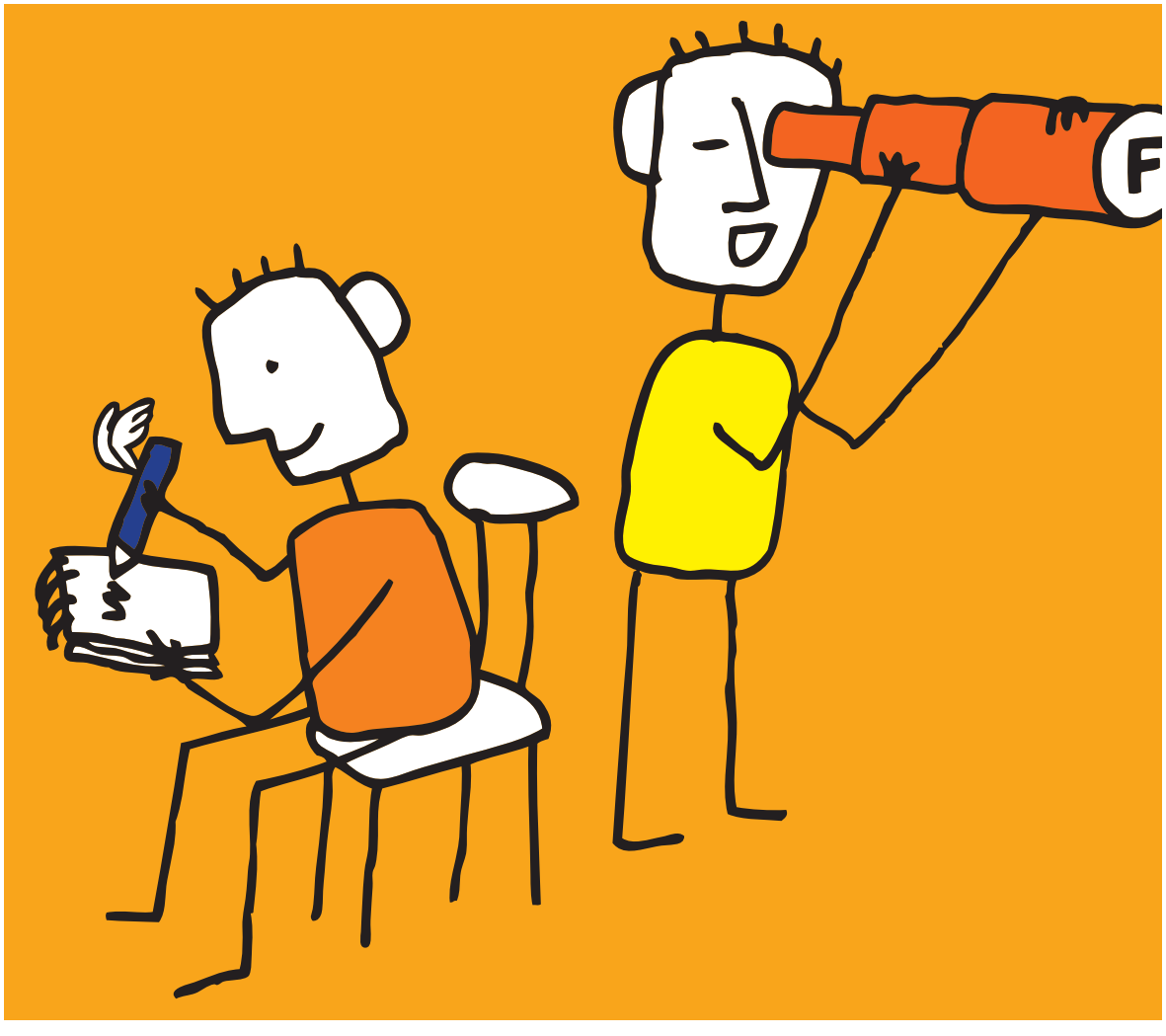
PDFlib GmbH

FontReporter

Version 1.11

Outdated version - only for demonstration!

A Plugin for analyzing fonts in PDF



Copyright © 2005-2017 PDFlib GmbH. All rights reserved.

PDFlib GmbH
Franziska-Bilek-Weg 9, 80339 München, Germany
www.pdflib.com
phone +49 • 89 • 452 33 84-0
fax +49 • 89 • 452 33 84-99

If you have questions check the PDFlib mailing list and archive at
groups.yahoo.com/neo/groups/pdflib/info

Licensing contact: sales@pdflib.com
Support for commercial PDFlib licensees: support@pdflib.com (please include your license number)

You can use PDFlib FontReporter free of charge; however, it is not in the public domain. This software cannot be sold or redistributed (whether for a fee or at no charge), either stand-alone or in combination with any other product, without the express written permission of PDFlib GmbH.

This publication and the information herein is furnished as is, is subject to change without notice, and should not be construed as a commitment by PDFlib GmbH. PDFlib GmbH assumes no responsibility or liability for any errors or inaccuracies, makes no warranty of any kind (express, implied or statutory) with respect to this publication, and expressly disclaims any and all warranties of merchantability, fitness for particular purposes and noninfringement of third party rights.

PDFlib FontReporter is provided »as is« without any warranty, express or implied, including but not limited to any implied warranties of merchantability and fitness for a particular purpose. In no event will PDFlib GmbH be liable for any damages, including lost profits, lost savings, or other incidental consequential damages.

Although PDFlib FontReporter is not a commercial product, we strive to provide high quality. If you run into problems you are encouraged to contact us at support@pdflib.com.

Adobe, Acrobat, PostScript, and XMP are trademarks of Adobe Systems Inc. AIX, IBM, OS/390, WebSphere, iSeries, and zSeries are trademarks of International Business Machines Corporation. ActiveX, Microsoft, OpenType, and Windows are trademarks of Microsoft Corporation. Apple, Macintosh and TrueType are trademarks of Apple Computer, Inc. Unicode and the Unicode logo are trademarks of Unicode, Inc. Unix is a trademark of The Open Group. Java and Solaris are trademarks of Sun Microsystems, Inc. HKS is a registered trademark of the HKS brand association: Hostmann-Steinberg, K+E Printing Inks, Schmincke. Other company product and service names may be trademarks or service marks of others.

Contents

1 Installing PDFlib FontReporter *5*

2 Working with FontReporter *7*

2.1 What can you do with FontReporter? *7*

2.2 Overview of PDF Font Formats *9*

2.3 Contents of a Font Report *11*

2.4 Investigate PDF Problems with FontReporter *14*

2.5 Error Messages *15*

A Revision History *17*

1 Installing PDFlib FontReporter

Requirements. PDFlib FontReporter works with the following Acrobat versions:

- Acrobat 8/9/X/XI/DC on Windows
- Acrobat X/XI/DC on macOS

The Plugin doesn't work with Adobe Reader/Acrobat Reader.

Installing FontReporter on Windows. To install PDFlib FontReporter in Acrobat, the plugin files must be placed in a subdirectory of the Acrobat plugin folder. This is done automatically by the plugin installer, but can also be done manually. A typical location of the plugin folder looks as follows:

```
C:\Program Files\Adobe\Acrobat XXX\Acrobat\plug_ins\PDFlib FontReporter
```

For 32-bit versions of Acrobat running on 64-bit Windows the first part should be

```
C:\Program Files (x86)\...
```

Installing FontReporter for Acrobat X/XI/DC on macOS. Proceed as follows to install the plugin for all users:

- Double-click the disk image to mount it. A folder with the plugin files will be visible.
- Copy the plugin folder to the following path in the system's *Library* folder:

```
/Library/Application Support/Adobe/Acrobat/XXX/Plug-ins
```

Alternatively you can install the plugin only for a single user as follows:

- Click the desktop to make sure you're in the Finder, hold down the *Option* key, and choose *Go, Library* to open the user's *Library* folder.
- Copy the plugin folder to the following path in the user's *Library* folder:

```
/Users/<username>/Library/Application Support/Adobe/Acrobat/XXX/Plug-ins
```

Multi-lingual interface. PDFlib FontReporter supports multiple languages in the user interface. Depending on the application language of Acrobat, FontReporter chooses its interface language automatically. Currently English and German interfaces are available. If Acrobat runs in any other language mode, FontReporter uses the English interface.

Troubleshooting. If PDFlib FontReporter doesn't seem to work check the following: Make sure that in *Edit, Preferences, [General...], General* the box *Use only certified plug-ins* is unchecked. The plugin is not loaded if Acrobat runs in Certified Mode.

2 Working with FontReporter

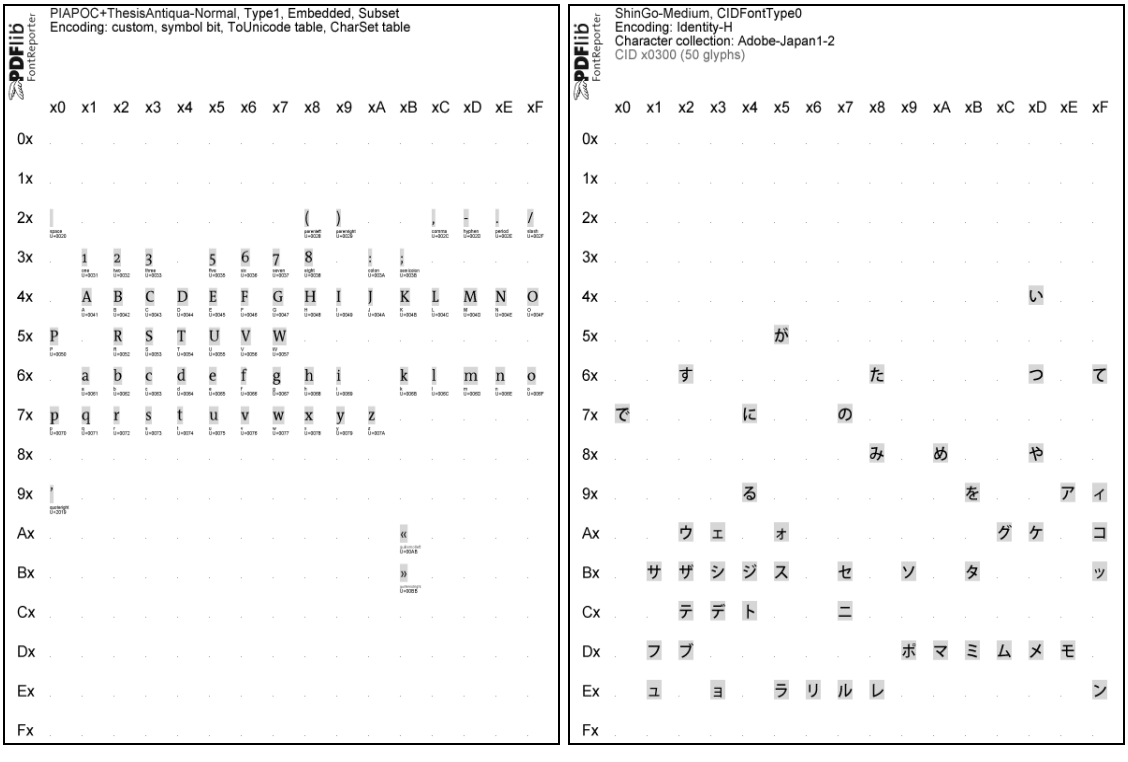
2.1 What can you do with FontReporter?

FontReporter is a useful tool if you are interested in fonts within PDF documents. It provides font- and encoding-related information which helps in a variety of situations:

- analyze printing problems (e.g. a particular font causes printing errors)
- investigate text extraction problems (e.g. copying text from a PDF results in garbage)
- visualize Unicode mappings for a font
- find flaws in the PDF creation workflow (e.g. printer driver converted a PostScript Type 1 font to Type 3)
- test whether ToUnicode mapping tables (required for PDF/A-1a) are present
- identify logos and symbols which are represented as text in a PDF
- learn which fonts are contained in a PDF, and which glyphs they contain (e.g. the file size is too large because some fonts ended up in the PDF unintentionally)
- check font subsets to see which glyphs are contained in the subset
- learn more about PDF font technology

Using FontReporter is as easy as bringing up the menu *Plug-Ins, PDFlib FontReporter..., Create Font Report* in Acrobat. This will create a font report for all pages of the current PDF document as a separate PDF. Two pages from typical font reports are shown in Figure 2.1.

Fig. 2.1
Sample font reports



Supported PDF and font formats. FontReporter supports all PDF versions up to Acrobat DC. All font and encoding formats in PDF are supported, as well as all types of embedded font data.

Advantages over Acrobat's font properties panel. All versions of Acrobat including Adobe Reader provide font information via *File, Document Properties..., Fonts*. However, Acrobat's font overview is limited in use; FontReporter provides the following advantages compared to Acrobat's font list:

- FontReporter provides much more information about each font
- FontReporter deals with CJK font names even on Western systems
- FontReporter provides glyph tables containing the glyphs of a font along with their widths, names, and Unicode values
- FontReporter presents the output as a PDF document so that you can save or print it
- FontReporter is guaranteed to process the full document, regardless of which pages have already been displayed in Acrobat

PDF text extraction with PDFlib TET. FontReporter is an auxiliary tool to our PDFlib Text and Image Extraction Toolkit (TET). TET is software for extracting the text and image contents of PDF documents. It is available both as a standalone program and a programming library/component which can be integrated into existing software. TET extracts text from all kinds of PDF documents and normalizes the text to Unicode. FontReporter can be used to create Unicode mapping tables for PDF documents which do not contain enough information for extracting text, or which contain wrong Unicode mapping tables. Fully functional evaluation versions of TET are available for download from *www.pdflib.com*.

TET PDF IFilter. TET PDF IFilter extracts text and metadata from PDF documents and makes it available to search and retrieval software on Windows. This allows PDF documents to be searched on the local desktop, a corporate server, or the Web. TET PDF IFilter is based on the patented PDFlib Text and Image Extraction Toolkit (TET). TET PDF IFilter is a robust implementation of Microsoft's IFilter indexing interface. It works with all search and retrieval products which support the IFilter interface, e.g. SharePoint and SQL Server. Fully functional evaluation versions of TET PDF IFilter are available for download from *www.pdflib.com*.

Free TET Plugin. The TET Plugin is a free companion to the FontReporter Plugin. It can be installed in Adobe Acrobat and allows interactive use of the Text and Image Extraction Toolkit (TET) with any PDF document that is currently open in Acrobat. Using the TET plugin you can access TET's functionality and experiment with TET options. The TET plugin can freely be downloaded from *www.pdflib.com*.

2.2 Overview of PDF Font Formats

PDF supports a large number of font formats, the details of which can get confusing. In order to help you interpret the reports created by FontReporter we provide a quick summary of PDF font formats and their most important properties.

While the format of a font in a PDF document depends on the format of the original font used to compose the document, this is not the only factor which plays a role here. Other factors include the configuration options in the PDF-creating software, the settings of the printer driver used to generate PostScript data for PDF conversion, the set of characters in the document, the overall number of used characters, and more.

A particularly important aspect is the distinction between simple fonts and composite (CID) fonts.

Simple fonts. Simple fonts comprise the PostScript Type 1 (including Multiple Master), TrueType, and Type 3 types, and are addressed with 8-bit codes. They are therefore limited a maximum of 256 characters. Simple fonts use a name-based encoding, which is a table for mapping the character codes to the glyphs in the font.

Composite (CID) fonts. Composite or CID (character ID) fonts come in PostScript and TrueType flavors. They can contain up to 65535 characters and are much more flexible than simple fonts. While CID fonts often use 2-byte codes for addressing the glyphs in the font, more complicated schemes with a variable number of bytes per character (1-4) are used for CJK fonts. Instead of an encoding table CID fonts require a CMap (Character Map) for providing the mapping from character codes to actual glyphs in the font. Dozens of predefined CMaps are available for common CJK fonts. The font's character collection specifies a particular set of Chinese, Japanese, or Korean characters. So-called Identity CMaps are used (mainly for Western fonts) in order to directly address the glyphs in a font without any intermediate mapping table.

Comparison of PDF font formats. Table 2.1 details the font formats supported in PDF, and explains which original font formats can be converted to these types by the PDF creation software.

Table 2.1 Font formats in PDF

Name	*Notes*
Type1	*Classic PostScript Type 1 fonts. In addition to the original Type 1 format they can also be embedded as CFF (Compressed Font Format) under the name Type1C (Type 1 Compressed).* *These fonts are the result of classic PostScript Type 1 fonts or OpenType fonts with PostScript outlines.*
MMType1	*(In Acrobat:* `MM`*) Multiple Master fonts are an extension of the Type 1 format, and are rarely used.* *These fonts are the result of PostScript Type 1 Multiple Master fonts.*
Type3	*User-defined fonts, i.e. the glyphs are described by raw vector or image operations instead of a ready-made font. Type 3 fonts are always embedded. They are mainly intended for bitmapped fonts and logo fonts.* *These fonts are often the result of a printer driver converting a PostScript Type 1 or TrueType font to a bitmap font. Some applications use Type 3 fonts for achieving special effects, such as filling an area with a pattern.*