

The comp.text.pdf Frequently Asked Questions

Michael Still

The comp.text.pdf Frequently Asked Questions
by Michael Still

Copyright © 2001 by Michael Still

Table of Contents

1. Introduction	7
2. Credits	9
3. Acrobat.....	11
4. Acrobat forms	25
5. Common questions about PDF	33
6. Conversion	37
7. Development: Acrobat plugins	43
8. Development: Examples.....	45
9. Development: General questions	53
10. Development: Microsoft Windows specific.....	71
11. Development: Problems with my PDF documents	73
12. Distiller	75
13. History.....	79
14. Internet stuff	81
15. Online tutorials	89
16. OS: Linux specific	91
17. OS: Microsoft Windows specific.....	93
18. Other.....	95
19. PDF tools from people other than Adobe	97
20. PDF/X	109
21. Postscript	111
22. Printing	115
23. Security	119
24. Specification things	123

Chapter 1. Introduction

Welcome to the comp.text.pdf frequently asked questions document. If you have any suggestions on how this document could be improved, better answers to questions included, or even a new question, please contact the maintainer, Michael Still at mikal@stillhq.com

Whenever the name and email address of a person who has answered a question is known, they have been credited within this document. It should be noted that some people chose to add extra text to their email addresses to confuse spam senders. These extra pieces of text have been retained in this document and will need to be removed manually before sending email.

Please note that questions and answers are sometimes edited for readability. I have started marking the answers which have been edited from the version labelled 2001-04. [] denotes a comment from the editor.

This document may also be found online at

<http://www.stillhq.com/cgi-bin/getpage?area=ctpfaq&page=index.htm>

which also includes past versions of the FAQ.

This version of the FAQ covers information up to and including Thu, 3 Jan 2002 10:04:46 GMT and is based on 9076 postings. I am currently quite behind with bringing the latest postings into the FAQ – there are some 2000 or so waiting in my queue. My apologies for the delay.

This edition introduces several new section names, and has moved some answers to more logical locations.

Version: 2002-02

Chapter 2. Credits

This FAQ is maintained by Michael Still, who can be contacted at mikal@stillhq.com
Many thanks to Catherine and Andrew for being supportive of Michael whilst he works on this document

The following people have provided answers which have been used in this FAQ:

- 44: Aandi Inston
- 25: Michael Still
- 11: Dan Sideen
- 7: Helge Blischke, Bryan Guignard
- 6: Rachel Burnsed
- 4: Thomas Merz
- 3: Bill Segraves, Scott Robert Ladd, Mark Anderson, Christian Koch, John Farrow, Aaron Watters, Matti Vuori, Dick Margulis, Ken Sharp
- 2: Alex Cherepanov, Giacomo, mikal, William Alexander Segraves, Louis Vosloo, Filiep Maes, Mark Gavin, Vladimir Katalov, Martin Bailey, Kent Isaksson, Frank M. Siegert, Bob Niland, Arne, Veli Holopainen
- 1: Jorma Heimonen, laird@gunsmoke.ecn.purdue.edu, Bruno Lowagie, Don Daniels, Rocker95, Klaus Gotthardt, U Berlin, H.Blischke@srz-berlin.de, Guy Vdh, Jim K, Edward Mendelson, Dale Rebgetz, Eckhard Henkel, Andreas Wall, MSD, Dinu Gherman, Don Lancaster, Michael Sweet, Tom Kacvinsky, Vi Wisdom, Dan Cogliano, GDCII, Michael Mantz, Doug Milliken, Miguel Tavares, Timothy L. Jordi, MicroPress, Inc, Mr T, Harvey Turnbull, Diana Diehl, Kevin Ashley, SANFACE Software, Dave Braze, Tore Hanson, Stefan Treutlein, JW, Dave, J.H.M. Dassen (Ray), Matthias Haeusser, Steve Cook, Ross Finlayson, Happy Go Lucky, Mark Lookabaugh, DesQuite, John Deubert, Marc Wieber, Ed Bomke, Dave Bloodgood, Jose Fernando Tepedino, Michael Hemmer, Steve Auerweck, Tiger, Jonathan Weiner, A. Sinan Unur, Ingemar Djurhuus, Jim Driscoll, Philippe Lauwers, Derek B. Noonburg, Jim Ley, John Doherty, Kurt Foss, Ivincent, Peter Flynn, Ron Teplitz, JoeZ, mschulz, Pierre Baillargeon, Sanjoy Mahajan, Craig Broadbear, Anders Thulin, John Freund, Mimi Burbank, Jean Piquemal, Paulo Soares, Gunther Schmidt, Robert Anderson, Ross, Mike Bernardo, Chaize Michaël, Frankie Lam, Jens-Uwe Mager, George Johnson, Massimo Brini

Chapter 3. Acrobat

1. Should I upgrade from 4.0 to 5.0?

Q: What are the new features/improvements? [Q24]

A: Too many to list here. Adobe's site presently has a ton of material listing all the new tools and features. Here are a few of my favorites. Workflow automation Database connectivity Online collaboration Much improved javascript editor Great new JavaScript objects and interface control Tools for adding PDF structure XML support

Mr T (nert@bobco.com) [A41]

A: We just got our copy of Acrobat 5. First impressions are: 1. If you use forms, you can now use all available fonts, not just the base 13. 2. You can export pdf's to jpeg, tiff, and rtf. The rtf means that you can now create Word documents from PDF documents. We tried it, and the results were OK for text, but poor for anything with complicated formatting, like tables and columns. And non existant for graphics. 3. There are numerous improvements for the colour pre-press market, but we haven't evalutated them yet

Dan Sideen (dansideen@home.com) [A42]

A: 4.05 has Paper Capture. It is an add-on in 5.x. Paper Capture is OCR. Use Acrobat to scan a doc and edit the doc in Acrobat subject to the rule that you must edit text one line at a time. Very useful when making extra copies of documents.

(REMOVEraindoll@ziplip.com) [A43]

A: Adobe have now announced that the Capture plug-in is coming back for v5. It will be a separate download (for v5.00 users anyway) and is slated to be available about June 2001.

Mark Anderson (mark@notmeyeardley.demon.co.uk) [A44]

A: It wasn't worth it a few months ago, especially since Adobe removed the OCR capability. I understand they have since added a patch to put this feature back, but limited it's range. I bought it thinking I could use some of the new "bookmark" features, however I had to go back to ver. 4 since I particularly needed OCR (capture). I have yet to reinstall it, however by now it might be worth part of the price of the "upgrade".

Tore Hanson (hansont@ameritech.net) [A222]

A: I have seen several eps-files, that Acrobat 4 cannot distill but version 5 can. It seems version 5 can correct several font problems which version 4 cannot handle.

Veli Holopainen (veli.holopainen@kaleva.fi) [A223]

2. Cutting text from a PDF and pasting into Word

Q: Is there any way to cut and paste from a pdf to a word document or to another document in the adobe suite? I want to copy text from a pdf and put it somewhere else, but not as an image. [Q26]

A: First put Acrobat into continuous view mode. That way you can copy more than one page at a time, up to the entire document, but subject to clipboard size limits.

Dave Braze (davebraze@yahoo.com) [A49]

A: Adobe Acrobat 5.0 also has a save as RTF (a format Word can open) option...

Michael Still (mikal@stillhq.com) [A50]

A: In acroreader or acrobat just use the text select tool to select your text, copy & paste in the normal way.

Steve Cook (steve.cook@spamulike.bigfoot.com) [A51]

3. Guiding your readers through your document

Q: I received in the mail a pdf that when over a page, the hand icon has a down arrow on it and it functions to advance forward through the document when you click. I would like to add this feature to my documents but can find no reference to it in the Acrobat User's Guide. Does anyone know how to add this? [Q27]

A: You can do this in Acrobat by using the article tool. It allows you to guide your readers through the document. It is described in the Acrobat User's Guide from page 247 onwards.

mschulz (mschulz@bigpond.net.au) [A52]

4. Printing from the command line

Q: Can anybody tell me if a PDF commandline-tool is available, ie. I would like to execute a line like:

```
PDFPRINT ACRO.PDF PRINTERXX [COPIES=4]
```

resulting in 4 copies of the ARCO.PDF-document on the printer called PRINTERXX [Q28]

A: You can use the Acrobat Reader for printing PDF files via the command line (although you can't set the number of copies). From the Acrobat Developer FAQ:

"Using Command Lines with Acrobat and Acrobat Reader under Windows

These are unsupported command lines, but have worked for some developers. There is no documentation for these commands other than what is listed below. You can display and print a PDF file using command lines with Acrobat and Acrobat Reader.

```
AcroRd32.exe filename - Executes the Reader and displays a file.
AcroRd32.exe /p filename - Executes the Reader and prints a file.
AcroRd32.exe /t path printername drivename portname - Initiates
Acrobat Reader, prints a file while suppressing the Acrobat print
dialog box, then terminates Reader.
```

The four parameters of the /t option evaluate to path, printername, drivename, and portname (all strings).

```
printername - The name of your printer.
drivename - Your printer driver's name. Whatever appears in the
Driver Used box when you view your printer's properties.
portname - The printer's port. portname cannot contain any "/"
characters; if it does, output is routed to the default port for that
printer.
```

If using Acrobat, substitute Acrobat.exe in place of AcroRd32.exe in the command lines."

Gunther Schmidt (g.schmidt@bigfoot.de) [A53]

A: It has been reported that the `acrord32 /t` command line option does not work with Acrobat 5.0 It would appear to be a genuine bug, and Adobe apparently doesn't care, because the command-line switches are officially "undocumented". I went back to Acrobat Reader 4 because the `/t` switch doesn't work in Acrobat Reader 5. [This answer has been edited by mikal@stillhq.com]

Edward Mendelson (edward_mendelson@ziffdavis.INVALID) [A54]

5. Editing existing PDF files

Q: Using Acrobat 4 or 5, or any other program, is it possible to edit an existing PDF file: I need to add 1000 pictures (logos) to 1000 different, existing PDF-files. [Q31]

A: Well, yes, you can edit. One page at a time. You will need patience, or else some product that goes far beyond Acrobat. If using Acrobat, the easiest way to add a logo is to use the Object Touch-up tool, and copy/paste the logo from one PDF (which has it on a blank page, in the right place) to each new page.

Aandi Inston (quite@dial.pipex.com) [A57]

6. Command line searching a PDF

Q: Is there a parameter for having acrobat reader search for a specific word? If anybody knows some other parameters, can you post them (or just point me to a page where they are listed)? [Q32]

A: No. The Acrobat Developer FAQ lists some of them, but specifically notes that they are unsupported. I would strongly recommend that they are not used, and a supported method used instead. (If you do choose to use them, please don't complain when they stop working on an upgrade.)

Aandi Inston (quite@dial.pipex.com) [A58]

7. Running multiple versions of Acrobat

Q: I am working with Acrobat 4.05 and we have purchased Acrobat 5.0 (not the upgrade, the new one). Do I have to uninstall Acrobat 4 before installing 5.0? [Q33]

A: No, you can keep both. But do NOT uninstall 4 later if you don't do it first. Or, if you do, you will have to reinstall 5.

Aandi Inston (quite@dial.pipex.com) [A59]

8. Identifying the author of a document

Q: Does pdf involuntarily store information relating to the serial number/author of the acrobat install that was used to create the document? [Q34]

A: No. You can determine (often) the name of the software that produced the document and at what time though.

Michael Still (mikal@stillhq.com) [A60]

9. Creating PDFs within MS Word

Q: I've just installed Acrobat 4 and now it seems I have 3 different ways to create a PDF from within Word: - Print to the PDF writer - Print to Distiller - Use the "Create Adobe PDF" command on the File menu. How do these differ? Which are the preferred methods in terms of time taken/quality/file size? [Q35]

A: PDFWriter is fast, but not terribly good at graphics. Distiller is general better, but slower. It gives more control to you, and you should spend time understanding its options. Neither will include links, bookmarks etc. Use Create Adobe PDF if you want these features. This will make a larger file of course.

Aandi Inston (quite@dial.pipex.com) [A61]

10. Searching in online PDFs

Q: We are producing a cd-rom with lots of pdf's, all indexed by Adobe Acrobat Catalog. The navigation is done in plain HTML. When viewing pdf's, it is done by the Acrobat Reader (including the search option) embedded in the web browser. The effect is a Acrobat Reader with reduced menus. Now there is no access to the Adobe Acrobat Search menu item (edit/search/query), which gives us access to the Catalog build index. Anyone out there who knows about some add-on product, which enable the >catalog index of Acrobat Reader called through an internet browser? [Q36]

A: There is unlikely to be any. Search does not work through a browser; it has no programming interface in Acrobat Reader at all; and the programming interface in Acrobat does not allow the way the result list is handled to be changed.

Aandi Inston (quite@dial.pipex.com) [A62]

11. Highlighting the contents of a PDF

Q: Suppose I have a PDF document and a keyword I want to search for in the document. I want to create a hypertext link from an HTML document to this PDF document and have it either go to the instance of the keyword or highlight every instance of this keyword. How can I do this? Will this require products other than the free version of acrobat Reader? [Q37]

A: You can read the highlighting file format first

<http://partners.adobe.com/asn/developer/acrosdk/docs/highlt.pdf>

There are some products which do PDF search and highlight like

<http://www.dtsearch2.com>

or you can do it yourself by grapping all text from the pdf file and find the position of words.

Dave (davelo@hotmail.com) [A63]

12. Raster vs Vector images

Q: As I understand it a PDF supports both raster and vector images. Is there any way in Acrobat 4 to tell what type an image is? As far as I can tell I can't even select an image in Acrobat and act on it. [Q38]

A: You could use Quite A Box Of Tricks. Even the free demo allows you to click on the page. If it is an image you will get information on the image. If it is text, you will get information on that too. If it is neither, nothing will happen.

Aandi Inston (quite@dial.pipex.com) [A64]

13. Paper capture vs Acrobat capture

Q: Could someone give me some idea what the difference is between using the paper capture feature of acrobat and the full Acrobat Capture software that costs several hundred dollars? I'm using acrobat 4 at present, and the paper capture feature seems to work terribly. [Q112]

A: Capture is designed for high volume, semi-automated use. But take care! Acrobat's OCR is perfectly capable of good results from good originals. If you are getting poor results it may reflect poor quality originals, "difficult" originals (e.g. maths, coloured text on backgrounds, faxes, many others), or a scanning or post-scanning technique that needs improvement (e.g. poor choice of resolution, saving as a JPEG).

Aandi Inston (quite@dial.pipex.com) [A161]

14. Acrobat error messages

Q: I have generated a document using *****insert name of widget***** and I get error messages when viewing it in Adobe Acrobat. The messages are only numbers. How do I find out what they mean? [Q114]

A: Try hitting Ctrl-Ok, which will display a text description of the message for you.

Michael Still (mikal@stillhq.com) [A164]

15. Access for vision impaired users

Q: Are there any tools to ease access for people who are vision impaired? [Q117]

A: Adobe offers a plugin

<ftp://ftp.adobe.com/pub/adobe/acrobatreader/win/4.x/plugins/accs405.exe>

"The Adobe Acrobat Access Plug-in 4.05 enables vision impaired users to read Adobe PDF documents in Acrobat 4.0x or Acrobat Reader 4.0x. The Access plug-in supplements the standard Acrobat and Reader display of PDF documents with an alternative view that supports screen-reading applications for Microsoft Windows 95, Windows 98, and Windows NT. This view presents the textual content of a PDF file in a separate window. The window contains the text in an approximated reading order. Standard Acrobat features, such as hypertext links, notes, bookmarks, sounds, and movies, are available in the Access view"

And it gives you as option export to html or export to txt. I found it far superior to the export capabilities of Omnipage 11 and it is for free

U Berlin (twassmann@t-online.de) [A169]

16. Acrobat is really slow to load

Q: How can I make Adobe Acrobat faster to load? [Q121]

A: Adobe recommends that you put unused plug-ins in the "Optional" folder. This will mean that they are not loaded automatically on startup.

Miguel Tavares (medalha@hotmail.com) [A176]

17. The Acrobat Reader ActiveX control

Q: I'm using the Adobe ActiveX from Reader. It has the methods SetPageMode and SetLayoutMode, which are handy. Problem is no documentation as to what values to set the parameters to. I'm sure someone amongst the experts here knows. [Q127]

A: I don't have any documentation, but it's worth mentioning perhaps that there is no documentation for a reason: it is not a developer tool. Adobe created it only to display PDF files in internet explorer, and they don't support any other use. Given that, and that it could stop working as a developer tool with any upgrade, I'd advise people to consider very carefully before starting any development that they have to commit to supporting.

Aandi Inston (quite@dial.pipex.com) [A187]


```

500 500 500 500 500 500 500 500 500 500 500 500 500 500 500 500 750
]
/Encoding /WinAnsiEncoding
/FontDescriptor 117 0 R
>>

12 0 obj
<<
/Type /Font
/Subtype /TrueType
/Name /F2
/BaseFont /AJIANI+CMR12
/FirstChar 31
/LastChar 255
/Widths [ 500 913 272 489 816 489 816 761 272 380 380 489 761 272 326
272 489
489 489 489 489 489 489 489 489 489 489 272 272 272 761 462 462
761 734 693 707 747 666 639 768 734 353 503 761 611 897 734 761
666 761 720 544 707 734 734 1006 734 734 598 272 489 272 489 272
272 489 544 435 544 435 299 489 544 272 299 516 272 816 544 489
544 516 380 386 380 544 516 707 516 516 435 489 979 489 489 500
500 500 500 500 500 500 500 500 500 500 500 500 500 500 500 500
500 500 500 500 500 500 500 500 500 500 500 500 500 500 500
500 611 816 761 679 652 734 707 761 707 761 500 500 707 571 544
544 816 816 272 299 489 489 489 489 489 734 435 489 707 761 489
883 992 761 272 489 500 500 500 500 500 500 500 500 500 500 500
500 500 500 500 500 500 500 500 500 500 500 500 500 500 500
500 500 500 500 500 500 500 500 500 500 500 500 500 500 500
500 500 500 500 500 500 500 500 500 500 500 500 500 500 500
]
/FontDescriptor 13 0 R
>>

```

The Arial definition is embedded by your Visio page, notice that the Tex font 'CMR12' has no /Encoding.....

The possible pre-defined Encodings are :

1. StandardEncoding
2. MacRomanEncoding
3. WinAnsiEncoding
4. PDFDocEncoding

I've had a quick try at inserting an Encoding for each of these in the file, and I'm afraid it still didn't work, so there may be a more fundamental problem here.

I did run it through a PostScript interpreter, and it printed correctly, which suggests that the default encoding is StandardEncoding, or at least is compatible with it.

My suggestion would be to take this to a Tex newsgroup, and see if anyone there can help you, it may be that updating to a newer version of Tex (or whatever it was you used to convert the Tex document to PDF) will solve your problem.

Ken Sharp (ken@spamcop.net) [A202]

A: Sorry, but this doesn't tell us much. Which TeX System? How did you create the PDF from the TeX or DVI source? Did you use DVIPS? Did you use Distiller? Did you use pdfTeX?

Looking at the file, one notes that it was made by PDFWriter. Strike one. Do not use PDFWriter - always use Distiller.

Then it seems that it has CM fonts in TrueType form. Don't use them. Several versions of these are incorrectly made (for example, marked as "text" fonts so they will be reencoded by Windows). They all show up as empty boxes in Acrobat Reader 5.

The only quality font in there is Adobe's Arial MT. And it is reencoded to Adobe Standard Encoding!

Use quality tools to make PDF. See discussion at <http://www.yandy.com/acrobat5.htm> for example also <http://www.yandy.com/compare.htm>

Once you strip out the bad CM fonts you get a PDF file that works just fine. I put it up at <http://www.yandy.com/download/intro.pdf> (it's also substantially shorter than the file it was derived from).

Take a look (of course, there are still other problems, like unequal rendering of equal width rules, but that is a problem of your TeX System too hard to fix by editing the PDF file).

Louis Vosloo (support@yandy.com) [A204]

19. Changing the author of an annotation

Q: I edit a journal and we have recently begun to send manuscripts to reviewers as PDF files. A reviewer sent back an extensively annotated PDF file with instructions to send it on to the author. Unfortunately, his annotations all identify him (his name is in the annotation title bar). Thus, we cannot send this marked MS to the authors because the reviewer's anonymity would be lost. I looked through the Acrobat documentation but can't find any obvious way to change the authorship (say, to "Reviewer #2") of the annotations. Is there any way of doing that? [Q140]

A: You could export the comments as an FDF file (File > Export), open the FDF in a text editor and perform a Search & Replace operation, and then import the FDF back into a virgin document.

George Johnson (gejohn@att.net) [A216]

A: Checked out your suggestion. Works very well! What's very dangerous, is I imported the edited *.fdk back into the source file, and it placed the original annotations back on top of the original ones! For this reason, you're correct in suggesting that the virgin document be used as the destination. Otherwise, the annotated re-ports will include mirrored notations!!!

Happy Go Lucky (haphazard@earthlink.net) [A217]

20. Issues with Office XP

Q: We're about to upgrade from Office 97 to XP and I'm wondering if there are any potential Acrobat 4 or 5 issues I should know about. [Q148]

A: Have a look at <http://support.microsoft.com/support/kb/articles/Q302/5/96.ASP>

Harvey Turnbull (harvey.turnbull@ntlworld.com) [A230]

21. Problems viewing PDFs created with Adobe Illustrator in Acrobat 4 and 5

Q: I know someone who is attempting to create PDFs using Illustrator 9. When they go into Illustrator and try to save the file directly as a PDF (using Acrobat 4), the images look fine, but the text that she's typed loses some quality (looks a little jagged). [Q149]

A: The same problem occurs with Acrobat 5, and it isn't all rasterization.

Here's what we know so far (there have been a few brief threads on this topic in various newsgroups, including this one, with three or four contributors, total):

If the type in the Illustrator drawing is in front of anything that includes either transparency or certain types of gradients (I've learned there are something like nine different kinds of gradients and they don't all show this effect), Illustrator will convert the type to outlines prior to the Distiller step.

If type has been converted to outlines, either because Illustrator did it on its own, pursuant to the preceding paragraph, or because the user did it for other reasons, all hinting information is naturally lost, and each character is just a filled bezier curve.

Acrobat looks at each node of the curve and makes the pixel in which that node falls type color (black, if your type is black, for example). When the zoom level is small, all those solid black, unaliased pixels make each character look quite jagged. However, as you zoom in, the shapes get smoother and smoother. At a high magnification, the letters look the same as if they were actually type instead of being rendered into outlines.

The upshot is that for printing on a PS printer or sending to a service bureau, most everything works out okay. For printing on a non-PS printer, all bets are off. **HOWEVER**, it seems that in some cases, parts of a type block do indeed get rasterized. I can't tell you exactly where, but it seems to have to do with the

number of layers involved at a given point on the page. So one line of type may begin perfectly normally and halfway along the line turn into a raster image.

Adobe is aware of the problem and it is on their worklist for Acrobat 6.

Dick Margulis (margulis@fiam.net) [A231]

22. Starting page

Q: How can I open a pdf-file and show at start a different page ? [Q151]

A:

```
http://<path>/<pdffile>#page=123
```

Just learned it two days ago from this newsgroup...

Michael Hemmer (mhemmer@no_spam_samson.de) [A234]

23. Font information

Q: I am interested in knowing what font is being used for a given piece of text in a PDF. How can I do that? [Q156]

A: If you've got the full version of Acrobat, select the text using the TouchUp Text Tool. Then, right-click and select "Attributes". It will show you the name of the font used.

Rachel Burnsed (burnsed@bcl-computers.com) [A239]

24. Acrobat 5 is slow drawing line art

Q: It's not so much the loading time that concerns me, it's the slowness of panning and scrolling in an already opened page.

Many of them do contain line art; stuff like mechanical drawings, electrical circuits, etc. [Q159]

A: This leads me to believe you are seeing the exact same problem we observed when moving from 4 to 5.

In our case it was electrical schematics which really highlighted the problem. While I haven't checked version 5 of the full product, I would expect it to behave exactly the same as I assume it shares the same drawing engine. We have stuck with 4.05 as its drawing speed for such line art is snappy - near instantaneous on a 750MHz PIII with a Matrox G400 Max.

Version 5 was terrible. I sent Adobe a sample file (single page A2 schematic, 70Kb size) where, when zoomed in, a redraw was taking about 3 seconds to complete.

I have not checked 5.05.

Dale Rebgetz (Dale.Rebgetz@setec.com.au) [A244]

Chapter 4. Acrobat forms

1. Methods of creating Acrobat forms

Q: We use Word for producing forms, which I then dutifully run thru Distiller to get pdfs. Now I'd like to go to the next stage, which is putting in form fields so people can fill them out onscreen rather than on paper. Now the only option that I can see is that either (a) I get busy with the Acrobat Exchange form tool (boring, and the forms are vveeerrry long) or (b) I get the people who create the forms to put in some marker in Word that gets converted by a Acrobat add-on to Acrobat fields or (c) install Acrobat on form producers' PCs and train them up (not more attractive than (a), really, and it's certainly expensive).

Can anyone advise on the possibilities of (b)? This must be a relatively common problem.

I've checked whether Exchange converts Word forms into Acrobat ones: no, it doesn't (in v3, not sure about v5). [Q69]

A: Here are your options, and there aren't many.

Learn all the intricacies of the Acrobat forms tools. It has many time saving shortcuts you can use. Learning how to duplicate and rename fields effectively can really cut down your development time. Also learn to "borrow" ready made stuff from other forms. You can easily cut your development time in half or less by using these shortcuts.

There's an excellent forms tutorial that explains these tricks. It's here:

<http://www.planetpdf.com/mainpage.asp?webpageid=1155>

Since you are using Word, you should seriously consider Caere Omni form. What it does is it opens a Word form and converts it to an Omni form, which can then be saved as a PDF form or HTML form. The conversion is not perfect, but you can expect about a 90% accuracy rate. It will even convert simple form functions and calculations for you. I highly recommend it, and the price is right.

Another option is to import your Word files into a PDF savvy app like FrameMaker or PageMaker, and let them perform the PDF conversion. FrameMaker has a feature called PostScript frames which allows you to manually enter the PostScript or pdfmark code required for generating PDF form fields. PageMaker has a third party plugin you can purchase. It simulates the Acrobat forms tool. I have yet to find a way to do this with Ventura.

And the last way I know, is to include pdfmarks directly in your Word documents, by inserting them through Word's field codes. Thomas Merz has an excellent pdfmark primer that explains how to do most of these procedures. It's free. Get it here: <http://www.pdflib.com/pdfmark/index.html>

Bryan Guignard (bryang@sympatico.ca) [A117]

2. Images inside form elements

Q: Can anyone confirm that it is not possible to put an image (jpg/gif) into a form element? I would be happy to convert the image to pdf and then placing it into the form if that needs be. I want to make a pdf document service on our website that provides our customized data and allows the user to upload their logo to place into the document. At this point, I think we need a form (FDF) which will act as a template container, into which we can populate from the database and HTML form, which will include images..... Can anyone provide some pointers, even if to say, its not possible. [Q70]

A: If you make the form of type Button, and select Icon only to display, you can choose a pdf as the icon. this effectively puts a pdf inside a pdf, and, if the pdf is an image, your problem is solved.

Its even easier in Acrobat 5, which allows you to select most bit mapped image formats as icons for buttons.

Dan Sideen (dansideen@home.com) [A118]

3. Using ASP to collect form values

Q: Can I use ASP to collect field values from a pdf form? [Q71]

A: Yes, you can. Put the field values into session variables, then assign the variable name to a form field using the FDFSetValue method available in the FDF toolkit from Adobe. You can code in VBScript (preferred) or Javascript.

Here is some sample code that does that, then displays the data in a "master" pdf

```
Set FdfAcX = Server.CreateObject("FdfApp.FdfApp")
Dim objFdf
Set objFdf = FdfAcX.FDFCreate
name = Session("name")
address1 = Session("street") & " " & Session("poboxorfloor")
address2 = Session("city") & " " & Session("province") & " " &
Session("postalcode")
title1 = Session("title1") title2 = Session("title2")
email = Session("email")
objFdf.FDFSetValue "name", name, Off
objFdf.FDFSetValue "title1", title1, Off
objFdf.FDFSetValue "title2", title2, Off
objFdf.FDFSetValue "address1", address1, Off
objFdf.FDFSetValue "address2", address2, Off
objFdf.FDFSetValue "email", email, Off
objFdf.FDFSetFile
"http://warehouse.informco.com/asp_projects/liberty/lib06_D.pdf"
Response.ContentType = "application/vnd.fdf"
Response.BinaryWrite objFdf.FDFSavetobuf

objFdf.FDFClose
```

```
Set objFdf = Nothing Set FdfAcX = Nothing
```

Dan Sideen (dansideen@home.com) [A119]

4. Onscreen completion

Q: Many forms are available in PDF format. I was wondering if there is a package that will allow the user to fill them out onscreen before printing them out? Note that I am not looking for the user to edit them, only to fill the blanks. [Q72]

A: Acrobat can make any (unprotected) PDF file into a fillable form. Try for instance <http://www.quite.com/box/qboxbuy1.pdf> (it's an order form, but don't worry, you'd have to print and fax it to order).

Aandi Inston (quite@dial.pipex.com) [A120]

A: Acrobat Reader, if the forms already have fillable form fields.

Acrobat (full distribution), if you need to add any fields, including submit/import methods. Once you've added the required fields, Acrobat Reader, operating as a plugin in your browser, as well as a live TCP/IP connection, is enough to fill, submit, and import data with your form.

If you've deduced this means you can save form data (not PDF, but FDF), using Acrobat Reader, you are correct. See <http://segraves.tripod.com/index3.htm> for a couple of simple examples.

You DO NOT have to be connected to the internet to do this. It works fine with a web server running on the same computer with the browser, i.e., "localhost".

Bill Seagraves (wsegrave@mindspring.com) [A121]

5. FDF form filling with CGI

Q: Could you please tell wheter and how it is possible to use CGI to fill in a PDF-form [Q73]

A: Yes. Most easily, by the CGI generating an FDF file and delivering that.

Aandi Inston (quite@dial.pipex.com) [A122]

A: No and Yes!

No: CGI can't do it, as it is a standard, i.e. Common Gateway Interface.

Yes: You could write a script in a CGI-compliant language, e.g., Perl, that would generate a FDF file with the field names and associated values in it. For an example, please see <http://segraves.tripod.com/index3.htm>. You can see the required format of the FDF by simply submitting a blank form.

Bill Segraves (wsegrave@mindspring.com) [A123]

A: Get the adobe FDF toolkit from www.adobe.com. It contains Activex controls and is useable in C++, VB, VBS, ASP and Java to automate the creation of separate fdf files. The populated forms can then easily be displayed inside a browser just by pointing to the fdf. This loads the "parent" pdf, and fills the fields.

Dan Sideen (dansideen@home.com) [A124]

6. FDF, Java, and special characters

Q: When I do a setValue("fieldname", a Java String, false) if the java string has a french character inside (like an accented character), they are transformed in the document and not property I use the Times Roman font in the fields. [Q74]

A: The Java version of the FDF toolkit is horribly broken with respect to special characters. It uses the Java Native Interface (JNI) for attaching C libraries to Java, and the JNI relies on UTF-7 encoding. This doesn't seem to be implemented in the FDF toolkit, so the Java FDF toolkit is unusable from a European point of view. I haven't yet tested the new version 5 of the FDF toolkit.

Thomas Merz (tm@pdflib.com) [A125]

7. Emailing completed forms

Q: When I create the form with the full version of Acrobat 5.0 using a java script – the email function works great. When I access the form with Acrobat Reader 5.0 the function does not work. Does anyone have any recommendations on the best way to email forms using the Reader? [Q75]

A: Use a CGI or ASP. Except in very limited circumstances (centrally controlled intranets) sending direct e-mail is completely impractical. This limitation is nothing to do with Acrobat.

Aandi Inston (quite@dial.pipex.com) [A126]

A: You should forget the idea of sending e-mail directly from the client for at least two reasons. 1. You cannot reliably send e-mail from user's machines. Millions of people do not have working e-mail programs, relying instead on ser-

vices like Hotmail. Others use e-mail programs that can't be automated. 2. The free Acrobat Reader can't e-mail. So, what you do instead is have your web process (typically an ASP or CGI) send the mail. However, it won't have access to the whole form, only an FDF file.

Aandi Inston (quite@dial.pipex.com) [A166]

8. Inner workings of submitting a PDF form

Q: In the submit button on my PDF form I reference the URL of my Uniface program. In my Uniface program I then expect the FDF data to be in the first parameter/input to the form. But I only get the first line of the FDF file! Could someone explain to me how the submit action works with the PDF file and how it parses data to a CGI program. [Q118]

A: The entire FDF file is generated and sent as a single stream of data to the web server. It is not sent as lines (http has no concept of line breaks in the data portion). If you don't get the whole thing, this must be a problem with your chosen interface language/system. Possibly, there's a way to get more data. Possibly, the line endings in the FDF upset the development system (they could be CR, CRLF or LF; a program must be capable of understanding all the combinations).

Aandi Inston (quite@dial.pipex.com) [A170]

9. What is FDF?

Q: What is FDF, what's it for and where does it come from? [Q126]

A: FDF = Forms Data Format It is an ASCII based file format. It's designed to hold the data a user fills into a PDF form so it can be submitted to a web server for processing. Also, it can contain processed data that a web server can send back to the user. The most common way of creating an FDF file is to fill out a PDF form and click the SUBMIT button. This is the only option if you are using Acrobat Reader. However if you have the full commercial Acrobat package you also have an option to save FDF file(s) to a local disk. A special version of FDF is also used to store PDF comments data. One big advantage FDF has over regular HTML forms data is that it can contain vector and bitmap graphics as well as text. The PDF specification document contains the entire FDF specification. It's a free download (7 megs) from the Adobe website.

Bryan Guignard (bryang@sympatico.ca) [A185]

10. Software to fill in forms

Q: There are a lot of fillable PDFs floating around now. The problem is, with Acrobat Reader, you can fill in the form but it doesn't allow you to save the data.

What's the minimum product that will allow the user to fill in a form, save the data to an FDF, retrieve it and print the filled-in form? [Q147]

A:

1. Acrobat Reader to fill in the form and submit the form data to a CGI server, which can be on the same computer as the User, if need be.
2. A one-line Perl script to convert the URL-encoded name=value pairs to FDF, returning same to the User's browser.

Although a one-line Perl script can do the job, it is hardly readable. At

<http://segraves.tripod.com/index3.htm>, you'll find samples of what can be done with this approach. The Perl scripts that do the job are a dozen or so lines of code, depending on whether null values are dropped or kept in the result, according to specific needs. This technology is used routinely for forms with hundreds of form fields.

William Alexander Segraves (wsegrave@mindspring.com) [A228]

A: Adobe Acrobat Approval. Just recently released.

(zevevet@hotmail.com) [A229]

11. FDF form creation with CGI - gotchas

Q: Are there are things I should be aware of when generating FDFs via a CGI on a webserver? [Q162]

A: A couple of suggestions:

You may wish to write a program/script to generate the FDF in the prescribed format. Get the script working from the command line before you try to implement it on a web server, fi that's your approach.

You may wish to use care in the choice of the Content-type: of what you send back to your browser from a CGI script on a web server.

MSIE5.X (and6) doesn't treat Content-type: text/plain kindly, as it ignores your specified Content-type: statement and attempts to guess the Content-type: from what's inside the data. This is bad because parts of the FDF look something like HTML tags. MSIE mangles it. See the form versions on <http://segraves.tripod.com/index2.htm> for a Content-type: text/html workaround for the MSIE problem using (obsolete) tags to wrap the plain text FDF data, forcing MSIE to display it as example text.

Of course, if all you wish to do is fill the PDF form, use the appropriate Content-type: application/vnd.fdf, together with, IIRC, the required "#FDF" at the end of the URL, if the URL-encoded form data originates from a PDF.

William Alexander Segraves (wsegrave@mindspring.com) [A249]

Chapter 5. Common questions about PDF

1. Document sizes

Q: I would know what is the size per page of a pdf document when using a scanner : a) text only b) text and pictures [Q84]

A: Impossible to say. It depends first and most importantly on the resolution. 400 dpi is 1600% larger than 100 dpi. Beyond that everything depends on the kind of original, other scanning settings, and how you make the PDF. DON'T scan directly to Acrobat UNLESS you run Paper Capture.

Aandi Inston (quite@dial.pipex.com) [A139]

2. The size of a scanned page

Q: I have tried to make a decent pdf but quality and size is hard to get into an acceptable range. I've tried to scan an original consisting of 5 pages clean quality paper document (A4) but the result becomes a 1.7MB PDF - readable but poor and ragged screen text. Acceptable when printing but the byte size is hardly acceptable for the web? [Q85]

A: We find an a4 page, once G4 fax compressed (CCITT Group 4) is about 50kb.

Michael Still (mikal@stillhq.com) [A140]

A: Pure scans are rarely suitable for the web. They are too large. That is a key reason to use Paper Capture. However, you may be OK with a pure black and white scan. If you are not going to run Paper Capture, do NOT scan direct to Acrobat. Use Distiller. Scan to Acrobat (and import images to Acrobat 4) ONLY as a preparation for Paper Capture.

Aandi Inston (quite@dial.pipex.com) [A141]

A: I've noticed a strange thing about scanning into Acrobat: Scanned (grayscale) the first of 5 A4 docs, saved it and it was 51KB. Then, not altering any settings, I scanned all 5 pages consecutively and saved the PDF was 2MB. I don't get the math here! Seems to be a waste of time to scan import directly into Acrobat.

Kent Isaksson (kentisaksson@netscape.net) [A142]

3. Reliability of Paper Capture

Q: How reliable is Paper Capture at OCRing scanned text? [Q86]

A: Paper Capture is good but it really doesn't produce a reliable result which is the case of all OCR programs I've seen. It mistakes 0 for O and 6 for 8 etc and production hours soars.

Kent Isaksson (kentisaksson@netscape.net) [A143]

A: This may be obvious, but OCR stands for "Optical Character Recognition" It will only deal with text, and then only text that is clean enough so it can recognize it and then only text that does not include math or non-alphabetical symbols. So the pictures will not be reduced in size, and the text will only be reduced if it managed to actually "OCR" it.

Louis Vosloo (support@YandY.com) [A144]

4. Customizing bookmarks

Q: I'm just tired of the bland icons that Adobe serves up for use when creating bookmarks. Is there a way to make different symbols or perhaps even colors for use in the bookmarks sections I use pagemaker and photoshop to create pdf files and the default bookmarks take away from the overall look and feel. [Q88]

A: Acrobat 5 allows you to specify text formatting such as Bold, Italic, and different colors for the text part of bookmarks. I'm not aware of any way to change the icon. it would be nice if Acrobat had some different bookmark icons to choose from, the way you can change text annot icons.

Bryan Guignard (bryang@sympatico.ca) [A146]

A: Versions of Acrobat prior to 5.0 do not understand color entry in the outlines (or text styles); they ignore them.

MicroPress, Inc (support@micropress-inc.com) [A147]

5. VB silent PDF printing from Acrobat Reader

Q: I'm trying to make a simple VB program that silent prints a given PDF file. [Q89]

A: Here is a VB routine that use DDE and works with Reader 4. Note that you must start Acrobat or Reader first, and that FormsHome is a global variable defined elsewhere.. Sub Printit(FileToPrint As String) 'Print a file using Acrobat Dim sCmd As String 'command string to pass via DDE Dim FullName As String ' fully qualified file name FullName = FormsHome + "" + FileToPrint sCmd = "[FilePrintSilent(" & FullName & ")]" frmFindForm.lblDDE.Caption =

```
"Printing..." & FileToPrint frmFindForm.lblDDE.LinkTopic = "Acroview | Control"  
frmFindForm.lblDDE.LinkItem = "" If frmFindForm.lblDDE.LinkMode <> 2  
Then frmFindForm.lblDDE.LinkMode = 2 End If Rem frmFindForm.lblDDE.LinkTimeout  
= 1 frmFindForm.lblDDE.LinkExecute sCmd frmFindForm.lblDDE.LinkMode  
= 0 frmFindForm.lblDDE.Caption = "" frmFindForm.lblDDE.LinkMode = 2
```

Dan Sideen (dansideen@home.com) [A148]

6. Word counts

Q: How can I get a count of the number of words in a PDF file? [Q91]

A: If you upgrade to Acrobat 5, you can then use a simple javascript to count words. It's very easy and works well for me.

Bryan Guignard (bryang@sympatico.ca) [A151]

A: There is an accessibility plugin from Adobe that can export the contents to text format, and that should be easy enough to count words from

Anders Thulin (Anders.X.Thulin@telia.se) [A152]

7. Blank PDFs

Q: How do I create a PDF with just a blank page in it? [Q92]

A: With your word processor, distill an empty page.

Matti Vuori (mvuori@koti.soon.fi) [A153]

A: You could Distill a job like this : %!PS showpage with Acrobat 5, but a dim memory suggests that this may not work in Acrobat 4. In which case, I think you need to start with a simple job like : %!PS /Helvetica findfont 20 scalefont setfont 100 100 moveto (delete me) show showpage And then delete the text in Acrobat. If you 'save as' then you should eliminate the deletion totally.

Ken Sharp (ken@spamcop.net) [A154]

A: <http://www.planetpdf.com/mainpage.asp?webpageid=480>

Bryan Guignard (bryang@sympatico.ca) [A155]

A: There is an example program in the Panda distribution (<http://www.stillhq.com>) that creates a PDF just containing a blank page.

Michael Still (mikal@stillhq.com) [A156]

8. Do WebTVs support PDF documents?

Q: I have a user who is complaining that she cannot view, print or save PDFs with WebTV. Is this true? [Q142]

A: VERY true. WebTV does NOT have an operating system that supports PDF's among other things. It was never designed as a computer. Here are some WEBTV links

<http://help.webtv.net/>

<http://developer.webtv.net/design/tools/viewer/instructions/Default.htm>

<http://www.geocities.com/ResearchTriangle/8795/links.html>

Don Daniels (viewelk@NOJOYINSPAMavci.net) [A219]

Chapter 6. Conversion

1. PDF to raster / bitmaps

Q: Is there an easy way to turn a PDF into a TIFF [or other raster / bitmap files]? I've tried exporting as an EPS then opening it in Photoshop, but Photoshop is giving me an error message. [Q64]

A: I just did it with Illustrator 9.

Jim K (jkajpust@concentricRATS.net) [A105]

A: Try using ghostscript from <http://www.ghostscript.com>

Michael Still (mikal@stillhq.com) [A106]

A: Try Konvertor_pdf2xxx (<http://www.logipole.com>)

Jean Piquemal (j.piquemal@wanadoo.fr) [A107]

A: BCL Computers' Freebird software is available as a command line program. Freebird converts PDF to TIFF, JPEG, and BMP. For more information, visit our website at <http://www.bcl-computers.com>

Rachel Burnsed (burnsed@bcl-computers.com) [A108]

A: Try ghostscript, e.g.: `gs -sDEVICE=jpeg -sOutputFile=temp.%03d.jpg -c save pop -f your_file.pdf` This will create one jpg file for every page, the "%03d" being replaced by a 3 digit page count.

Helge Blischke (H.Blischke@srz-berlin.de) [A109]

A:

```
gs command line options:  
ad b: -r300  
ad c,d: -sDEVICE=png... -sOutputFile=filepg%03d.png  
        (where the "... " denote one of the variants of the png devices)  
ad a: -c save pop -f your_pdf_file  
        (the "-c save pop" recommended but nor essentially necessary)  
ad e: not possigle by command line switches. You'd have to write a  
        PostScript wrapper that determines the current page size and
```

orientation (using the BeginPage approach).

Helge Blischke (H.Blischke@srz-berlin.de) [A175]

2. TIFF to PDF

Q: I'm trying to convert TIFF files to PDF files. Preferably, I'd like to do this on a Unix (Solaris) server, and it definitely has to be something that can run in the background, unattended. [Q65]

A: PDFlib does this. And yes, we will answer your support mails :-). There's also a dedicated PDFlib mailing available.

Thomas Merz (tm@pdflib.com) [A110]

A: Panda is a free (as in GNU GPLed) API which runs on various Unices, Linux and Windows and can do this sort of conversion work. Have a look at <http://www.stillhq.com> for more details

Michael Still (mikal@stillhq.com) [A111]

A: 'convert' (part of the ImageMagick tools) can translate almost any image format to almost any other. So 'convert your.tiff your.pdf' will do what you want. I just tried it on a tiff file lying around and it worked fine. Convert is tremendously useful.

On Linux it is part of the ImageMagick RPM package. For Solaris I guess you have to get the source and compile. If rpm runs under Solaris, you might be able to get the src rpm and do `rpm --rebuild ImageMagic...src.rpm` where the ... depend on what version you get.

Go to and you'll be redirected to page pointing to tons of versions of the RPM (for various releases of many different linux distributions). They list

<http://www.imagemagick.org>

as the source for the software but it seems to be down right now.

Sanjoy Mahajan (sanjoy@skye.ra.phy.cam.ac.uk) [A112]

A: You can use "DaVince Tools" at <http://www.davince.com>. The tiff2pdf converter does exactly what you specified, except for the hidden text (are you referring to OCR?). It does recursion, 1 to 1 and many to 1 PDF conversion and thumbnail generation.

Dan Cogliano (dan@davince.com) [A208]

3. PDF to office document

Q: How can I convert my PDF documents to something I can edit in a text editor? [Q66]

A: You could convert the PDF file to another format. BCL Computers' Drake software will convert PDF to RTF for editing in Word; however, it will also preserve page breaks. If you use BCL Magellan for converting PDF to HTML, and select the "HTML3" option, you should get an output file which can be opened and edited in Word. Free demos of these Acrobat plug-ins are available at <http://www.bcl-computers.com>.

Rachel Burnsed (burnsed@bcl-computers.com) [A113]

A: You can convert the PDF file to test with ghostscript.

<http://www.cs.wisc.edu/~ghost/>

A. Sinan Unur (asu1@cornell.edu) [A196]

A: Try xpdf, "pdftotext" is your friend. <http://www.foolabs.com/xpdf/>

Frank M. Siegert (frank@this.net) [A198]

A: Acrobat 5 does a nice job of this. Also create RTF files.

Dan Sideen (dansideen@home.com) [A200]

A: There are several ways to get PDF content into Word. A few options: Acrobat 5 has a "Save as RTF" feature. This exports the text with little formatting.

BCL Computers' Jade and Drake software allow you to convert PDF to RTF. Jade requires that you manually select the areas of text, tables, and graphics to convert; Drake automatically converts the entire file to RTF, preserving the layout and text formatting from the PDF. Free demos are available online at <http://www.bcl-computers.com>.

PDFZone (<http://www.pdfzone.com>) should list other PDF conversion tools as well.

Rachel Burnsed (burnsed@bcl-computers.com) [A240]

4. HTML to PDF

Q: My question is how to convert a html file into pdf file. Are there any free-soft for doing this? [Q67]

A: HTMLDOC does this: <http://www.easysw.com/htmldoc/>

Michael Sweet (mike@easysw.com) [A114]

5. MS Office to PDF

Q: How can I convert my MS Office documents to PDF? [Q68]

A: [Have a look at Adobe Distiller, and the Distiller like products listed in the Products not from Adobe section]

Michael Still (mikal@stillhq.com) [A115]

A: You should try the Sowedoo Easy PDf converter that enables you to convert a batch of Office files. Have a look at www.sowedoo.com

Chaize Michaël (mc@sowedoo.com) [A116]

A: ImageToPDF for Excel version 1.0 has been tested on Win 95/98SE and Excel 97/2000.

This macro is free and will/is:

1. Batch convert image files supported by Adobe Acrobat versions 4 and 5.
2. Fill Document Info fields.
3. Create Thumbnails
4. Set the file to Standard Open, Bookmarks or Thumbnails.
5. Information placed on the spreadsheet is PDFWorkshop Friendly.

http://freepdf.homestead.com/files/ImageToPDFver1_0.zip

Tiger (saccuzzo@home.com) [A209]

6. Tex to PDF

Q: I have a problem in distilling postscript documents produced with LATEX. The pdf doesn't look good on the screen, although the printed document is perfect. [Q116]

A: Take a look at "Creating quality Adobe PDF files from TeX with DVIPS" by Kendall Whitehouse/EMERGE

<http://www.math.hawaii.edu/~ralph/MathOnWeb/TeXPDF.html>

Alex Cherepanov (alexcher@erols.com) [A168]

7. XML to PDF

Q: How can I convert from XML to PDF? [Q123]

A: Check out <http://www.xmlpdf.com/overview.html>

John Farrow (john.farrow@xmlpdf.com) [A180]

A: `java -cp $classpath org.apache.xalan.xslt.Process -IN myfile.xml -XSL mystyle.xsl -OUT newfile.fo` MSXML will allow you to do the exact same thing for except it is a COM object. The transformation would be performed by XSL.

GDCII (miles@netset.com) [A183]

8. PS to PDF

Q: How do I convert a PS file into a PDF? [Q128]

A: Try `ps2pdf` if you are using a Unix iteration...

Michael Still (mikal@stillhq.com) [A192]

A: If you want to convert more than one PS file into a single PDF, then try:

```
gs -dNOPAUSE -dBATCH -sDEVICE=pdfwrite -sOutputFile=your.pdf -c save pop -f file1.ps file2.ps
```

Helge Blischke (H.Blischke@srz-berlin.de) [A193]

9. PDF to BMP

Q: How do I convert PDF documents into BMP files? [Q141]

A: Try pdf2bmp.

See <http://www.tinaja.com/psutils/pdf2bmp.psl> for the sourcecode and <http://www.tinaja.com/psutils/pdf2bmp.bmp> for the demo.

Don Lancaster (don@tinaja.com) [A218]

10. RTF to PDF

Q: I am in search of a server side solution for converting RTF or MS Word docs to PDF. The proposed solution is to be deployed on an UNIX platform. [Q161]

A: For RTF, you could use unrtf (<http://freshmeat.net/projects/unrtf/>) to convert to PostScript, and run ps2pdf on the result to get PDF.

For MS Word, try wvWare (<http://wvware.sourceforge.net/>).

J.H.M. Dassen (Ray) (usenet@zensunni.demon.nl) [A248]

Chapter 7. Development: Acrobat plugins

1. Toolbar icons

Q: I've created my own ToolButton in the default ToolBar but I'm not able to load the icon I've designed. I'm using the AVToolButtonNew method. [Q25]

A: What are the attributes of the icon: - type (bitmap, icon, ...) - size - colours?

Aandi Inston (quite@dial.pipex.com) [A45]

A: The icon type is Icon the size is 18x18 with 256 colours.

Giacomo (x-ray69@usa.net) [A46]

A: In my project it is a Bitmap, not Icon. This may be critical. Mine is also 20 x 20 with 16 colours. This may be less important. Refer to (I think) the ImageSel project; the documentation on this point is poor.

Aandi Inston (quite@dial.pipex.com) [A47]

A: It displays just a question mark insted of my icon. I've also tried to convert the .ico file into a bitmap and load it using the same procedure as shown in ImageSel project. No way. I'm banging my head on the wall... I need it to work out!

Giacomo (x-ray69@usa.net) [A48]

2. Embedding a font into a PDF

Q: I have a plug-in that adds a text banner to the foot of each page in a PDF. How do I correctly add a font resource, e.g. Arial, to the PDF so that the banner prints correctly. The PDFs are straight bitmap conversions so have not font info. [Q78]

A: If you are using the Cos layer this is an interesting and serious challenge. If you are using the PDFEdit layer, this does include options to create and embed a font from a "system" (i.e. installed) font. Since Arial is a virtual clone of Helvetica, that particular example would not be worthwhile. The base fonts like Helvetica are guaranteed to print without embedding.

Aandi Inston (quite@dial.pipex.com) [A132]

Chapter 8. Development: Examples

1. Masks for images (PDF 1.3)

Q: PDF 1.3 has documented a new "Mask" key that provides explicit masking ability. This is supposed to allow you to embed an image (say, a jpeg) along with a 1 bit-per-sample mask that indicates where the jpeg should and should not appear on the page.

I haven't been able to get this to work – the image just renders without respecting the mask.

Does anyone have any experience with this feature? The PDF reference does not have any sample code that does this, so I would be interested in any examples you could throw my way. [Q4]

A: It would be easier if you provided an excerpt of the PDF you generated. Don't forget to put the ImageMask dictionary entry in the mask image, the mask's BitsPerComponent to 1, and no color space.

Pierre Baillargeon (pb@artquest.net) [A9]

A: And another thing I noticed about legal masks is that the mask image XObject appears in the resource dictionary for the page it appears on.

Ed Bomke (edb@discryptic.com) [A10]

A: Here are some fragments from one that works:

```
25 0 obj <<
  /Type /XObject
  /Subtype /Image
  /Width 317
  /Height 299
  /BitsPerComponent 1
  /ImageMask true
  /Length 524
  /Filter /CCITTFaxDecode
  /DecodeParms <<
    /K -1
  /Columns 317
  >> >>
stream
```

```
26 0 obj <<
  /Type /XObject
  /Subtype /Image
```

```
/Mask 25 0 R
/Width 317
/Height 299
/BitsPerComponent 8
/ColorSpace /DeviceRGB
/Length 95613
/Filter /DCTDecode
>>
stream
```

Aandi Inston (quite@dial.pipex.com) [A11]

2. ASCII85 encoding / decoding

Q: I'm looking for a piece of code to do ASCII85 encoding/decoding. Does anyone know where to get this? [Q6]

A: <ftp://ftp.webcom.com/pub/haahr/src/encode85.c>

<ftp://ftp.webcom.com/pub/haahr/src/decode85.c>

Tom Kacovinsky (tjk@ams.org) [A15]

A: The text of the file <ftp://ftp.webcom.com/pub/haahr/src/encode85.c>:

```
/* encode85 - convert to ascii85 format */

#include <stdio.h>
#define atoi(s) strtol(s, 0, 0)

static unsigned long width = 72, pos = 0, tuple = 0;
static int count = 0;

void init85(void) {
    printf("<~");
    pos = 2;
}

void encode(unsigned long tuple, int count) {
    int i;
    char buf[5], *s = buf;
    i = 5;
    do {
        *s++ = tuple % 85;
        tuple /= 85;
    } while (-i > 0);
    i = count;
    do {
        putchar(*-s + '!');
    }
```

```

        if (pos++ >= width) {
            pos = 0;
            putchar('
');
        }
    } while (i- > 0);
}

void put85(unsigned c) {
    switch (count++) {
        case 0: tuple |= (c << 24); break;
        case 1: tuple |= (c << 16); break;
        case 2: tuple |= (c << 8); break;
        case 3:
            tuple |= c;
            if (tuple == 0) {
                putchar('z');
                if (pos++ >= width) {
                    pos = 0;
                    putchar('
');
                }
            } else
                encode(tuple, count);
            tuple = 0;
            count = 0;
            break;
    }
}

void cleanup85(void) {
    if (count > 0)
        encode(tuple, count);
    if (pos + 2 > width)
        putchar('
');
    printf("~>
");
}

void copy85(FILE *fp) {
    unsigned c;
    while ((c = getc(fp)) != EOF)
        put85(c);
}

void usage(void) {
    fprintf(stderr, "usage: encode85 [-w width] file ...
");
    exit(1);
}

extern int getopt(int, char *[], const char *);
extern int optind;
extern char *optarg;

```

```

int main(int argc, char *argv[]) {
    int i;
    while ((i = getopt(argc, argv, "w:?")) != EOF)
        switch (i) {
            case 'w':
                width = atoi(optarg);
                if (width == 0)
                    width = ~0;
                break;
            case '?':
                usage();
        }

    init85();
    if (optind == argc)
        copy85(stdin);
    else
        for (i = optind; i < argc; i++) {
            FILE *fp = fopen(argv[i], "r");
            if (fp == NULL) {
                perror(argv[i]);
                return 1;
            }
            copy85(fp);
            fclose(fp);
        }
    cleanup85();
    return 0;
}

```

Michael Still (mikal@stillhq.com) [A246]

A: The text of the file `ftp://ftp.webcom.com/pub/haahr/src/decode85.c`:

```

/* decode85 - convert from ascii85 format */
#include <stdio.h>

static unsigned long pow85[] = {
    85*85*85*85, 85*85*85, 85*85, 85, 1
};

void wput(unsigned long tuple, int bytes) {
    switch (bytes) {
        case 4:
            putchar(tuple >> 24);
            putchar(tuple >> 16);
            putchar(tuple >> 8);
            putchar(tuple);
            break;
    }
}

```

```

        case 3:
            putchar(tuple >> 24);
            putchar(tuple >> 16);
            putchar(tuple >> 8);
            break;
        case 2:
            putchar(tuple >> 24);
            putchar(tuple >> 16);
            break;
        case 1:
            putchar(tuple >> 24);
            break;
    }
}

void decode85(FILE *fp, const char *file) {
    unsigned long tuple = 0;
    int c, count = 0;
    for (;;)
        switch (c = getc(fp)) {
            default:
                if (c < '!' || c > 'u') {
                    fprintf(stderr, "%s: bad character in ascii85 re-
gion: %#o
", file, c);

                    exit(1);
                }
                tuple += (c - '!') * pow85[count++];
                if (count == 5) {
                    wput(tuple, 4);
                    count = 0;
                    tuple = 0;
                }
                break;
            case 'z':
                if (count != 0) {
                    fprintf(stderr, "%s: z inside ascii85 5-
tuple
", file);

                    exit(1);
                }
                putchar(0);
                putchar(0);
                putchar(0);
                putchar(0);
                break;
            case '~':
                if (getc(fp) == '>') {
                    if (count > 0) {
                        count--;
                        tuple += pow85[count];
                        wput(tuple, count);
                    }
                    c = getc(fp);
                    return;
                }
        }
}

```

```

        fprintf(stderr, "%s: ~ without > in ascii85 section
", file);
        exit(1);
    case '
': case '
': case ' ': case ' ':
        case "": case 'f': case "'": case 0177:
            break;
        case EOF:
            fprintf(stderr, "%s: EOF inside ascii85 section
", file);
            exit(1);
    }
}

void decode(FILE *fp, const char *file, int preserve) {
    int c;
    while ((c = getc(fp)) != EOF)
        if (c == '<')
            if ((c = getc(fp)) == '~')
                decode85(fp, file);
            else {
                if (preserve)
                    putchar('<');
                if (c == EOF)
                    break;
                if (preserve)
                    putchar(c);
            }
        else
            if (preserve)
                putchar(c);
}

void usage(void) {
    fprintf(stderr, "usage: decode85 [-p] file ...
");
    exit(1);
}

extern int getopt(int, char *[], const char *);
extern int optind;
extern char *optarg;

int main(int argc, char *argv[]) {
    int i, preserve;
    preserve = 0;
    while ((i = getopt(argc, argv, "p?")) != EOF)
        switch (i) {
            case 'p': preserve = 1; break;
            case '?': usage();
        }

    if (optind == argc)
        decode(stdin, "decode85", preserve);
}

```

```

else
    for (i = optind; i < argc; i++) {
        FILE *fp = fopen(argv[i], "r");
        if (fp == NULL) {
            perror(argv[i]);
            return 1;
        }
        decode(fp, argv[i], preserve);
        fclose(fp);
    }
return 0;
}

```

Michael Still (mikal@stillhq.com) [A247]

3. Where can I find samples of xxx?

Q: I need to obtain samples of streams that are encoded with each of the types of filters that PDF 1.3 supports. Is there a "test bank" anywhere? Ideally, the streams would be fairly small to facilitate debugging. Some of them, like FlateDecode and CCITTFaxDecode are easy to find. But for example, I haven't been able to locate an ASCIIHexDecode stream anywhere. [Q11]

A: Have you tried the spec? There is an examples appendix, as well as many examples spread throughout the document. I can't check your specific example for you, as my copy is at work, but you should be fine.

Michael Still (mikal@stillhq.com) [A20]

Chapter 9. Development: General questions

1. APIs not developed by Adobe

Q: What APIs are available, and under what terms? [Q3]

A: PDFLib, free for non-commercial use, <http://www.pdflib.com>

ClibPDF, free for non-commercial use, <http://www.fastio.com>

Panda, open source / free (GPL), <http://www.stillhq.com>

ReportLab, open source / free (BSD license), <http://www.reportlab.com>

Michael Still (mikal@stillhq.com) [A8]

A: We are pleased to announce the release of version 1.0 of the XMLPDF library for Java.

The preview period for this product has been completed and the first production version has been released.

This library converts XML to PDF including support for:

- complex table formatting including nested tables
- JPEG and PNG images
- automatic pagination of text and tables
- text kerning
- Type 1 and TrueType fonts including embedding
- defining a document template in XML and merging of data from a separate XML source
- server-side use in Web and Application servers

A quick overview of capabilities is at

<http://www.xmlpdf.com/overview.html>.

XMLPDF retails for US\$ 99 per developer seat. For information on source licences contact sales@xmlpdf.com.

John Farrow (john.farrow@xmlpdf.com) [A194]

A: I've made an inexpensive (149 USD), royalty free component to create PDF files with bookmarks, images (jpg), standard base 14 fonts, vector drawing...

It has a very small footprint; the Delphi version compiles in the exe and the OCX is only 350 Kb in size.

Developers can distribute freely, royalty free applications with PDF creation features.

Very fast (1000 pages = 15 secs on a old Celeron 400), no DLL Hell, no acrobat needed.

If interested visit www.dreamscape.it

Massimo Brini (brinim@tiscalinet.it) [A214]

2. How can I determine the compression used in a PDF file?

Q: I got a few documents looking good and at a reasonable size for the final use. But months later I have to re-issue and I can't repeat the results. Other than trial and error, is there a utility or steps to determine the compression options used in a PDF? [Q5]

A: Open the PDFs using a txt editor which can afford binary data (like UltraEdit on Winxx) and search for the "/Filter" keywords.

Helge Blischke (H.Blischke@srz-berlin.de) [A12]

A: <http://www.enfocus.com/plugins.htm> and look for the free Enfocus Browser plug-in. Enfocus Browser allows you to navigate the low-level object hierarchy in a PDF file, and view the PDF page description for a particular page.

Filiep Maes (filiepm@enfocus.be) [A13]

A: Our Quite A Box Of Tricks product will tell you the compression used, image by image. This includes the level of JPEG used. This can be done with the free demo.

Aandi Inston (quite@dial.pipex.com) [A14]

3. Discussion areas for PDF developers

Q: What mailing lists are available for PDF developers? [Q7]

A: You could try some of the following:

- comp.text.pdf

- PDFDev (<http://www.pdfzone.com>)

- PlanetPDF developer's forum (<http://www.planetpdf.com>)

Aandi Inston (quite@dial.pipex.com) [A16]

4. What is that stuff inserted into PostScript by Acrobat when printing encrypted PDFs?

Q: Encrypted PDFs when printed have the following in them:

```
% Removing the following eight lines is illegal, subject to the Digital
Copyright Act of 1998.
```

```
mark currentfile eexec
54dc5232e897cbaaa7584b7da7c23a6c59e7451851159cdbf40334cc2600
30036a856fabbl96b3ddab71514d79106c969797b119ae4379c5ac9b7318
33471fc81a8e4b87bac59f7003cddaeb2a741c4e80818b4b136660994b
18a85d6b60e3c6b57cc0815fe834bc82704ac2caf0b6e228ce1b2218c8c7
67e87aef6db14cd38dda844c855b4e9c46d510cab8fdaa521d67cbb83ee1
af966cc79653b9aca2a5f91f908bbd3f06ecc0c940097ec77e210e6184dc
2f5777aacfc6907d43f1edb490a2a89c9af5b90ff126c0c3c5da9ae99f59
d47040belc0336205bf3c6169b1b01cd78f922ec384cd0fcab955c0c20de
00000000000000000000000000000000000000000000000000000000000000
```

Which stops Distiller from converting to PDF. What is it? [Q8]

A: The eexec'd code reads in cleartext:

```
/currentdistillerparams where {
pop /pdfmark where {
pop (This PostScript file was created from an encrypted PDF file.)
print (Redistilling encrypted PDF is not permitted.)
print userdict /quit get exec }if} if currentfile
closefile
```

That means, if either your printer knows about currentdistillerparams and pdfmark or the PostScript job itself defines these operators (even as dummies, see note below), this code assumes you are going to re-distill the PS job which is forbidden.

NOTE: The PostScript driver you use might insert statements like

```
/currentdistillerparams where {pop} {userdict/currentdistillerparams{1
dict}put}ifelse /pdfmark where {pop} {userdict/pdfmark{cleartomark}put}ifelse
```

or the like (see the recommendations in Adobe's Pdfmark Reference Manual).

Helge Blischke (H.Blischke@acm.org) [A17]

5. Why are large GETs (PDF forms) truncated in Microsoft Internet Explorer?

Q: Sometimes GETs are truncated for web pages (including PDFs), why is this? [Q9]

A: See <http://www.networkice.com/Advice/Intrusions/2000608/default.htm> for a security discussion on GET Data Overflow, which might explain why MSIE-transmitted URL-encoded strings from PDF "submit" are sometimes truncated at something a bit less than 4KB length.

Bill Segraves (wsegrave@mindspring.com) [A18]

6. How machine intensive is generating PDF documents dynamically?

Q: Is PDF like Postscript? What I'm trying to get an idea of is whether all those PDFlib function calls that are generating parts of the file are doing massive amounts of processing or are they just doing simple things like writing PS-like markup tags wrapped around data.

PDF is conceptually similar to PostScript. It isn't hugely complex to generate, but there is an additional (small) overhead of not only generating the graphical information, but also generating a file structure and index.

In general, the time taken to make PostScript and PDF directly should be comparable. I wouldn't describe either of them as at all like HTML, but in the sense that you aren't rendering to a bitmap or anything like that, they are similar. [Q10]

A: This is very true. There are also the aspects of compression of data within the PDF to be considered as well... These can be quite processor intensive.

PDF is also pretty thingie about the format that bitmaps take, and it take take a fair bit of memory and time to convert the bitmaps to the right format. The operations aren't slow, it's more the fact that there can be millions of them.

I would say that the similarity ends when you say that both formats (HTML and PDF) have structure. PDFs are pretty funky in their object layouts. I am not sure about the timing statement comparing PS and PDF though. I have generated large PS files (containing images), which have been quite slow, but the PDF has been much faster because of better compression support.

I would think the best route with your ISP is to just do some trials and logging and see what happens. Also, I don't think pdflib does linearisation, which might cause problems with large documents online.

Michael Still (mikal@stillhq.com) [A19]

7. Problems with linking to web pages from within a PDF

Q: I would like to create a link to an external document with a remote go-to action. However using a URL file specification (chapter 3.10.4 of the PDF Reference) does not work. This is the created object in the pdf file:

```

1 0 obj <<
  /Type /Annot /Subtype /Link /A << /S /GoToR /F <<
    /Type /Filespec /FS /URL
    /F (http://www.tug.org/applications/pdftex/calculat.pdf) >>
  /NewWindow true >>
/Rect [124.802 706.129 266.534 791.168] >> endobj

```

[Q12]

A: It's clear from experiments that Acrobat does not support the full generality of what might theoretically be possible. If you create something Acrobat wouldn't do, then you may be stuck, especially with indirect file references, which tend to work only if Acrobat would expect them there.

An Acrobat weblink would have an /A field more like

```
/A << /S /URI /URI (http:...) >>
```

Aandi Inston (quite@dial.pipex.com) [A21]

8. Bezier curve approximation

Q: I'm trying to draw a pie slice, given the pie's center and the start and end angles. As far as I can tell, I need to use bezier curves to do the arc. Is there an easier way, or if not, what do you use for the bezier's control points? [Q15]

A: You can calculate them like:

```

$alpha = ($alpha * 3.1415 / 180); $beta = ($beta * 3.1415 / 180);

my $bcp = (4.0/3 * (1 - cos(($beta - $alpha)/2)) / sin(($beta -
$alpha)/2));

my $sin_alpha = sin($alpha); my $sin_beta = sin($beta); my $cos_alpha =
cos($alpha); my $cos_beta = cos($beta);

my $p0_x = $x + $a * $cos_alpha;
my $p0_y = $y + $b * $sin_alpha;
my $p1_x = $x + $a * ($cos_alpha - $bcp * $sin_alpha);
my $p1_y = $y + $b * ($sin_alpha + $bcp * $cos_alpha);
my $p2_x = $x + $a * ($cos_beta + $bcp * $sin_beta);
my $p2_y = $y + $b * ($sin_beta - $bcp * $cos_beta);
my $p3_x = $x + $a * $cos_beta;
my $p3_y = $y + $b * $sin_beta;

$x,$y ...
center point of arc $alpha,$beta ...
start/end angle of arc $a,$b ...

```

```
x/y extents of fitting elipsis (for circle $a=$b)

$p0_x,$p0_y ...
start-point of bezier $p1_x,$p1_y ...
control-point 1 of bezier $p2_x,$p2_y ...
control-point 2 of bezier $p3_x,$p3_y ...
end-point of bezier
```

Mind that you cannot calculate correct arcs for $\text{abs}(\beta - \alpha) > 180$ using this bezier approximation, so if your arcs span more than 180 degrees split it into two using a middle angle.

(alfredreibenschuh@yahoo.com) [A25]

9. JBIG2 compression support

Q: The version 1.4 PDF specification allows JBIG2 compression to be used in PDF files. Has anybody been able to get Acrobat 5.0 create a PDF file that contains JBIG2-compressed images? [Q16]

A: I don't think it will. I couldn't do it programmatically. There is certainly nothing in the user interface. The support may be read-only at the moment (and I have no way to confirm even that).

It may be that Adobe have done something unusually sensible with regard to changes: add support for reading the files, then wait another year before making anyone. This could make the transition a lot smoother.

Aandi Inston (quite@dial.pipex.com) [A26]

10. Rotated text

Q: I'm trying to draw rotated text, given the point where the text starts, and an angle of rotation (e.g., 90 degrees). So far, the text does not show up, so I guess my transformation matrix is incorrect, or maybe it's something else. If you know how to do this, could you show an example?

To be more specific, what I'm trying to do is rotate text 90 degrees without "changing" the page's coordinate system. In essence, with x,y being a point in the page's "normal" coordinate system, the text is to start at x,y , except be rotated 90 degrees. What I tried was

```
0 1 -1 0 0 0 Tm
```

[Q17]

A: The text matrix is the way to move text. The best approach in debugging the matrix is to do the mathematics yourself. Since the text starts at 0,0 in user space, calculate the transformation of 0,0 with your matrix and see where it is. (This assumes that cm has not also been used).

Aandi Inston (quite@dial.pipex.com) [A27]

A: OK, what you want to do is save the graphic state ("q") then do the rotation, then restore the old state ("Q"). The rotation itself is also a bit more complex than what you have. For a single line of text it will be 0 1 -1 0 h 0, where "h" is the height of the text. Think about rotating a rectangle about the origin and you can see where the "h" comes from.

Arne (arnet@hpcoplhx.cv.hp.com) [A28]

11. Centering text

Q: I am a fairly competent PostScript programmer but now need to generate some PDF. I have the basic tree structure down, but I need to accurately position some text, to be specific I need to center some text. In PostScript it is easy to figure out the size of the string to be set then translate appropriately. This seems impossible in PDF, so how do you estimate the font metrics programmatically to set the text. To make it more difficult I am generating pDF from a Java Applet so small code size is a definite factor. [Q18]

A: You need to include a table of font metrics in the code - so you can calculate the width and height, and then use those to translate (just like postscript). Adobe provides the data for the builtin fonts, otherwise you are on your own.

Dave Bloodgood (dabldgd@home.com) [A29]

12. Enumerating the images in a PDF

Q: I'm new to PDF's and am wondering how I might go about programatically enumerating all embedded images in a PDF, extracting them, then inserting a modified version back into the PDF? Basically I am looking to create a Windows program that will make a small change to all images in a PDF but otherwise leave the PDF alone. Are there decent PDF SDK's out there that will let me programatically manipulate a PDF? [Q19]

A: The Acrobat SDK includes the Core API. This can be used to write plug-ins in C or C++. Provided that you don't need to run the result on a server this may

be suitable. The Acrobat SDK mostly requires a full copy of Acrobat installed and licensed per machine. Acrobat SDK: <http://partners.adobe.com/> You may find the PDFEdit layer does what you want, though you might be surprised at the learning curve, the amount of coding required, and how much you have to understand PDF internal formats. Indeed, you should start this task by reading the 600 page PDF Specification.

Aandi Inston (quite@dial.pipex.com) [A30]

13. Modifying elements in a PDF

Q: Is it possible to extract an embedded image, modify it and put it back without creating a whole new PDF? [Q20]

A: It depends what you mean. The SDK could edit a PDF file and save it. The changes would be added to the end of the file - so you are storing both the original and the new image. The SDK can also do a Save As, rewriting the file.

Aandi Inston (quite@dial.pipex.com) [A31]

A: Most information inside a PDF document is stored in a data structure known as an object. These objects have two major identifiers, and object number, and a generation number. It is therefore possible to create a new version of an object without having to recreate the entire PDF structure – you simply increment the generation number. Then again, you still need to update the xref table and other areas of the document to refer to the new version of the object. Hence, it probably is faster just to recreate the entire file programatically.

Michael Still (mikal@stillhq.com) [A32]

14. Embedding WMF in PDF

Q: Is there a possibility to easiliy embed WMF in a PDF file while creating the PDF using a PDF library? Normally the libs only support JPG, PNG and TIFF. [Q21]

A: Easily? No. WMF is a Windows-specific format, containing graphical concepts with no easy or direct equivalent in PDF. If you are on Windows you can play the metafile to get the graphical constructs, and try to convert them to PDF equivalents. But on other platforms this is a huge development - probably larger than all of the rest of PDFLib.

Aandi Inston (quite@dial.pipex.com) [A33]

A: The WMF file format is discussed in the "Graphics File Format FAQ", which can be found at <http://www.faqs.org/faqs/graphics/fileformats-faq/part3/> – this includes a link to the specification and some discussion. This should give you enough information to be able to implement changes to the PDF library of your choice (assuming you have the source code).

Michael Still (mikal@stillhq.com) [A34]

A: There is a library to support the WMF file format at:

http://sourceforge.net/project/showfiles.php?group_id=10501

Michael Still (mikal@stillhq.com) [A35]

A: This is not helped by the fact that the file format is not fully documented by Microsoft. Only the header format has documentation; the contents are basically renderings of GDI calls, and the documentation for them only comes with MS-platform compilers.

To the original poster (and those saying 'me too!'):

If you are a reasonable programmer, you might well find it useful to look at `wmflib2` at <http://www.wvware.com/libwmf.html>

As well as documenting those parts of the WMF file format that have been successfully reverse-engineered, there is software to convert WMF to other formats such as `fig` which you might find more tractable.

Aandi's comment that WMF contains graphical concepts with no easy or direct equivalent in PDF is true, but you may also find that `_many_` of the most commonly-used primitives in WMF have pretty direct analogues in PDF, postscript and many other drawing languages.

So no, it isn't easy. But neither is it impossibly difficult. It's somewhere in-between. YMMV.

Kevin Ashley (K.Ashley@ulcc.ac.uk) [A36]

15. Unicode support

Q: I have a problem concerning special Polish characters which I write down as Unicode characters but which don't show up correctly in a PDF document when using the Helvetica standard Pdf font. The problem scenario is the following: I have implemented a PDF Report Generator in Java. To describe it briefly, the generator is a piece of Java code that reads an XML layout specification and a stream of Java business objects and generates a Pdf document from it. For the

PDF low-level stuff, I am relying on using Bruno Lowagie's Java library 'iText', version 0.37. Generally, only Pdf standard fonts are used. [Q22]

A: I think that's the problem right there. PDF standard fonts include only ISO-Latin1Encoding. That indeed doesn't include Polish characters. Helvetica does NOT include "Z with a dot accent". Standard fonts are 1-byte, not Unicode. Generating a PDF with Unicode characters is possible, but the program to do it has to be written COMPLETELY differently and is generally an order of magnitude more complex. Check if iText offers this option. Another possibility, again if iText offers it, is to use a non-Unicode font containing Polish characters, and embed it in the PDF.

Aandi Inston (quite@dial.pipex.com) [A37]

A: Yes, you can change the font-encoding so you can use the regular codepage with the embedded font, or with the base 14 fonts, or with Type 1 fonts with definitions for the glyph names, or the WGL Truetype fonts with the glyphs implemented for the codepage.

Thus, depending on what codepage it is, then there are font-encodings for that codepage in the PDF so that the unchanged text character data is used with the fonts.

Converting it to Unicode would be the step if the PDFs were ever used to be sources of Unicode PDF data in readers that were unaware how to reorder the glyphs in the glyph map from a non-standard encoding of a font in the PDF to Unicode, or to some other codepage. That is to say, software meant to search through them would have to understand the glyph mappings to the character values outside the regular printable ASCII characters.

Where the string elements' data like the glyphs can be in the native codepage, the elements in the navigational interface for the PDF reader applications require the PDFDocEncoding or UnicodeEncoding, those are predefined encodings.

For example, C code to write a font encoding differences from the predefined WinAnsiEncoding for codepage 1250 is as this:

```
"extern long writePDFFontEncodingDifferencesCentralEurope(FILE* fd){
//CP1250

    long written=0;

    written+=fwrite("128 /Euro ", 1, 10, fd);
    written+=fwrite("140 /Sacute /Tcaron /Zcaron /Zacute ", 1, 36, fd);
    written+=fwrite("156
/sacute /tcaron /zcaron /zacute ", 1, 37, fd);
    written+=fwrite("161 /caron /breve /Lslash ", 1, 26, fd);
    written+=fwrite("165 /Aogonek ", 1, 13, fd);
    written+=fwrite("170 /Scedilla ", 1, 14, fd);
    written+=fwrite("
175 /Zdotaccent ", 1, 17, fd);
    written+=fwrite("178 /ogonek /lslash ", 1, 20, fd);
```

```

        written+=fwrite("185 /aogonek /scedilla ", 1, 23, fd);
        written+=fwrite("188 /Lcaron /hungarumlaut
/lcaron /zdotaccent /Racute ", 1, 55, fd);
        written+=fwrite("195 /Abreve ", 1, 12, fd);
        written+=fwrite("197 /Lacute /Cacute ", 1, 20, fd);
        written+=fwrite("200 /Ccaron ", 1, 12, fd);
        written+=fwrite("202 /Eogonek ", 1, 13, fd);
        written+=fwrite("
204 /Ecaron ", 1, 13, fd);
        written+=fwrite("207 /Dcaron /Dslash ", 1, 20, fd);
        written+=fwrite("209 /Nacute /Ncaron /Ohungarumlaut ", 1, 15, fd);

        written+=fwrite("216 /Rcaron /Uring ", 1, 19, fd);
        written+=fwrite("219 /Uhungarumlaut ", 1, 19, fd);
        written+=fwrite("
222 /Tcedilla ", 1, 15, fd);
        written+=fwrite("224 /racute ", 1, 12, fd);
        written+=fwrite("227 /abreve ", 1, 12, fd);
        written+=fwrite("229 /lacute /cacute /ccedilla /ccaron
", 1, 39, fd);
        written+=fwrite("234 /eogonek ", 1, 13, fd);
        written+=fwrite("236 /ecaron ", 1, 12, fd);
        written+=fwrite("239 /dcaron /dstroke ", 1, 21, fd);
        written+=fwrite("241 /nacute /ncaron ", 1, 20, fd);
        written+=fwrite("245
/ohungarumlaut ", 1, 20, fd);
        written+=fwrite("248 /rcaron /uring ", 1, 19, fd);
        written+=fwrite("251 /uhungarumlaut ", 1, 19, fd);
        written+=fwrite("254 /tcedilla /dotaccent ", 1, 25, fd);

        return(written);
}

```

Your data might be in a different codepage, on UNIX systems probably one of the ISO 8859 codepages or ASCII codepages, on Windows the Windows codepages, and on Macintosh the Mac codepages. The PDF has the predefined encodings for those types of codepages, basically, then writing the font encoding differences remaps the characters back to their glyphs in the font in the native codepage.

That way, you can use the extended Western glyphs in the base 14 fonts. The extended CJK fonts are using the composite Type 0 fonts. For example, the Big 5 and JIS encodings are for much of Chinese or Japanese, respectively, where their non-Western characters are not in the Western codepages. Basically that is the difference between a single byte codepage which has enough characters for almost all of the glyphs of a Western script, and the multiple byte codepages to support more than 256 glyphs. The single byte codepages are mostly having around 220 glyphs, in the codepages, not of their script.

Overall, it is probably best to convert it to Unicode, what is required then is the knowledge of the Unicode fonts. The Unicode fonts can be very large, for example a single Unicode font has tens of thousands of characters in it, where

a regular font file has less than a thousand. So, feasibly, it might be better to write the font encoding differences into the PDF generated from the Java. In a different case, it might be more feasible to convert the data to Unicode and then embed a subset of the font into the PDF, where it would only have as many characters as are glyphs in the data, which is for Western languages probably less than a hundred, not the complete font.

So, for Western languages and single-byte codepages, the most direct way to put the text data into a PDF is using the base 14 fonts with the font encoding differences. This is very convenient, for example, to have the Euro glyph without Unicode, with the font encoding different to map it to 128, and to remap Zcaron and zcaron to 142 and 158, where those are names of the glyphs that are used in the font encoding.

The names of the glyphs are in the Adobe PostScript glyph lists, and there are representations for each of those I think in Unicode. Information about this is from Unicode and Adobe.

Then, there are the strings that have to be the predefined PDFDocEncoding or UnicodeEncoding, those as part of navigational elements.

Ross Finlayson (apex@alpha.com) [A38]

16. Raster vs Vector images?

Q: I've been looking at the PDF file in a text editor and found a number of /Subtype /Image objects. Are these raster or vector images? [Q23]

A: These are raster images, but all the vector images are held in the page stream, as are text and some bitmap images, and you won't be able easily to see what they are.

Aandi Inston (quite@dial.pipex.com) [A39]

A: You are on the right track. Vector "images" are not images, but consist of PDF (Postscript) drawing commands. In an uncompressed PDF you can recognise them easily. In a compressed PDF I do not know of a way to tell text sections from vector sections, as they both consist of PS drawing commands. Maybe someone else knows more.

Arne (arnet@hpcoplhx.cv.hp.com) [A40]

17. Required font tables

Q: Can anyone verify which of these tables are required when embedding a true type font? I've got 3 sources that say 3 different things. OS/2 cmap cvt fpgm glyf head hhea hmtx loca maxp name post prep [Q79]

A: Depends on what you means by embedding. You need to include these tables in your embedded file `String[] tables = { "head", "hhea", "loca", "maxp", "cvt ", "prep", "glyf", "hmtx", "fpgm", "cmap" };` but you need to read some of the others (such as OS/2) to get values you need during the embedding process.

John Farrow (jfarrow@visualprogramxming.co.nz) [A133]

A: cvt, fpgm, and prep are definitely not required by Acrobat, although the PDF reference may specify otherwise.

Thomas Merz (tm@pdflib.com) [A134]

18. Custom fields in the info dictionary

Q: Has anybody out there:- added there own custom Info Dictionary fields? does Acrobat Readers "doc info" command show them? does Exchange etc interfere with them? I noticed that that the spec recomends adding your own custom fields to the "catalogue" object. But I intend adding them to the Info Dictionary in the hope that Verity will index them. [Q134]

A: You may add any key value pairs to the DOCINFO dictionary, and Verity will index them (if properly configured). But Acrobat (reader) only shows the predefined ones.

H.Blischke@srz-berlin.de (Helge Blischke) [A205]

19. CMYK to RGB conversion

Q: What is the correct algorithm for converting between colour spaces? [Q137]

A: The problem is that GhostScript uses Adobe's documented method of converting CMYK to RGB. In this, for instance, 100% magenta + 100% yellow simply becomes 100% red. This is optically correct, but not a good approximation of a printer, and not what Acrobat does. Adding colour management, perhaps to match a profile, and ideally using embedded source profiles from the PDF, would not be a simple undertaking. It may be worth posting a new topic specifically asking about Color management of CMYK conversion in GhostScript to see this question reach the largest number of people. I know you're not the first to be interested in this area.

Aandi Inston (quite@dial.pipex.com) [A211]

20. JavaScript popup menus on hot spots

Q: I want to create WWW links in my PDF documents. The URL will start a JAVA request servlet on my webserver, and the result may be more than one document. Is it possible to create a list menu behind the hotspots, which displays the documents in a list so that the reader can select which he/she wants to open? [Q138]

A: I have been playing with Named Destinations and the popupmenu method in PDF javascript. Perhaps this javascript will lead you to a solution.

Here I had set up some named destinations based upon a variable name appended to a standard set of suffixes.

Grab the name of the hotspot and use it to build a destination name. Use popupmenu to have the user select the suffix they want (in this case it's always "removal").

Call gotoNamedDest to go there

```
var itemname = event.target.name;

var xreply
var reply = app.popupMenu( [ itemname+"removal",
    "Removal",
    "Installation",
    "Operation",
    "Part Info",
    "Loc on Dwg",
    "Loc on Schematic",
    "Detail Dwg or Picture" ] );

if (reply!= null)
{
switch ( reply ) {
case "Removal":
    xreply = "removal";
    break;
case "Installation":
    xreply = "removal";
    break;
case "Operation":
    xreply = "removal";
    break;
case "Part Info":
    xreply = "removal";
    break;
case "Loc on Dwg":
    xreply = "removal";
    break;
case "Loc on Schematic":
    xreply = "removal";
    break;
case "Detail Dwg or Picture":
    xreply = "removal";
    break;
}
}
```

```

var args = new String();
args = itemname+xreply;

this.gotoNamedDest(args );

```

John Freund (jfreund@freundassociates.com) [A212]

21. Internationalization of accented characters

Q: Does anybody has information about using german umlauts and other european special characters in pdf documents? [Q144]

A: I think that using fonts with WinAnsiEncoding will solve your problem. It seems to correspond to the ISO Latin 1 character set (could somebody please confirm this?). So you can use accented characters directly in PDF text commands. I'm using them for Portuguese words. For example:

```

6 0 obj
<<
  /Type /Font
  /Subtype /Type1
  /Encoding /WinAnsiEncoding
  /BaseFont /Helvetica
>>
endobj
8 0 obj
<<
  /Type /Font
  /Subtype /Type1
  /Encoding /WinAnsiEncoding
  /BaseFont /Helvetica-Bold
>>
endobj
10 0 obj
<</Length 11 0 R>>
stream
0 g
BT
1.0204 0 0 1.0237 368.5677 641.0251 Tm
/ HB 4.081 Tf      % set Helvetica-Bold font
(CLASSIFICAÇÃO)Tj
ET
BT
181.548 703.3136 TD
(Freqüência)Tj
ET
endstream
endobj

```

Jose Fernando Tepedino (jose@wiser.com.br) [A224]

22. Internationalization of fonts

Q: I'm writing a program that has pdf as output format. now I need to display a text in cyrillic letters. i have no idea how to do that. Please, can you help me or at least point me to some documentation. [Q154]

A: It's all in the pdf reference document that I assume you already have. The main problem you have is that the built in fonts in Acrobat don't cover cyrillic so, you'll have to embedd an external font or at least refer to it. If you are not yet discouraged I must tell you that you'll also have to peek at the font to get the character widths, which is rather easy with an type1 afm file and not so easy with a true type file.

Here's a couple of dictionaries created with iText for a cyrillic true type font without embedding:

```
6 0 obj
<<
/Ascent 728
/Flags 32
/StemV 80
/ItalicAngle 0
/Type /FontDescriptor
/FontName /ArialMT
/CapHeight 699
/FontBBox [-222 -324 1071 1037]
/Descent -210
>>
endobj
3 0 obj
<<
/FirstChar 32
/BaseFont /ArialMT
/FontDescriptor 6 0 R
/Encoding <<
/Differences [32 /space /exclam /quotedbl /numbersign /dollar /percent
/ampersand /quotesingle /parenleft /parenright /asterisk /plus /comma
/hyphen /period /slash /zero /one /two /three /four /five /six /seven
/eight /nine /colon /semicolon /less /equal /greater /question /at /A
/B /C /D /E /F /G /H /I /J /K /L /M /N /O /P /Q /R /S /T /U /V /W /X
/Y /Z /bracketleft /backslash /bracketright /asciicircum /underscore
/grave /a /b /c /d /e /f /g /h /i /j /k /l /m /n /o /p /q /r /s /t /u
/v /w /x /y /z /braceleft /bar /braceright /asciitilde /.notdef
/afiii0051 /afiii0052 /quotesinglbase /afiii0100 /quotedblbase
/ellipsis /dagger /daggerdbl /Euro /perthousand /afiii0058
/guilsinglleft /afiii0059 /afiii0061 /afiii0060 /afiii0145 /afiii0099
```

```

/quoteleft /quoteright /quotedblleft /quotedblright /bullet /endash
/emdash /.notdef /trademark /afiii10106 /guilsinglright /afiii10107
/afiii10109 /afiii10108 /afiii10193 /space /afiii10062 /afiii10110
/afiii10057 /currency /afiii10050 /brokenbar /section /afiii10023
/copyright /afiii10053 /guillemotleft /logicalnot /hyphen /registered
/afiii10056 /degree /plusminus /afiii10055 /afiii10103 /afiii10098 /mu
/paragraph /periodcentered /afiii10071 /afiii61352 /afiii10101
/guillemotright /afiii10105 /afiii10054 /afiii10102 /afiii10104 /afiii10017
/afiii10018 /afiii10019 /afiii10020 /afiii10021 /afiii10022 /afiii10024
/afiii10025 /afiii10026 /afiii10027 /afiii10028 /afiii10029 /afiii10030
/afiii10031 /afiii10032 /afiii10033 /afiii10034 /afiii10035 /afiii10036
/afiii10037 /afiii10038 /afiii10039 /afiii10040 /afiii10041 /afiii10042
/afiii10043 /afiii10044 /afiii10045 /afiii10046 /afiii10047 /afiii10048
/afiii10049 /afiii10065 /afiii10066 /afiii10067 /afiii10068 /afiii10069
/afiii10070 /afiii10072 /afiii10073 /afiii10074 /afiii10075 /afiii10076
/afiii10077 /afiii10078 /afiii10079 /afiii10080 /afiii10081 /afiii10082
/afiii10083 /afiii10084 /afiii10085 /afiii10086 /afiii10087 /afiii10088
/afiii10089 /afiii10090 /afiii10091 /afiii10092 /afiii10093 /afiii10094
/afiii10095 /afiii10096 /afiii10097]
/Type /Encoding
>>
/LastChar 255
/Subtype /TrueType
/Widths [277 277 354 556 556 889 666 190 333 333 389 583 277 333 277
277 556 556 556 556 556 556 556 556 556 277 277 583 583 583 556
1015 666 666 722 722 666 610 777 722 277 500 666 556 833 722 777 666
777 722 666 610 722 666 943 666 666 610 277 277 277 469 556 333 556
556 500 556 556 277 556 556 222 222 500 222 833 556 556 556 556 333
500 277 556 500 722 500 500 500 333 259 333 583 0 864 541 222 364 333
1000 556 556 556 1000 1057 333 1010 582 854 718 556 222 222 333 333
350 556 1000 0 1000 906 333 812 437 556 552 277 635 500 500 556 488
259 556 667 736 718 556 583 333 736 277 399 548 277 222 411 576 537
277 556 1072 510 556 222 666 500 277 666 656 666 541 677 666 923 604
718 718 582 656 833 722 777 718 666 722 610 635 760 666 739 666 916
937 791 885 656 718 1010 722 556 572 531 364 583 556 668 458 558 558
437 583 687 552 556 541 556 500 458 500 822 500 572 520 802 822 625
718 520 510 750 541]
/Type /Font
>>
endobj

```

Paulo Soares (psoares@ip.pt) [A237]

23. Signature sample code

Q: Can I get the wrapper code that builds a PKCS#7 object from the Raw Signature format? [Q160]

A: It's buried at the Adobe site at <http://support.adobe.com/devsup/devsup.nsf/docs/50752.htm>

Thomas Merz (tm@pdflib.com) [A245]

Chapter 10. Development: Microsoft Windows specific

1. Control over OLE automation

Q: I have a general question in regard to OLE automation. I've built an application that controls Acrobat, and I'm trying to make it more robust. The problem I'm having is what to do when I try to open a non-existent file (a fdf pointing to a missing pdf file). Acrobat then displays a dialog box, and waits for user input. Since there's no user to click on the cancel or ok button, my program freezes waiting for Acrobat to respond. Is there any way to pass a "cancel" or something similar to Acrobat via OLE automation? [Q30]

A: You can get the name of the pdf file that the fdf points to using the FDFGetFile method from the Acrobat forms toolkit. Then use code to verify that the pdf exists (and display a suitable message if it doesn't) before trying to open it via the pdf.

Dan Sideen (dansideen@home.com) [A56]

2. Acrobat and DDE

Q: I want to access the Acrobat Reader with DDE by using VB. The DDE-commands are described in <http://partners.adobe.com/asn/developer/acrosdk/docs/iac/IACRefer>

I am searching for a working example. [Q155]

A: It took me a long time to get this to work, but it does, at least with Acrobat 4.

```
Sub Printit(FileToPrint As String)
    'Print a file using Acrobat
    Dim sCmd As String
    Dim FullName As String    'defined as string elsewhere
    FullName = FormsHome + " " + FileToPrint
    sCmd = "[FilePrintSilent(" & FullName & ")]"
    frmViewForm.lblDDE.Caption = "Printing..." & FileToPrint
    frmViewForm.lblDDE.LinkTopic = "Acroview|Control"
    frmViewForm.lblDDE.LinkItem = ""
    If frmViewForm.lblDDE.LinkMode <> 2 Then
        frmViewForm.lblDDE.LinkMode = 2
    End If
    frmViewForm.lblDDE.LinkTimeout = 1000
    frmViewForm.lblDDE.LinkExecute sCmd
    frmViewForm.lblDDE.LinkMode = 0
    frmViewForm.lblDDE.Caption = ""
    frmViewForm.lblDDE.LinkMode = 2
End Sub
```

Dan Sideen (dansideen@home.com) [A238]

Chapter 11. Development: Problems with my PDF documents

1. "The file is damaged but is being repaired."

Q: I have created a PDF file and get the following error when trying to view the file in Acrobat Reader 4.0:

"The file is damaged but is being repaired."

No other information is given. The file then opens and I can view the document. The file has not been zipped or e-mailed. As a test, I created a PDF example from the Adobe Portable Document Format Manual Version 1.3 and I received the same error. In both cases, the result is messed up. Is there a way to either truly fix such a file, or at least "extract" the good parts? [Q13]

A: This normally means you have the offsets in the XREF table wrong, so they are regenerated. Make sure you are using an editor that will show you binary content, and remember to count each newline as two bytes if you are running on Windows.

Michael Still (mikal@stillhq.com) [A22]

A: When I see this message it means "something is wrong" and it could be just about anything. The first thing to think of is: remember that PDF is indexed with exact byte offsets – so be sure you are using the right sort of line ending and counting all the bytes correctly.

Aaron Watters (aaron@at.reportlab.dot.com) [A23]

Chapter 12. Distiller

1. Licensing

Q: We want to use Distiller Server to create PDF versions of customer annual statements. These will then be made available to the customer over the internet.

Is this a violation of the Distiller Server license agreement?

I would normally have said not, but the FAQ says

"Using Distiller Server, can I create an Adobe PDF file from one of my company's internal documents and publish it on the internet for someone outside my company? For example, if a customer requests a bank statement over the Internet, can I publish it in Adobe PDF?. No....[snip]"

This sounds exactly like what we are doing, except that the document is not generated at the customer's request, but periodically.

Also, it seems to be saying that no company documents can *ever* be published on the internet, which can't be right!!

Any advice or clarification? [Q39]

A: Tricky, it depends I believe on whether the customer will request these statements, or you will simply publish them anyway. The first is a violation, the second is (I think) not. It certainly appears that you can read it that way. This would certainly restrict the usefulness of the format.

I *thought* the idea of the licence was to prevent you making a 'distillation service' available, whereby people could send you files and have them converted and sent back. Creating your own documents in PDF form would certainly seem reasonable, and if you used the standard Distiller, and paid someone to sit and convert the files, this would certainly be permitted under that licence.

I think you may have to seek a legal opinion, or better, a clarification from Adobe. If you get one, I'd be interested in hearing it.

Ken Sharp (ken@spamcop.net) [A65]

2. Colour in distilled PDFs

Q: Should I be able to get the jpeg to appear in color in my pdf doc created through Distiller 3.01? [Q40]

A: Sure, use a colour driver.

Aandi Inston (quite@dial.pipex.com) [A66]

A: Yes. Do not use the HP driver, but the one Acrobat installed in your system: Acrobat Distiller 3.01. Using a black and white driver makes your images black and white.

Matti Vuori (mvuori@koti.soon.fi) [A67]

3. Distiller page size problems

Q: Why is it the whenever I create a pdf file from Word I get a page size of 8x11? I have changed the paper size in the advanced properties of the Distiller printer driver to Postscript Custom Page size and specified the new size but it makes no difference. It doesn't matter whether I print to file and then use Distiller or use the Create Adobe PDF button in Word and Use Distiller that way. [Q41]

A: I had encountered the same problem and thought it was only on a German Windows2000/NT4 platform, since it didn't occurred on an English version that was at my disposal. May be the solution works for you as well. If you have a user defined format in Word and converting it to pdf via Acrobat Destiller does not output your format, try the following (only Win2000 or NT4): In the printer dialog (- start - settings - printers) in the uppermost menu-list you'll find the "file" menu. In the file menu you'll find the entry "server properties". In the "server properties" you can add a new format (like A4 or US-Letter). Name it "myformat" (or what ever you like). Fill in the your paper size (in the lower part of the dialog). Press the "save format" button. Now every printer that is listed in the printer-dialog-window has the "myformat"-format. Switch to Word. Look in the - file - "set up paper" (or whatever it's called in the English Word) dialog. The "myformat"-format should be displayed instead of "custompapersize". The new format is only local on your workstation (independent from your computer being connected to your LAN). If you want to convert doc to pdf on another machine, you must first add your self-defined paper size to the printer server properties as described above. No convert your Word document to pdf via the destiller. It should work correctly now.

Andreas Wall (Andreas.Wall@shinkatech.com) [A68]

4. Output filenames

Q: Is there a possibility to automatically generate PDF files with Acrobat Distiller without typing in the name for the output file? I heard that in Acrobat 3 patching some INI files works fine, but is there any chance to do the same in Acrobat 4? [Q42]

A: With a watched folder you can create automatically a pdf with the same name as the original postscript file.

Veli Holopainen (veli.holopainen@kaleva.fi) [A69]

5. Driver recommendations

Q: What is the best colour printer driver for rendering graphics to a pdf file? I have one graphic that renders very badly using hp5/6ps. [Q143]

A: The AdobePS driver (free download from Adobe, of course) works well. Set up your Distiller printer to default to that driver.

Dick Margulis (margulis@fiam.net) [A220]

6. Missing fonts

Q: I generate PDF's from PostScript using Distiller 3 on a Windows NT4 machine which has an old PostScript printer directly attached. My PS files use the built-in PS Helvetica & Helvetica Narrow. I have pfb files for both font families on my disk but not installed in Windows. When I tried generating the same pdf's on a client's NT machine (PostScript printer on network) using Distiller 4 I got error messages, sample:

```
<program listing> % %[ Warning: Helvetica not found, using Font Substitution.  
Font cannot be embedded. ] % %
```

I assume that there is a font licensing question involved but I thought that Helvetica itself (though not Helvetica Narrow) was included in the pdf basic fonts. [Q152]

A: No, it is not font licensing. Font licensing problems say that is the problem. The problem is exactly as stated: Helvetica is not found. Now, Distiller 4 does not include a copy of Helvetica. This does not stop Acrobat from processing the PostScript, because it knows all about Helvetica. But your job options also ask to embed it. If you want to do that, Distiller needs the actual font. No, it is not font licensing. Font licensing problems say that is the problem. The problem is exactly as stated: Helvetica is not found. Now, Distiller 4 does not include a copy of Helvetica. This does not stop Acrobat from processing the PostScript, because it knows all about Helvetica. But your job options also ask to embed it. If you want to do that, Distiller needs the actual font.

This is not an issue with Distiller 3, since it would not embed Helvetica at all.

The chances are that you don't need to embed Helvetica either. If you don't, then add it to the Never Embed list. If you do want to embed it, purchase a copy and add it to a folder which Distiller searches for PostScript fonts. This is not an issue with Distiller 3, since it would not embed Helvetica at all. The chances are that you don't need to embed Helvetica either. If you don't, then add it to the Never Embed list. If you do want to embed it, purchase a copy and add it to a folder which Distiller searches for PostScript fonts.

Aandi Inston (quite@dial.pipex.com) [A235]

7. Maximum number of named destinations reached

Q: While distilling a rather large file, Distiller popped up an error message saying

```
% [ Warning: Maximum number of named destinations reached. Ignoring future  
named destinations ]
```

Does anyone know what the limit is and perhaps how I can get some info on which dests have not been created. [Q157]

A: Do not select compatibility with Acrobat 2.1, and the limit should vanish. (I think that's right; if not, you need Distiller 4.0 with 4.0 compatibility).

Aandi Inston (quite@dial.pipex.com) [A241]

A: I don't think you'll see it with Distiler 4 and Acrobat 3 compatibility either.

Ron Teplitz (ronald.teplitz@alcatel.com) [A242]

Chapter 13. History

1. How old is PDF?

Q: Does anybody know when PDF was inveted or introduced? [Q47]

A: The first PDF specification was introduced in 1993. PDF's roots go back much earlier, though, to the invention of Postscript.

Scott Robert Ladd (scott.ladd@maximal.com) [A74]

A: Here is some information from the preface of Adobe's Pdf Specification.

<http://partners.adobe.com/asn/developer/acrosdk/docs/PDFRef.pdf>

Hope this helps.

"THE ORIGINS OF THE Portable Document Format and the Adobe Acrobat product family date to early 1990. At that time, the PostScript page description language was rapidly becoming the worldwide standard for the production of the printed page. PDF builds on the PostScript page description language by layering a document structure and interactive navigation features on PostScript's underlying imaging model, providing a convenient, efficient mechanism enabling documents to be reliably viewed and printed anywhere. The PDF specification was first published at the same time the first Acrobat products were introduced in 1993. Since then, updated versions of the specification have been and continue to be available from Adobe via the World Wide Web. This book is the first version of the specification that is completely self-contained, including the precise documentation of the underlying imaging model from PostScript along with the PDF-specific features that are combined in version 1.3 of the PDF standard."

DesQuite (desquite@hotmail.com) [A75]

Chapter 14. Internet stuff

1. Problems with IE displaying PDFs

Q: Does anyone have any suggestions that could help us understand why IE intermittantly will not get the PDF from the web server. We just get a blank screen occasionally, with 'Done' on the status bar. We have tried unoptimizing the DPF so as not to use byte serving. Web server is Apache 3.1.12 HTTP 1.1 on solaris UNIX. The client is NT4 + IE5. [Q43]

A: I've noticed that it depends (somewhat, not always) on how your serving up your PDF's. I have some software on my site that creates PDFs on-the-fly which wasn't working in IE (got the white screen or Adobe complained it wasn't a valid PDF) but worked fine in Netscape. From your message I take it you have the PDFs on disk already?

Instead of writing the PDF stream out to the user's browser I started putting the PDFs on disk and doing a redirect. It seems to have alleviated the problem (server is NT with PDFs being created from Perl scripts). What was really strange was that when I sent the data directly to the browser, and Adobe complained in IE, I did a save to my disk to inspect the contents. Parts of the PDF were missing and even some of the internal structure, like what object the page info was stored at, was changed. I suspect that the plug-in for IE was modifying it (not sure what else could have). Perhaps the same thing is happening on your end?

Even some static PDFs I have out there don't always come up properly in IE although in Netscape I don't think I've ever had a problem.

Mike Bernardo (mbernardo@chartermi.net) [A70]

A: The problem is that MS fixes a PDF bug in every new version of MSIE, but everytime they succeed in introducing new PDF bugs. So it is very hard to find a workaround that works with EVERY version of MSIE. For what it's worth: I experienced 2 different problems with MSIE and DYNAMICALLY generated PDFs: 1. MSIE sends multiple requests to the server: so if the users asks for 1 dynamically generated PDF file, MSIE sends the same request 2 or 3 times to the server (if you don't anticipate, a PDF document is generated 2 or 3 times). 2. MSIE gives a blank page. This last problem can be solved by sending a server header indicating the exact length of the PDF file. In iText we solved this like this: <http://www.lowagie.com/iText/faq.html#msie> We are not happy with this situation, because if you have large documents, you want to send data to the browser bit by bit, not the whole file at once when it's finished: if the timeout is reached, you get 'connection reset by peer' errors.

Bruno Lowagie (bruno.lowagie@rug.ac.be) [A186]

2. Optimised, compressed, linearised? Arrrrgh! My brain hurts!

Q: What is the difference between an optimised PDF, a compressed PDF, a linearised PDF, and a cheese stick? [Q44]

A: These terms are confusing because many people mis use them (to a certain extent Adobe didn't pick very helpful terms either). The PDF specification uses the terms to mean:

Optimised: the PDF has been laid out in the most graceful manner possible. For instance, you have saved a black and white image as a colour or grayscale image, which would take a lot more space.

Compressed: some elements in the PDF are compressed. The whole document is not required to be compressed however.

Linearised: the document has had it's internals rearranged so that byte serving will work. Byte serving is that thing you get on some web sites when only the page you are currently reading is downloaded... This means that you can flick through large documents without having to first download the entire thing. It is often called optimised by people who haven't read the PDF specification.

Michael Still (mikal@stillhq.com) [A71]

3. Encrypted PDFs and search engines

Q: I noticed that in Windows, PDF files have an extra 'PDF Properties' tab in file properties. It shows the 'General Info' (ctrl+D in Acrobat) of the PDF file. However, if the PDF is encrypted (eg. changing the document is not allowed), this info is not showed. Does anyone know if encrypting affctcs the indexing of PDFs (Acrobat Catalog, search engines)? [Q45]

A: Certainly, Catalog used not to be able to index encrypted PDFs. With Acrobat 5.0 it is now a plug-in, so it probably can. Many, perhaps most, search engines will not be able to index files that are encrypted.

Aandi Inston (quite@dial.pipex.com) [A72]

4. Document corruption in older web caches

Q: I'm having trouble. I've tested with Acrobat Reader 5.0.1 27.3.2001 & W2k (both SP1 &SP2) and IE 5.5SP1 & IE 6.0beta.

Error message is "File does not start with '%PDF-'".

I get the error message about 99% of the pdf viewing attempts, also with www.airtug.com/brochure.pdf [Q46]

A: When PDF files are fetched from the web, this is often in small pieces - a process called byteserving.

Some proxy servers don't recognise that the pieces are separate, and so they give back the wrong pieces. This leaves the PDF files in quite a mess, as if they are broken up and glued together wrong.

I don't know if that proxy has problems, but if a proxy didn't understand byte-serving your symptoms are exactly what I'd expect.

Aandi Inston (quite@dial.pipex.com) [A73]

5. Problems linking to PDF documents

Q: Why don't my PDF documents display properly when I link to them from a web page? [Q125]

A: There are several threads about this. I'd like to add my experience and my fix. First, the problem: we added several Adobe PDF documents to a new web site. They had formerly been linked to successfully on another page. The documents had recently been revised. When we published the new site, a group of PDF files from one page came up blank when a user tried to link to them. Sometimes, using the Reload button would then display the page. The problem occurred with Netscape 4.6 on Win 2000 and with NS 4.76 on an unknown Windows platform (user did not report OS). I got a blank page with NS 6.1 on NT4, but after I changed a configuration setting to stop asking me if I wanted to save or display when I clicked on a PDF link, the blank page no longer occurred. The pages displayed first time on this same machine with NS 4.61 and NS 4.78 (However, I installed these older browsers after making the config change on 6.1.)

All these machines were using the Adobe Acrobat Reader plug-in 4.0 or 4.05. I had some problems with the pages locking up the entire browsers from home with NS 4.7 and IE 5.5. on Win ME; I upgraded to Acrobat Reader 5.0 and the files displayed fine in this configuration. (Considering the locking up of the browsers, this problem may be a different one than our users experienced.)

The documents in question were created with MS Word and converted to Acrobat via a plug-in. They were modified with Adobe PDFWriter 4.05 for Windows NT, PDF Version 1.3. However, other documents that had been modified by the same version, but were linked on another web page, displayed fine on all platforms and browsers. In addition, when we recreated one of the PDF documents by converting in Word with an older Adobe PDFWriter (PDF Version 1.2—which documents had displayed correctly on the old site) the PDFs still would not link from the problem page. Here's the final clue: when we linked a supposedly bad PDF file from another web page, it could be viewed fine.

So we presumed that some problem in the HTML code was causing the display glitch. Consequently, we moved all the HTML from the "bad" page to a new page. However, we did it in chunks—layout separate from border code, etc. (We use NetObjects TeamFusion Authoring Server 2000 for Web publication)

Voila! The PDFs displayed fine when linked to from the new page.

While I have not examined all the code on the "bad" page, the contents of the <a href> tags themselves are identical to those on the new page. I conclude

that some hidden control code has caused an oddity in interpretation of this link by certain configurations of browser and Adobe plug-in. Obviously, it is not consistent across the machines. Other threads on Usenet have mentioned that it can happen on ASP pages, on IE, and in UNIX in addition to multiple versions of Netscape. The fixes offered in Usenet (reformat your hard drive - LOL - or change your Adobe Reader from a Plug-in to a helper application) have worked for some, but not others. I add to those suggestions another less hard intervention—recreate your web page.

(Please compare this hypothesis and scenario to what happens in some cases where Word text is copied and pasted to an HTML page in Team Fusion; or in SQL server code when pasting from one screen to another. Occasionally, the code will not work until you delete all spaces that might contain hidden codes or retype the code. Or the case of dropped close-table tags that display correctly in IE, but show a blank page in some versions of Netscape)

Hoping this helps someone who runs across this problem.

Diana Diehl (ddiehl@uillinois.edu) [A184]

6. Problems with accessing linearized PDFs from an Apache web server

Q: There have been some reports that some versions of Apache don't byte serve properly, which means that linearized PDFs don't serve properly... Is there any fact to this? [Q130]

A: Apache 1.3.14 had a serious bug affecting byteserving, which is the method Acrobat uses to fetch pages. I don't believe it affected any other version. Still you could try upgrading to the latest.

Aandi Inston (quite@dial.pipex.com) [A197]

7. Problems with IE caching PDF documents

Q: Hi, we have the following problem and I hope someone can help me. We are creating a web application using pdf forms. These forms are updated with data from a database via a web server through servlets. We are using TOMCAT as a web server and the FDF toolkit to create the form data.

When a user presses the submit button, the servlet handles the request, builds an updated FDF file and sends it back to the client. The problem is that although the servlet response gets to the client, the pdf file does not get updated with the newly generated FDF data. We do not have this problem with Netscape communicator this is why I think it is a problem with IE5.5's caching.

Any help would be greatly appreciated. Thank you in advance. [Q131]

A: The problem appears to be that IE checks for files with newer dates on the server, and the pdf does not have a newer date, since it hasn't changed. You

need to disable caching in IE. Try Tools:Internet Options: Temporary Internet Files. The bad news is that this disables caching for all pages, and will slow down all web work.

Alternatively, try using a different file name for the "master" PDF file using FDFSetFile. This will force a new version to be downloaded.

Dan Sideen (dansideen@home.com) [A199]

8. Printing files within browsers using Javascript

Q: We have an application which generates dynamically pdf files on the server side. This pdf's can be viewed in the browser (which of course uses the acrobat reader for displaying the contents.) In a special convenient function I would like to print these docs directly without popping up an acrobat reader. eg the user presses simply a button called 'print'.

In the newsgroups I found solutions like this

```
..
<EMBED name="pdf" SRC="rg.pdf" WIDTH=85 HEIGHT=115>

<script>
  window.document.pdf.print();
</script>
..
```

which DON'T work for me. [Q133]

A: Yep that was an oldone which relied on Adobe's ActiveX control being wholly not safe for scripting, and violating basic MS plug-in rules... on Win32,

<http://www.meadroid.com/scriptx/> would be a solution (combined with Neptune from the same site for non IE.)

Jim Ley (jim@jibbering.com) [A203]

9. Linking to named destinations within a PDF

Q: I'm normally a lotus-notes-developer (so not much of a pdf-expert). For a project in domino.doc, I should be able to create URL-links that open pdf-documents that are stored in domino.doc to a certain location.

I can get this to work when I specify a certain page (you know, <http://.../filename.pdf#page=7>), but it doesn't work when I want to go to a named destination (as in

<http://.../filename.pdf#nameddest=destination>). It opens the document, but always the first page.

I have created a named destination in my test-document (using the destination-window in acrobat 4.0), and tested it (using go to destination in the destination-window) so I'm sure there is a named destination.

What am I doing wrong, or is this a browser(IE 5.0)-related problem ? [Q135]

A: I found it myself, apparently there's a mistake in the acrobat-help-files (leave out the #nameddest) :

<http://www.adobe.com/support/techdocs/a17e.htm>

Further I was using blanks in the name of my destination, and this also seemed to cause trouble (browser translates them to %20 and this doesn't seem to match the name of the destination anymore).

Philippe Lauwers (eb.xeruces@srewual.eppilihp) [A206]

10. Font changes when byte serving PDF documents

Q: This is hardly a serious problem... I'm just curious. We're in a WinNT environment (Acrobat 4) and are serving PDF documents over the Web. The documents were Distilled with an embedded LucidaSans Type 1 font. When looking at the font characteristics within the document the embedding looks like it took just fine. This is what we are experiencing...

The end-user is browsing away (Netscape 4.08), reading PDF documents with the Reader 4 plugin. They bring up a PDF document and when it displays, for a fraction of a second, the text that is displayed is rendered in a different font. Very quickly it then shifts to the LucidaSans font. Although it happens very fast, I think the font that is initially displayed is the AdobeSansMM font. At least that's what it looks like to me. [Q153]

A: That's right - it's exactly what is supposed to happen. When displaying a "byteserved" PDF Acrobat will download the page in stages. First the text itself, which is rendered with substitute fonts. Then bitmap graphics. Finally, any embedded fonts come down and the text will, if necessary be redrawn.

This can take a long time, not just happen in a flash, and that's the point of it - to give the user something to read as quickly as possible. This enormously improves the perceived performance (which is much more important in most cases than the real performance).

Aandi Inston (quite@dial.pipex.com) [A236]

Chapter 15. Online tutorials

1. What online PDF tutorials are available?

Q: Are there any tutorials online which deal with PDF related issues? [Q76]

A: I've a number of tutorials on my site (see www.yeardley.demon.co.uk/pdf.html and www.yeardley.demon.co.uk/morpdf.html) that might help folk & your FAQ. 4 spring to mind:

1. PDF on how to set up Reader v4.x to run from PC
2. PDF explaining the v4.x Acrobat/ABT/Reader toolbar and tool shortcuts.
3. PDF showing how to create Doc level JavaScripts.
4. PDF / Word97 DOC for troubleshooting Reader v4.05 (Win) installs (icons, double-click link, etc.)

Mark Anderson (mark@yeardley.demon.co.uk) [A127]

Chapter 16. OS: Linux specific

1. Acrobat on Linux character set problems

Q: Sometimes, when trying to display a PDF file with Acrobat Reader (version 4.0 for Linux, from the Debian distribution), I get this message: Unable to extract the embedded font 'AHCAAI+CMSS17'. Some characters may not display or print correctly. [Q29]

A: You may set the locales for Acrobat Reader to english. You must edit the start script "acroread" which is normally located in /usr/local/Acrobat4/bin :

```
-----  
#!/bin/sh  
# LC_ALL=C  
# <-- Add these export LC_ALL  
# <-- two lines ver=4.0  
install_dir=/usr/local/Acrobat4/Reader  
-----
```

Christian Koch (christian_koch@gmx.de) [A55]

Chapter 17. OS: Microsoft Windows specific

1. Automatically displaying a PDF when a CD inserted (Windows)

Q: Is it possible to have an autorun.inf file on a CD that would cause Acrobat to open a PDF file on the CD when the CD was inserted? [Q90]

A: Try looking at <http://www.yeardley.demon.co.uk...>

JoeZ (jzach@swbell.net) [A149]

A: I wrote the document when v4.0 was current but it works fine for v4.05c. IMO the jury's out on using Reader v5.01 and CD as it does unwanted things like ask (by unchangeable default) whether you want to integrate Reader with your browser. Integrate a CD-based app with your browser - crazy! Unless you really need v5.x features I'd use v4.05c for the while on CD. That said apart from the filelists changing a bit, the v4 procedure as described in the PDF should be OK for v5.0.

Mark Anderson (mark@notmeyearley.demon.co.uk) [A150]

Chapter 18. Other

1. Can people trace my PDF document back to me?

Q: I have a pdf file with no document properties, so I am able to tell who compiled it. I was hoping that the unique serial number of the Acrobat compiler might be embedded in the PDF somewhere. [Q129]

A: No, the document properties are probably all there is. If the file is created with Acrobat tools, the document properties are filled in.

Unfortunately, most PDF creators that were written in the bathtub on a Friday afternoon do NOT do this, though these are the creators that one most needs to identify, because the PDF files are most likely to be buggy.

Note that I am referring here to the creation software; I don't particularly care what person made it.

Aandi Inston (quite@dial.pipex.com) [A195]

2. Adobe PDFWriter stops ODMA from working in MS Office

Q: Some people have noted that ODMA stops working within their Microsoft Office suites because they have the Adobe PDFWriter loaded. [Q139]

A: This is because the PDFWriter is written in VBA, and when you have VBA loaded some office applications incorrectly identify themselves as "MS VB" to the ODMA DMS. The solution is to unload the PDFWriter macro.

Michael Still (mikal@stillhq.com) [A213]

3. Dynamic views of data using PDF

Q: I'm a college student and don't have much experience with PDFs, but perhaps someone here can direct me.

I wonder if PDF supports multiple views (or any other software). That is, I want to create a document in LaTeX (a math document) and convert it to a PDF. Yet, I want to allow the user to choose how they want to see it. The document will be arranged in sections, I want it to hide certain sections unless a user wants to see them. Basically, the user can choose to see a more in depth explanation if needed. Is there a way to do this perhaps in the menu? [Q163]

A: Take a look at "Stratify PDF" from <http://www.lantanarips.com/>

Robert Anderson (bobanderson@mac.com) [A250]

Chapter 19. PDF tools from people other than Adobe

1. Truetype capable PDF generators (not APIs)

Q: I am looking – mainly for my Word-Docs – for a free- or a share-ware pdf-creator (Win98) which can embed TrueType and Type1. Any idea? [Q48]

A: http://www.this.net/~frank/pstill_win.html

Doug Milliken (bd427@freenet.buffalo.edu) [A76]

2. Page extraction

Q: Is there a way to automate the extracting of pages from a PDF document by using a script or batch process? [Q49]

A: I think we might have exactly what you're looking for. Our TK40 toolkit lets you define criteria (how pages break, what indexes exist, etc.) for extracting logical documents from a larger compound PDF file. <http://www.maximal.com/products>

Scott Robert Ladd (scott.ladd@maximal.com) [A77]

A: <http://www.reportlab.com/pageCatcher/index.html>

Dinu Gherman (dinu@reportlab.com) [A78]

A: Here are a few more (that aren't vaporware like TK40):

Ari's PDF Splitter Pro (ask for the extraction version) (<http://www.dionis.com>)

AppendPDF (<http://www.appligent.com>)

Glance PDF CLT tools for Win/Unix (pdsel) (<http://pdf.glance.ch>)

JW (18tni7m5x001@sneakemail.com) [A79]

3. Distiller equivalents

Q: Are there any equivalents of Adobe distiller? [Q50]

A: GhostScript 7.0 supports almost all of the features of distiller.

Alex Cherepanov (alexcher@erols.com) [A80]

A: <http://www.ctrlp.com/freepdf.asp?st=pdf>

Ingemar Djurhuus (djurhuus@mail1.stofanet.dk) [A81]

A: Jaws Creator http://www.jawssystems.com/products/products_fs.html at \$120 US, is an enhanced, commercial alternative to the original 5D program, which can still be downloaded for free at <http://www.ctrlp.com/freepdf.asp?st=pdf>. According to the feature comparison chart, Jaws does almost everything Adobe does, including recognizing Word 97 and 2000 subcomponents.

Steve Auerweck (steven.auerweck@verizon.net) [A82]

A: The offer of www.ctrlp.com to download a free copy of "5D PDF Creator" (a version from 1999) exists not any longer. The successor product "Jaws PDF Creator 2.0" can be downloaded as an evaluation copy from www.jawssystems.com. For Ghostscript look at www.ghostscript.com or www.cs.wisc.edu/~ghost or www.aladdin.com (the URL <http://aladdin.com> doesn't work)

Eckhard Henkel (eckhard.henkel@t-online.de) [A178]

A: There is also my PStill converter which is quite inexpensive for private use on Windows systems (and free for private and edu use on Linux, Solaris, IRIX, HPUX and AIX).

You can download the program here

<http://www.wizards.de/~frank/pstill.html>

It works also unregistered but will draw a small text line near the bottom of each generated PDF page but is otherwise uncrippled.

Frank M. Siegert (frank@this.net) [A179]

A: 5D PDF Creator has now been replaced with Jaws PDF Creator at <http://www.jawssystems.com> in 2 forms: trial version and full version. I was able to get the FREE 5D version which allowed me to get the full version of Jaws. That FREE version is no longer available from the link below. The trial version at Jaws allows you to try it out but your PDF files will have a ghosted message in the background until you pay a fee to get the full version. I don't remember how much it is but it is well worth it in my opinion.

On the other hand, Jaws is a good PDF program and makes them easily from any application. BUT Adobe Acrobat gives you the editability that I don't think you will get with any other PDF creation program. It seems to be the industry-leader at this point.

Vi Wisdom (viwiz@worldnet.att.net) [A233]

4. Extracting images from PDF documents

Q: How can I extract images from a PDF file? [Q51]

A: You could try xpdf's pdffimages (<http://www.foolabs.com/xpdf>)

Jorma Heimonen (Jorma.Heimonen@kone.com) [A83]

A: Several ways... and I have used all of these, often just because it's easier than figuring out where I stored the original source imagery (or when I actually want the presented composite image, and not the original raw image).

* Some applications, like PhotoShop and Illustrator can open an individual PDF page and edit it.

Illustrator is ideal for this, because it preserves all of the page elements as individually selectable entities, and further, preserves their vector or raster nature (and probably the color model).

PhotoShop may rasterize the whole PDF page as a single object. Rasterizing entire pages at high resolutions results in HUGE data objects. And unfortunately, unless you can calculate precisely what dpi to use to get 1:1 source:captured pixels, you want to oversample by at least 2x. Run the numbers.

Neither Adobe app, however, will open the page if it has any kind of security (open or admin security). Further, embedded fonts may be incorrectly rendered.

* Reprint to .eps

If you can configure your PostScript driver to a. print-to-file, and b. print in EPS format you can print the page of interest to page.eps, then edit it with whatever EPS-capable image editor you have.

Some apps, like Adobe FrameMaker, can import an entire PDF page as a referenced graphics object. You can then reprint it as .eps. FM won't import pages from secured PDFs.

* Acrobat and Acrobat Reader can select rectangular sections of a page - if (big if) selecting graphics is allowed in the file (see File:DocInfo:Security).

On Windows, place the mouse cursor over the [T] text selection tool, hold the left mouse button down, and let up over the "graphics select tool" icon.

Outline the image desired. Zoom to the screen resolution desired. Edit:Copy or [Ctrl][c] Selected image is now on clipboard.

The selected area can be off-screen, and even off logical desktop, but you will be limited on some systems to a maximum graphics object size - 32MB for Windows. If you get a nastygram dialog box, zoom out until you don't.

* Screen copy

The last resort is to use whatever tools the OS provides, or are available in the freeware, shareware or aftermarkets to perform a full-screen-copy or window-only-copy. On MS Windows, the [PrntScrn] and [Alt[PrntScrn]] keys do this. Zoom to desired size - but the entire desired image must [usually] be on-screen. Generally, unless you can run the numbers and match the object raster res to 1 pixel per screen pixel, you want to zoom as large as possible and over-sample to minimize re-sampling damage.

If your graphics card supports large logical desktops, as many Matrox cards do, the image can be partially off-screen as long as it is entirely in the card's on-board RAM, i.e. is entirely on the logical desktop.

The 32MB Windows GDI limit applies.

All of the above assumes that you own or have permission to re-use the image(s) in question.

Bob Niland (rjn@fc.hp.com) [A84]

A: Yes. If a PDF file permits you to retrieve text, it also permits you to retrieve graphics.

Acrobat has three ways to do this. I'm not sure whether the free Acrobat Reader supports any of these. Someone else will have to check, as I don't even have it installed.

1. There is an object touch-up tool (a solid diagonal black arrow icon on a button). In Acrobat 5 this is accessed by pulling out the button strip attached to the text touch-up tool (outline capital T). Press and hold the T button to access the pullout strip. Then select the arrow button.

Using this tool, you can select one or more objects (graphics as well as text blocks). The tool selects every object it *_touches_* as you drag it, not every object it *_surrounds_*. You can deselect unwanted objects with shift+click.

Once you select an object, you can move it, copy it, cut it, or delete it. If you copy it or cut it, you can paste it into another PDF but not into any other application that I'm aware of.

2. There is a graphics select tool (dotted box with a white and a black circle inside it). You can drag a box around any area of the page, which may include both graphic and text material. You get everything inside the rectangular box you draw, as a bitmap (screen capture style). If you don't draw the box exactly right the first time, you cannot adjust it. Just drag a new box. After selecting the area you want, you can increase the magnification on the view. You can zoom until the selected area is much larger than your monitor if you want to, in order to increase the resolution of the bitmap you are capturing. Of course this causes a geometric increase in the file size, and you may not need such high resolution, depending on what your ultimate purpose is.

Once you have the area selected and zoomed, you can copy it to the Windows clipboard (Ctrl-C) and then paste it in any application that accepts a raster image from the clipboard. (The dimensions in pixels of the copied area travel with

the clip. If you are in PhotoShop and open a new image with the graphic on the clipboard, the default size of the new image will match the piece you clipped.)

3. You can export an EPS of a single page (File > Export) and select the graphic in a program that can parse the EPS.

Finally, you can open the PDF in Adobe Illustrator and select the graphic easily there.

Dick Margulis (margulis@fiam.net) [A215]

5. Why can't I just and paste the images

Q: I am astonished that this is (apparently) such a complex job. If I want to take a pdf graphic and move it to say, Word, I simply select what I want with the graphic tool (in Exchange), and paste into my Word doc. And I get the graphic - some or all of it depending on what is selected ... Am I missing something here? [Q52]

A: Possibly.

What you describe has some serious limitations:

1. If the document has inhibited "selecting text and graphics", it won't work at all. You'll be limited to screen capture using host OS tools.

2. Even when it does work, you get a RASTER image of the entire selection area. You can't easily separate elements or even eliminate overlay text.

Further, if the original graphic was vector, you still get raster - not as scaleable - and often vastly larger.

3. You get that raster at screen resolution. This means that even in the case of raster originals, you are either under- or over-sampling the original, with potential damage to the image.

Normally, you need to select the area, then zoom until the "copy" fails due to the object size (32mb in Windows), then down-size the resulting object in your image editor. This minimizes re-sampling artifacts.

Bob Niland (rjn@frii.com) [A85]

A: Yes. What you describe is equivalent to faxing yourself a copy of the image. It's not the image itself. Not only is it no longer in PDF format but it's limited to the dot-density of your screen, so while it's OK for a Web site, the low resolution will be horribly obvious if you try to print it.

Adobe went out of their way to make it *_difficult_* to extract a real image from a PDF file, under pressure from publishers who don't want their expensively-generated imaged being pirated.

Peter Flynn (peter@silmaril.ie) [A86]

6. Inserting watermarks over pages in a PDF document

Q: What tools can help me insert a watermark over pages in a PDF document? [Q53]

A: <http://www.reportlab.com/pageCatcher/index.html>

Aaron Watters (aaron@at.reportlab.com) [A87]

A: StampPDF Batch will allow you to do this easily. More information on StampPDF Batch; including documentation and online demos, can be found on our web site at <http://www.appligent.com>

Mark Gavin (mgavin@appligent.com) [A88]

7. Linearization tools

Q: What non-Adobe linearization tools are available? [See earlier questions for a description of linearization itself] [Q54]

A: http://www.pdfzone.com/products/software/tool_pdlinearize.html

Bryan Guignard (bryang@sympatico.ca) [A89]

8. Tools to concatenate PDFs

Q: What tools are available to concatenate PDFs? [Q55]

A: <http://www.appligent.com/>

Aandi Inston (quite@dial.pipex.com) [A90]

A: I'm working on a PDF Append application (Freeware), but it has yet to be beta tested. You're welcome to try it out, but keep in mind I'll assume no legal responsibilities (normal legal stuff). Anyway, it can be downloaded at: <http://www.northlandpublishing.net/append.zip>

Timothy L. Jordi (jordi@northlandpublishing.net) [A91]

9. HTML to PDF conversion

Q: Is there somebody who knows a solution to convert HTML files to PDF files on a linux system? [Q56]

A: Try HTMLDOC from <http://www.easysw.com/htmldoc/>

Matthias Haeusser (matthias.haeusser@t-systems.de) [A92]

10. PDF permissions tools

Q: What tools are available to manipulate permissions on PDF documents? [Q57]

A: PDF Crypt What is pdfcrypt? pdfcrypt is a very flexible and powerful program. pdfcrypt allows you to set permissions to a PDF-file. For example you can publish a document without to allow to print it. The button to print the file will be disabled in Acrobat Reader application. It's simple to use it like a batch application to set permissions to every PDF in you archive. It's simple to use it like a pipe application. It's simple to use it inside your cgis. We distribute only executable versions (if you need ask us the original PERL code). Download it and test it! <http://www.sanface.com/pdfcrypt.html>

SANFACE Software (sanface@yahoo.com) [A93]

11. Inserting images onto existing PDFs

Q: Using Acrobat 4 or 5, or any other program, is it possible to edit an existing PDF file: I need to add 1000 pictures (logos) to 1000 different, existing PDF-files. [Q58]

A: For the logos, if you know where you want to put them you can use Page-Catcher for this. <http://www.reportlab.com/pageCatcher/> Regarding searching usenet: if you go <http://www.deja.com> you will be redirected to google's usenet searching service which will save you some time (try the advanced search option where you can control the search with a lot of options – list, keyword, sort order, phrase, etc.)

Aaron Watters (aaron@at.reportlab.dot.com) [A94]

A: Have a look at www.enfocus.com for PitStop Pro or PitStop Server. PitStop Pro is a plug-in for Acrobat 4/5. PitStop Pro has many edit capabilities. If you want to automate a task then you can make use of Action lists and apply it to each file in Acrobat by opening each file or do it in a batch using PitStop Server. PitStop Server is a standalone product. All our products have a 30 trial period.

Filiep Maes (filiepm@enfocus.be) [A95]

A: Cheapest way (beside using an illegal copy) is to print from Acrobat Reader into a PostScript file. Interpret this file with CorelDraw or Illustrator or Freehand and add the logo. Export as PDF from the applications or print to a PostScript file again. Use Ghostscript to convert the PostScript file to a PDF. Don't ask me about how to do it step-by-step or which settings you should use. The cheap way will cost you also (your time).

Marc Wieber (mgww@exmail.de) [A96]

A: StampPDF only costs \$179 and is a plug-in for Adobe Acrobat. StampPDF can easily add "Evaluation Copy" across every page of your document.

"StampPDF Batch", costing \$2995, is a stand alone, server based application designed for high production, on-demand web based applications.

More information on StampPDF can be found on our web site at

<http://www.appligent.com>.

StampPDF can be purchased through us or a variety of resellers and online purchases can be done at PlanetPDF at

<http://www.planetpdf.com>.

Mark Gavin (mgavin@appligent.com) [A201]

12. PDF Viewers

Q: Does anybody knows if there is another program which reads pdf files besides Adobe Acrobat? I ask this because my Acrobat reader doesn't start, it says that there is an error in the file msvcr7.dll. [Q59]

A: You can read them with GSview , which requires GhostScript . Install GhostScript first, then GSView.

Matti Vuori (mvuori@koti.soon.fi) [A97]

A: Freshmeat is your friend...

<http://freshmeat.net/search/?site=Freshmeat&q=pdf+viewer>

laird@gunsmoke.ecn.purdue.edu (Kyler Laird) [A98]

13. Problems with appending or extracting data from large documents

Q: <http://www.appligent.com/> - AppendPDF. <http://www.dionis.com/> - PDF Splitter Pro I have tried both of these tools, as well as pdsel (pdf.glance.ch), and none of them works. AppendPDF eats up memory and dies, Ari's PDF Splitter can only split into single-page fragments, and pdsel doesn't work on files greater than several hundred pages. What other solutions are there? [Q60]

A: You may have about covered it. Did you contact the tech support, especially for AppendPDF? It is designed for very heavy duty use, so a major memory leak is a little surprising.

Aandi Inston (quite@dial.pipex.com) [A99]

A: You might try the PDF Handshake product from HELIOS (www.helios.de), that includes a pdfcat utility that can concatenate as well as extract arbitrary page ranges from an Unix command line.

Jens-Uwe Mager (jum@anubis.han.de) [A100]

14. LaTeX to PDF eps graphics support

Q: I have a LaTeX file with some included *.eps graphics. If I create a *.ps document of this everything works fine. Creating a *.pdf document, all the graphics are not shown - only the captions are displayed. Does anybody know a solution for this problem? Can you recommend me a package to make the *.pdf files? [Q61]

A: `usepackage[pdftex]{graphicx}` <- you *really need to use the `graphicx` pkg)
`usepackage[epstopdf]` if you put these two commands before your `egin{document}` and run the file (barring the fact that there are no other complications) it will not only create pdf files of your EPS graphics files, but include them in your final PDF document... really neat too ;-)) you can get the `epstopdf.sty` from CTAN in `macros/latex/contrib/supported/oberdiek`

Mimi Burbank (mimi@csit.fsu.edu) [A101]

15. Printer drivers that produce PDF

Q: What non-Adobe PDF generating printer drivers are available? [Q62]

A: I thought you might like to know about pdfFactory, our new PDF printer driver for Windows that creates PDFs. There is a free trial version available at <http://www.fineprint.com>. The registration fee is \$39.95.

Among it's features:

- ability to seamlessly combine multiple print jobs (even from separate applications) into a single PDF file - pdfFactory has a Send button to email a PDF to

someone without having to Save As and type a file name - easy font embedding UI - preview jobs before freezing into PDF, ability to delete pages

- all print jobs automatically saved for later retrieval if you need them again (especially useful for transient content like Web page searches, ecommerce receipts, emails that get deleted, etc.) - smaller files than PDFWriter (usually, but not guaranteed)

The current version does not support non-ansi charset fonts or security. We will be adding those features shortly. When use with FinePrint, our other product, you can create PDF booklets, multi-up renderings, watermarks and much more.

Check it out at <http://www.fineprint.com>

Jonathan Weiner (jonathan@singletrack.com) [A102]

A: BCL Computers just released the Beta version of their new PDF printer driver for Windows 2000. To download a copy, visit <http://www.bcl-computers.com>.

Rachel Burnsed (burnsed@bcl-computers.com) [A103]

A: The pdfMachine is a Windows print driver that produces great quality PDF files. It integrates with MAPI compliant mail programs such as Outlook, Outlook Express to ease the sending of PDF's via email. pdfMachine is easy to install and use. If you know how to print a document then you already know how to use pdfMachine. It works with virtually any windows application - just print to the pdfMachine. The trial version of pdfMachine creates a faint watermark of "pdfMachine by BroadGun Software" on every page. This watermark is not present if the software is purchased and registered. For more information or to download a copy goto <http://broadgun.com/pdfmachine/index.htm>

Craig Broadbear (craig@broadgun.com) [A165]

16. Online PDF generators

Q: What online PDF generators are available? [Q63]

A: goBCL is a free service for converting your files to PDF. PDF to HTML conversion is also available. <http://www.gobcl.com>

Rachel Burnsed (burnsed@bcl-computers.com) [A104]

17. Stripping references to external files

Q: What tools can help me extract references to external files inside my PDF documents? [Q122]

A:

APStripFiles is a command line application that removes attached or embedded files from PDF documents. It enables you to protect your systems from malicious unwanted PDF file attachments.

APStripFiles offers you the option to remove the files from the original PDF document or make a copy of the PDF without the attachment, leaving the original PDF untouched. It can be used on the desktop, a web server or directly on an e-mail server to avoid the transfer of viruses that can be carried by a PDF file attachment. APStripFiles supports the removal of attachments from multiple PDF files using file names or wildcards.

APStripFiles for Mac OS X can be downloaded free from,

http://www.appligent.com/newpages/freeSoftware_Mac.html

APStripFiles for AIX, HP-UX, Sun Solaris and Red Hat Linux can be downloaded free from,,

http://www.appligent.com/newpages/freeSoftware_Unix.html

APStripFiles for Windows 95/98/NT/2000 can be downloaded free from,

http://www.appligent.com/newpages/freeSoftware_Win.html

lvincent (lvincent@digapp.com) [A177]

18. Server side PDF generation tools (webby)

Q: What tools will allow me to generate PDF documents from my web server? [Q124]

A: For server side generation, you may want to check out Cocoon, part of Apache Jakarta project. <http://jakarta.apache.org>

Frankie Lam (franky@mindless.com) [A181]

A: you can use the free tool DBtoPDF from <http://www.kgo.de/dbtopdf.html>

Klaus Gotthardt (k.gotthardt@em.uni-frankfurt.de) [A182]

19. Document info extractors

Q: I need a package which can extract information about the document for me – for instance the name of the author of the document. What is available? [Q145]

A: The "pdinfo" application from my Xpdf package does this.

<http://www.foolabs.com/xpdf/>

The web page has source code and various binaries, licensed under the GPL.

Derek B. Noonburg (derekn@foolabs.com) [A225]

20. Linux PDF printer driver

Q: How can I setup a virtual printer on linux that produces PDF documents?
[Q150]

A: Here's something I've done in my linux-box ; so I have a printer named "pdf", and if I choose it, a pdf-file is created in my home-directory with a name based on the current date. Maybe you'll have to adapt it, and you must of course have a ghostscript version with the output-device pdfwrite available (try "gs -help | grep pdfwrite"). Most applications with a GUI give a menu where you can select the name of the printer.

```
new section in /etc/printcap/
#####
pdf:
    :sd=/var/spool/lpd/pdf:
    :mx#0:
    :sh:
    :lp=/dev/null:
    :if=/var/spool/lpd/pdf/filter:
#####
the input filter /var/spool/lpd/pdf/filter :

#!/bin/bash
# set > /tmp/print_env.txt #only for debugging
HOME="/home/$USER"
OUTPUT_FILE=$(date +%x-%X | tr / .).pdf
/usr/bin/gs -dNOPAUSE -q -sDEVICE=pdfwrite -sPAPERSIZE=a4
    -sOutputFile="$HOME/$OUTPUT_FILE" -
```

(last 2 lines on the same line, in fact!)

(gauthier-vdm@ibelgique.com) [A232]

Chapter 20. PDF/X

1. What are PDF/X, PDF/X-1 and PDF/X-2?

Q: What are PDF/X, PDF/X-1 and PDF/X-2? [Q120]

A: [Please note that the following answer has been shamelessly stolen from the PDF/X FAQ which can be found at http://www.ddap.org/solutions/pdf-x_faqs.html]

Michael Still (mikal@stillhq.com) [A172]

A: "PDF/X" can be thought of as a shorthand way of specifying most of what you need to tell somebody in order for them to create a file that's likely to print correctly when they send it to you, even if they don't understand the details of what it's doing for them. That's looking ahead a little to a point where applications include a 'PDF/X' check-box on their PDF export dialogs, but signs are that's not going to be too far off.

Phrased slightly differently, think of all file formats used for file transfers as being compromises between flexibility and reliability (where reliability is defined as the final printed piece looking like you expected it to).

At one end of the scale are application files like QuarkXPress documents. You can change those in whatever way you like if you have the application. Unfortunately the receiver of a file can also change them accidentally rather too easily, and the results you get when printing are dependent on many factors in the environment in which that copy of XPress is running.

At the other end of the scale you have copydot scans. Those will print absolutely as you expect, given the necessary provisos about resolution and calibration & that's part of why they are inflexible.

In between, in order of decreasing flexibility and increasing reliability, other options at other positions on the scale include PostScript, EPS, PDF, PDF/X and TIFF/IT. When I use a name like 'PostScript' in that list I mean the format in an otherwise unspecified way. It's always possible to push such a format towards the reliable end of the scale by using appropriate software to create it. In northern Europe many people use ProScript, which limits the options used in EPS files. A 'ProScript EPS' file might be placed on the scale somewhere between PDF and PDF/X. In the same way, appropriate use of pre-flight tools on PDF can give you a 'reliable PDF' much closer to where PDF/X is on that scale. The point of PDF/X is that it gives you a convenient and well specified label to use when asking for such a 'reliable PDF' file.

Martin Bailey (http://www.ddap.org/solutions/pdf-x_faqs.html) [A173]

A: The standard has been split into two. PDF/X-1 is a file format for what's known as "blind interchange", where all technical information and content is held within one single file and nothing needs to be supplied alongside it.

It's expected that publishers will make their specifications available to ad agencies and prepress houses. These will include page sizes, bleed allowances, number and types of hard copy proofs to be sent alongside digital files and file formats that will be accepted. No further technical discussion will be required in order to make a suitable PDF/X-1 file for delivery. Obviously commercial discussions are a separate matter.

PDF/X-2 will address exchanges where there is more discussion between the supplier and receiver of the file - the receiver may have certain fonts available which would therefore not need embedding, or maybe the receiver already holds high resolution images to replace FPOs (low resolution previews) in the supplied file.

Martin Bailey (http://www.ddap.org/solutions/pdf-x_faqs.html) [A174]

Chapter 21. Postscript

1. Using epilogue.ps to rotate pages

Q: Somebody could explain me how to use and write the files epilogue.ps and prologue.ps with adobe distiller to carry out a rotation of 90 degrees of the whole of a document postscript (or only of one page) [Q136]

A: Try this for your prologue.ps file:

```
<<
  /BeginPage {   pop           % Throw away the showpage count
                 612 0 translate % translate by the page width
                                     % (your width may differ from
                                     % mine, of course.)
                 90 rotate      % Rotate by 90 degrees
               }
>> setpagedevice
```

Note that this won't work on a Level 1 printer. (setpagedevice was added in Level 2.)

John Deubert (john@acumentraining.com) [A207]

A: Or better, to be independent of the page size of the document being processed, use the following: <prorgamlisting> << /EndPage { exch pop 2 ne { [ThisPage] <> /PUT pdfmark true} {false}ifelse } >>setpagedevice

Helge Blischke (H.Blischke@srz-berlin.de) [A210]

2. Postscript fonts

Q: What fonts should exist in a postscript device? [Q146]

A: There are 35 official PostScript fonts (they should be resident in the firmware of a postscript-capable printer):

1. AvantGarde-Book
2. AvantGarde-BookOblique
3. AvantGarde-Demi
4. AvantGarde-DemiOblique
5. Bookman-Light
6. Bookman-LightItalic

7. Bookman-Demi
8. Bookman-DemiItalic
9. Courier
10. Courier-Oblique
11. Courier-Bold
12. Courier-BoldOblique
13. Helvetica
14. Helvetica-Oblique
15. Helvetica-Bold
16. Helvetica-BoldOblique
17. Helvetica-Narrow
18. Helvetica-Narrow-Oblique
19. Helvetica-Narrow-Bold
20. Helvetica-Narrow-BoldOblique
21. NewCenturySchlbk-Roman
22. NewCenturySchlbk-Italic
23. NewCenturySchlbk-Bold
24. NewCenturySchlbk-BoldItalic
25. Palatino-Roman
26. Palatino-Italic
27. Palatino-Bold
28. Palatino-BoldItalic
29. Symbol
30. Times-Roman
31. Times-Italic
32. Times-Bold
33. Times-BoldItalic
34. ZapfChancery-MediumItalic
35. ZapfDingbats

Jim Driscoll (jdriscoll@trentu.ca) [A226]

A: That may be true about resident fonts but most fonts can be written to postscript. It is just math The base fonts are just resident so everyone does not have to shuffle so much. However - again Any graphic designer worth their

salt will supply ALL the fonts they use in a file to the printer. DON'T count on anyone else. One of the magics of PDF is the embedding of fonts.

MSD (mikedodds1@home.com) [A227]

3. Conditional postscript and PDF

Q: I am looking to to embed postscript code into a pdf so that on a particular printing device x will happen and on another device y will happen.

In essence I would like the embedded code to determine if the printer is in fact capable of producing my pdf file the way I planned ie. postscript level, trapped, cmyk, blah blah blah.

Can such code be put into the postscript file before I distill it? [Q158]

A: Yes. But it won't help one iota because all the questions you ask will be answered by Distiller (level 3, no trapping, CMYK) and the single PDF generated will be based on that one answer.

You cannot embed conditional PostScript into a PDF.

Aandi Inston (quite@dial.pipex.com) [A243]

Chapter 22. Printing

1. Subverting no-print

Q: How can I print a pdf of which the security setting is changed so that it cannot be printed? Should it be converted or is there a crack to enable that function? [Q1]

A: Email the owner of the document and ask their permission to be able to print the document. One must presume that the author has chosen to not let you print for a reason.

Michael Still (mikal@stillhq.com) [A1]

A: <http://www.elcomsoft.com/apdfpr.html>

Vladimir Katalov (vkatalov@elcomsoft.com) [A2]

A: <http://www.password-crackers.com/crack/guapdf.html>

Christian Koch (christian_koch@gmx.de) [A3]

A: <http://www.ecn.purdue.edu/~laird/PDF/>

Christian Koch (christian_koch@gmx.de) [A4]

A: <http://www.cs.wisc.edu/~ghost/>

Michael Mantz (michael.mantz@de.bosch.com) [A5]

A: [Please note that there is more discussion on this topic in the security section of this document]

Michael Still (mikal@stillhq.com) [A6]

2. Can I print a PDF directly to a printer?

Q: I'd like to print a pdf file directly to a printer (or at least without human intervention) from a program running in the background on a win32 type of machine. [Q2]

A: If you have a Postscript level 3 printer, its easy. Just send the file to the printer, since PS 3 devices can directly print pdf files. Acrobat is not required.

To sent it to the printer you can use lpr on Win, or any Unix, PRINT on any Win platform, or COPY on any Win platform.

On the Mac, you probably need to use a font download utility and Applescript.

Dan Sideen (dansideen@home.com) [A7]

3. Printing PDF documents with various page sizes

Q: Is there any Windows utility to have a mixed sizes PDF document printed correctly? [Q113]

A: Adobe Acrobat 5.0 appears to offer this choice. It is not in the free Reader. I don't know what kind of printers it works with.

Aandi Inston (quite@dial.pipex.com) [A162]

A: I don't think you need a utility. After you've pressed the print command, a dialog box appears with three options in the field "Copies and Adjustments": shrink oversized pages to paper size; exapnd small pages to paper size; and auto-rotate and centre pages. By default, all options are activated. At least you should deactivate the first option.

Stefan Treutlein (streutlein@icqmail.com) [A163]

4. Why do my documents print slowly?

Q: I have a 26K pdf file that I created that contains 2 tif images and I'm using /Filter[/CCITTFaxDecode] to interpret the compression. What happens is the pdf file takes a long time to print (4 minutes on a HP LaserJet 2100 with the HP LaserJet 2100 print driver), but actual print job is at 1.83mb, whereas when I use a HP5si Laserjet network printer, my print job is only 300K and it prints immediately). The computer is a pentium 200 with 64mb of RAM and WIN95. Does anyone know why the print job is so large? Or why it takes so long? Is it the commpressed images being converted to pcl that could be taking a long time? [Q119]

A: We learnt the hard way at work that postscript doesn't deal with biggish bitmap images. This is what the PDF is producing with your CCITT Group 4 images. You'll probably find that the problem is that your print engine is not coping with the processing. There might also be issues with spooling the large bitmaps.

Try PCL on the problem printer and things should go better – you might find that the network printer is already PCL.

Michael Still (mikal@stillhq.com) [A171]

Chapter 23. Security

1. Is it possible to have viruses inside a PDF document?

Q: It doesn't seem possible to me, as I didn't know there could be executable code in a PDF. While there is no code inside a PDF file, there cannot be any kind of virus. AFAIK. [Q87]

A: No, this is incorrect. PDF can contain embedded executable code (in a wide variety of ways), which can in turn contain malicious code like viruses, worms and trojans. However, executable code must first be extracted from the PDF file with Acrobat's built in tools or with some other tool, before it can be executed so the malicious code can activate itself.

That is why when you extract file annotations with Acrobat there is a warning window that appears letting users know of the potential for malicious code. Non Acrobat may or may not offer such warnings.

This is fully documented in the Acrobat online guide.

MacAfee now has virus detecting software specifically for PDF files.

Bryan Guignard (bryang@sympatico.ca) [A145]

2. eBook security

Q: How secure are ebooks? [Q93]

A: ElcomSoft Co. Ltd. has released Advanced eBook Processor, a Windows ME/98/95/NT4/2000/XP program that makes it easy to remove both password encryption and usage restrictions from Adobe Acrobat PDF files and eBooks. The latest addition to ElcomSoft's family of password recovery software allows business managers to deal with lost and destroyed passwords, as well as with employees who, intentionally or unintentionally, are unable to edit and print password-protected PDF files.

Advanced eBook Processor lets users make backup copies of eBooks that are protected with passwords, security plug-ins, various DRM (Digital Rights Management) schemes like EBX and WebBuy, enabling them to be readable with any PDF viewer, without additional plug-ins. In addition, the program makes it easy to decrypt eBooks and load them onto Palm Pilot's and other small, portable devices. This gives users - especially users who read on airplanes or in hotels - a more convenient option than using larger notebooks with limited battery power to read their eBooks.

PDF protection can prevent users from changing or printing information, adding or changing annotations and form fields, or even selecting and copying text or graphics. With Advanced eBook Processor, these PDF files can be decrypted, opened, and used without any of these restrictions. Once protection has been removed, PDF files created with Adobe's Acrobat program can be opened in any PDF viewer, including Adobe's Acrobat Reader.

Advanced eBook Processor protects businesses from losing control of their eBooks, technical articles, documentation manuals, presentations, and all PDF documents that could be rendered unusable by improperly managed passwords and licenses.

Advanced eBook Processor costs \$99(US) and may be purchased securely online at <http://www.elcomsoft.com/aebpr.html>. You can download a free trial version of the software at the same web address.

Vladimir Katalov (kitten@elcomsoft.com) [A157]

A: The Anti-Piracy Enforcement Team at Adobe Systems has notified ElcomSoft, a software company based in Russia, that its Advanced eBook Processor software program violates the copyright of materials – primarily PDF-based eBooks – published by Adobe. ElcomSoft has five days to meet Adobe's demand that the commercial product be removed, after which the matter will be "pursued aggressively" by the San Jose-based company. Watch for updates.

<http://www.planetebook.com/mainpage.asp?webpageid=157&nl>

Kurt Foss (kfoss@planetpdf.com) [A158]

3. PDF based viruses

Q: Are there any PDF based viruses? [Q115]

A: Daily News 'Peachy' mass-mailing worm hides in Adobe's PDF files By Steven Bonisteel, Newsbytes. August 09, 2001

Virus-hunting companies are warning users of Adobe's Portable Document Format technology to beware of a new Internet worm that can launch a mass-mailing attack after arriving in such a PDF file. However, researchers at McAfee.com Corp., the virus-software subsidiary of Network Associates, say that the worm they are calling "Peachy" is a threat only to users of what's known as the full version of Adobe's Acrobat software - the version that can also create PDF files. Users of the ubiquitous - and feely available - Adobe Acrobat Reader software are not affected by the worm, McAfee.com said. The researchers said the worm is written in Microsoft's VBScript language and, once activated, seeks out Microsoft's Outlook e-mail software in an attempt to send copies of infected PDFs to contacts in a victim's address book and to addresses found in the software's various e-mail folders. McAfee.com said the PDF arrives attached to an e-mail message whose randomly generated subject line may contain text like "Find the peach" or "Joke." The body of the message, also randomly selected from a variety of phrases, may contain text such as "try finding the peach" and "I don't usually send this things, but..." The attached PDF files may have names like "peach.pdf," "joke.pdf," or "search.pdf." Recipients who open the PDF file will see a document with the message: "You have one minute to find the peach." Says McAfee.com: "A collage containing images of naked female buttocks is displayed, one of which is actually the image of a peach." Users who follow the instructions to "double-click the icon to show the solution" will launch the

VBScript worm. Staff at Symantec's Anti-virus Research Center (SARC), the research arm of the company behind the Norton virus-scanning software, said in a bulletin that the distribution of "Peachy" is likely to be low because of the worm's reliance on the full version of Adobe Acrobat. Reported by Newsbytes.com, <http://www.newsbytes.com>

Rocker95 (Rocker95@excite.com) [A167]

Chapter 24. Specification things

1. Is there a maximum document size for PDF?

Q: Is there a size limit, (either in terms of filesize or page numbers) for pdf files? We are looking at creating a document of around 1000 pages and would like to know if it's feasible before we start. [Q77]

A: Check out the Acrobat SDK on <http://partners.adobe.com/> - not because you need the contents, but because it has a file (Core API Reference) which is close to 3000 pages and heavily hyperlinked. It works well for me.

Aandi Inston (quite@dial.pipex.com) [A128]

A: I have personally created 10,000 page PDF documents, admittedly on fairly large machines with a lot of RAM (4gb plus). Whilst the PDF specification does not specify a maximum filesize (that I can find), the maximum byte offset for an object reference is a 10 digit number, which means that I would imagine that the absolute maximum file size would be 9,999,999,999 bytes or about 10gb (page 74, edition 2). I think you would have to have a pretty meaty machine to be able to open the document anyway at that point.

Michael Still (mikal@stillhq.com) [A129]

A: We regularly deal with PDFs containing 50,000 to 100,000+ pages, so I think your 1,000-page document will be fine... That said, be aware that Acrobat itself can sometimes have trouble with very large PDF files, especially those a gigabyte or more in size.

Scott Robert Ladd (scott.ladd@maximal.com) [A130]

A: We have created Acrobat documents of up to 20,000 pages (mutual fund statements for archiving) without problems. Files can be several Gb in size, so make sure that you have lots of disk space.

Dan Sideen (dansideen@home.com) [A131]

A: The Count attribute of a Pages tree is a signed 32-bit integer, which implies that it cannot be greater than 2,147,483,647. I doubt you'll hit that limit, though. :-). Distiller will not produce a PDF with more than 32,767 pages, though some other PDF producers do not seem to be similarly limited.

John Doherty (jdoherty@gstyp.com) [A221]

2. Multiplatform cross book links

Q: Does anyone know why cross-books links (from an index file to another file) will not work on a Solaris machine while they will work on a PC? The naming of the PDF files and the links are all lowercase to avoid upper/lowercase problems. [Q80]

A: To work cross-platform, the file names of the destination files must be stored as "generic" names in the PDF file. If they are platform-specific, they won't work on other platforms.

Helge Blischke (H.Blischke@srz-berlin.de) [A135]

3. Converting from user space coordinates into millimeters

Q: What is the best method of converting user space units to millimetres. [Q81]

A: I multiply by 25.4 and divide by 72.

Aandi Inston (quite@dial.pipex.com) [A136]

4. Large paper sizes

Q: I am trying to understand if there is any paper size limitations for the PDF files. If so, what? My objective is to create postscript files (PS) from all the corporate applications, and then from PS, create a PDF. The corporate applications will include AutoCAD, Word, Images, some Graphics apps and so on. I searched Adobe and many other web sites regarding this, but I could not get any info. Is there anyone out there that has generated PDF's for larger sizes like D or E drawings sizes? [Q82]

A: The PDF 1.2, it has the 200 inches x 200 inches, with the PDF 1.2 having the 40 inches x 40 inches, where it might be 1.3 and 1.2. So, the PDFs can be the size of the drawings.

Ross (apex@calpha.com) [A137]

5. Where can I find the specification?

Q: I am looking for the pdf file description, because I want to create / generate pdf files from a program. [Q83]

A: The PDF Reference Manual is available at:

<http://partners.adobe.com/asn/developer/technotes/acrobatpdf.html>

Mark Lookabaugh (mlookaba (at) telepath.com) [A138]

