

Система pdfTeX вместе с неразрывно с ним связанным макропакетом CONTEXТ — это на сегодняшний день наиболее бурно растущая с 2001 года ветвь на почти 30-летнем дереве TeXнологии. За 2005 год вышло 7 обновлений pdfTeX и 33 CONTEXТ (в 2006 уже вышло несколько очередных обновлений). В каком-то смысле — это революционные разработки — они меняют почти все: цели системы, ее фундаментальные основы, географические координаты центров поддержки. Все нововведения, однако, стоят на крепком фундаменте 100% совместимости с традиционными TeX-системами.

Изначально TeX — это система, ориентированная на работу только с печатающими устройствами, — до середины 90-х даже предпросмотр результатов ее работы на экране дисплея был либо слишком неудобен, либо требовал дорогостоящей немасовой аппаратуры. До появления pdfTeX были лишь две попытки несколько “сменить ориентиры”.

Система Texinfo — стандарт для документации GNU — позволяет из одного документа получать либо распечатку, либо гипертекст, в частности, html. Сама идея получать из одного документа данные в форматах совершенно разного назначения присутствует в TeX изначально — сам TeX и его основные компоненты написаны на языке WEB, тексты на котором преобразуются либо в программы на Паскале, либо в документ-описание в формате plain TeX. Код pdfTeX связан с WEB.

С середины 90-х предпросмотр результатов работы TeX в формате dvi стал обычной практикой. Программы для такого просмотра простейшим образом — вызовом браузера Internet — стали поддерживать гиперссылки. Помимо Texinfo www-ссылки поддерживаются, в частности, популярнейшим макропакетом L^ATeX.

Но радикальный поворот к интерактивным электронным документам произошел только в связке pdfTeX+CONTEXТ. Именно в ней создание, например, электронных учебников (CONTEXТ создавался для нужд сферы образования) или справочников, стало не побочной плохоразвитой возможностью, но одним из приоритетов.

pdfTeX сохраняет совместимость с форматом dvi, но позволяет также вместо него использовать pdf, что, с одной стороны, ставит его в зависимость от разрабатываемого в коммерческой фирме Adobe формата, а с другой, предоставляет ряд дополнительных возможностей. В pdfTeX, однако, произошел полный отказ от расширений на языке Postscript, встречающихся в некоторых макропакетах.

Центры разработки pdfTeX и CONTEXТ расположены в Европе, в основном, в Нидерландах. Участие в этих проектах принимают и разработчики из стран бывшего соцлагеря, в частности, из Чехии.

TeX и dvi

Результат работы изначального TeX Дональда Кнута имеет формат dvi, который превосходно отображает документы, содержащие только текст с масштабируемыми шрифтами, математические формулы и таблицы. Реализация работы с математическими формулами в TeX до сих пор является наилучшей среди *всего* существующего

программного обеспечения. Таблицы также реализованы на достаточно высоком уровне, но несколько хуже, чем в ряде коммерческих программ. В представлении текстов уже появляются ряд проблем: нет возможности менять межбуквенные интервалы при работе с одним шрифтом; текст можно размещать только строго по горизонтали без каких-либо трансформаций, например, вращения; неплохой алгоритм разбиения абзаца на строки был все-же несколько улучшен, что поддерживается не всеми реализациями Т_ЕX. Для работы с графическими изображениями есть только одно примитивное средство — резервировать на бумаге место для последующего помещения туда графики. Можно работать только с двумя цветами — черным и белым.

Все перечисленные недостатки Т_ЕX обозначились не сразу. В начале 80-х и даже 90-х — они были естественны для любого программного обеспечения такого класса. Отказ от продолжения разработки METAFONT, а затем и самого Т_ЕX их автором и быстрое наращивание ресурсов персональных компьютеров не могли вместе не привести к возникновению этих очевидных на сегодняшний день недостатков. Нужно, конечно, не исключать из виду, что Т_ЕX создавался не как система общего назначения, а скорее как средство для написания Кнудом знаменитого “Искусства программирования для ЭВМ”^{*}.

Все перечисленные и некоторые прочие недостатки Т_ЕX обусловлены примитивностью формата dvi — передового для 80-х, но явно устаревшего уже для 90-х. Наращивание возможностей этого формата или даже простое “поддержание на плаву” является чрезвычайно трудоемкой задачей. Драйверы dvi напрямую работают с устройствами вывода, которые весьма разнообразны, число которых огромно и непрерывно растет. Вместо этого постоянного сопровождения dvi сообществом пользователей Т_ЕX был выбран другой путь — путь разработки универсального драйвера, конвертирующего dvi в стандартный для полиграфических систем язык Postscript. Параллельно с развитием GNU системы Ghostscript, обеспечивающей трансляцию текстов Postscript в форму, непосредственно воспринимаемую устройствами вывода, это решило проблему совместимости с устройствами печати.

Postscript и dvips

Postscript — это весьма интересный универсальный язык программирования. Передовое использование постфиксной (польской обратной) записи, ассоциативных массивов и поддержка на *базовом* уровне работы со сложной графикой выгодно отличают его от многих других языков программирования. К сожалению, этот язык программирования высокого уровня используется в абсолютном большинстве случаев, как язык непосредственно для аппаратуры, что, с одной стороны, делает наличие в нем обычных программных конструкций, типа циклов, присваиваний и т. п. — поведение совокупности которых непредсказуемо, нежелательным, а с другой требует внесения дополнительной информации (в виде комментариев) для устройств вывода. Специфическая узкая направленность Postscript обусловила также почти полное отсутствие в нем средств для интерактивного взаимодействия.

^{*} Кстати, эту работу рекомендует прочитать всем, кто считает себя хорошим программистом “неизвестный” Билл Гейтс и добавляет: “Если вы сможете прочесть этот труд, то вам определенно следует направить мне резюме”.

Система dvips, которая существует в нескольких вариантах, конвертирует файлы в формате dvi в Postscript. Возможности Postscript значительно превосходят dvi, поэтому естественно, что были разработаны макропакеты для использования этих возможностей (вставка картинок, различные трансформации текста, цвет) с T_EX. Формат dvi, благодаря наличию в T_EX команды \special, позволяет вставлять команды, специфические для выбранного драйвера, в частности, программы на Postscript.

Файлы в формате Postscript/PDF получаются совершенно независимыми от программного обеспечения компьютера, в отличие от dvi, для использования которых необходимо наличие всех шрифтов, использованных при создании dvi.

PDF

Этот формат (portable document format) — прямое развитие Postscript, но это уже не язык программирования. Значительная часть данных в этом формате — бинарные, но даже их текстовая часть связана абсолютными ссылками, что не дает ее просто изменять.

В pdf присутствуют развитые возможности для организации диалога с пользователем. Фактически pdf имеет те же возможности, что и html (гиперссылки; формы, которые могут обрабатываться самим документом или сервером http; Javascript) и может использоваться вместо него. В отличие от html, языка обобщенной разметки, pdf жестко фиксирует способ показа каждого элемента документа. Документ pdf предназначен только для чтения и может иметь несколько степеней защиты.

Структура pdf-файла весьма проста — это список занумерованных объектов, после которых следует каталог ссылок на них, после которого располагается ссылка на этот каталог. Часть данных в этом формате — содержимое объектов, как правило, упакована по известным алгоритмам сжатия информации.

Одним из недостатков pdf, отсутствующим у Postscript, является необходимость считывания всего документа перед показом его любой части.

Система pdfT_EX создает основу для работы макропакетов типа CON_TE_XT, производя результат в виде pdf. Она, с одной стороны, позволяет использовать разнообразные полиграфические средства — такие же как в Postscript, а с другой стороны, использовать все диалоговые возможности pdf. Поддерживаются работа с графическими изображениями, цветом, мультимедиа, трансформациями текстов и изображений, элементами управления типа кнопок, полей ввода данных и т. п., WWW. Среди не столь “выпуклых” возможностей — поддержка всплывающих аннотаций к любому участку документа, возможность размещения статьи в виде несвязанных между собой фрагментов текста, простота — одной командой — изготовления брошюр, когда, например, на лист А4 печатается две страницы А5. Ссылки помимо URI могут быть на страницу используемого или другого документа с опциональным уточнением позиции на странице. Для pdf-ссылки характерно указание точного способа показа указываемой страницы.

Эта система может быть напрямую использована с большинством T_EX-макропакетов, в частности, L^AT_EX. Однако, pdfT_EX запрещает прямо использовать Postscript, следовательно, таких макропакетов как T_EXdraw, PSTricks, и др. Кроме того, отсекаются ряд расширений к некоторым макропакетам, например, X_u-pic. В документ стало невозможным вставлять картинки в формате eps. Если последнюю проблему легко

решить, перековертировав картинку в формат png, jpeg или pdf, которые поддерживаются pdf \TeX , то другие в рамках pdf \TeX не имеют решения, т. к. команда `\special` игнорируется. Система pdf \TeX в не dvi-режиме производит только pdf и позволяет вставлять только фрагменты данных в форме pdf при помощи команды `\pdfliteral`.

Использование макропакетов, использующих Postscript, с форматом pdf возможно — для этого можно использовать стандартные средства: конвертор ps2pdf из пакета программ Ghostscript или сразу драйвер dvi2pdfm — аналог dvips. Однако, использование в одном документе средств как Postscript, так и pdf \TeX почти невозможно. Единственная существующая возможность обойти это ограничение очень громоздка — это прямая постраничная вставка содержимого одного документа pdf в другой.

В pdf \TeX большое внимание уделяется работе с объектами — более мощными разновидностями традиционных теховских “боксов”.

К сожалению, поддержка кириллицы в pdf \TeX находится на очень примитивном уровне. Документы, произведенные этой системой, на русском, украинском и других языках, использующих кириллицу, получаются несколько дефективными. Хотя их внешний вид будет соответствовать высшим оценкам, их содержимое будет практически невозможно использовать как текст — в этих документах невозможно проводить поиск нужных слов, а также невозможно копировать фрагменты текста в другие документы. Эта проблема вызвана отсутствием должной поддержки кириллицы в разработках фирмы Adobe — отсутствует таблица для связи фирменных названий для букв кириллицы с их кодами в Unicode, т. е. это не проблема pdf \TeX , а проблема самого формата pdf. К pdf \TeX добавлена *недокументированная* команда `\pdffontattr`, которая позволяет использовать созданные пользователем такого рода таблицы. Владимиром Воловичем разработаны и помещены в CTAN такие таблицы для стандартных кодировок L \TeX , но только для *невиртуальных* шрифтов. К сожалению, многие кириллические шрифты — виртуальные. Автором статьи обозначенная проблема частично решена — большинство виртуальных шрифтов (*все* кириллические) можно связать с Unicode, использовав соответствующие таблицы для всех неvirtуальных шрифтов, входящих в состав виртуальных.

Другая проблема связана с тем, что в закладках — это, как правило, оглавление в отдельном окне справа — нельзя напрямую использовать что-либо кроме символов из стандартных кодировок Adobe, среди которых кириллицы нет до сих пор. Проблема решается лишь прямой вставкой кодов Unicode (UTF-16BE) для букв кириллицы, что, конечно, очень неудобно — нужен соответствующий макрос. Проблема усложняется тем, что в большинстве макропакетов \TeX символ “\” (backslash) очень трудно использовать самостоятельно, а в формате pdf все Unicode-коды удобнее начинать именно с него.

Помимо собственно \TeX , созданного Кнудом, существует ε - \TeX — развитие базовой системы уже без участия Кнута, для которой существует аналог — pdf $\varepsilon\TeX$, который используется CON \TeX T по-умолчанию.

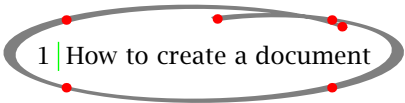
CON \TeX T

Это ориентированное на европейское языковое многообразие средство преимущественно для создания качественных электронных изданий. Руководство к нему доступно на 6 языках: английском, немецком, итальянском, голландском, чешском и

румунском. Более того, для перечисленных языков существует свой набор макросов, например, вместо английского названия макроса `\appendix`, чешский пользователь может использовать `\dodatek`. Помимо этих языков менее масштабно поддерживаются и другие, в частности, вьетнамский. К сожалению, ситуация с кириллицей в этом макропакете даже хуже, чем в `pdfTeX`, хотя базовая необходимая поддержка имеется.

В качестве иллюстрации возможностей `ConTeXt` приводится страница текста из электронного руководства к этому макропакету.

How to create a document



Let's assume you want to create a simple document. It has some structure and contains a title page, a few chapters, sections and sub sections. Of course there is a table of contents and an index.

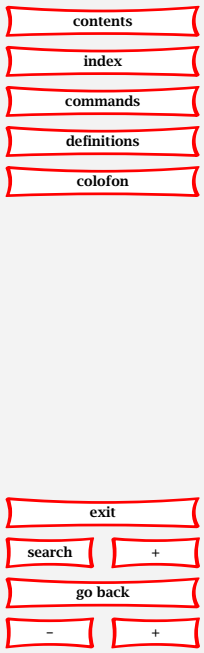
`CONTEXt` can create such a document automatically if you offer the right input by means of a file. So first you have to create an input file. An input file consists of a name and an extension. You can choose any name you want but the extension has to be `tex`. If you create a file with the name `myfile.tex` you will find no difficulties in running `CONTEXt`.


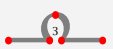
An input file could look like this:

```

\starttext
\startstandardmakeup
  \midaligned{How to make a document.}
  \midaligned{by}
  \midaligned{The Author}
\stopstandardmakeup
\completecontent
\chapter{Introduction}
... your text\index{indexentry} ...
\chapter{One Chapter}
\section[firstsection]{The first section}
... your text ...
\section{The second section}
\subsection{the first sub section}

```



`CONTEXt` содержит все необходимые средства для создания документов со сколько угодно сложной структурой. Большое внимание уделено удобству работы со ссылками на различные компоненты документа, идентификации этих компонент и способам размещения их друг относительно друга. Можно сказать, что `CONTEXt` включает в себя: почти все базовые возможности макропакета `LATEX`; полиграфические расширения `Postscript/PDF`; интерактивные возможности `pdf`. В отличие от имеющего сотни различных стилевых настроек `LATEX`, `CONTEXt` пока не имеет простых средств для быстрой смены внешнего вида своих документов.

Для построения перекрестных ссылок, оглавления, списка таблиц, фигур и т. п. используется отдельная вспомогательная программа `texutil` на языке `Perl`. Кроме того, `CONTEXt` может “незаметно” использовать систему `MetaPost` в процессе работы.

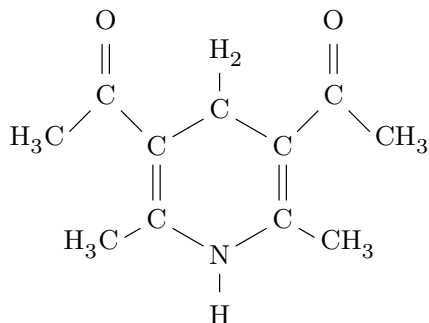
CONTEXТ, хотя и отличается внешне от первого макропакета plain TEX Дональда Кнута меньше, чем Texinfo, но все же гораздо больше, чем L^ATEX. В CONTEXТ практически не используются макросы plain TEX. Даже запись математических формул выглядит несколько иначе.

Все соответствующие элементы документа (номера страниц в индексах, элементы оглавления и т. п.) *автоматически* получаются интерактивными. Эта возможность, однако, приводит к существенному росту размера документа, поэтому ее можно отключать.

Далее следует пример простого законченного 7-страничного документа CONTEXТ из трех глав с подразделами, с титульным листом, указателем с индексами и оглавлением.

```
\starttext
\setupbodyfont[cyr]
\startstandardmakeup
\midaligned{Как сделать документ}
\midaligned{самим автором}
\stopstandardmakeup
\completecontent
\chapter{Введение}
... ваш текст\index{элемент индекса} ...
\chapter{Глава 1}
\section[firstsection]{Первый раздел}
... ваш текст ...
\section{Второй раздел}
\subsection{первый подраздел}
... ваш текст\index{другой элемент индекса} ...
\subsection{второй подраздел}
... ваш текст ...
\section{третий подраздел}
... ваш текст ...
\chapter{Другая глава}
... ваш текст ...
\chapter[lastchapter]{Последняя глава}
... ваш текст ...
\completeindex
\stoptext
```

Для CONTEXТ существует пока только одно отдельное значительное расширение — пакет PRCSTEX для печати структурных формул химических соединений. Например, изображение следующей формулы порождается 4 строчками текста описания ее структуры.



Кроме того, существует около двух десятков небольших расширений-модулей для решения сравнительно несложных задач, например, подключения разного рода шрифтов или печати нормальных форм Бэкуса-Наура (НФБН).

В заключении необходимо отметить гибкость основ CONTEXT , отсутствие в нем жесткой привязки как к $\text{pdf}\text{T}\text{E}\text{X}$, так и к самому формату pdf. CONTEXT *всегда* может производить результат в dvi-формате с разделами `\special` для разных драйверов `dvips`, `dvipdfin`, ... Конечно, в таких “dvi” может отсутствовать все, кроме такого раздела для команд конкретного драйвера. CONTEXT может также работать и с XML.

Где найти дополнительную информацию по этим системам?

Абсолютно *вся* информация по $\text{pdf}\text{T}\text{E}\text{X}$ и CONTEXT доступна в электронной форме на ряде сайтов сети:

- www.pdfTeX.org;
- www.pragma-ade.com;
- mirror.contextgarden.net.

Copyright © 2006 Владимир Лидовский.

Для подготовки статьи использовалась системы plain $\text{pdf}\text{T}\text{E}\text{X}$ и CONTEXT

Опубликована в разных вариантах в журналах “КомпьюТерра” №14, 2006 и “Информационные технологии” №11, 2006.