

Лекция 10. Создание документов в форматах Postscript и PDF

Введение

Распространение документа на уровне \TeX -файла требует установки \TeX 'а с тем же набором пакетов.

Для распространения электронных документов обычно используют форматы Postscript и PDF.

Кроме того, использование этих форматов для вывода документа расширяет графические возможности \TeX 'а и позволяет подключать новые шрифты.

Язык управления принтером

Лазерный принтер представляет собой специализированную вычислительную машину со своим процессором и памятью.

Каждое семейство лазерных принтеров имеет свой язык описания страниц — специализированный язык программирования.

Интерпретатор этого языка зашит в ПЗУ принтера. Он носит название RIP (Raster Image Processor).

При выдаче команды печати Print происходит следующее:

- Драйвер принтера, установленный в операционной системе, преобразует документ в код на входном языке принтера.
- Этот код обрабатывается RIP-интерпретатором постранично. В результате получается поточечное описание страницы — битовая карта, которая зависит от разрешения принтера (dpi — dots per inch).
- Полученный битовый образ можно отобразить на принтере.

Язык управления принтером (продолжение)

Наиболее распространены следующие языки:

- Postscript — разработан фирмой Adobe для принтеров Apple LaserWriter, исторически первый язык описания страниц,
- PCL (Printer Command Language) — разработка фирмы Hewlett Packard, узко специализирован.

Различают Postscript принтеры (встроен интерпретатор Postscript) и PCL принтеры.

Дорогие модели Hewlett Packard имеют встроенный эмулятор Postscript.

Postscript

Postscript — полнофункциональный низкоуровневый язык со стековой организацией.

На нем можно писать программы, но он очень редко используется для непосредственного программирования (в основном, профессиональными дизайнерами).

Кроме команд документ Postscript обычно содержит битовые образы (двоичный код).

Отсутствие встроенного PostScript не означает, что с ним нельзя работать на данном принтере — просто нужно использовать внешний транслятор языка Postscript.

Postscript-файл

Файл на Postscript может иметь расширение ps или eps (encapsulated postscript).

Формат EPS имеет следующие особенности.

- В EPS-формате можно сохранять только одностраничный документ, поэтому его используют исключительно для хранения отдельных рисунков.
- В EPS разрешается использовать только ограниченное подмножество команд языка Postscript.

Postscript-файл (продолжение)

- В начале EPS-файла хранятся дополнительные встроенные данные. Наиболее важные из этих данных — информация о размерах изображения (оно может включать как текст, так и рисунки) вида:

```
%% BoundingBox 50 100 210 450
```

Первые два числа — координаты левого нижнего угла изображения на странице, вторые — правого верхнего угла.

Вертикальная ось направлена вверх. Единица измерения — большой пункт (0.351 мм).

Как прочесть или напечатать Postscript-файл

- Для просмотра и печати PS или EPS файлов можно использовать программу Adobe Acrobat или ее усеченную версию Adobe Reader. Допустимы и некоторые другие программы фирмы Adobe: Adobe InDesign, Adobe Photoshop.
- Можно использовать бесплатно распространяемый интерпретатор языка Postscript, который называется Ghostscript.

Ghostscript

Ghostscript можно бесплатно скачать с сайта

<http://www.ghostscript.com/download/>

Точнее, следует скачать две программы:

- Ghostscript — программа-компилятор. Ее следует установить. Если вы не будете программировать на Postscript, вызывать ее не нужно. Ghostscript необходим для просмотра и печати PS файлов и для работы некоторых программ ЛАТ_EX.
- GSview — программа просмотра и печати. С помощью GSview можно выполнить преобразование PS→EPS.

Как получить PostScript файл

- Использовать специальную программу, входящую в MiKTeX: dvips.exe. Создаем файл name.dvi, а затем выдаем команду

```
dvips name
```

При этом в системе должен быть установлен Ghostscript. Можно регулировать разрешение:

```
dvips -D 600 name
```

(разрешение 600 dpi).

- Рисунки из MATLAB можно сохранять с формате eps (см. тему «Рисунки»).

Замечания

- Некоторые оболочки содержат кнопку dvips.
- Использование Postscript позволяет использовать Postscript рисунки и шрифты. Если шрифт отсутствует на вашем компьютере, могут возникнуть проблемы с его чтением.
- PS файл обычно достаточно большой и при пересылке его обычно архивируют.

Формат файла PDF

Формат PDF (portable document format) — разработка фирмы Adobe для распространения документации, в частности через Internet.

- В отличие от Postscript, PDF — не язык программирования, а файловый формат.
- PDF файл довольно компактен (практически несжимаем), поэтому архивировать его не нужно.
- В него можно встроить (полностью или частично) описания используемых шрифтов. Обычно файл PDF занимает существенно меньше места, чем файл PS, но при встраивании большого количества шрифтов он может оказаться и больше.

Формат файла PDF (продолжение)

- Для его просмотра и печати можно использовать программы Adobe Reader (freeware), Adobe Acrobat (коммерческий продукт), GSview (freeware) и некоторые другие.
- Файл в PDF формате слабо редактируем (Adobe Acrobat и plugins к нему).
- PDF может содержать не только информацию по описанию страницы, но и другие типы информации — шрифты, изображения, ключевые слова для поиска и индексации файла, звук, видео, активные формы, гипертекстовые ссылки и т.д. (Для встраивания используется Adobe Acrobat.)

Как получить PDF файл (1-й способ)

Вместо программы latex используем программу pdflatex:

```
pdflatex name
```

При этом dvi не образуется, а сразу получаем pdf файл.

В последних всегда используется частичное встраивание шрифтов — subset embedding, т.е. встраиваются только те символы, которые используются в документе. Это важно, так как полное встраивание нарушает права фирм-производителей шрифтов.

Замечание: при использовании pdflatex рисунки должны быть в формате PDF, а не EPS. Преобразование можно сделать автоматически, подключив пакет:

```
\usepackage{epstopdf}
```

Как получить PDF файл (2-й способ)

Создаем PS файл как указано ранее.

Обрабатываем его программой Acrobat Distiller, которая входит в состав Adobe Acrobat. В настройках

Settings/Edit Adobe PDF Settings/Fonts/Embed All Fonts

В инструкциях конференций часто указывают точные настройки Distiller'a (критичные настройки — встроенные шрифты, версия PDF).

Эти настройки сохраняются в файле *.joboptions. Часто такой файл можно скачать с сайта конференции. Файл нужно подсоединить к Distiller'у с помощью меню:

Settings/Add Adobe PDF Settings/

Как получить PDF файл (2-й способ — продолжение)

Иногда PDF файл получается со слишком узким верхним или левым полем.

В этом случае текстовое поле на старнице можно подвинуть — вызываем Adobe Acrobat (не Adobe Reader!).

Настраиваем так, чтобы страница была на экране целиком.

Tools/Advanced Editing/TouchUp Object Tool

Выделяем всё текстовое поле мышкой и двигаем стрелками.

Как получить PDF файл (2-й способ — продолжение)

Если не предпринять дополнительных мер, будут использованы TeXовские шрифты. Они будут встроены как шрифты Type 3 (растровые). Такой файл будет некрасиво выглядеть на устройстве, у которого разрешение не совпадает с заданным при создании. Лучше использовать векторные шрифты.

Современные версии dvips используют шрифты Type 1 по умолчанию.

Для старых версий dvips нужно вызвать со специальным ключом Ppdf:

```
dvips -Ppdf name
```

Как получить PDF файл (3-й способ)

Преобразование можно выполнить из PS файла с помощью функции Ghostscript gswin32c.exe. Ее можно вызвать явно с большим количеством сложных ключей.

В MiKTeX можно использовать команду

```
ps2pdf имя.ps
```

Для встраивания шрифтов к ней нужно добавить ключи.

Растровые и векторные шрифты

Шрифты подразделяются на растровые и векторные (или близкие к ним шрифты True Type).

Векторные шрифты называют также Type 1, растровые — Type 3.

Когда Кнут разрабатывал TeX, существовали только растровые шрифты. Кнут разработал программный пакет, предназначенный для генерации растровых шрифтов — Metafont.

Высокое качество печати достигается тем, что для каждого разрешения экрана или принтера (dpi) генерируется свой набор описаний шрифтов. Современные программы TeX делают это автоматически, на лету. Если dvi файл вывести на устройство с другим dpi, то качество может быть утрачено.

Растровые и векторные шрифты (продолжение)

Для передачи в редакцию обычно требуется использование векторных шрифтов (шрифтов Type 1). Небольшой набор таких шрифтов входит в дистрибутив MiKTeX.

Проблема состоит в том, что шрифты являются объектом авторского права и для их профессионального использования нужно покупать лицензию. Во многих случаях дистрибутив TeX содержит только пакеты для интерфейса программ TeX и внешних шрифтов, но сами шрифты отсутствуют. (Их можно купить отдельно.)

Ряд ведущих полиграфических фирм предоставляют ограниченный набор своих шрифтов в бесплатное пользование.

Форматы вывода документов TeX

При подготовки документа в TeXе нужно четко представлять, каким образом вы собираетесь его использовать. Возможны следующие варианты.

- Вывод документа на домашний или офисный принтер.

При домашнем использовании документа его формат обычно не имеет значения (обычно используют DVI или PDF).

- Передача документа в издательство или типографию для печати в сборнике, журнале, книге и пр. Здесь возможны варианты.

Физико-математические журналы и издательства обычно просят исходный файл в L^ATeX. При этом нужно избегать сложных приемов форматирования, указания длин в физических единицах и пр. — работники издательства их все равно переделают, возможно с ошибками.

Форматы вывода документов Т_EX (продолжение)

Если текст редактироваться не будет, поинтересуйтесь, каким способом буде печататься материал. Наиболее распространенный способ — ризограф или офсет.

Ризограф используют при небольших тиражах (300-400 экз) и качество печати ниже. Так печатают небольшие сборники и труды конференций.

По сути это копировальная машина. В качестве оригинал-макета служит ваш бумажный документ. Никакой файл здесь обычно не нужен. Документ должен быть напечатан на лазерном принтере. Струйный, даже хороший, даст плохой результат.

Форматы вывода документов Т_EX (продолжение)

Офсетная печать — основная при печати физ.-мат. литературы. В качестве оригинал-макета служат специальные офсетные пленки (кальки), напечатанные в зеркальном отображении. Обычно их готовит само издательство. Пленки выводятся из файла.

В редких случаях файл может быть получен сканированием бумажного оригинала — это дорого и используется в тех случаях, когда автор не в состоянии подготовить качественный файл для вывода.

Наиболее удобная форма для издательства — файл в формате Postscript или PDF со шрифтами Type 1.

Форматы вывода документов Т_EX (продолжение)

- Электронная публикация на CD или флешке.

Сейчас это фактический стандарт для трудов конференций. Кроме того, так распространяются сборники старых сканированных книг. Обычно на крупную международную конференцию требуется представить текст доклада в формате PDF, оформленный по заданным правилам.

На диске эти файлы объединяются в единой оболочке (которая создается в формате PDF или HTML) и получается электронная книга. В электронных публикациях можно использовать цветные рисунки и фотографии (при обычной печати они сильно удорожают издание).

Форматы вывода документов Т_EX (продолжение)

Основное требование к PDF файлам — все используемые шрифты должны быть встроены. При это требуется частичное встраивание (subset embedding). При этом встраиваются только символы, используемые в тексте. Частичное встраивание не нарушает авторских прав владельца шрифта.

На многих конференциях при загрузке PDF-файла на сайт автоматически контролируются его характеристики — для этого используют специальные плагины к Adobe Acrobat (обычно Enfocus PitStop). В случае нарушения правил файл просто не будет загружен.

Форматы вывода документов Т_EX (продолжение)

- Публикация в интернете.

Сейчас большинство журналов и книг имеют электронные версии. Читатель может скачать текст статьи, которая его заинтересовала.

В одних журналах скачивание платное (подписку оплачивает учреждение), в других — бесплатное (Open Access).

Кроме того, имеются чисто электронные журналы.

Публикации в тех из них, которые имеют государственную регистрацию, учитываются при защите кандидатских диссертаций.

Форматы вывода документов TeX (продолжение)

Полные тексты статей обычно доступны в формате PDF.

Реже используется zipped Postscript.

Аннотации и тексты с простыми формулами могут быть созданы в формате HTML. Для этой цели имеется несколько конвертеров из LaTeX в HTML. В состав MiKTeX входит конвертор tex4ht. Из наиболее популярных отметим конвертор tth. Обычно он обрабатывает только протейшие команды и результат получается не очень качественный.

Для сканированных документов иногда используют формат DejaVu.

Пример: правила конференции CDC

Only PDF files will be accepted. A maximum file size of 2MB for all uploaded papers is enforced. Click [here](#) for information on creating IEEE Xplore-compatible PDFs.

Caution: US Letter and margins settings below are required for the final submission.

		Letter paper
Paper Size	(8.5x11.0)in	(21.59x27.94)cm
Top margin (1st page)	1.0in	2.54cm
Top margin (pages 2+)	0.75in	1.9cm
Left margin	0.75in	1.9cm
Right margin	0.75in	1.9cm
Bottom margin	0.75in	1.9cm

Пример: правила конференции CDC

Note: Initial submissions may be up to 8 pages. However, for final submission, paper should be 6 pages long. Any page over 6 pages will be charged EUR 200.00/page up to a maximum of two extra pages.

Caution: Style files and templates provided here have been created to ensure that margin requirement is adequately met.

MS WORD	
MS Word template:	ieeeconf_letter.dot
Sample Documents:	sample_letter.doc
LaTeX	
LaTeX style file:	ieeeconf.cls
LaTeX sample document:	sample.tex

REMINDER: Authors are responsible for submitting their paper in the required format. All papers that are accepted will be published as submitted by

the Author. The Conference is NOT responsible for editing or correcting errors in the paper.

Пример: правила конференции CDC

Create IEEE Xplore-compatible PDF

- The Paperplaza CDC 2019 system supports compliant conversion of pdf files. Upload the pdf file of your submission and the system converts it to compliant pdf, satisfying all technical requirements regarding pdf version and font embedding.
- Be sure to use the Word template or LaTeX style file provided by the conference and follow the conference instructions concerning paper size, number of pages and page margins while preparing your contribution.
- PDF compliance can also be tested, here before uploading.

Пример: правила конференции CDC

Manuscript requirements

- All fonts must be embedded in the file.
- Fonts that require non-English language support are not allowed.
- Two-column format in the IEEE style is required for final submissions and recommended for initial submissions.
- For IEEE conferences, the paper size for final submission must be US Letter. For non-IEEE conferences, please refer to the conference website for the requirement.
- The document margins must satisfy the printer's requirements.
- The number of pages and the file size must satisfy the conference requirements.
- The document should not have any password protection.
- The final PDF should not have any page numbers, footers or headers.
- If your manuscript does not pass the compatibility requirements then you will not be able to upload it.

Пример: правила конференции CDC

Test your pdf file

- “Font” Not embedded error: Encountered in both LaTeX and Word due to incorrect options in creating PDF.
- Bitmapped Font error: Typically encountered using LaTeX. See also Bitmapped Font in Figures.
- Oriental Language Font error: Typically encountered using Word.