

# TEX В ЭПОХУ ЭЛЕКТРОННЫХ ПУБЛИКАЦИЙ\*

Е. В. ПАНКРАТЬЕВ, О. Б. ХРИСТЕНКО, В. А. МИТЮНИН

*Московский государственный университет, Россия*

e-mail: pankrat@shade.msu.ru

The present article represents the extended version of the report made by the authors on the Workshop EIPub2001. It is based on experience of the natural-scientific publications preparation with use of various publishing systems, first of all, system TEX. The students, post-graduate students and fellows of Mechanics and Mathematics Department of MSU have introduced the powerful contribution to development and distribution of TEXrussified version.

## Введение

Последнее десятилетие XX века ознаменовалось быстрым проникновением компьютерных технологий в издательскую деятельность. При этом разные задачи, стоящие перед издательствами, налагают различные, часто противоречивые, требования к используемым программным продуктам.

В нашей статье мы обратим основное внимание на проблемы подготовки естественнонаучной литературы. Основная ее особенность — наличие в тексте значительного количества сложных формул.

## 1. До TEX'а

Внедрение компьютерных технологий в издательский процесс проходило в нашей стране достаточно болезненно. На начальном этапе основные сложности носили чисто административный характер. Так, в 1988 году один из авторов статьи принес в издательство МГУ учебное пособие, подготовленное на компьютере (если можно назвать таковым YAMANA) и распечатанное на матричном принтере. В Издательстве долго отказывались брать рукопись. В ректорате автору показали инструкцию, в которой было написано: “Рукопись должна быть напечатана на пишущей машинке с очком-литерой, с черной лентой средней жирности . . .” И только благодаря активной поддержке главного редактора Издательства МГУ, с которым автор был знаком, рукопись была в конце концов принята. И хотя полиграфическое качество издания было не слишком высоким, путь компьютерному набору был открыт.

---

\*Работа выполнена при поддержке Российского фонда фундаментальных исследований, грант 99-07-90224.

© Е. В. Панкратьев, О. Б. Христенко, В. А. Митюнин, 2003.

Первой программой, получившей широкое применение при подготовке к изданию литературы по естественным наукам на персональных компьютерах типа IBM PC, был редактор Chiwriter, разработанный Каем Хорстманом. Этот редактор позволял набирать достаточно сложные формулы, поскольку в нем имелось несколько шрифтов, содержащих буквы латинского, греческого и кириллического алфавитов, а также различные математические и другие специальные символы. Кроме того, в нем имелись простые средства для создания и подключения новых шрифтов. В этих возможностях, однако, таились и существенные недостатки данного редактора. Как правило, российские пользователи начинали освоение этого редактора с создания своих шрифтов. При этом открывались богатые возможности проявить фантазию не только при создании дизайна шрифта, но и в установлении соответствия между клавишами клавиатуры и символами шрифта. В частности, существовали два подхода к раскладке букв русского алфавита по клавиатуре: один использовал раскладку, принятую на пишущих машинках, другой — устанавливал соответствие между русскими и латинскими буквами, при этом русская буква “а” ставилась в соответствие той же клавише, что и латинская “a”. При этом “лишние” символы (такие как ю, ч, ъ, э и др.) распределялись произвольным образом. В результате замена имени шрифта приводила к искажению текста. Более того, под одним и тем же именем у разных авторов могли существовать совершенно различные шрифты (некоторые авторы даже меняли авторские шрифты Хорстмана), часто шрифты создавались только для одного конкретного типа монитора или принтера. В результате возникали проблемы переноса текстов с одного компьютера на другой: часто было недостаточно просто скопировать шрифты или даже полностью весь редактор, поскольку периферийные устройства компьютеров были различными.

Качество печати также оставляло желать лучшего. Фактически Chiwriter был реализован как пишущая машинка с достаточно большим набором сменных головок: расстояние между строками измерялось в интервалах, длина строки задавалась в символах. При этом шрифты были не только равноширинные; даже в равноширинных шрифтах ширина символов в разных шрифтах существенно различалась. В результате при фиксированной длине строки (в символах) попытка “выключки” (выравнивания как левой, так и правой границ) приводила к тому, что некоторые строки оказывались длиннее, а в других появлялись неестественно большие пробелы.

## 2. С T<sub>E</sub>X’ом

В начале 90-х годов отечественные издательства начали использовать для подготовки естественно-научной литературы систему T<sub>E</sub>X, которая к настоящему времени стала фактически стандартной системой для подготовки, хранения и передачи естественно-научной информации, выигрывая в этой области у текстовых процессоров типа Word. Успех системы T<sub>E</sub>X обеспечен достаточно высоким качеством выходных документов, хорошей переносимостью (в том числе на различные платформы), компактностью представления информации.

Наш опыт общения с T<sub>E</sub>X’ом начался с того, что в 1990 году Издательство МГУ предложило нам организовать подготовку оригинал-макетов английской версии журналов “Вестник Московского университета”. В то время английская версия журналов готовилась в США с минимальными затратами (текст распечатывался на матричном принтере, формулы, даже небольшие, вырезались из русских журналов и клеивались). Нам предложи-

ли использовать для подготовки оригинал-макетов какой-либо редактор типа Chiwriter. Но мы решили готовить журнал в TeX'e, хотя опыта работы с ним не имели. В течение года требовалось подготовить более 40 выпусков (серии: Математика, Механика, Вычислительная математика, Физика, Химия и др.). В нашем распоряжении была только группа студентов 2–3 курсов.

При первом знакомстве TeX поразила нас не только высоким качеством выходных документов, но и высокой совместимостью с различными системами. В то время в обращении находилось несколько систем, работавших под управлением операционной системы MS-DOS: PCTEX, ARBORTEX, BSTEX и др. Совместимость различных систем осуществлялась не только на уровне исходных текстов, но и на уровне .dvi-файлов: документы, подготовленные в какой-либо системе, могли просматриваться и распечатываться в любой другой, и результаты оказывались тождественными. Тексты готовились в латинском алфавите, возможности использования кириллицы были очень скромные. Высокая степень переносимости — один из основных принципов, заложенных Д. Кнудом в разрабатываемую им систему. Хорошо известна его резкая реакция [?] на появление в ОС LINUX cm-шрифтов с измененными метрическими характеристиками.

Использование TeX'a в России началось еще в 80-е годы. Пионерами в этом деле выступили сотрудники Института физики высоких энергий, не только освоившие TeX, но и разработавшие, по-видимому, первую русифицированную систему на его базе. К 1990 году первое знакомство с TeX'ом состоялось у нескольких коллективов. Часть из них (например, сотрудники Издательства "Мир") ездили в США на курсы, организованные Американским математическим обществом, но в основном люди осваивали TeX самостоятельно (в том числе и наша группа). В 1990 году была организована Ассоциация пользователей кириллического TeX'a (CyrTUG), просуществовавшая до декабря 1999 г. В состав инициативной группы входили представители Ленинградского и Московского университетов, издательства "Мир", ИФВЭ, ИПМ, ЦЭМИ. Среди основных целей Ассоциации были пропаганда и распространение TeX'a в России и координация работ по созданию русифицированной системы TeX. И если с первой задачей Ассоциация справилась весьма успешно, хотя успех TeX'a обусловлен, в первую очередь, его несомненными достоинствами, то с созданием единой русифицированной версии дело обстояло не столь гладко. В течение нескольких лет активно развивались различные, часто несовместимые версии русифицированного TeX'a. Отличия возникали и на уровне .tex-файлов, в частности, одни системы допускали использование в командах русских букв, а другие — нет. Проблемы переносимости на уровне .dvi-файлов гораздо сложнее: различные шрифты далеко не всегда могут быть объединены в одной системе.

Распространение TeX'a в России имело свои особенности. Прежде всего требовалось создание качественных кириллических шрифтов. Ситуация со шрифтами для TeX'a отличается от аналогичной ситуации с редактором Chiwriter, поскольку одним из преимуществ системы TeX является высокое полиграфическое качество шрифтов, что, естественно, предъявляет дополнительные требования к их разработчикам. Тем не менее несколько семейств кириллических шрифтов были разработаны и используются в различных издательствах. Они различаются метрическими свойствами, в результате чего .tex-файл, отформатированный в одной системе, в другой системе требует нового форматирования (если даже эти системы совместимы по используемым макрокомандам). При этом не всегда возможно подключить одновременно шрифты различных семейств. Большинство русификаций для подключения кириллических шрифтов использовали механизм виртуальных шрифтов, когда латинская и кириллическая части шрифта размещались в различных файлах. В то

же время в русификации, выполненной А. Шенем и С. Львовским, семибитный латинский шрифт был дополнен до восьмибитного введением букв кириллического алфавита. При этом шрифты сохранили имена, которые соответствующим латинским шрифтам присвоил Д. Кнут. Это существенно затрудняло поддержание в единой системе шрифтов А. Шеня и каких-либо других шрифтов, использовавших оригинальные *cm*-шрифты Д. Кнута, и создавало серьезные проблемы совместимости. Такое решение имело неожиданный побочный эффект.

В рекламных материалах, посвященных этой системе, почеркивалась ее универсальность, говорилось, что нет необходимости различать кириллические и латинские шрифты и что на текстах, которые не содержат русских букв, система работает идентично стандартной западной. Однако в действительности отечественному пользователю нужно, чтобы система работала идентично западной на текстах, которые *не должны* содержать русских букв. Основной поток таких документов получается при авторской подготовке статей для зарубежных издательств, переводимых с русского языка на английский. В таких переводах авторы забывали перевести отдельные слова, чаще всего предлоги и союзы. Так как базовый шрифт был восьмибитным, система не выдавала сообщений об ошибках в случае, когда в тексте остались отдельные русские слова. При компиляции текста в классическом  $\TeX$ 'е возникали ошибки. Кроме того, условия, на которых Д. Кнут предоставил систему в свободное пользование, грубо нарушались. Однако система Шеня — Львовского получила в свое время достаточно широкое распространение. Не в последнюю очередь этому способствовало то, что наиболее распространенная в России книга по  $\LaTeX$ 'у [?] была подготовлена с использованием этой системы.

Первые кириллические шрифты были созданы в США Д. Вулисом. Они предназначались для набора коротких цитат или библиографических ссылок. Соответственно их качество было не самым высоким. В России первые кириллические шрифты для системы  $\TeX$  были созданы в ИФВЭ Самариным и Глонти, в настоящее время они известны как “протвинские”. Сегодня эти шрифты используются многими пользователями и любой желающий может найти их на CTAN. Кириллическая часть шрифтов А. Шеня представляет собой модификацию протвинских шрифтов. Другая модификация протвинских шрифтов используется А.И. Роженко в созданной им системе  $\text{NCS-}\TeX$ . Также весьма распространенными кириллическими шрифтами являются шрифты LN-семейства (шрифты Лапко — Ходулева). Они также имеются на CTAN. В некоторых академических журналах (“Математические заметки”, “Успехи математических наук” и др.) применяют шрифты С.В. Знаменского.

Итак, использование  $\TeX$ 'а в МГУ началось с подготовки оригинал-макетов английской версии журналов “Вестник Московского университета”. Подготовив в течение года более 40 выпусков журнала, коллектив приобрел значительный опыт работы с  $\TeX$ 'ом. На следующий год приступили к подготовке книг на русском языке. Появились новые проблемы. Во-первых, в нашем распоряжении имелась коммерческая версия системы  $\text{RSTeX}$  с “вашингтонскими” шрифтами, не поддерживающая механизм виртуальных шрифтов. В своей работе мы опирались на пакет макрокоманд  $\text{AMS-}\TeX$  и долгое время практически не пользовались  $\LaTeX$ 'ом. Это обусловлено тем, что в набор поступали уже готовые рукописи, а преимущества  $\LaTeX$ 'а сказываются при авторской подготовке текста.

Для подготовки русскоязычной литературы  $\TeX$  нужно было русифицировать. Нами были написаны препроцессор, постпроцессор и некоторый комплекс вспомогательных программ. Сложности начались при общении с работниками Издательства МГУ. Нам удалось уговорить их снять некоторые требования, выполнение которых в  $\TeX$ 'е затруднительно

(например, дублирование знаков арифметических операций при переносах формул). Но осталось условие, что каждая книга должна выглядеть по-своему. Таким образом, для каждой книги приходилось готовить свой стилевой файл, часто очень нетривиальный.

Взаимодействие программистов с работниками издательств приводило, с одной стороны к созданию все более изоциренных стилевых файлов, а с другой — к изменению требований издательств к оформлению книг. Однако труднее всего было изменить эти требования в издательствах, выпускающих учебную литературу для школьников. Одно из таких требований — поддержание четырех типоразмеров шрифтов в математических формулах (TeX использует три типоразмера). В этой связи особо хочется выделить разработку М.И. Гринчука ARTeX [?], наиболее полно учитывающую отечественные полиграфические традиции. Его система выполнена на базе коммерческого пакета Vector TeX с использованием специального формата шрифтов, открывающего широкие возможности для их модификации. Начат перенос системы ARTeX для использования с системой emTeX.

Развитие Интернета привело к значительной унификации существующих русификаций системы TeX. В частности, используемые в среде Windows системы MikTeX и frTeX базируются на “легальных” кириллических шрифтах, хранящихся на CTAN, в первую очередь, семейства LN. Более того, система frTeX полностью совместима с системой teTeX, эксплуатируемой под ОС LINUX. В разработке русификаций, представленных на CTAN, большую роль сыграл В.В. Волович — один из главных консультантов конференции по русифицированному TeX’у, организованной в Интернете.

Одним из недостатков, препятствующих повсеместному распространению TeX’a в России, является относительная сложность (по сравнению, к примеру, с MS Word) работы с системой. Для большинства пользователей, начинающих осваивать TeX, требуется некоторое время на адаптацию к системе. В рамках проекта для решения этой проблемы в Протвино была разработана система справочной информации по пакетам AMS-TeX и L<sup>A</sup>TeX на базе HTML. Система обеспечивает интерактивный доступ к справочной информации через Интернет, также существует версия системы на CD-ROM. С момента создания справочной системы прошло достаточно много времени, и, несомненно, она нуждается в обновлении, в частности, это касается команд пакета L<sup>A</sup>TeX2e.

К сожалению, далеко не каждый пользователь может позволить себе “скачать” через Интернет систему типа frTeX, занимающую более 200 Мб. Мы записали эту систему на компакт-диск, добавили туда русифицированную версию системы emTeX и справочную систему, подготовленную в Протвино и пользующуюся очень большим успехом у пользователей. В течение месяца к нам обратились за этим диском более десяти пользователей.

Использование систем на базе TeX’a для подготовки электронных публикаций может предоставить ряд преимуществ, в частности: сохранение технологии подготовки документов, возможность “синхронизировать” электронные и обычные публикации при сохранении высокого полиграфического качества последних. Одно из главных преимуществ — возможность перевода в электронную форму существующих материалов — статей, книг и др.

Одним из решений, позволяющих сочетать преимущества TeX’a и возможности, предоставляемые более современными форматами, например PDF, является разработка В.К. Малышева ВАКОМА-TeX. Это интегрированная система, предоставляющая пользователю возможности подключения шрифтов в форматах PostScript, TrueType и OpenType, а также достаточно мощные средства интерактивной отладки. Альтернативный подход — преобразование в PDF документов, подготовленных в “классических” версиях системы TeX, — применен в программе DVIPDF, разработанной С. Лесенко.

В связи с перечисленными выше особенностями использования системы  $\text{\TeX}$  в России возникает вопрос о необходимости существования наряду с СТАН'ом российского информационного ресурса в Интернете, содержащего разработки на базе  $\text{\TeX}$ 'а, адаптированные к российским стандартам, а также стилевые и форматные файлы, необходимые как для оформления научной документации (диссертаций, отчетов в РФФИ) по общероссийским стандартам, так и для подготовки публикаций в научной периодике. В последнем случае многое зависит от желания руководства журналов или издательств. При этом, как показывает опыт, весьма полезным является дублирование информации на CD, что обеспечивает возможность использования ресурса теми пользователями, кто подключен к сети Интернет через низкоскоростной канал или вообще не имеет доступа в сеть.

Разработка такого ресурса в настоящий момент ведется при поддержке РФФИ (см. [ftp://ftp.ict.nsc.ru/pub/TeX/CD\\_MSU/](ftp://ftp.ict.nsc.ru/pub/TeX/CD_MSU/)).

### 3. Что дальше?

Сегодня единого стандарта публикации математических документов в сети Интернет не выработано. Существует несколько подходов к решению этой проблемы. Например, используются программы для конвертации исходников  $\text{\TeX}$ 'а в HTML-странички. Этот подход не очень удобен, так как в результате часто появляется достаточно большое количество вспомогательных файлов даже на документ небольшого размера. Другие подходы подразумевают установку специальных расширений для программ просмотра, которые позволяют просматривать документы определенного формата, как правило, полученного из исходников  $\text{\TeX}$ 'а. Оба этих подхода обладают недостатком: в обоих случаях затруднена интеграция фрагмента математического документа в другой документ. Есть однако еще один вариант решения данной проблемы, который претендует на то, чтобы стать стандартом публикации математических документов в Интернете, — это использование языка XML. Существует проект определения типа документа (DTD), направленный на публикацию математических документов, называемый MML — Mathematical Markup Language. Он уже поддерживается в качестве выходного формата в ряде ведущих систем компьютерной алгебры наряду с  $\text{\TeX}$ 'ом и позволяет избавиться от указанных выше проблем. Не исключено, что на просторах WWW именно этот подход окажется всеобщим стандартом, вытеснив  $\text{\TeX}$ . Следует отметить, однако, что этот подход имеет определенные преимущества только при электронной публикации документов, но не при ручном их наборе.

До недавнего времени проблеме публикации и просмотра математических документов в сети Интернет не уделялось должного внимания, и сложилась до некоторой степени парадоксальная ситуация: только благодаря науке стало возможным существование самой глобальной сети, а ведь математика — язык науки. Изменить состояние дел в настоящий момент взялось немалое количество разработчиков, в число которых входят все компании, производящие браузеры, однако задача далеко не тривиальна и ее полного решения еще никем не было предложено. Обзор текущего состояния средств создания и публикации математических документов на страницах сети Интернет, подготовленный В.А. Митюниным, находится по адресу <http://hash.dorms.spbu.ru/mathmag/other/mml/>.

В последние годы возник большой интерес к XML-технологиям. К наиболее важным их преимуществам можно отнести легкость понимания формата и возможность внесения исправлений с помощью обычного текстового редактора (human-readable), стандартные средства для обработки XML-документов, более строгие по сравнению с HTML правила

для составления документов, возможность создания своего собственного типа документа, наличие языка запросов, стилей и трансформаций. Безусловно, есть и недостатки, но вряд ли можно найти технологию, полностью их лишенную. Появилось большое количество программных продуктов для работы с XML. Однако их качество сильно различается. Очень непросто выбрать из этого разнообразия то, что лучше всего подходит для тех или иных целей. Нами была проведена работа по тестированию процессоров XML и XSLT на производительность и соответствие спецификациям W3C. С результатами можно ознакомиться на страницах сети Интернет по адресам

[http://tex.msu.ru/archives/xml\\_xslt\\_test.zip](http://tex.msu.ru/archives/xml_xslt_test.zip)

[http://tex.msu.ru/archives/xml\\_xslt\\_restest.rtf.gz](http://tex.msu.ru/archives/xml_xslt_restest.rtf.gz)

Из вышесказанного можно сделать вывод о том, что на самом деле, несмотря на развитие XML, Т<sub>Е</sub>X еще значительное время будет сохранять ведущие позиции, поскольку ничего лучше Т<sub>Е</sub>X'а пока не создано. Наличие большого объема уже подготовленных в системе Т<sub>Е</sub>X работ делает задачу создания доступной для рядового пользователя системы преобразования из Т<sub>Е</sub>X'а в более современные форматы достаточно актуальной. Именно в направлении интеграции Т<sub>Е</sub>X'а и более современных форматов представлений данных должно пойти в ближайшее время развитие систем на базе Т<sub>Е</sub>X'а.

## Список литературы

- [1] KNUTH D. E. Important Message to all Users of Т<sub>Е</sub>X // СурTUG'97: Тр. конф. 8–11 сент. 1997, Петергоф (Санкт-Петербург), Протвино, 1997. С. 184–187.
- [2] ЛЬВОВСКИЙ С. М. Набор и верстка в пакете Л<sup>A</sup>T<sub>Е</sub>X. 2-е изд. М.: Космоинформ, 1995.
- [3] ГРИНЧУК М. И. Проект АРТ<sub>Е</sub>X: русифицировать так русифицировать, или Пошто шевелить Кнотов Компилятор // СурTUG'97: Тр. конф. 8–11 сент. 1997, Петергоф (Санкт-Петербург), Протвино, 1997. С. 74–81.

*Поступила в редакцию 26 ноября 2001 г.,  
в переработанном виде — 9 сентября 2002 г.*