на реферативные и библиографические ресурсы).
5. Удобный просмотр с экранов мобильных устройств.
6. Добавление обновляемых статистических данных, списка цитирования и других данных, могут изменяться во времени.

Примеры расширенного функционала HTML версии можно увидеть в статьях математических журналов, представленных в информационной системе IOPScience [8].

Какую версию предпочитают читатели?

В работе [9] описан опрос, проведенный компанией Elsevier среди читателей научных журналов, представленных на платформе ScienceDirect, какую версию статьи они предпочитают. Выяснилось, что читатели предпочитают бегло просмотреть HTML версию для того, чтобы понять, соответствует ли публикация их ожиданиям, и какую научную ценность она представляет. Для внимательного прочтения они предпочитают скачать и даже распечатать PDF файл. Среди преимуществ PDF файла отмечалось то, что его можно сохранить и переслать. HTML формат удобен для доступа online и быстрого изучения материала. При этом если публикация содержит интерактивные и мультимедиа материалы, HTML формат становится предпочтительней PDF, но если таких материалов нет и статья "традиционная", читатели выберут PDF. Также читатели отмечали расширенный функционал литературных ссылок HTML версии, когда ссылку можно просмотреть, не покидая текущий фрагмент публикации.

Таким образом, можно сделать вывод, что научное сообщество в целом консервативно и предпочитает иметь дело со "старым добрым" PDF файлом, но при этом готово изучать новый формат, если это необходимо.

Создание HTML версии полного текста

Выше отмечалось, что создание PDF файла не представляет особого труда, его можно экспортировать из любой современной издательской системы. С HTML версий готовых решений не существует. HTML версия полного текста - это часть информационной системы (веб-сайта) издателя, она должна соответствовать общему дизайну и функционалу системы. Логично предположить, что информационная система использует стандартный шаблон для формирования HTML файлов всех статей, а сами статьи хранятся в формате XML. Для преобразования XML в HTML, который отображается на сайте издателя, используется XSLT преобразование. Таким образом, для каждой статьи необходимо подготовить ее полный текст в формате XML, совместимом с используемым XSLT преобразованием, то есть XML файл, стандартный для информационной системы издателя.

Издатели продолжают производить PDF файлы статей, процесс производства статьи включает такие этапы как набор и верстка, редакторская правка, авторская правка, внесение исправлений, верстка выпуска журнала. Для того чтобы избежать двойной работы на каждом этапе, необходимо объединить процессы получения PDF и XML файлов, сделать так, чтобы они получались из одного источника, и таким источником должен стать файл конечной версии статьи, подготовленный в используемой издательством системе верстки, то есть LaTeX, InDesign или MS Word файл. Как InDesign, так и MS Word имеют встроенные возможности экспорта в XML, поэтому основной задачей, которую необходимо решить, является создание соответствующего шаблона для экспорта документа в XML. В случае верстки в LaTeX требуется написание конвертора из LaTeX файла в XML.

HTML версии полных текстов российских математических журналов

Математические журналы Математический сборник, Известия РАН, серия математическая и Успехи математических наук, издателем русской версии которых является Математический институт им. В.А. Стеклова РАН, а также их английские версии Sbornik: Mathematics, Izvestiya: Mathematics и Russian Mathematical Surveys, издаваемые Turpion Ltd. (<http://www.turpion.org>) и распространяемые IOP Publishing (<http://iopscience.org>), верстаются в LaTeX. Уже несколько лет английские версии этих журналов представлены в системе IOPScience в обоих форматах - как PDF, так и HTML. Для подготовки XML файлов используется специальным образом написанный конвертор, затем XML файлы одновременно с PDF файлами загружаются в информационную систему. Исходным файлом для производства как PDF, так и XML является LaTeX файл,

подготовленный редакцией. Разметка LaTeX файла, в том числе блоки `\section`, `\subsection` и т.п. переносятся в соответствующие блоки XML файла. Математические формулы оформляются в виде картинок (gif файлов), конвертор их производит автоматически, запуская команду `pdflatex` для каждой формулы отдельно, а затем конвертируя pdf файл формулы в формат gif. Также для отображения формул используется библиотека MathJax [7], для этого в XML сохраняется исходный текст формулы в LaTeX. Ниже приведены фрагменты XML файла, содержащие строчную и выносную математические формулы, XML содержит и ссылку на картинку и исходный текст формулы в LaTeX. На странице полного текста в формате XML есть переключатель между режимами отображения формул в виде картинок или MathJax (рис. 1).

```

<inline-formula>
<tex-math>
<?CDATA $p\geqslant 1$?></tex-math>
<inline-graphic xlink:href="MSB_209_5_605ieqn8.gif"/>
</inline-formula>

<disp-formula>
<label>1.6</label>
<tex-math><?CDATA
\begin{equation}
M_n(f(\Gamma))\leqslant KM_n(\Gamma),
\end{equation}
?></tex-math>
<graphic xlink:href="MSB_209_5_605eqn6.gif"/>
</disp-formula>

```



Рисунок. 1. Переключатель “Turn on/off MathJax” на странице полного текста статьи в формате HTML на сайте IOPScience.

Структура LaTeX файла русской и английской версий идентична, а указанный конвертор можно применять как для конвертации английских, так и русских версий статей. Для портала Math-Net.Ru, распространяющего русские версии указанных журналов, подготовлен шаблон для вывода HTML версий полных текстов статей математических журналов. Пример приведен на рис. 2.

Math-Net.Ru
Математический сборник

ЖУРНАЛЫ ПЕРСОНАЛИИ ОРГАНИЗАЦИИ КОНФЕРЕНЦИИ СЕМИНАРЫ ВИДЕОТЕКА ЛИЧНЫЙ КАБИНЕТ

Матем. сб., 2018, том 209, номер 1, страницы 37–57

О структуре функций, универсальных для классов L^p , $p \in (0, 1)$

М. Г. Григорян^а, А. А. Саргсян^б

^а Ереванский государственный университет, Армения
^б Российско-Армянский (Славянский) государственный университет, г. Ереван, Армения

Аннотация:

Работа проливает свет на структуру функций, универсальных для классов L^p , $p \in (0, 1)$, относительно знаков коэффициентов Фурье–Уолша. Доказано существование такого измеримого множества $E \subset [0, 1]$ со сколь угодно близкой к-1 мерой, что надлежащим изменением значений любой функции $f \in L^1[0, 1]$ вне E можно получить функцию $\tilde{f} \in L^1[0, 1]$, универсальную для каждого класса $L^p[0, 1]$, $p \in (0, 1)$, относительно знаков коэффициентов Фурье–Уолша. Библиография: 28 названий.

Ключевые слова:

универсальная функция, коэффициенты Фурье, система Уолша, сходимость в-метрике.

Section § 1.

Введение

Вопросы существования функций, универсальных в том или ином смысле в различных функциональных классах, заинтересовали математиков начиная с первой половины XX века. В 1929 г. Дж. Биркофф (см. [1]) доказал существование целой функции $g(z)$, которая универсальна относительно сдвигов, а именно, для любой целой функции $f(z)$ и числа $r > 0$ существует возрастающая последовательность натуральных чисел $\{n_k\}_{k=1}^{\infty}$ такая, что последовательность сдвигов $\{g(z + n_k)\}_{k=1}^{\infty}$ равномерно сходится к $f(z)$ на круге $|z| \leq r$. В 1952 г. Дж. Маклейн (см. [2]) доказал аналогичный результат для другого типа универсальности. Он показал, что существует целая функция $g(z)$, которая универсальна относительно производных, а именно, для любой целой функции $f(z)$ и числа $r > 0$ можно найти возрастающую последовательность натуральных чисел $\{n_k\}_{k=1}^{\infty}$ такую, что последовательность производных $\{g^{(n_k)}(z)\}_{k=1}^{\infty}$ равномерно сходится к $f(z)$ на круге $|z| \leq R$. Далее, в 1975 г. С. Воронин (см. [3]) доказал теорему универсальности дзета-функции Римана, а именно, если $0 < r < 1/4$ и функция $f(s)$, аналитическая внутри круга $|s| \leq r$ и непрерывная вплоть до его границы, не имеет нулей внутри этого круга, то для всякого $\varepsilon > 0$ существует вещественное число $T = T(\varepsilon)$ такое, что

$$\max_{|s| \leq r} \left| f(s) - \zeta \left(s + \frac{3}{4} + iT \right) \right| < \varepsilon.$$

Рисунок 2. Полный текст статьи в формате HTML на сайте Math-Net.Ru

Заключение

PDF файл остается основным способом представления полного текста научной публикации. Рассмотренный в данной работе функционал HTML версии статьи в настоящее время является ценным дополнением к "стандартному" PDF файлу, но пока не может заменить его. В ближайшем будущем западные издатели будут продолжать выпускать обе версии статей. К этому непременно придут и издатели русских версий научных журналов.

Работа подготовлена при поддержке программы Президиума РАН № 01 "Фундаментальная математика и ее приложения" (грант PRAS-18-01).

Литература

1. Горбунов-Посадов М. М., Ролдугин Д. С., Слепенков М.И., Тузов И.В. Анимация и видео в научной публикации // Препринт ИПМ № 104, Москва, 2014.
2. E. Zudilova-Seinstra. Designing the Article of the Future. // Elsevier connect. 16 Jan 2013. — URL: <https://www.elsevier.com/connect/designing-the-article-of-the-future>
3. Article evolution starts here. 12 Sep 2011. — URL: <http://iopublishing.org/article-evolution-starts-here/>
4. APS Launches Full-Text HTML. — URL: <https://journals.aps.org/edannounce/aps-launches-full-text-html>
5. M. Roncevic. E-book Platforms for Libraries. // Library Technology Reports, 2013, 49(3), p. 28.
6. М.М. Горбунов-Посадов. Живая публикация // Открытые системы, 2011, № 4, С. 48–49. — URL: <http://keldysh.ru/gorbunov/live.htm>
7. MathJax Project. — URL: <https://www.mathjax.org>
8. Article evolution. IOPScience. — URL: <http://iopscience.iop.org/info/page/articleevolution>
9. I.J. Aalbersberg. PDF versus HTML - which do researchers prefer? // Elsevier connect. 9 Jul 2013. — URL: <https://www.elsevier.com/connect/pdf-versus-html-which-do-researchers-prefer>

References

1. M.M. Gorbunov-Posadov, D.S. Roldugin, D.S. Slepenkov, I.V. Tuzov. Animation and video in the scientific publication // Preprint IPM № 104, Moscow, 2014.
2. E. Zudilova-Seinstra. Designing the Article of the Future. // Elsevier connect. 16 Jan 2013. — URL: <https://www.elsevier.com/connect/designing-the-article-of-the-future>
3. Article evolution starts here. 12 Sep 2011. — URL: <http://iopublishing.org/article-evolution-starts-here/>
4. APS Launches Full-Text HTML. — URL: <https://journals.aps.org/edannounce/aps-launches-full-text-html>

5. M. Roncevic. E-book Platforms for Libraries. // Library Technology Reports, 2013, 49(3), p. 28.
6. M.M. Gorbunov-Posadov Live publication // Otkryte systemy, 2011, № 4, C. 48–49. — URL: <http://keldysh.ru/gorbunov/live.htm>
7. MathJax Project. — URL: <https://www.mathjax.org>
8. Article evolution. IOPScience. — URL: <http://iopscience.iop.org/info/page/articleevolution>
9. I.J. Aalbersberg. PDF versus HTML - which do researchers prefer? // Elsevier connect. 9 Jul 2013. — URL: <https://www.elsevier.com/connect/pdf-versus-html-which-do-researchers-prefer>