

Putting documents on the Web as PDF files (for Windows users)

Francis Wright

This document can be accessed on the web in the directory http://centaur.maths.qmw.ac.uk/Generating_PDF/ in several formats: Microsoft Word, the original source format; HTML¹ generated by Word; PostScript (PS) generated by Word; and PDF generated from the PS using GhostScript. The former two formats contain copious hyperlinks; generation of the latter two formats is explained below.

Appropriate software

I will assume Windows 95/8 or NT 4.0. The main issue is to install appropriate software; that which I will consider is not part of Windows but is *all freely available* and relevant URLs are listed in

<http://centaur.maths.qmw.ac.uk/Links.html>.

I recommend that T_EX users install MiKTeX. I will also describe the use of GhostScript. (The versions I will describe are MiKTeX 1.20 and GhostScript 6.0.) Provided MiKTeX and GhostScript are properly installed, then everything relating to the generation of PDF files described by Peter Cameron in <http://www.maths.qmw.ac.uk/~pic/MAS999/instr.pdf> should work under the Windows command interpreter (MS-DOS command prompt) *exactly as illustrated*. You can then use FTP (or HTTP, NFS, Samba, floppy disc, etc.) to copy files from Windows to UNIX. I describe below some alternative ways to generate PDF files.

Software installation

MiKTeX and GhostScript are available as self-extracting archives (or equivalent) that are very easy to install, but they may or may not set the Windows execution path correctly. (I have forgotten, and I think that the precise details change between releases anyway.) If the installation process does not do so automatically then you need to set it yourself before you can conveniently use the tools I will describe. One general way to do this is to add to your C:\AUTOEXEC.BAT file the following (but with the precise directories as appropriate for your machine, and then restart Windows 95/8):

```
set path=%path%;E:\texmf\MikTeX\bin
```

```
set path=%path%;E:\Aladdin\gs6.0\bin;E:\Aladdin\gs6.0\lib
```

Note that you can check the current execution path by opening an MS-DOS command prompt window and typing the command: `path`

One-click conversion

I primarily generate PDF files in two ways: from DVI files and from PS (PostScript) files. I normally do this via the Windows Explorer context menu on the right mouse button, so I just point at a DVI or PS file, right click and select "To PDF". This

¹ This is HTML generated by Microsoft Word 2000 and stripped of the XML that carries the full Word content; note that it uses standard HTML 4.0 CSS (Cascading Style Sheet) facilities, which seem to be displayed best by Microsoft Internet Explorer.

generates “file.pdf” from “file.dvi” or “file.ps”, in the same folder. The Explorer context menu can be programmed by selecting the “Folder Options...” dialogue from the View menu (or directly by editing the Registry, which is not recommended!).

Converting DVI to PDF

This is convenient if you start from (L^A)T_EX source format. DVI files can be converted to PDF, without using PDF(L^A)T_EX or generating an intermediate PS file, but using dvipdfm (included with MiKTeX). In my experience, this often generates smaller PDF files than other methods, and is currently my preferred approach.

The command currently installed in my context menu for .dvi files is

```
E:\texmf\miktex\bin\dvipdfm.exe -p a4 "%1"
```

The full path of the executable is required here, and Explorer automatically replaces the “%1” by the full name of the DVI file. The `-p` option is to set the right paper size. Alternatively, this command could be issued in a command interpreter:

```
dvipdfm -p a4 file.dvi
```

Converting PS to PDF

Almost *any* document can be converted to PDF by first generating a PS file. This is very easy: in any application select “Print...” from the File menu, in the Print dialogue select a PostScript printer driver and click the “Print to file” checkbox, then print. This will probably generate a file named “file.prn”; if so then I recommend renaming it to “file.ps”. It is not necessary to have a PS printer; Windows comes with a selection of PS printer drivers, one (or more) of which you can install from the Printers Control Panel accessory. You can set it to print to file by default, which is particularly useful if you don’t have a PS printer!

PS printer drivers may support “n-up” printing, which can be selected from the Properties dialogue when printing. If none of the PS drivers supplied with Windows supports this feature then you can probably download one that does from the Adobe web site; they are free!

Generating PDF via PS has the advantage that there are many (free) programs available to manipulate PS files. A particularly useful one is “psnup”, which converts a PS file to “n-up” format (n logical pages reduced onto 1 physical page). MiKTeX includes a selection of such programs. In this way, I provide my Computational Mathematics II lecture notes as PDF files in 1, 2 and 4-up formats.

To convert the PS file to PDF, given the path setup as described above, open a command prompt window on the directory containing the file (select the right drive and `cd` to the right directory if necessary) and run the command

```
ps2pdf file.ps file.pdf
```

Before doing this, you may want to issue the command

```
set GS_OPTIONS=-sPAPERSIZE=a4
```

to ensure that the PDF file is formatted for A4 paper. Or you can put this command in your `C:\autoexec.bat` file (and then restart Windows 95/8).

One-click conversion under Windows NT (and probably Windows 2000) can use essentially the standard GhostScript command file, because the NT command processor (`cmd.exe`) provides the necessary facility to convert the extension, so that `ps2pdf` can be run with only the PS file name as argument. I use the command file shown in the appendix to automate everything including setting the required path. However, this will not work under Windows 9x because the command processor (`command.com`) does not provide the necessary facilities.² Nevertheless, the commands in this file show exactly what set-up is necessary.

To PDF or not to PDF ...

Document formats that allow sophisticated formatting include HTML, PDF, T_EX, DVI, PostScript. HTML has the advantage that it does not require any software beyond a web browser, and Internet Explorer has been included with Windows for several years. All the other formats require either a browser plug-in to be installed or a helper application to be set up.

HTML is probably the best choice for simple text documents that do not contain mathematics or diagrams; this may change when formats such as MathML and SVG are better developed and have built-in browser support.

PDF is currently a good choice for documents containing mathematics or diagrams. The Adobe Acrobat Reader is free and easy to install, and provides good quality display and printing of most documents; it also provides good searching and quite good text extraction (by copying to the clipboard). PDF files should not be huge and can provide hyperlinks and navigation facilities.

Other formats are less standard: T_EX source (and MathML) can be displayed by the IBM techexplorer plug-in (a version of which is free); DVI can be displayed by a previewer as a helper application or by various Java plug-ins but requires access to the right set of fonts; PostScript can be displayed via GSVIEW (or GhostView) as a helper application (but does not seem to be displayed well under Windows).

Where (else) to put it?

Files in any format can be put on the web, and an arbitrary directory structure can appear as exactly the same web structure (except for the location of the root). It is not *necessary* to use any HTML files. I use a combination of HTML files to explain the content of my website, together with directories that give direct access to the files that I want to publish. For example, it is convenient to provide a directory listing of program files so that potential users can see the file sizes and last-modified dates. It is also convenient to be able to put CW files into a directory on the web for access by students, without the chore of updating an HTML file of links. See my website at <http://centaur.maths.qmw.ac.uk/> for lots of real examples.

² The real problem is attempting to port a UNIX approach to Windows, which is much harder with 9x than with NT/2000. The right approach under Windows is probably to use the Windows Scripting Host, but I have not explored that.

Appendix: Batch file for one-click conversion of PS to PDF under Windows NT

Using the following batch file ([ps2pdf.bat](#), available in the same directory as this document), the command to use in the context menu for .ps files is just the full pathname of the batch file.

```
@echo off
REM FJW: GhostScript 6 PS2PDF with A4 paper
REM and SINGLE file argument under Windows NT

set opath=%path%
set path=E:\Aladdin\gs6.0\lib;E:\Aladdin\gs6.0\bin
set GS_OPTIONS=-sPAPERSIZE=a4
set PS2PDF=%1
set PS2PDF=%PS2PDF%.ps=.pdf%

REM Convert PostScript to PDF 1.2 (Acrobat 3-and-later).
call ps2pdfwr -dCompatibilityLevel#1.2 %1 %PS2PDF%

REM Convert PostScript to PDF 1.3 (Acrobat 4-and-later).
REM call ps2pdfwr -dCompatibilityLevel#1.3 %1 %PS2PDF%

set PS2PDF=
set GS_OPTIONS=
set path=%opath%
set opath=
```