Appligent

Recognizing Corrupt and Malformed PDF Files

Mark Gavin
Chief Technology Officer
Appligent, Inc.

PDF CONFERENCE

Appligent

Introduction

Appligent

Introduction

As the number of PDF document creation and manipulation tools has increased; there are many more PDF documents in circulation which were simply not created correctly; thus, they are malformed or corrupt.

Appligent

Introduction (continued)

- Properly formed PDF is constructed by closely following the PDF file format specification as defined in the PDF Reference Manual.
- Following the PDF Reference Manual is not always an easy task.
  - PDF is a moving target.
  - The PDF Reference is sometimes ambiguous and/or incomplete in its description.

PDF Reference
third edition
Adobe Portable Document Format
Version 1.4
Adobe Systems Incorporated

Library of Congress Cataloging-in-Publication Data
PDF reference : Adobe portable document format version 1.4

Adobe Systems Incorporated. — 3rd edition

ISBN 0-201-75839-3 (alk. paper)

Appligent

Introduction (continued)

In this presentation we will discuss the following:

- What are the mistakes being made by developers.
- How to recognize some corrupt and malformed PDF files.
- Simple steps to try to correct problems.

Appligent

Agenda

- Background
- Types of Problems
- Problem Diagnostics
- Some Common Errors
- Correcting Problems
- Other Malformed PDF Issues
- Questions & Answers

Appligent

Background

Where We are and How We Got Here

Appligent

The PDF Reference

- The Portable Document Format specification is a moving target
- 1.0, 1.1, 1.2, 1.3 and 1.4
- Updated almost every two years
- Currently 945 pages
- Features Rarely Obsoleted
- Requires other references
  - Postscript Language Reference
  - Data Structures and Algorithms; Aho, Hopcroft & Ullman

PostScript Language Reference, Third Edition, Addison-Wesley, Reading, MA, 1999.

Aho, A. V., Hopcroft, J. E., and Ullman, J. D., Data Structures and Algorithms, Addison-Wesley, Reading, MA, 1983. Includes a discussion of balanced trees.

Adobe Type 1 Font Format. Explains the internal organization of a PostScript Type 1 font program. Also see Adobe Technical Note #5015, Type 1 Font Format Supplement.

Apple Computer, Inc., TrueType Reference Manual. Available on Apple's Web site at http://developer.apple.com/fonts/TTRefMan/index.html.

Please see the PDF Reference Manual Bibliography for more references.

Appligent

Well-Formed and Valid

- There is no concept of a Well-Formed and Valid PDF file like there is in XML.
- PDF is not really amenable to the use of a scanning tool (lint) to check the validity of the file.
- Acrobat is considered the benchmark for conformance to the PDF Specification.
- Unfortunately; Acrobat will display many types of invalid and corrupt PDF files.

Appligent

History

- Long ago in a galaxy far away; Distiller and PDF Writer were the only tools available to create PDF files.
- This gave the end user community PDF files; which, if not perfect, were at least consistent in form and quality.

http://www.pineapplesoft.com/newsletter/archive/19980501_xml.html

"XML documents come in two flavors: well-formed and valid. Well-formed documents are the least stringent: they simply require that all elements are cleanly nested. Valid documents, on the other hand, must include a DTD and adhere to it! A variety of XML tools, known as validating parsers, check the conformance of documents against their DTDs."

"Clearly, well-formed XML documents are similar to HTML documents. Indeed HTML documents never include a DTD. There is HTML DTD (published as part of the HTML standard) but, as an HTML user or author, you will never see it. The HTML DTD is supposed to be universally available and is therefore not included in documents, if only to reduce download times."

"Valid documents, on the other hand, are akin to full-blown SGML documents. They carry the bulk of the DTD with them and this makes it possible to validate them."

Appligent

Today Anyone Can Create PDF Documents

- Distiller, Global Graphics, PStill, Amyuni, Sanface
- Photoshop, Illustrator, Corel Draw
- Appligent, Arts, MapSoft, Quite
- PDFLib, Zeon, Glance, iText, FOP
- Others

Lists of PDF Tools can be found at the following:

http://www.planetpdf.com

http://www.pdfzone.com

http://www.pdfzone.de

http://www.pdfworker.com

http://www.adobe.com/store/plugins/acrobat/main.html

Appligent

"True Adobe PDF"

- There is No "True Adobe PDF" Anymore
- If there is a printing problem one of the first questions asked is "Are you using Adobe Postscript or a Postscript Clone?"
- This concept simply doesn't exist in the world of PDF anymore.
- Every Application, including Adobe's own applications, produce very different PDF with various degrees of quality.

Even within Acrobat itself; compare the PDF produced by Acrobat Distiller against the PDF produced when you request summary of comments.

Then, compare the PDF produced by InDesign against Illustrator.

Appligent

Types of Problems

Does a Problem Exist

Appligent

Obvious Problems

Acrobat tells you something is wrong

- File is Corrupt; Being Repaired Dialog.
- Error dialog Displayed when opening a PDF file.
- Asks to save the file on closing; even though no changes were made.
- A Blank Page is Displayed.

Appligent

Hidden Problems

- Incorrect Object Streams
- Incomplete Object Definitions
- Malformed Page Content Streams
- Incorrect Object Type

Appligent

Problem Diagnostics

Rummaging Through the Internal's of a PDF File

Appligent

Does a Problem Exist

Acrobat as a diagnostic tool

- Can Acrobat Open the PDF File
  - Is the "Corrupt" dialog displayed briefly
  - Is an error dialog displayed
- Close the file
  - Does Acrobat ask you to save
- Step through the document a page at a time
  - Does Acrobat complain about any individual pages

Appligent

Acrobat Error Dialogs

Many times, Acrobat can give you more information about why an error dialog has been displayed

- To display more information; press & hold a modifier key while clicking on the OK button in the error dialog
  - Option-Click on Macintosh
  - Control-Click on Windows

Appligent

Browsing the PDF File

- Text Editor
  - BBEdit
  - Note Pad
  - Showing Invisible Characters is a Plus
  - Soft Wrapping is another Plus
- Enfocus Browser
  - Hard to Find; but, still available

Appligent

Check for Basic Required information

- Header - %PDF
- Trailer
- xref table
- Document Information Dictionary
  - Creator & Producer
  - Creation Date and Modification Date

Appligent

Carriage Returns and Line Feeds

Side Note: Carriage Returns and Line Feeds behave differently

- Carriage Returns - Macintosh
- Line Feeds - Unix
- Carriage Returns and Line Feeds - Windows
- They are different:
  - Carriage Return - ASCII 13 Decimal, 0x0D Hex
  - Line Feed - ASCII 10 Decimal, 0x0A Hex

This slide on line endings is more of a side note; but, line endings do cause a significant amount of confusion. Thus, I felt it was important to place here in the presentation.

PDF Reference, Third Edition

3.4 File Structure

As a matter of convention, the tokens in a PDF file are arranged into lines; see Section 3.1, “Lexical Conventions.” Each line is terminated by an end-of-line (EOL) marker, which may be a carriage return (character code 13), a line feed (character code 10), or both. PDF files with binary data may have arbitrarily long lines. However, to increase compatibility with other applications that process PDF files, lines that are not part of stream object data are limited to no more than 255 characters, with one exception: beginning with PDF 1.3, an exception is made to the restriction on line length in the case of the Contents string of a signature dictionary (see “Signature Fields” on page 547). See also implementation note 11 in Appendix H.

Appligent

Header

Only Two Lines
How Hard Can It Be?

```
%PDF-1.2
%„Œ”

%PDF-1.3
%JetForm PDF Support Version 2.3.000
%EncodingObject=0
%˚¸˝˛ 1 0 obj << /Type /Catalog /Pages 3 0 R /Outlines 4 0
R >> endobj
```


PDF Reference, Third Edition

3.4.1 File Header

The first line of a PDF file is a header identifying the version of the PDF specification to which the file conforms.

Note: If a PDF file contains binary data, as most do (see Section 3.1, “Lexical Conventions”), it is recommended that the header line be immediately followed by a comment line containing at least four binary characters—that is, characters whose codes are 128 or greater. This will ensure proper behavior of file transfer applications that inspect data near the beginning of a file to determine whether to treat the file’s contents as text or as binary.

Some developers cheat by omitting the binary data in the second line.

Appligent

Header (continued)

A Quirk being Exploited by some PDF Creators

- Creating a PDF file without the binary data in the second line causes Acrobat to enter a mode where it will not report all errors.
- So, missing binary data is a good sign that something else if wrong with the given PDF file.

Appligent

Trailer

/Info and /ID will be part of any well formed document trailer

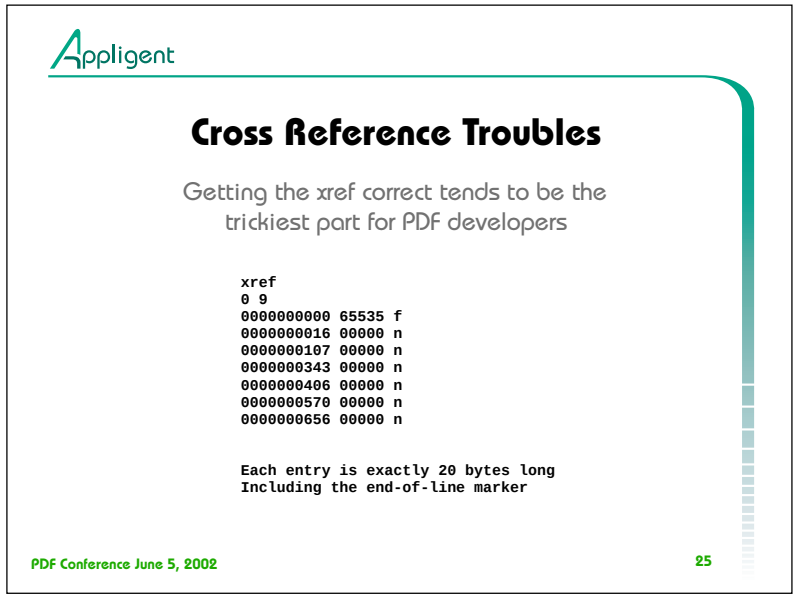
```
trailer
<<
/Size 9  <- Size actually does need the correct object count
/Root 1 0 R
/Info 2 0 R
/ID[<22fe617fe156d37892dd946294182028><22fe617fe156d37892dd9
46294182028>]
>>
startxref
51347
%%EOF
```

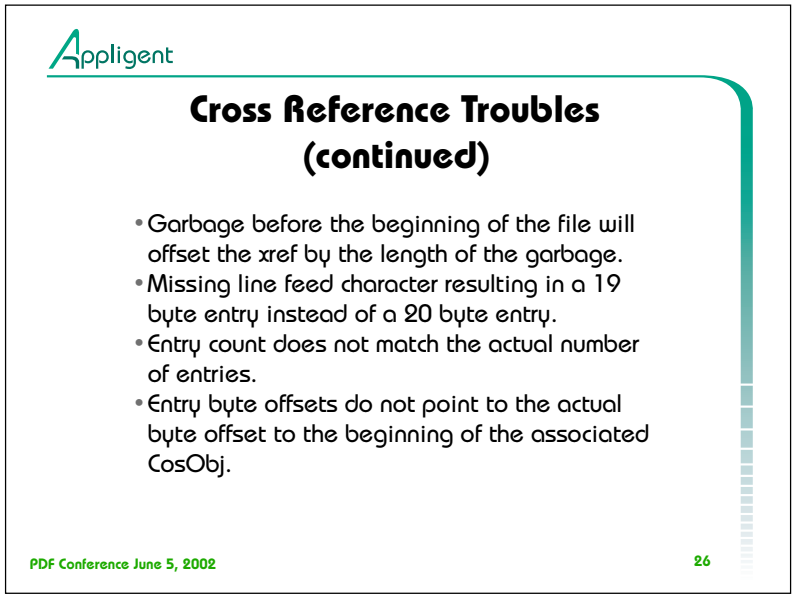

PDF Reference, Third Edition

3.4.4 File Trailer

The trailer of a PDF file enables an application reading the file to quickly find the cross-reference table and certain special objects. Applications should read a PDF file from its end. The last line of the file contains only the end-of-file marker, %%EOF. (See implementation note 14 in Appendix H.) The two preceding lines contain the keyword startxref and the byte offset from the beginning of the file to the beginning of the xref keyword in the last cross-reference section. The startxref line is preceded by the trailer dictionary, consisting of the keyword trailer followed by a series of key-value pairs enclosed in double angle brackets (<<…>>). Thus, the trailer has the following overall structure:

trailer
<< key1 value1
key2 value2
…
keyn valuen
>>
startxref
Byte_offset_of_last_cross-reference_section
%%EOF





PDF Reference, Third Edition

3.4.3 Cross-Reference Table

The cross-reference table contains information that permits random access to indirect objects within the file, so that the entire file need not be read to locate any particular object. The table contains a one-line entry for each indirect object, specifying the location of that object within the body of the file.

The cross-reference table is the only part of a PDF file with a fixed format; this permits entries in the table to be accessed randomly.

Each cross-reference section begins with a line containing the keyword xref.

Following this line are one or more cross-reference subsections, which may appear in any order.

The subsection begins with a line containing two numbers, separated by a space: the object number of the first object in this subsection and the number of entries in the subsection.

Appligent

Garbage Before, After and In Between

- The PDF Reference allows up to 1K of garbage before the beginning of the PDF file.
- Acrobat will accept an almost unlimited amount of garbage after the %%EOF marker at the end of the file.
- Binary garbage can also exist between the end of an object and the beginning of the next object.

Binary data before the %PDF is used in some prepress workflows.
However, binary data before the %PDF could also be a sign of a file transfer error.

Binary data after the %%EOF could be caused in several different ways.
1. File system error.
2. File transfer error.
3. Programmer error.

Binary data between objects within a PDF file, that does not overwrite data or invalidate the cross reference table, is almost always caused by programmer error.

```
%!PS-Adobe-3.0 PDF-1.3
%KDKChargeNumber: AVIREPORTS
...
%%Title: BBJ JULY RUN-B73729054 (YG005)
%%Emulation: pdf
%KDKOutputMedia: stapler
%KDKChaptersAreSets: on
%%EndComments%PDF-1.3
%„Œ”
```


Appligent

Document Information Dictionary

Sometimes called Doc Info Fields

- The Info Dictionary is officially Optional; but, since this is a presentation on recognizing corrupt and malformed PDF files; a missing or incomplete Info Dictionary is a sign of a poorly built file.
- Creator & Producer
- Creation Date and Modification Date

PDF Reference, Third Edition

9.2.1 Document Information Dictionary

The optional Info entry in the trailer of a PDF File (see Section 3.4.4, “File Trailer”) can hold a document information dictionary containing metadata for the document.

Example 9.1 shows a typical document information dictionary.

Example 9.1
1 0 obj
<< /Title (PostScript Language Reference, Third Edition)
/Author (Adobe Systems Incorporated)
/Creator (Adobe® FrameMaker® 5.5.3 for Power Macintosh)
/Producer (Acrobat® Distiller™ 3.01 for Power Macintosh)
/CreationDate (D:19970915110347-08’00’)
/ModDate (D:19990209153925-08’00’)
>>
endobj

Appligent

Creator and Producer

Document Information Dictionary

- One of the first items to check
- For some unknown reason; the PDF Reference still considers these items optional.
- Many PDF files are being created without the Creator and Producer information
- A well formed PDF file will always have a Creator and Producer

PDF Reference, Third Edition

9.2.1 Document Information Dictionary

Creator text string (Optional) If the document was converted to PDF from another format, the name of the application (for example, Adobe FrameMaker®) that created the original document from which it was converted.

Producer text string (Optional) If the document was converted to PDF from another format, the name of the application (for example, Acrobat Distiller) that converted it to PDF.

CreationDate date (Optional) The date and time the document was created, in human-readable form (see Section 3.8.2, “Dates”).

ModDate date (Optional; PDF 1.1) The date and time the document was most recently modified, in human-readable form (see Section 3.8.2, “Dates”).

Appligent

Incorrect Date Format

Acrobat Date format is
D:20010605110739

```
2 0 obj
<<
/Creator(Adobe Photoshop 5.0)
/CreationDate( Tue Jun 05 11:07:39 2001
)
/Producer(Adobe Photoshop for Windows)
```


PDF Reference, Third Edition

3.8.2 Dates

PDF defines a standard date format, which closely follows that of the international standard ASN.1 (Abstract Syntax Notation One), defined in ISO/IEC 8824 (seethe Bibliography).

A date is a string of the form (D:YYYYMMDDHHmmSSOHH’mm’)

For example, December 23, 1998, at 7:52 PM, U.S. Pacific Standard Time, is represented by the string D:199812231952-08’00’

Appligent

Same Common Errors and What They Actually Mean

At least some that are fairly easy to find

Appligent

Types of Common Errors

- Expected a Name Object
- Expected a Number Object
- The font 'X' contains bad /Widths
- Unable to find or create the font 'X'
- Bad Parameter

Appligent

Expected a Name Object

Typically occurs when the CosDict item contains a CosString instead of a CosName

- CosString - /MyName (The Cos String)
- CosName - /MyName /TheCosName

Appligent

Expected a Number Object

Searching a CosDict for a CosNumber and found something else

- Common to find an indirect reference instead of a number.
- When working on an international system; some software will use the local numeric delimiter within numbers; for example; comma instead of period.
  - /BBox [0 0 11,4651 10,8281]
  - /BBox [0 0 11.4651 10.8281]

Appligent

The font 'X' contains bad /Widths

```
Programmer Error
11 0 obj
<<
/Type /Font
/Subtype /TrueType
/Name /Fo1
/BaseFont /CourierNew,Bold
/Encoding /WinAnsiEncoding
/FirstChar 255
/LastChar 0
/Widths [  ]
>>
endobj
```

PDF Conference June 5, 2002 35

A Font Object with a /Widths array should look like the following:

10 0 obj
<<
/Type /Font
/Subtype /TrueType
/Name /Fo8
/BaseFont /TimesNewRoman+000008
/FontDescriptor 6 0 R
/Encoding /WinAnsiEncoding
/FirstChar 32
/LastChar 255
/Widths [294 332 460 498 498 885 774 332 332 332 498 885 332 332 332332 498 498 498 498 498
498 498 498 498 498 332 332 995 885 995 442 885 719 626 682 737 626 590 719 737 332 386 700
626 903 737 719 571 719 664 534 626 737 719 940 719 719 645 332 332 332 498 498 332 442 498
442 498 442 332 498 498 276 276 498 276 774 498 498 498 498 368 386 276 498 498 719 498 498
442 442 498 442 995 995 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 294 332 498
498 498 498 332 498 498 498 498 406 498 885 737 498 498 885 498 498 498 552 498 332 332 442
498 406 885 885 885 442 719 719 719 719 719 719 886 882 626 626 626 626 332 332 332 332 737
737 719 719 719 719 719 664719 737 737 737 737 719 571 498 442 442 442 442 442 442 664 442
442442 442 442 276 276 276 276 571 498 498 498 498 498 498 885 498 498498 498 498 498 498
498]
>>

35

Appligent

Unable to find or create the font 'X'

Some characters may not display or print correctly.

- Typically Acrobat is trying to find a font which it can not locate in the PDF file or on the local system
- In the case in the notes, an obscure font is used by was not embedded
- Lesson Learned: Embed Fonts

PDF Conference June 5, 2002 36

7 0 obj <<
/Type /Font
/Subtype /Type1
/Name /G1F9
/BaseFont /TerrapinCode3of9HD
/FirstChar 0
/LastChar 255
/Widths [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 319 0 0 0 319 319 0 0 0 0 319
319 0 319 319 319 319 319 319 319 319 319 319 319 319 319 0 0 0 0 0 0 0 319 319 319 319 319
319 319 319 319 319 319 319 319 319 319 319 319 319 319 319 319 319 319 319 319 319 0 0 0 0 0
0 319 319 319 319 319 319 319 319 319 319 319 319 319 319 319 319 319 319 319 319 319 319
319 319 319 319 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 319 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 319 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
/FontDescriptor 8 0 R
>> endobj
8 0 obj <<
/Type /FontDescriptor
/Ascent 0
/CapHeight 0
/Descent 0
/FontBBox [0 0 0 0]
/FontName /TerrapinCode3of9HD
/Flags 6
/ItalicAngle 0
/StemV 0
>> endobj

36

Appligent

Bad Parameter

Another Photoshop Quirk

- 11 0 obj
- << /Length 12 0 R >>
- stream
- endstr eam
- endobj
- 12 0 obj
- 0
- endobj

Appligent

Correcting Problems

Some Simple, and Many Times Effective,
Ways of Repairing PDF Files

Gives a bad parameter error.

2 0 obj
<<
/Creator (Adobe Photoshop 6.0)
/CreationDate (D:20010605144636)
/Producer (Adobe Photoshop for Macintosh)
/ModDate (D:20010605151707-04’00’)
>>
endobj

11 0 obj
<< /Length 12 0 R >>
stream
endstream
endobj
12 0 obj
0
endobj

This pdf was created by Photoshop 6.0 for Mac. When the pdf is opened in a text editor, it has multiple spaces before the first xref. The pdf is linearized.

The file contains a zero-length CosStream

Appligent

Correcting Problems

These work for many corrupted PDF files

- Acrobat Save
  - To Save Acrobat's Background Correction
- Acrobat Save As...
  - Turn Off Optimize for Fast Web View
- Acrobat Save As...
  - Turn On Optimize for Fast Web View
- Acrobat Distiller
  - Use Distiller 4 or 5
  - Form Fields can be pasted in the new file
  - Distiller is Your Friend - It's amazing how many problems can be corrected by simply re-distilling a corrupted PDF file

Appligent

Other Malformed PDF Issues

Appligent

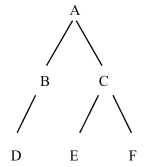
Other Malformed PDF Issues

- Balanced Page Trees
- Required Items
- CosStrings
- Linearization
- Losing Form Field Lengths

Appligent

Balanced Page Trees

Not all PDF Creators are Building Balanced Page Trees



- To optimize performance of viewer applications, Acrobat Distiller constructs balanced trees.
- Search speed for a balanced tree is O(log n)
- Search speed for a completely unbalanced tree can approach O(n)
- Unbalanced page trees are slower

PDF Reference, Third Edition

3.6.2 Page Tree

The simplest structure would consist of a single page tree node that references all of the document's page objects directly; however, to optimize the performance of viewer applications, the Acrobat Distiller and PDF Writer programs construct trees of a particular form, known as balanced trees. Further information on this form of tree can be found in Data Structures and Algorithms, by Aho, Hopcroft, and Ullman (see the Bibliography).

http://www.nist.gov/dads/HTML/balancedbitr.html
balanced binary tree (data structure)
Definition: A binary tree where no leaf is more than a certain amount farther from the root than any other.

Appligent

Balanced Page Trees

A Kids Array will typically have six entries

```
2959 0 obj
<<
/Producer (Kenda DSSERVER IMG to PDF)
>>
endobj
3 0 obj
<<
/Type /Pages
/Count 985
/Kids [ 4 0 R 7 0 R 10 0 R 13 0 R 16 0 R 19 0 R 22 0 R 25 0
R 28 0 R 31 0 R 34 0 R 37 0 R 40 0 R 43 0 R 46 0 R 49 0 R 52
0 R 55 0 R 58 0 R 61 0 R 64 0 R ...
```


Appligent

Balanced Page Trees (continued)

```
1 0 obj
<<
/Title (ReportPrinter Report)
/Producer (Amyuni PDF Converter)
/Version (Version 1.51 - Developer Licence N∞ 44B65202-B23E)
/CreationDate (20/8/2001 11:5:22)
>>
3 0 obj
<<
/Type /Pages
/Count 2192
/Kids [6 0 R 9 0 R 12 0 R 15 0 R 18 0 R 21 0 R 24 0 R 27 0 R
30 0 R 33 0 R 36 0 R 39 0 R 42 0 R 45 0 R 48 0 R 51 0 R 54 0
R 57 0 R 60 0 R 63 0 R 66 0 R 69 0 R 72 0 R 75 0 R 78 0 R 81
0 R 84 0 R 87 0 R 90 0 R 93 0 R 96 0 R 99 0 R 102 0 R 105 0
R 108 0 R 111 0 R 114 0 R 117 0 R 120 0 R 123 0 R 126 0 R
129 0 R 132 0 R 135 0 R 138 0 R 141 0 R 144 0 R 147 0 R 150
0 R 153 0 R 156 0 R 159 0 R 162 0 R 165 0 R 168 0 R 171 0 R
174 0 R 177 0 R 180 0 R 183 0 R 186 0 R 189 0 R 192 0 R ...
```


PDF Pages are referenced using a binary tree mechanism. Unfortunately, not all PDF producers have read that part of the PDF reference.

Appligent

Required Items

When the PDF Reference specifies that an item is (Required); the item actually is Required

- Bookmarks require the following items:
  - Title
  - Parent - must be an indirect reference
  - Prev - for all but the First item at each level
  - Next - for all but the Last item at each level
  - First - if the item has descendants
  - Last - if the item has descendants
  - Count - if the item has descendants

Appligent

Required Items (continued)

P270 - IEE - Interrogatories
P270 - IEE - Part I
IEE - Part II
P270 - IEE - Part III
IEE - Part III
P270 - IEE - Overflow Page

```
52 0 obj
<<
/Count -1    <- missing /Parent
/First 53 0 R
/Last 53 0 R
/Prev 51 0 R
/Next 54 0 R
/Title (P270 - IEE - Part II)
/Dest [24 0 R /XYZ 0 594.96 0]
>>
endobj
53 0 obj
<<
/Title (IEE - Part II)
/Dest [27 0 R /XYZ 0 594.96 0]
>>
```



```
51 0 obj <- none of the bookmark objects have the required /Parent
<<
/Prev 50 0 R
/Next 52 0 R
/Title (P270 - IEE - Part I)
/Dest [21 0 R /XYZ 0 990.96 0]
>>
endobj
56 0 obj
<<
/Prev 54 0 R
%%ext 66 0 R      <- another quirk; there are only 57 objects in this
file
/Title (P270 - IEE - Overflow Page)
/Dest [36 0 R /XYZ 0 594.96 0]
>>
endobj
1 0 obj
<<
/Producer (Amyuni PDF Converter)
/Version (Version 1.58 - Developer Licence N∞ 09D80350-60BA)
/CreationDate (28/3/2001 14:13:26)
>>
endobj
```

Appligent

CosStrings

The object that consistently causes more problems then any other object type

- "A string is a sequence of characters, enclosed in parentheses."
  Well, not always
- (This is a String)
- A CosString can also be a sequence of hexadecimal data enclosed in <>
  <546869732069732061205374726696E67>

Appligent

Common CosString Problems

- Missing Line Continuation Character \
- Unbalanced Parentheses ()
- Missing Escape Sequences
  - \(, \), \\

Appligent

Linearization

There are Actually Degrees of Linearization

- A file which Acrobat thinks is linearized may not actually be linearized.
- Linearization is not fully documented in the PDF Reference Manual.
- Many Linearized PDF files are only linearized for the first page.
- Acrobat itself does not add or support secondary hint tables.

Appligent

Losing Form Field Lengths

Form File Issue

- MaxLen property is being ignored in 5.0
- /MaxLen key is in the Annot dictionary instead of the field dictionary
- Caused by a bug in 4.0x

PDF-Forms Email List

Subject: Acrobat 5.0 losing form field lengths
From: “Roberto”

Just to add to Max’s response: This is a known issue and is being fixed in the next dot release of Acrobat. The reason that the MaxLen property is being ignored in 5.0 is because the PDF is malformed. The /MaxLen key is in the Annot dictionary instead of the field dictionary. This is caused by a bug in 4.0x which occurs when the following steps are taken.

1. Create a form field.
2. Copy the field or ctrl-drag the field to create two fields of the same name.
3. Delete one of the fields
4. Set the character limit via the properties dialog.

Opening the file in 5.0 and resetting the character limit for the field will “repair” the PDF file, however, simply opening and saving the file in 5.0 will not.

Appligent

Questions & Answers

Appligent

www.appligent.com

Appligent, Inc.
60 South Lansdowne Avenue
Lansdowne, PA 19050
(610) 284-4006