

CIRCUIT CELLAR **ONLINE**

THE MAGAZINE FOR COMPUTER APPLICATIONS

Circuit Cellar Online offers articles illustrating creative solutions and unique applications through complete projects, practical tutorials, and useful design techniques.

[This Month](#)[Archive](#)[About Us](#)[Contact](#)[Looking for More?](#)

RESOURCE PAGES

RESOURCE LINKS

A Guide for Online Information About:

The Making of PDF and PostScript Files How to Get from Here to There using Print

by [Bob Paddock](#)

I once had the misfortune to be involved in a contract with a quasi-government agency. The type that is entrenched in a bureaucracy, with the attitude "we're big, so we're correct". Things like this would be laughable if you and I were not paying for it with the high cost of products and taxes.

The agency allowed us access to its data via a modem but only to view or print it. I wasn't permitted to save data, which the agency knew I needed in electronic form to fulfill our contract obligations. The agency said I could print the data and have someone type it back in. Of course, the agency wouldn't pay for a typist.

I needed the data in a couple different formats, one for human consumption via the document management system, and the other format in a database-style that I could manipulate the numbers.

The agency wouldn't pay for the necessary software, either. I'm sure you've experienced "tell us what you need and we'll tell you why you don't" from your management. Such was the case when I suggested purchasing a copy of [Adobe's Acrobat](#) or [Acrobat Capture](#).

For now, the defacto standard for the exchange of data, like datasheets, has become [Adobe's](#) Portable Document Format, at least until [DjVu](#) catches on, because of its significant amount of compression for documents. Portable Document Format

files are known as PDF files. Using PDF files seemed like the way to go for the human readable format, but how do you make PDFs without Acrobat?

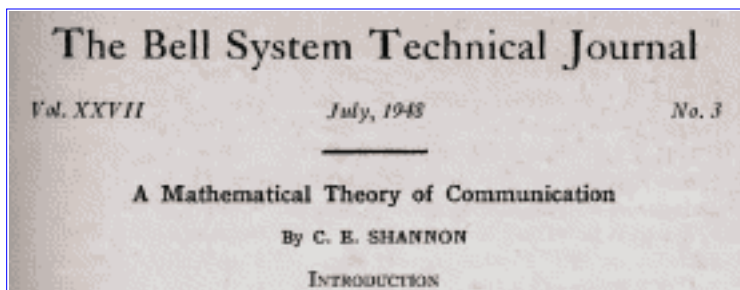
I remembered [Don Lancaster](#) often extolled the virtues of [PDF](#) and [PostScript](#), so I started at his site. His site has many interesting PDF and PostScript applications. From there, I discovered [Ghostscript](#) and a interesting utility called [Redmon](#).

This month, I noticed a few graphics illustration packages. I found things like [Mayura Draw](#) useful to illustrate test procedures.

[DjVu \(pronounced "dij`vu"\)](#), developed by [AT&T Labs](#), is image compression technology that allows the distribution of high resolution versions of scanned pages and compact versions of digital documents on the Internet. With DjVu, content developers can scan high-resolution color pages of books, magazines, catalogs, manuals, and historical or ancient documents, and make them available on the Internet. DjVu image compression is up to 10 times more compact than JPEG compression.

The [DjVu plug-in](#) is used with an Internet browser. It allows you to display and navigate through DjVu documents easily.

What better example to demonstrate [DjVu](#) than [the 1948 article that founded the original field of Information Theory](#)? Without this seminal paper by AT&T Labs Scientist Claude Shannon, data



compression would not be what it is today. The whole article, compressed with DjVu v2.0, is 1.4 MB. The same document compressed with the experimental DjVu v3.0 compressor is only 1.0 MB, even though it contains the searchable text (plug-in 3.0 required).

Interestingly, I was able to scan a old Intel datasheet. Four pages in [DjVu](#) format took only a total of 79 KB, but conventional images like TIFF and JPEG were over 300 KB, per page.

On-line data books anyone?

The name [PostScript](#) is a registered trademark of Adobe Systems Inc. All instances of the name PostScript in this text are references to the PostScript language as [defined by Adobe Systems Inc.](#) unless otherwise stated.

The official PostScript definition can be found [here](#).

The official definition of portable document format can be found [here](#).



[Ghostscript](#) is an open technology that has evolved over 10 years. It is a fully functional PostScript language interpreter. By installing a PostScript print driver and redirecting it via Ghostscript, it is possible to save documents in many different formats, such as PostScript, PDF, [TIFF](#), JPEG and several others.

[RedMon](#) (Redirection Port Monitor) redirects a printer port to Ghostscript to provide transparent PostScript printing from Windows 95 and NT, and has the ability to redirect printer output to a file or *StdOut*.

I manipulated the data into a format I could use with the language, Awk, when printing via RedMon.

[Awk \(named after the authors' initials\)](#) is an interpreted language for massaging text data developed by Alfred Aho, Peter Weinberger, and Brian Kernighan in 1978. It is characterized by C-like syntax, declaration-free variables, associative arrays, and field-oriented text processing.

[Gawk](#) is a version of Awk available from the [GNU Project](#).

[Gawk](#) is upwardly compatible with the latest POSIX specification of Awk. It also provides several useful extensions not found in other Awk implementations.

[Gawk](#) is an excellent language to use when you need to manipulate text, based on [regular expressions](#).

The associative arrays can make short work of many problems, such as counting the number of unique words in a document or looking for duplicate [CRC's](#) in a massive file of data.

```
awk '
# Print list of word frequencies
{
  for (i = 1; i <= NF; i++)
    freq[$i]++
}

END {
  for (word in freq)
    printf "%s\t%d\n", word, freq[word]
}'
```

Never being some one who settles with only one solution, I wanted to see what else is available for printing files.

The simplest way to generate a PostScript file under Windows is to install the *Apple LaserWrite II NT* driver that comes with Windows.

You can also download a PostScript Printer Driver [AdobePS for Windows 95 and Windows 98](#) from Adobe's site. It is available in several languages.

The [Net Distillery Service](#) is a server that runs Aladin Ghostscript to convert postscript into PDF files. You can put postscript documents into one directory on the server (using FTP) and pick up the resulting Acrobat Portable Document (TM) format file (PDF) in another.

[*Simple-PDF is simple. It creates PDF files.*](#)

"Simple-PDF is a full-fledged software to create a PDF publication from a set of text files."

[Simple-PDF](#) comes with Simple-PDF Composer to layout PDF publications. All the hard work of maintaining a .SIM file (which is the file-format by which Simple-PDF does its magic) is done by The Composer. You don't need anything else to produce a PDF file. It even has a simple, useful text editor.

[DocuCom PDF Driver](#) which is implemented as printer driver, is a PDF producer under Windows 95/98. You can use a Windows application such as MS Word, to create an original document. Then print to DocuCom PDF Driver to create a PDF file.

[Adobe PDFMaker](#) for [Microsoft](#) Word 97 provides enhanced features when creating PDF files from Word 97 documents.

PDF Maker for Word 97 automatically converts the following features in Word 97 documents to corresponding features in PDF files:

[Download PDFMaker 1.0 \(1.3MB - Windows 95/NT\).](#)

[There are easy software](#) releases [HTMLDOC](#).

[HTMLDOC](#) Can:

- convert HTML files to PDF or PostScript
- generate a table-of-contents for books
- generate indexed HTML files

It is Available for Unix and Windows.

PDFlib GmbH Releases On-Tthe-Fly PDF Generator

[PDFlib](#) is the software component you need if you want to generate PDF on your server, convert text and graphics, or implement PDF output in your own products.

The [pdfmark primer](#) gives an easy-to-use example, leveraging a powerful technique for generating hypertext elements in PDF. It's available in English and German.



[retepPDF](#) is a Java library to create PDF files from Java Applications and Servlets.

[ClibPDF](#) is a library of ANSI C functions, distributed as source code, for creating PDF (Acrobat) files directly via C language programs without relying on any Adobe Acrobat tools and related products.

[Graeme Dykes Prepress Software](#) offers a interesting utilities, couple [autility for rearranging pages in PDF files](#) and a [utility for placing background images behind pages in PDF files](#).

[Etymon™ PJ](#) is a developer tool kit for parsing, modifying, and creating PDF documents. The main part of the tool kit is a Java class library that provides software developers with an object representation of a PDF document.

[Xpdf](#) is a viewer for PDF files. These are also sometimes also called Acrobat files Xpdf runs under the X Window System on Unix, VMS, and OS/2. The non-X components of the package (pdf tops, pdf to text, and so on.) also run on Win32 systems and should run on any system with a decent C++ compiler.



[PCL File Viewing](#) allows you to view LaserJet PCL print files on-screen in Windows.

[PCL to Acrobat PDF](#) converts LaserJet PCL print files to Adobe Acrobat PDF.

[Encrypted PDFs](#)

When you give a PDF document an owner password, Adobe Acrobat encrypts it so that you can't remove the password from the file. Unfortunately, because of [U.S.regulations](#) governing exporting cryptographic software, the encryption cannot be added to programs like [Ghostscript](#), which are distributed with source code from the U.S.

This web page, however, is not in the U.S., so I can provide the modifications necessary to [view encrypted PDFs with Ghostscript](#).

Of course, to do this you must know the user or owner password of the file. If you forgot the password, take whatever documents you generated the PDF file from, and

re-generate it.

[Kyler Laird discusses PDF security.](#)



[ARTS](#) specializes in PDF software development and integration. It develops PDF-related software and provides online PDF services like [Planet PDF](#), PDF Store, and AcroBuddies Forum.



For the most comprehensive PDF resources on the Internet, PDFzone.com sponsors the Acrobat PDF [WebRing](#). The member sites showcase uses and users of Adobe Acrobat software and PDFs. Sites that are involved in the development and use of PDF-based documents are invited to join the Ring.



**The online authority for Acrobat, PDF
and Document Management Professionals**



Looking for a solution to a PDF problem? Why not search the new [Planet PDF](#) list which is a veritable plethora of PDF tools. It has tools for creation conversion, print, management, and developer libraries.

Digital Applications, Inc.

Tools for On-Demand Generation, Modification and Delivery of PDF Documents

Commercial Plug-Ins and Applications
Cross Platform Development
Macintosh - Windows - Solaris - Linux - AIX - Digital UNIX

SVG2PDF is a tool to convert [Scalable Vector Graphics \(SVG\)](#) documents into PDF files.

PlaceHolder, a plug-in for Adobe Acrobat, remembers the last page displayed in a document and returns to that page when the document is reopened.

The ReversePages plug-in for Acrobat Exchange reverses the order of all pages in a PDF document.

TriState replaces the three viewing (page mode) buttons and the three page view (zoom) buttons in the Acrobat toolbar with single TriState buttons. TriState thus reduces the number of icons on the Acrobat toolbar by four.

The ShowBookmarks plug-in displays the bookmark pane if the PDF file contains bookmarks, and the bookmark pane is not displayed.

PostScript

Programmers interested in learning PostScript will find [PostScript Language Reference, third edition \(January 1999\)](#) valuable resource.



[First Guide to PostScript](#) is a simple introduction to programming in PostScript.

[David Maxwell](#), has more to say for an [introduction to PostScript](#).

Thinking in PostScript is an important book about programming in PostScript. Now out of print, the author has made it available on <http://www.rightbrain.com/> - just click on the Books link. The book is in Acrobat PDF format, and covers beginner to intermediate PostScript programming experience.

[ToastScript](#) is a viewer for PostScript page descriptions similar to Ghostscript, but written entirely in Java. It is based on Java2D and requires JDK1.2 installed on your machine. It can be used as a standalone application or embedded in a HTML page

as an applet.

[Lout](#) is a document formatting system designed and implemented by [Jeffrey Kingston](#) at the Basser Department of Computer Science, University of Sydney, Australia.

The system reads a high-level description of a document similar in style to LaTeX and produces a PostScript file that can be printed on most laser printers and graphic display devices. Plain text and PDF (starting from version 3.12) output are also available.

Lout is multilingual. Adding new languages is easy. The following languages are currently supported: Czech, Danish, Dutch, English, Finnish, French, German, Hungarian, Norwegian, Italian, Russian, Slovenian, Spanish, and Swedish.



Software

It's difficult to find much information about [PostScript errors](#), so the [Quite Site](#) presents something about them. It is aimed at the non-programmer trying to print a file and getting PostScript errors. It includes details about what each PostScript error is likely to mean in a real-world problem.

Also by *Quite Site*:

- encapsulated PostScript
 - EPS in 10 easy stages.
 - making EPS files.
 - DCS, OPI, and so on.
-

[SANFACE Software](#) is a [PERL](#) specialist. The company uses the power of PERL (including DBI, DBD, LWP, and CGI.pm modules) combined with its products to develop scripts and CGI that create PDF files dynamically.

Can't print? Won't print? Do you want to view PostScript™ files but can't afford a PostScript printer? [RoPS](#) understands the common format for PostScript files on the Internet, level one, and has numerous level two enhancements, too. [RoPS](#) detects all PostScript errors, making it ideal for pre-flight checks. So, if a job can be viewed in [RoPS](#), you'll be able to print it. Plus [RoPS](#) handles process color, use [RoPS](#) to proof on screen and print composite pages and separations! [RoPS](#) is smaller, easier to configure, and faster than Ghostscript.



[The New Zealand Digital Library](#) system comprises several demonstration collections—computer science technical reports and bibliographies, literary works, humanitarian and development information, magazines—and makes them available on the 'Net through full-text interfaces.

[PreScript](#) offers PostScript conversion to plain ASCII or HTML.

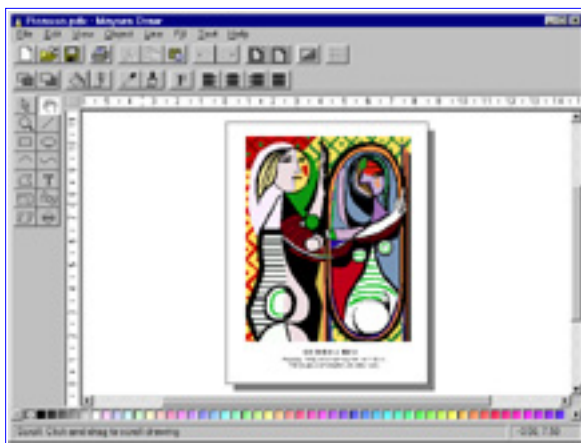
PreScript is really a PostScript- to- plain text converter, but rudimentary HTML can also be produced. Tags are inserted to mark paragraphs (<p>), short lines (
), page breaks (<hr>), and headers and footers (italicized with <i>...</i>).

PreScript determines the line spacing of a document and uses this (and also indentations) to determine paragraph boundaries.

Hyphenated words are de-hyphenated.

Most ligatures used by TeX document are detected. PreScript doesn't track font changes making it impossible to reliably detect all ligatures.

OTHER POSTSCRIPT CONVERTERS



[Mayura Draw](#) (formerly PageDraw) can export drawings in PDF format. You can use Adobe Acrobat 3.0 to publish PDF files on the Internet. Unlike GIF and JPEG formats, PDF format preserves the full resolution of your drawings. Using Acrobat Reader 3.0, you can view and zoom in on PDF graphics embedded in HTML files. You can also print at full resolution. Mayura Draw has direct support for PDF format.

With help from Ghostscript, Mayura Draw can edit any PostScript file. First convert the PostScript file using Ghostscript- to- AI format, and then open the AI file in Mayura Draw.

Mayura Draw can export drawings in EPS format for easy inclusion in LaTeX, Microsoft Word, and other word processors.

On PostScript printers, Mayura Draw outputs custom PostScript for the best results. The resulting quality is unmatched by ordinary drawing programs. For example, Mayura Draw produces high resolution vector pattern fills, whereas the patterns produced by ordinary drawing programs look jagged because they use low

resolution bitmaps.

Mayura Draw can export your drawings in Adobe Illustrator (AI) format. It also can import AI files created by other applications, such as Mathematica and GNUplot.

SVG is a new industry standard for vector graphics on the 'Net. Information about this new format is available [here](#). SVG files were tested using IBM's SVG viewer available at www.alphaworks.ibm.com.



The [GIMP](#) is the GNU Image Manipulation Program. It is a freely distributed piece of software suitable for such tasks as photo retouching, image composition, and image authoring.

[Sketch](#) is an interactive vector drawing program for Linux and other Unix compatible systems. It is a flexible and powerful tool for illustrations, diagrams and other purposes.

[E.G.S. Quick Vector](#) is a program for quick raster to vector conversion. In other words, conversion of a raster image into a set of primitive vector objects (such as lines, arc, circles, polylines, arrows, filled polygons, etc.).

[pstoedit](#) translates PostScript and PDF graphics into other vector formats. You can download [pstoedit](#) source code, which is binary for Windows 9x/NT, and Linux.

[PStill](#) is a PostScript-to-PDF converter written by Frank Siegert and available for a multitude of platforms.



[Simple
Document
Format](#)

[SDF](#) is a freely available documentation system designed and developed by Ian Clatworthy, with help from many others. Based on a simple, readable markup language, SDF generates high quality output in multiple formats, all derived from a single document source. Supported output formats include HTML, PostScript, PDF, man pages, POD, LaTeX, SGML, MIMS HTX and F6 help, MIF, RTF, and Windows help and plain text.

Related Information

[PostScript FAQ](#) [Internet resources for PostScript and Ghostscript](#)
[comp.lang.postscript newsgroup](#)
[Internet PostScript resources](#)

From the [Simtel.Net Windows 95/98 Collection Printing Utilities](#):

[pd14.zip](#)—bypasses printer drivers, prints unattended.

[pdfd40e.zip](#)—generates PDF files from Windows applications.

[sdp14.zip](#)—sets the default printer. Free command line utility.

[sprint13.zip](#)—command line that prints driver for html files.

[ptw_100.zip](#)—print to window, Windows 95/98 printer driver.

All product names and logos contained herein are the trademarks of their respective holders.

The fact that an item is listed here does not mean we promotes its use for your application. No endorsement of the vendor or product is made or implied.

If you would like to add any information on this topic or request a specific topic to be covered, contact [Bob Paddock](#).

Circuit Cellar provides up to date information for engineers, www.circuitcellar.com for more information and additional articles.

©Circuit Cellar, the Magazine for Computer Applications. Posted with permission.

For subscription information, call (860) 875-2199 or e-mail

subscribe@circuitcellar.com

Copyright ©1999 ChipCenter

[About ChipCenter](#) ■ [Contact Us](#) ■ [Hot Jobs at ChipCenter](#) ■ [Privacy Statement](#) ■ [Advertising Information](#)