



Лидовский В.

Современная Т_ЕХнология

Прошло уже более 20 лет с тех пор, как профессор Станфордского университета Дональд Э. Кнут сделал доступной всем свою систему для подготовки научных публикаций Т_ЕХ¹. Но ее популярность продолжает расти и сегодня. Например, число посетителей главного Т_ЕХ-сайта Великобритании www.tex.ac.uk за последние 6 лет (с 1996 по 2002) увеличилось в 10 раз, достигнув более миллиона в месяц. Значительная часть современных научных трудов по математике, физике и программированию и некоторым другим создается на Т_ЕХе. Хотя Т_ЕХ используется прежде всего в научных кругах — это универсальная система, сложность которой, однако, делает ее практически недоступной для среднестатистического пользователя. Но для последнего есть широкий выбор возможностей: Microsoft Word, Sun Star Writer, AbiWord для Linux, ... Относительно редкая частота использования Т_ЕХ для научных публикаций в России однозначно обозначает плохое знакомство с возможностями этой системы, во многом идеальной для своей области применения. Официальный российский Т_ЕХ-сайт www.cemi.rssi.ru/cyrtug влачит незаметное существование — он давно не обновлялся и с января 1998 по начало 2003 его посетили только около 27000 раз. А ведь Т_ЕХ произносится не как “текс”, а именно как “тех”, т. е. по-русски.

По сравнению с программами, подобными Microsoft Word, от не-

¹ Т_ЕХ — это торговая марка Американского математического общества — American Mathematical Society.

лицензионной копии которого автор отказался более пяти лет назад, \TeX -система не создает иллюзий относительно того, что сложные проблемы могут быть решаемы просто, дает полный контроль над обстановкой, имеет открытые и несложные для машинной обработки форматы данных, позволяет максимально быстро и качественно вводить математические формулы и, будучи созданной и развиваемой учеными в процессе работы над конкретными проектами, обеспечивает базу для здоровой научной прагматики.

Здесь стоит привести официальное мнение пользователей \TeX из США (цитата из документа “Just What Is \TeX ?” главного сайта TUG — \TeX Users Group). “Вопрос использования \TeX может послужить основанием для религиозных войн, т. к. те, кто используют его возможности для подготовки математических книг и статей, считают, что никакое сочетание других имеющихся средств не сможет обеспечить производство высококачественных полиграфических изданий при подготовке их самим автором. Мы с радостью предоставим материал для проверки тому, кто захочет доказать, что мы неправы. (Пользователей \TeX относительно мало и их проблемы слишком сложны, поэтому Quark, Microsoft или Adobe не проявляют заинтересованности). С другой стороны, если вам нужно средство для создания газеты, книги-романа, эффектного объявления или письма тете Генриете, то, если вы не используете его для чего-нибудь еще (скажем, диссертации), то \TeX — это средство не для вас.”

В силу того, что любая \TeX -система очень сложна (едва ли найдется хоть один человек, который знает все возможности ее современных вариантов) и состоит из большого числа разнотипных компонент, беглое знакомство с которыми требует нескольких лет ее интенсивного использования, для начинающих обычно необходимо иметь возможность получать консультацию от опытных \TeX нологов и иметь под рукой всю необходимую литературу, большая часть которой доступна только по-английски. Но не все так драматично — если использовать правильно сконфигурированный, наиболее популярный пакет \LaTeX , то начинающему нужна лишь минимальная стартовая поддержка, частью которой может послужить эта статья.

Ее история

Однажды, во-второй половине 1970-х Дональд Кнут получил гранки второго тома своего многотомника “Искусство программирования для ЭВМ”, посмотрел на него и сказал что-то похожее на “hex!”. Это была реакция на первые образцы новой компьютерной полиграфии, качество которой было настолько низкой, что он решил отказаться от издания с ее использованием. Подумав немного, он сказал приблизительно-

но следующее: “Я ученый-компьютерщик. Мне следует суметь что-то сделать с этим”. Это “что-то” превратилось в разработку системы, которая дает все возможности настоящей типографии кроме ощущения, может быть, запаха краски. Кнуту пришлось изучить традиционные типографские методы для качественной печати математических формул и подходы к разработке рисунков знаков, т. к. нужных шрифтов не существовало. Он предполагал, что ему понадобится 6 месяцев, чтобы справиться. В целом же для полного завершения работы понадобилось почти 10 лет и значительная помощь других специалистов.

Приблизительно через год после начала работы Кнут был приглашен Американским математическим обществом прочитать лекцию на свободную тему, возможно лишь косвенно связанную с чистой математикой. Он решил говорить о компьютерных науках на службе математики и представил основы своих новых разработок $\text{T}_{\text{E}}\text{X}$ и $\text{M}_{\text{E}}\text{TAFONT}^2$ (для создания шрифтов для $\text{T}_{\text{E}}\text{X}$). Тогда же были отмечены ряд привлекательных свойств $\text{T}_{\text{E}}\text{X}$:

- он предназначен для прямого использования авторами, которые точно знают, что им нужно;
- он пришел из академического источника;
- за него не нужно платить деньги (математики традиционно не богаты — карандаши, бумага, мел и доска — вот и все их наиболее важные исследовательские приборы до появления систем компьютерной алгебры);
- его входные и выходные данные одинаковы для компьютеров разных архитектур.

При создании своих собственных книг Кнут должен был иметь дело со всеми мелкими тонкостями академических публикаций: сносками, вставками и т. п. Для этого он создал язык для описания документа, пригодный не только для компьютера, но и для обмена информацией между людьми, и использовал алгоритмы для форматирования текста, которые остаются неулучшенными до сих пор.

Изначально в $\text{T}_{\text{E}}\text{X}$ использовалось 7-разрядное кодирование, но с 1990 года используется 8-разрядное и обеспечивается работа с документами, содержащими фрагменты на нескольких разных языках, что сделало его, в частности, весьма популярным в восточной Европе.

Кнут обещает вознаграждение всякому, нашедшему ошибку в его программах ($\text{T}_{\text{E}}\text{X}$ и $\text{M}_{\text{E}}\text{TAFONT}$) и соответствующих им пяти книгах [2–6]. Текущая цена ошибки — 327.68 \$.

² $\text{M}_{\text{E}}\text{TAFONT}$ — торговая марка компании Addison Wesley Publishing.

Состав и структура

Фундаментальные компоненты любой \TeX -системы — это программы `tex` (собственно \TeX), `mf` (METAFONT), `bibtex` (BIB \TeX), `mpost` (MetaPost) — их исходники на языке литературного программирования WEB (транслируется в Паскаль и затем, если нужно, в C++ перед компиляцией), доступны свободно в STAN:/systems/knuth. Используя WEB, можно получить одновременно программу и отличную документацию к ней [3, 5].

\TeX — это компилирующийся универсальный язык программирования. Энтузиасты даже написали на нем транслятор с Бэйсика.

Алгоритм на \TeX записывается, используя средства, подобные пре-процессору C++, т. е. макроподстановки, условную трансляцию, включение файлов и т. п., к которым добавлены возможности для работы с переменными и файлами, т. е. можно организовывать ветвления и, используя рекурсивные макросы, циклы, но не переходы. Файловые операции включают в себя только последовательные чтение и запись. Синтаксически \TeX -текст состоит из собственно текста, макросов и примитивных команд, причем последние внешне не различаются и выглядят за исключением нескольких особых случаев как последовательности, начинающиеся со специального escape-символа, за которым следуют либо буквы, либо одиночный символ. В Plain \TeX и L \TeX escape-символом является знак обратной наклонной черты (backslash), `\`, а в Texinfo — `@`. Макросы не обязательно вызывать явно — они могут вызываться сами в определенных контекстах, подобно конструкторам/деструкторам C++ или процедурам-демонам фреймовых систем.

\TeX имеет 256 регистров для хранения каждого из множества типов данных для работы со специфически полиграфическими объектами, т. е. наборных коробок для шрифтов, различных видов заполнителей вертикальных и горизонтальных промежутков и т. п. В \TeX имеется также 256 целочисленных регистров (для чисел из диапазона от -2^{31} до 2^{31}) и столько же регистров для хранения размеров до ≈ 5.7 метров величиной с точностью до ≈ 54 нанометра. Оба типа регистров различаются только по назначению и совместимы по присваиванию. С размерами и числами можно использовать четыре обычные арифметические операции. Однако \TeX дает возможность использовать гораздо более 256 переменных каждого типа. Дело в том, что числом 256 ограничивается только число переменных в одной области программы, а областей может быть много: одна, глобальная, есть всегда и можно сделать нужное число локальных, образуемых простым их заключением в фигурные скобки. Кроме того, локальные области могут вкладываться в друг в друга. Ситуация очень похожа на использование фигурных

скобок в C++. Есть в T_EX и 256 логических регистров и регистры для хранения разного рода специальной информации. Работа с логическими выражениями реализована по-минимуму. Нет, например, операции отрицания.

Недостатком T_EX является отсутствие средств для вызова других программ. Например, при подготовке указателей к книге список строк нужно отсортировать. Выполнять сортировку средствами самого T_EX слишком громоздко и неэффективно. Гораздо лучшим решением, было бы вызвать внешнее средство сортировки, например, программу `sort`, но это невозможно сделать изнутри T_EX-программы. На практике подобные задачи (среди которых стоит выделить работу с библиографией при помощи `ВivTEX` и построение указателей, например, при помощи `makeindex`) решаются двумя прогонами T_EXа, между которыми к данным применяются внешние средства.

Внесение в простой текст (`plain text`) дополнительной информации об его оформлении или структуре осуществляется при помощи разметки текста (`markup`). T_EX после подключения макроопределений задает разметку текста входных документов. Различают физическую (или процедурную) и логическую разметки.

При физической разметке точно указывается, что нужно сделать с выбранным фрагментом текста: показать курсивным, приподнять, центрировать, сжать, подчеркнуть и т. п. При логической разметке указывается структурный смысл выбранного фрагмента: примечание, начало раздела, конец подраздела, ссылка на другой фрагмент и т. п.

Логическую разметку всегда можно преобразовать в физическую, используя выбранный стиль. Имея наборы документов в логической разметке можно всегда при печати придавать им наиболее привлекательный вид, своевременно получая от специалистов-дизайнеров новейшие стили. Преобразование физической разметки в логическую формальными средствами практически невозможно.

Макропакеты `Plain` и `Eplain` T_EX задают процедурную разметку, а `LATEX`, `ConTEXt` и `Texinfo` — логическую. Последнее деление весьма условно: в `Plain` T_EX можно определить формат, соответствующий хорошей логической разметке, а, используя `LATEX`, можно применять в любых количествах средства физической разметки.

`METAFONT` используется для создания высококачественных растровых (`bitmap`) шрифтов. Шрифты для него описываются на специальном языке со значительной долей декларативности в семантике: рисунки знаков на нем задаются, как правило, системами уравнений, решаемых автоматически. Шрифты на этом языке — масштабируемые, растр знаков создается для конкретных устройств вывода, но не в ре-

альном времени интерпретацией как для шрифтов Adobe PostScript или TrueType, а предварительной компиляцией, т. е. METAFONT — это еще и растеризатор (raster image processor, RIP). Компиляция позволяет описывать сколь угодно сложные формы и проводить высококачественную растеризацию с коррекцией, но плата за это — время. Результат работы METAFONT — это не только двоичные матрицы для символов в файле формата pk (packed), но и файл метрик шрифта формата tfm (TeX font metric), в котором описываются размеры каждого символа шрифта. Кнут “заморозил” развитие METAFONT и поэтому эта программная система, имеющая возможности для развития и модификации не меньшие, чем сам TeX, практически не имеет вариантов.

ВивTeX — система, разработанная к концу 1980-х Ореном Поташником, для удобной работы с библиографией в TeX-документах [7]. Она является примером системы, задачи которой не могут быть решены простыми средствами.

MetaPost создан в AT&T Джоном Д. Хобби последним в серии необходимых компонент любой TeX-системы уже в первой половине 1990-х. Эта программа используется для подготовки графических иллюстраций. Она практически идентична METAFONT, отличаясь только в трех деталях: 1) создает не растровые картинки, а масштабируемые в формате PostScript; 2) не имеет поэтому возможностей для растеризации; 3) не имеет встроенных возможностей для показа рисунков знаков на экране. Ее можно использовать для преобразования шрифтов METAFONT в шрифты Adobe Type 3. Возможности MetaPost во многом перекрываются созданным к 1998 году графическим макропакетом Xy-pic.

Упрощенно взаимосвязи основных компонент классической TeX-системы представлены на схеме, из которой видно, что помимо собственной системы шрифтов можно использовать масштабируемые шрифты Adobe Type 1 и 3, а также TrueType. Использование фирменных шрифтов с TeX осложнено тем, что в них часто не определен полный диапазон нужных знаков. Остроту последней проблемы во многом снимают виртуальные шрифты, позволяющие объединять в себе знаки из разных шрифтовых файлов и задавать нужную кодировку. Программа Dvips обеспечивает подключение к TeX возможностей языка PostScript, которые значительно превосходят возможности языка dvi-файлов³ Д. Кнута, а программа Ghostscript (gs) позволяет как печатать ps-файлы на не PostScript принтере, так и конвертировать их в другие форматы, например, в pdf или растровые картинки. Работа с масштабируемыми не mf-шрифтами основана на Dvips и Ghostscript.

³ dvi — Device Independent — мобильный бинарный формат для результата компиляции документа TeX.

Возможности конкретной $\text{T}_{\text{E}}\text{X}$ -системы определяются подключаемым стандартным макропакетом. Кнут в своей книге “The $\text{T}_{\text{E}}\text{X}$ book” [2] описал макропакет Plain. В 1990-е Карл Берри и Стивен Смит значительно расширили его функциональные возможности своим макропакетом $\text{E}_{\text{r}}\text{plain}$ (Expanded Plain), доведя их до уровня $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ [7, 12]. Сам же $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ Лесли Лэмпорта, появившись еще в 1980-е и постоянно развиваясь, является самой значительной работой после той, которую проделал Кнут. К нему созданы сотни разнообразных стилей. Известный макропакет $\mathcal{A}\mathcal{M}\mathcal{S}\text{-T}_{\text{E}}\text{X}$ [10, 12], созданный Американским математическим обществом, ныне является частью $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$. Весьма интересным, появившимся в конце 1990-х, макропакетом является $\text{C}\text{o}\text{nT}_{\text{E}}\text{Xt}$ [8]. Он позволяет, в частности, создавать цветные интерактивные pdf-документы и печатать сложные химические структуры. Пока $\text{C}\text{o}\text{nT}_{\text{E}}\text{Xt}$ не поддерживает в полном объеме работу с текстами на русском. Особым макропакетом является $\text{T}\text{e}\text{xinfo}$ — это стандарт для ведения документации в GNU. Он позволяет из одного исходника получать как печатные издания, так и гипертекстные документы GNU info или html.

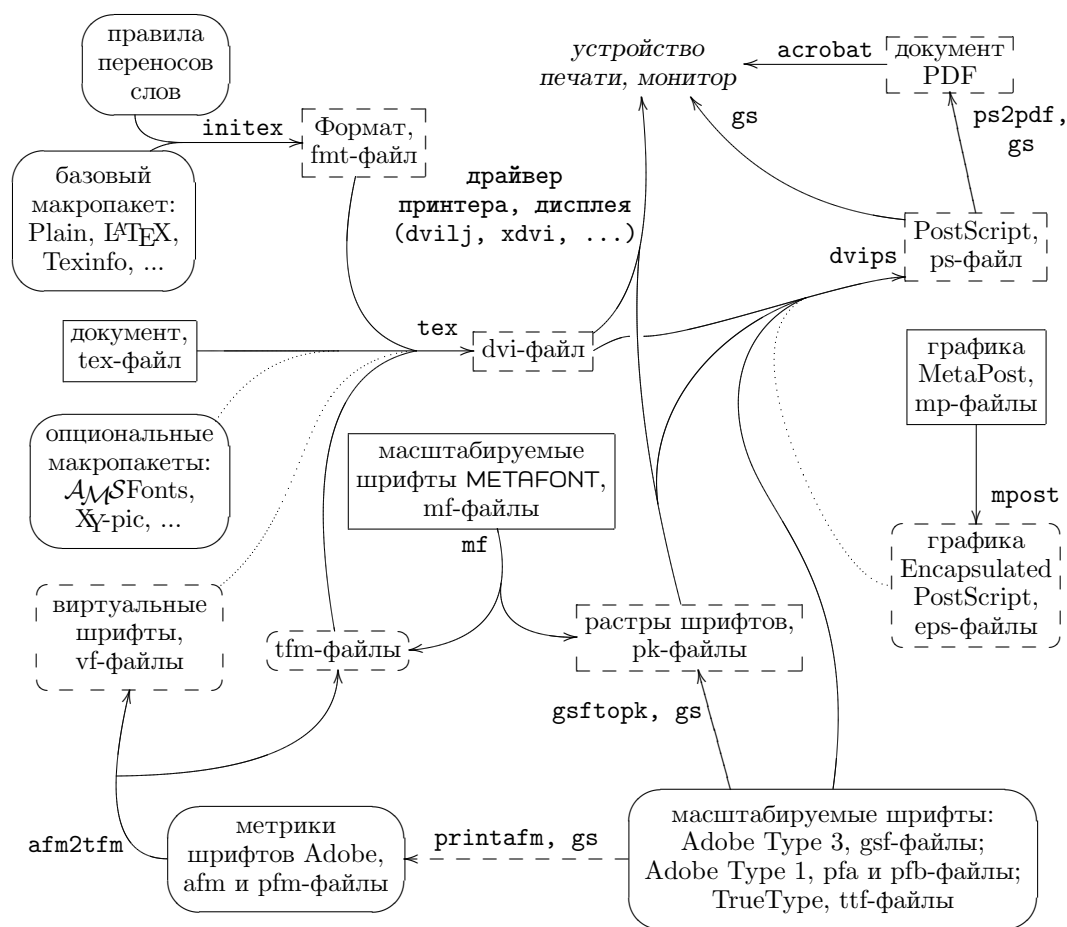


Схема.

Кроме базовых макропакетов к любой $\text{T}_{\text{E}}\text{X}$ -системе можно подклю-

чать макропакеты для расширения ее отдельных возможностей. Среди них стоит отметить графические $\text{P}\text{T}\text{E}\text{X}$ [11] и гораздо превосходящий его $\text{X}\text{Y-pic}$ [9], интересный еще и своим объектно-ориентированным языком описания рисунков, *Vabel* для поддержки работы с многоязычными текстами в $\text{L}\text{A}\text{T}\text{E}\text{X}$ и $\text{T}\text{r}\text{e}\text{e}\text{T}\text{E}\text{X}$ для рисования бинарных деревьев. Для TEX есть специальные макропакеты для нотной записи, описания принципиальных электрических схем, импорта графики и др.

Существуют программы конверторы SGML, HTML и RTF в $\text{L}\text{A}\text{T}\text{E}\text{X}$ и $\text{L}\text{A}\text{T}\text{E}\text{X}$ в HTML. Хотя идеальной средой для TEX является Unix, его варианты существуют для практически всех операционных систем, включая, конечно, Microsoft Windows и Mac OS.

Всего в каталоге STAN⁴ насчитывается около 1200 документированных компонент TEX . Кроме того, ряд программ доступны помимо STAN.

Конкуренты, перспективы и дистрибутивы...

В 1992 году началась, после отказа Кнута развивать свои системы, реализация проекта $\mathcal{N}\mathcal{T}\mathcal{S}$ (New Typesetting System) с целью разработки системы, лучшей TEX , в рамках которого разрабатывается $\varepsilon\text{-T}\text{E}\text{X}$. Работа идет очень медленно. Единственным заметным достижением по сравнению с другими подобными проектами стала возможность использования обычной математической нотации при расчетах, в частности, в логических конструкциях.

Наиболее перспективными ныне являются разрабатываемые TEX -системы Omega (Ω) и $\text{PDF}\text{T}\text{E}\text{X}$. Ω значительно расширяет рамки ограничений классической системы, оставаясь максимально совместимой с ней: позволяет использовать 16-разрядное кодирование, в частности, Unicode, количество регистров каждого типа увеличивается до 65536, дает возможность работать с фрагментами текстов разной ориентации: европейской — слева-направо и сверху-вниз, арабской — справа-налево и сверху-вниз, японской — сверху-вниз и справа-налево, старомонгольской — сверху-вниз и слева-направо. Система $\text{PDF}\text{T}\text{E}\text{X}$ может производить файлы в формате pdf вместо dvi и позволяет использовать дополнительные возможности этого формата (цвет, все типы шрифтов и шрифтовых манипуляций, интерактивность и прочее). Перспективными также являются разработки подключения TEX как заключительного звена к системам SGML и XML. Хотя нельзя не отметить, что разработки новых компонент TEX в последние несколько лет несколько

⁴ STAN — Comprehensive TEX Archive Network — полная сеть TEX -архивов, поддерживаемая на более 60 ftp-серверах по всему миру, среди которых, например, <ftp://ftp.dante.de/tex-archive>.

замедлилось.

Автор интенсивно использует \TeX уже более семи лет и системы WYSIWYG⁵ вызывают у него теперь лишь чувства, похожие на соблазн, который испытывает уставший человек, видящий удобную инвалидную коляску с мотором...

Но это последнее мнение не бесспорно для всех пользователей \TeX , поэтому существуют ряд мощных WYSIWYG редакторов, лучшие из которых — это LuX и $\text{\TeX}MACS$, причем последний по мнению его автора позволяет сконцентрироваться именно на логической структуре текста, а не на визуальной.

В настоящее время используются еще две системы во многом схожие с \TeX — это Troff и Lout . Первая является одной из самых древних средств компьютерной полиграфии, корни которой уходят в разработки AT&T конца 1960-х. Это весьма гибкая и мощная система, но крайне громоздкая и уже неразвивающаяся. Она используется до сих пор, т. к. на ней написана большая часть документации к Unix — страницы руководства (man pages). Сейчас она вытесняется Texinfo . Система Lout , созданная в 1990-е в университете Сиднея в основном Джеффри Х. Кингстоном, является потомком неудачливого конкурента \TeX , системы Scribe Брайана К. Рейда. Она целиком основана на фундаменте языка PostScript , что до появления программ Dvips и, особенно, $\text{PDF}\TeX$ давало ей некоторые преимущества...

Есть несколько известных дистрибутивов \TeX . В практически каждом дистрибутиве Linux есть пакет $\text{te}\TeX$, содержащий все необходимое. Можно заказать диски CD-ROM \TeX-Live или Dante STAN . Для MS-DOS и OS/2 есть хорошо себя зарекомендовавший дистрибутив $\text{em}\TeX$. Для среды Microsoft Windows неплохим выбором будет либо дистрибутив $\text{Mik}\TeX$, либо $\text{fp}\TeX$ — версия $\text{te}\TeX$. Есть еще дистрибутив $\text{BaKoMa}\TeX$ с улучшенной поддержкой русского языка. На Apple Macintosh можно использовать дистрибутив $\text{Oz}\TeX$. Перечислены только бесплатные или условно-бесплатные (shareware) программы.

В процессе работы над статьей использовались Plain \TeX с макропакетами Xy-pic и epsf . Список литературы, особенно той, что доступна в Internet, сознательно сокращен до минимума, т. к. в STAN есть поисковые серверы, например, на главной web-странице \TeX www.tug.org, да и в Internet в целом тоже. В последнем случае рекомендую www.altavista.com.

⁵ WYSIWYG — What Your See Is What Your Get — системы типа Microsoft Word, которые позволяют редактировать документ в виде, приближенном к напечатанному.

Литература

1. J. D. Hobby. *A User's Manual for MetaPost*. — AT&T Bell Laboratories, Murray Hill, NJ 07974, 1998.
2. D. E. Knuth. *The T_EXbook*. — Addison-Wesley, Reading, Massachusetts, 1986. Volume A of *Computers and Typesetting*. Русский перевод: Д. Е. Кнут. *Все про T_EX*. — АО RDT_EX, Протвино, 1993.
3. D. E. Knuth. *T_EX: The Program*. — Addison-Wesley, Reading, Massachusetts, 1986. Volume B of *Computers and Typesetting*.
4. D. E. Knuth. *The METAFONTbook*. — Addison-Wesley, Reading, Massachusetts, 1986. Volume C of *Computers and Typesetting*.
5. D. E. Knuth. *METAFONT: The Program*. — Addison-Wesley, Reading, Massachusetts, 1986. Volume D of *Computers and Typesetting*.
6. D. E. Knuth. *Computer Modern Typefaces*. — Addison-Wesley, Reading, Massachusetts, 1986. Volume E of *Computers and Typesetting*.
7. L. Lamport. *L^AT_EX: A Document Preparation System, User's Guide and Reference Manual*. — Addison-Wesley, 1994.
8. T. Otten, H. Hagen *ConT_EXt: an excursion*. — PRAGMA, Ridderstraat 27, 8064GH Hasselt, The Netherland, 1998.
9. К. Н. Rose, R. Moore *X_y-pic Reference Manual* — в электронном виде доступно по адресу CTAN:/macros/generic/diagrams/xypic.tar.gz, 1999.
10. M. Spivak. *The Joy of T_EX. A gourmet guide to typesetting with the A_MS-T_EXmacro package*. — American Mathematical Society, Providence, RI, 1990. Русский перевод: М. Спивак. *Восхитительный T_EX: руководство по комфортному изготовлению научных публикаций в пакете A_MS-T_EX*. — М., Мир, 1993.
11. M. J. Wichura *The P_TT_EX Manual*. — The University of Chicago, 1987.
12. С. М. Львовский *Набор и верстка в пакете L^AT_EX*. — М., Космосинформ, 1995.

Copyright © 2002 Лидовский Владимир Викторович.

Для подготовки материалов использовались Plain T_EX, X_y-pic и epsf
Опубликована в разных вариантах в журналах “Информационные технологии” №1, 2003; “КомпьюТерра” №6, 2003; “Вычислительные технологии” Том 8, №3, 2003; “Научно-техническая информация. Серия 2” №3, 2003.