



PDF Junction
Standard and Professional
Editions
Reference Guide Version 3.1

November 2011

© Copyright 2007-2011 Aquaforest Limited

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

CONTENTS

| | | |
|----------|---|-----------|
| 1 | INTRODUCTION | 3 |
| 1.1 | INSTALLATION | 3 |
| 1.2 | SYSTEM REQUIREMENTS | 3 |
| 1.3 | CONFIGURATION FOR CONVERSION TO PDF – SYSTEM REQUIREMENTS..... | 3 |
| 1.3.1 | <i>Native Application</i> | 3 |
| 1.3.2 | <i>Service Configuration</i> | 3 |
| 1.4 | LICENCING AND PURCHASING..... | 4 |
| 1.5 | CONVENTIONS | 4 |
| 2 | PDF JUNCTION WINDOWS INTERFACE..... | 5 |
| 2.1 | MAIN WINDOW..... | 5 |
| 2.1.1 | <i>Job Definition Section</i> | 5 |
| 2.1.2 | <i>Options Section</i> | 7 |
| 2.1.3 | <i>Split Options Section</i> | 7 |
| 2.1.4 | <i>PDF To TIFF Section</i> | 8 |
| 2.1.5 | <i>Extract Text From PDF Section</i> | 8 |
| 2.1.6 | <i>Task Log Section</i> | 8 |
| 2.2 | SECURITY SETTINGS | 9 |
| 2.2.1 | <i>Source Document Security</i> | 9 |
| 2.2.2 | <i>Target Document Security</i> | 9 |
| 2.3 | DOCUMENT METADATA | 10 |
| 2.4 | DOCUMENT DISPLAY OPTIONS | 10 |
| 2.4.1 | <i>Display Modes</i> | 11 |
| 2.4.2 | <i>Settings</i> | 11 |
| 2.5 | CUSTOM SETTINGS | 12 |
| 2.5.1 | <i>Custom Script and Log Files</i> | 12 |
| 2.5.2 | <i>Run Custom Script</i> | 12 |
| 3 | USING PDF JUNCTION FROM THE COMMAND LINE..... | 13 |
| 3.1 | COMMAND LINE SYNTAX | 13 |
| 3.2 | EXAMPLE COMMANDS..... | 14 |
| 3.2.1 | <i>Example jobid Commands</i> | 14 |
| 3.2.2 | <i>Example Split Commands</i> | 14 |
| 3.2.3 | <i>Example Merge Commands</i> | 15 |
| 3.2.4 | <i>Example Set Commands</i> | 15 |
| 3.2.5 | <i>Example Get Command</i> | 15 |
| 3.3 | JOBOPTION SPECIFICATIONS | 16 |
| 3.3.1 | <i>Job Options for Split Operations</i> | 16 |
| 3.3.2 | <i>Job Options for Merge Operations</i> | 17 |
| 3.3.3 | <i>Job Options for Set Operations</i> | 17 |
| 3.3.4 | <i>Job Options for Get Operations</i> | 18 |
| 3.4 | METADATA SPECIFICATION..... | 18 |
| 3.5 | SECURITY SPECIFICATION..... | 18 |
| 3.6 | DOCOPTIONS SPECIFICATION | 19 |
| 4 | CONVERSION TO PDF – GUI AND COMMAND LINES INTERFACES (PROFESSIONAL EDITION) | 20 |
| 4.1 | GUI INTERFACE | 20 |
| 4.2 | COMMAND LINE INTERFACE..... | 23 |
| 4.3 | CONFIGURATION FILE | 24 |
| 5 | PDF TO TIFF CONVERSION COMMAND LINE INTERFACE..... | 26 |
| 5.1 | COMMAND LINE INTERFACE..... | 26 |
| 6 | PDF TEXT EXTRACTION COMMAND LINE INTERFACE..... | 27 |

| | | |
|-----------|--|-----------|
| 6.1 | COMMAND LINE INTERFACE..... | 27 |
| 7 | WEB OPTIMIZATION (LINEARIZATION) COMMAND LINE INTERFACE | 27 |
| 7.1 | COMMAND LINE INTERFACE..... | 27 |
| 8 | SET XMP PROPERTIES (PROFESSIONAL EDITION ONLY)..... | 28 |
| 8.1 | SETTING METADATA | 28 |
| 8.2 | COMMAND LINE INTERFACE..... | 29 |
| 9 | XML JOB TICKET FILES | 30 |
| 10 | CUSTOM SCRIPTS..... | 31 |
| 11 | XML “INFORMATION” FILES..... | 32 |
| 12 | PDF JUNCTION FILES AND DIRECTORIES | 33 |
| 13 | SUPPORT | 33 |
| 14 | ACKNOWLEDGEMENTS | 33 |

1 INTRODUCTION

PDF Junction provides a set of processing tools for PDF files, including splitting, merging, setting of security and other attributes, and generation of XML metadata files. In addition, the Professional version of the product also supports conversion of Microsoft Office and other documents to PDF, and the setting of XMP properties.

The product has both a GUI and command-line interface. Batch processes can be defined using the GUI to create an XML Job Ticket file which can be run using the GUI, or using the command line. In addition, all options can be used from the command line without having used the GUI.

1.1 Installation

The installation MSI will guide you through the process of setting up PDF Junction.

1.2 System Requirements

- Windows XP, Windows 2003, 2008, Vista, Windows 7 (32/64 bit)
- Native Applications (such as Microsoft Office) are required for PDF Conversion.
- From version 3.1 PDF Junction may be configured to allow Open Office may be
- 50Mb of disk space is required.

1.3 Configuration for Conversion to PDF – System Requirements

The “Convert any file to PDF” process makes use of a print driver to generate the PDF which has the following pre-requisites to successfully perform a conversion :

1.3.1 Native Application

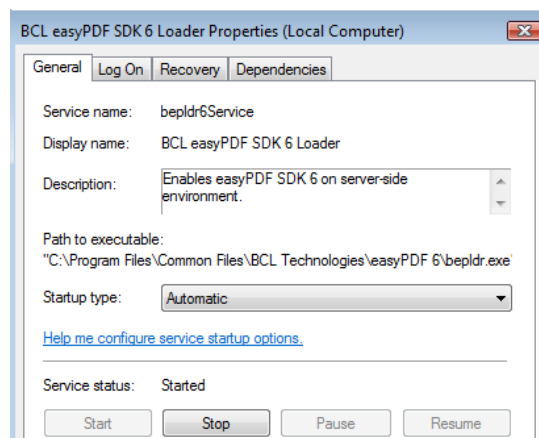
An application must be installed which is capable of printing the file type to be converted to PDF. Most commonly this will be applications or office suites such as Microsoft Office or Open Office. In addition, the application must be able to be automated as the print driver process occurs as a background automated task via a service.

Section 4.3 of the reference guide details how PDF Junction can be configured to make use of specific applications for certain file types – for example how to use Open Office to assist converting Microsoft Word documents and therefore eliminate the need for Microsoft Office to be installed.

1.3.2 Service Configuration

The conversion print driver is called via the BCL easyPDF SDK 6 Loader service which can be configured via Control Panel | Administrative Tools | Services.

As the conversion is automated it is important that the Log On user for the service has permission to access the required native applications without any pop-up dialog boxes. It may be convenient to configure a specific user – eg “bcluser” and configure the service to run as this user - having first logged in as the user and launch each of the relevant applications to ensure that there are no popup dialog boxes that would remain when accessed in batch mode.



1.4 Licencing and Purchasing

Unlicensed versions are fully functional and not time-restricted, but all generated PDF files have an alert message which indicates that the document was generated with an unlicensed versions of the product.

When a license is purchased (see <http://www.aquaforest.com/en/buy001.asp>) a permanent license key will be issued for use with the product which should be entered using the Help | License option of the GUI. There is no need to download another version of the software.

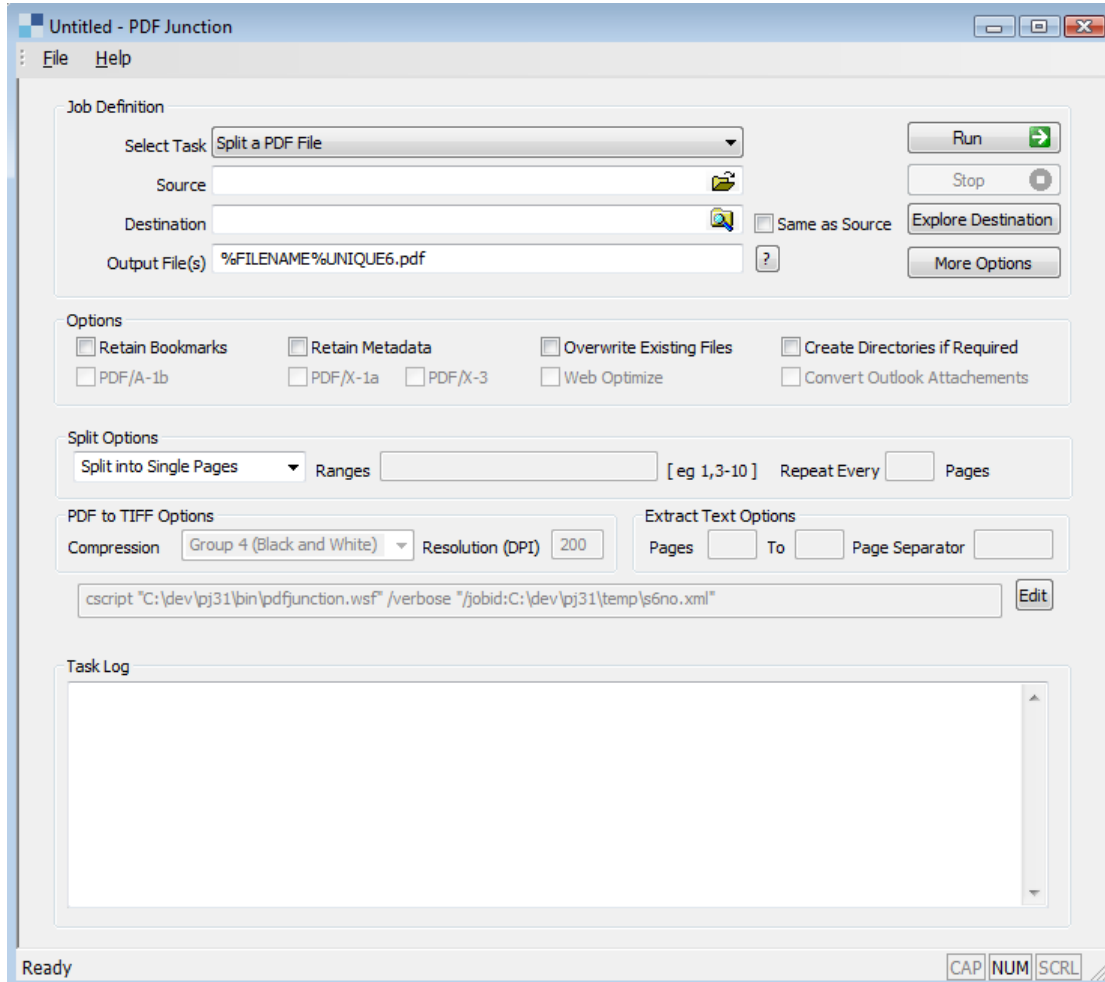
1.5 Conventions

Throughout this guide type in **bold courier** is used to represent operating system commands and *PDFJUNCTIONDIR* represents the directory in which PDF Junction was installed, by default C:\Aquaforest\PDF Junction.

2 PDF JUNCTION WINDOWS INTERFACE

2.1 Main Window

When PDF Junction is launched, the main window is displayed as shown below. This allows the definition and running of jobs, as well as saving and loading job definitions to and from files.



2.1.1 Job Definition Section

| Screen Field / Button | Description |
|-----------------------|---|
| Run | Runs the current job. Job output will appear in the "task log section of the screen". |
| Stop | Halts processing of the currently running job. |
| Explore Destination | Launches Windows Explorer using the destination as the starting point. |
| More Options | This launches a set of property sheets which can be used to set properties of the generated PDF files, such as security and metadata values. See sections 2.2 through 2.5 for more details. |

| Screen Field / Button | Description | | | | |
|-----------------------|---|-----------|----------------|----------------|--|
| Select Task | <p>This defines which task is to be run and may be one of :</p> <p>Split a PDF file Split a Folder of PDF Files Split a Folder Tree of PDF Files</p> <p>Merge a Folder of PDF Files Merge each Folder in Folder Tree</p> <p>Set Properties of a PDF File Set Properties of a Folder of PDF Files Set Properties of a Folder Tree of PDF Files</p> <p>Create Properties XML File of a PDF File Create Properties XML File for each PDF file in a Folder Create Properties XML File for each PDF file in a Folder Tree</p> <p>Convert a PDF File to TIFF Convert a Folder of PDF Files to TIFF Convert a Folder Tree of PDF Files to TIFF</p> <p>Extract Text from a PDF File Extract Text from each PDF File in a Folder Extract Text from each PDF File in a Folder Tree</p> <p>Convert a File to PDF Convert a Folder of Files to PDF Convert a Folder Tree of Files to PDF</p> <p>Set XMP Properties of a PDF File* Set XMP Properties for each PDF file in a Folder* Set XMP Properties for each PDF file in a Folder Tree*</p> <p>Web Optimize a PDF File Web Optimize a Folder of PDF Files Web Optimize a Folder Tree of PDF Files</p> <p>*See section 7 for further details</p> | | | | |
| Source | The source PDF file or directory. | | | | |
| Destination | The location where the generated PDF file(s) will be placed. This can be set to be the same as the source location. Note that PDF Junction cannot overwrite source files, a new file is always made. | | | | |
| Output File(s) | <p>This defines the template output file name.</p> <table border="1" data-bbox="671 1794 1342 2009"> <tbody> <tr> <td data-bbox="671 1794 895 1854">%BOOKMARK</td> <td data-bbox="895 1794 1342 1854">Bookmark title</td> </tr> <tr> <td data-bbox="671 1854 895 2009">%PAGE<i>n</i></td> <td data-bbox="895 1854 1342 2009">The first page number from source file, zero padded to <i>n</i> digits. If <i>n</i> is not supplied or is zero, then zero padding is not used.</td> </tr> </tbody> </table> | %BOOKMARK | Bookmark title | %PAGE <i>n</i> | The first page number from source file, zero padded to <i>n</i> digits. If <i>n</i> is not supplied or is zero, then zero padding is not used. |
| %BOOKMARK | Bookmark title | | | | |
| %PAGE <i>n</i> | The first page number from source file, zero padded to <i>n</i> digits. If <i>n</i> is not supplied or is zero, then zero padding is not used. | | | | |

| | | |
|--|---------------|---|
| | %FILENAME | Source file name without .pdf |
| | %UNIQUE n | Unique number starting at 1. If n is supplied, then zero padding to n digits is used. If n is not supplied or is zero, then zero padding is not used. |
| | %DIRNAME | Source directory name of the directory currently being processed. |
| | <i>String</i> | (Any string) |

The .pdf extension is automatically added if not specified.

For example, a split operation using %FILENAME-part%UNIQUE6% which would give *filename.pdf* split into *filename-part000001.pdf filename-part000002.pdf* etc.

2.1.2 Options Section

| Screen Field / Button | Description |
|-----------------------------------|--|
| Retain Bookmarks | If checked, generated files will include bookmarks from the original file. |
| Retain Metadata | If checked, generated files will include metadata (such as Author and Title) from the original file. |
| Overwrite Existing Files | If checked, generated PDF files will overwrite existing files of the same name (but not source files) |
| Create Directories where required | If checked, when processing a tree, subdirectories will be created in the target directory tree when required. |
| PDF/A-1b | The generated PDF file will be PDF/A-1b compliant. |
| PDF/X-1a | The generated PDF file will be PDF/X-1a compliant. |
| PDF/X-3 | The generated PDF file will be PDF/X-3 compliant. |
| Web Optimize | The generated PDF file will be Web Optimized (Linearized) |
| Convert Outlook Attachments | If selected (when converting .msg files) the generated PDF file will comprise the email plus each page of each attachment. |

2.1.3 Split Options Section

| Screen Field / Button | Description | | | | | | | | |
|-----------------------|---|---|-------------------------|-----|----------------------------|-----------|---------------------------------------|----------|--------------------------------------|
| Split Type | <i>One of:</i> Split into single pages Split by top-level bookmark Split by ranges (See below) Split by repeating ranges (See below) | | | | | | | | |
| Page Ranges | Set of page ranges separated by commas that defines which pages from the original should be extracted. The following types of page ranges are allowed : <table border="1" data-bbox="662 1881 1348 2016"> <tr> <td>1</td> <td>Specifies a single page</td> </tr> <tr> <td>1-3</td> <td>Specifies a range of pages</td> </tr> <tr> <td>10-20even</td> <td>Specifies even pages within the range</td> </tr> <tr> <td>10-20odd</td> <td>Specifies odd pages within the range</td> </tr> </table> | 1 | Specifies a single page | 1-3 | Specifies a range of pages | 10-20even | Specifies even pages within the range | 10-20odd | Specifies odd pages within the range |
| 1 | Specifies a single page | | | | | | | | |
| 1-3 | Specifies a range of pages | | | | | | | | |
| 10-20even | Specifies even pages within the range | | | | | | | | |
| 10-20odd | Specifies odd pages within the range | | | | | | | | |

| | | |
|-----------------|---|---|
| | 20-lastpage | Note that <i>lastpage</i> can be used to represent the last page in the document. |
| Repeating Range | <p>Apply the page range to each set of <i>Page Ranges</i> pages within the document. For example if 2-4 is specified for page ranges, and 4 is specified as the repeating range, then the range is re-applied every 4 pages.</p> <p>Hence the file is split such that the first output file contains pages 2-4 from the original document, the second contains pages 6-8 and so on.</p> | |

2.1.4 PDF To TIFF Section

| Screen Field / Button | Description |
|-----------------------|--|
| Compression | This specifies the compression that will be used in the TIFF version of the PDF file. The choices are Group 4 (black and white only) or LZW (which also supports color). The default is Group 4. |
| Resolution | This specifies the resolution in DPI of the result TIFF file. The default is 200. A higher figure of 300 can give better results in some cases, at the cost of a larger file size. |

2.1.5 Extract Text From PDF Section

| Screen Field / Button | Description |
|-----------------------|--|
| From, To | This defines the range of pages that will have text extracted. By default all pages are processed. |
| Page Sep | This allows the definition of an optional page separator string in the output text file. |

2.1.6 Task Log Section

| Screen Field / Button | Description |
|-----------------------|--|
| Command | This is the operating system command that is used to run the job. It can be edited after using the "Edit" button, although there is normally no need for this. |
| Task Output | The output of the job appears here. |

2.2 Security Settings

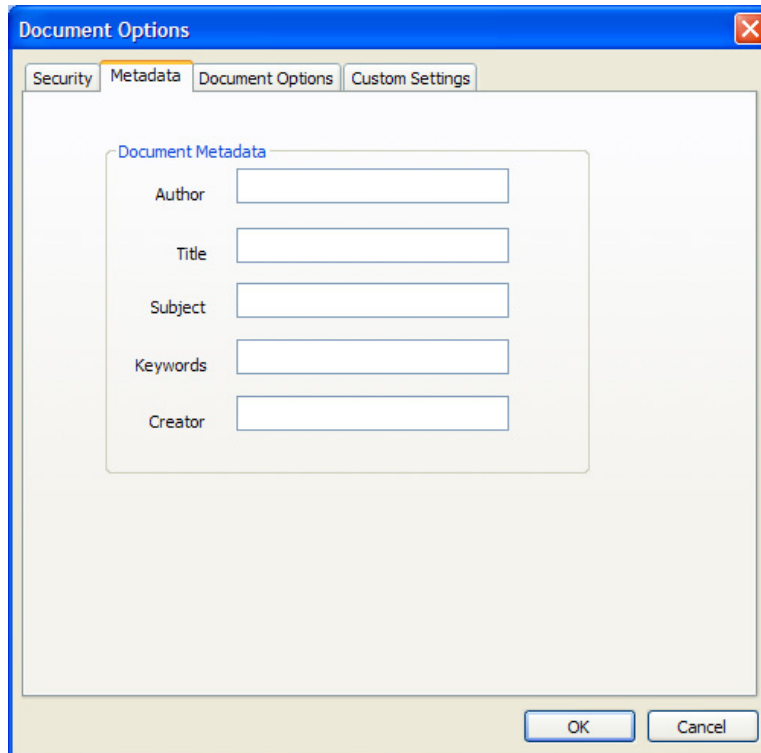
2.2.1 Source Document Security

| Screen Field / Button | Description |
|-----------------------|---|
| Password | If the source document requires a password it should be entered here. |

2.2.2 Target Document Security

| Option | Description | | | | | | | | | | | | | | | | | | | | |
|-------------------|--|--|-------------|-----------------------|--------------------------------|----------|-----------------------------|-----------------|---|------|---|--------------------|-----------------------------------|--------|------------------------------|----------------|--|----------|---|-------------------|----------------------------|
| User Password | A password that will be required to open the document. | | | | | | | | | | | | | | | | | | | | |
| Owner Password | A password that will be required to change the document permissions. | | | | | | | | | | | | | | | | | | | | |
| Permissions | <table border="1"> <thead> <tr> <th>Permission</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>Allow ALL Permissions</td> <td>All all the permissions below.</td> </tr> <tr> <td>Printing</td> <td>Allow high-quality printing</td> </tr> <tr> <td>Modify Contents</td> <td>Allow assembly (see below) and other document medications</td> </tr> <tr> <td>Copy</td> <td>Allow text and graphic copying and extraction</td> </tr> <tr> <td>Modify Annotations</td> <td>Allow modification of annotations</td> </tr> <tr> <td>Fillin</td> <td>Allow filling of form fields</td> </tr> <tr> <td>Screen Readers</td> <td>Allow extraction of text and graphics in support of accessibility.</td> </tr> <tr> <td>Assembly</td> <td>Allow rotation, insertion or deletion of pages.</td> </tr> <tr> <td>Degraded Printing</td> <td>Allow low-quality printing</td> </tr> </tbody> </table> | Permission | Description | Allow ALL Permissions | All all the permissions below. | Printing | Allow high-quality printing | Modify Contents | Allow assembly (see below) and other document medications | Copy | Allow text and graphic copying and extraction | Modify Annotations | Allow modification of annotations | Fillin | Allow filling of form fields | Screen Readers | Allow extraction of text and graphics in support of accessibility. | Assembly | Allow rotation, insertion or deletion of pages. | Degraded Printing | Allow low-quality printing |
| | Permission | Description | | | | | | | | | | | | | | | | | | | |
| | Allow ALL Permissions | All all the permissions below. | | | | | | | | | | | | | | | | | | | |
| | Printing | Allow high-quality printing | | | | | | | | | | | | | | | | | | | |
| | Modify Contents | Allow assembly (see below) and other document medications | | | | | | | | | | | | | | | | | | | |
| | Copy | Allow text and graphic copying and extraction | | | | | | | | | | | | | | | | | | | |
| | Modify Annotations | Allow modification of annotations | | | | | | | | | | | | | | | | | | | |
| | Fillin | Allow filling of form fields | | | | | | | | | | | | | | | | | | | |
| | Screen Readers | Allow extraction of text and graphics in support of accessibility. | | | | | | | | | | | | | | | | | | | |
| Assembly | Allow rotation, insertion or deletion of pages. | | | | | | | | | | | | | | | | | | | | |
| Degraded Printing | Allow low-quality printing | | | | | | | | | | | | | | | | | | | | |

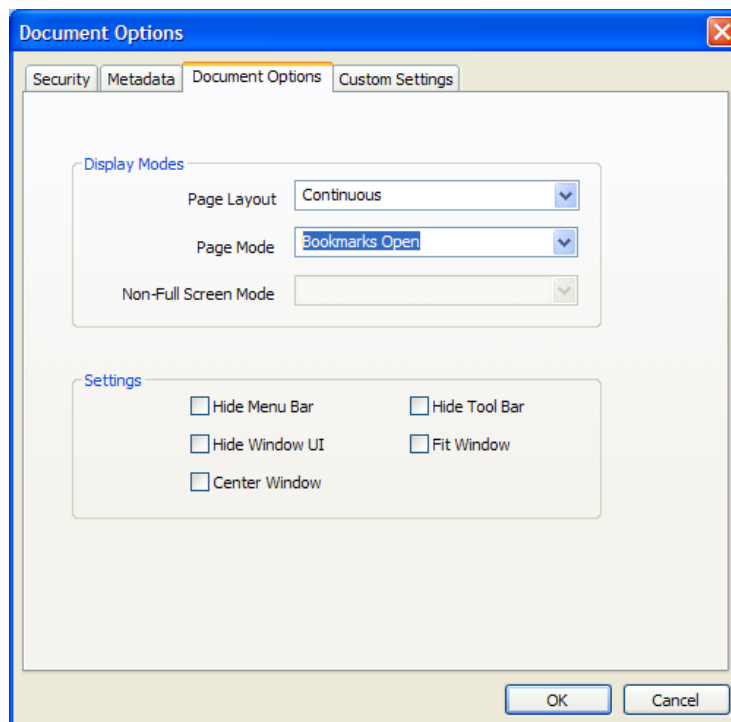
2.3 Document Metadata



The screenshot shows the 'Document Options' dialog box with the 'Metadata' tab selected. The 'Document Metadata' section contains five text input fields: Author, Title, Subject, Keywords, and Creator. The 'OK' and 'Cancel' buttons are located at the bottom right of the dialog.

This page allows setting of the generated PDF Document Properties. Custom and XMP metadata may also be set – see section 7 for further details.

2.4 Document Display Options



The screenshot shows the 'Document Options' dialog box with the 'Document Options' tab selected. The 'Display Modes' section contains three dropdown menus: Page Layout (set to 'Continuous'), Page Mode (set to 'Bookmarks Open'), and Non-Full Screen Mode. The 'Settings' section contains five checkboxes: Hide Menu Bar, Hide Tool Bar, Hide Window UI, Fit Window, and Center Window. The 'OK' and 'Cancel' buttons are located at the bottom right of the dialog.

2.4.1 Display Modes

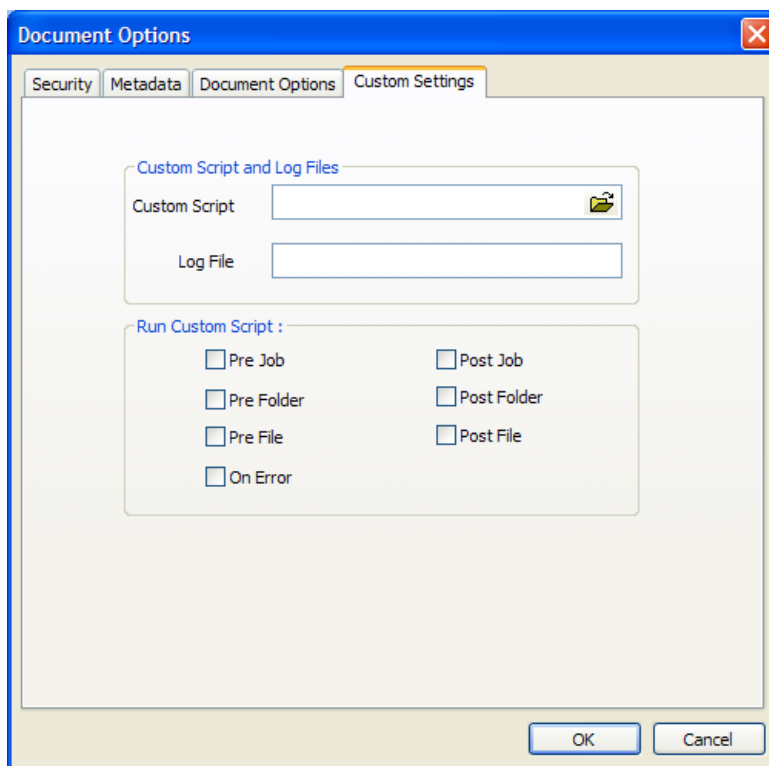
| Option | Description |
|----------------------|---|
| Page Layout | The setting for initial document page display. One of : <ul style="list-style-type: none">• Single Page• Continuous• Continuous Facing (odd pages left)• Continuous Facing (odd pages right) |
| Page Mode | The setting for initial viewer mode. One of : <ul style="list-style-type: none">• Neither Bookmarks nor Thumbnails Open• Bookmarks Open• Thumbnails Open• Bookmarks & Thumbnails Open• Full Screen |
| Non-Full Screen Mode | Only applicable where Page Mode=Full Screen. The setting for document page display when exiting Full Screen mode. One of : <ul style="list-style-type: none">• Neither Bookmarks nor Thumbnails Open• Bookmarks Open• Thumbnails Open |

2.4.2 Settings

| Option | Description |
|----------------|---|
| Hide Tool Bar | The viewer's tool bar will be hidden |
| Hide Menu Bar | The viewer's menu bar will be hidden |
| Hide Window UI | The viewer's UI elements (scrollbars etc) will be hidden |
| Fit Window | The viewer will resize the document's window to fit the size of the first displayed page. |
| Center Window | The document window will be positioned in the center of the screen. |

2.5 Custom Settings

See section 5 for full details of running Custom Scripts.



2.5.1 Custom Script and Log Files

| Option | Description |
|---------------|--|
| Custom Script | The path of a custom script to be used with the Job. |
| Log File | The name of a log file (which will be placed in <i>PDFJUNCTIONDIR</i> \logs). The log file will contain the same output as is written to the output panel. |

2.5.2 Run Custom Script

| Option | Description |
|-------------|---|
| Pre Job | If checked, the custom script will be called at the start of the job |
| Post Job | If checked, the custom script will be called at the end of the job |
| Pre Folder | If checked, the custom script will be called at the start of processing each folder |
| Post Folder | If checked, the custom script will be called at the end of processing each folder |
| Pre File | If checked, the custom script will be called at the start of processing each file |
| Post File | If checked, the custom script will be called at the end of processing each file |
| On Error | If checked, the custom script will be called when a processing error occurs. |

3 USING PDF JUNCTION FROM THE COMMAND LINE

3.1 Command Line Syntax

All operations supported by PDF Junction can be accessed via the command line as covered in this section and sections 4, 5 and 6.

Usage:

```
cscript pdfjunction.wsf [/jobid:value] /operation:value [/sourcetype:value] [/target:value]
[/joboptions:value] [/metadata:value] [/security:value] [/docoptions:value] sourcefiles1
[sourcefiles2...] [/version[+|-]] [/verbose[+|-]] [/debug[+|-]] [/onerrorcontinue[+|-]] [/silent[+|-]]
```

| Parameter | Notes |
|-------------------|---|
| /jobid:value | If used, specifies the full path of an XML Job Ticket file saved from the GUI or generated by some other means. Job files in the PDF Junction jobs directory may be referred to by a shorthand method of <i>jobname</i> where the Job Ticket file is <i>jobname.xml</i> . |
| /operation:value | Must be split, merge, set or get |
| /sourcetype:value | Source type (file, folder, tree). The default value is file. |
| /target:value | Target directory for result files. If not specified, the source directory will be used as the target. |
| /joboptions:value | Set of options for the specified operation. Described in detail in section 3.3. |
| /metadata:value | Metadata settings for result PDF files. By default metadata fields are left blank. Described in detail in section 3.4. |
| /security:value | Security settings for result PDF files. By default files are not secured. Described in detail in section 3.5. |
| /docoptions:value | Document option settings for result PDF files. Described in detail in section 3.6. |
| Sourcefiles | Source PDF file, folder or tree. May be multiple files for merge operations. |
| /version | Provide version and license information |
| /verbose | Verbose progress messages |
| /debug | Debug progress messages |
| /silent | No progress messages |
| /onerrorcontinue | Allow job to continue after errors |

Notes

1. cscript.exe is required to run the command. This is usually included in a standard PATH, but if not would need to be specified explicitly

Eg C:\WINDOWS\SYSTEM32\cscript.exe ...

2. Options that contain spaces will need to be quoted. For example `"/target:c:\output files"`

3.2 Example Commands

3.2.1 Example jobid Commands

Run the job specified in the job file C:\jobs\job001.xml :

```
cscript pdfjunction.wsf /jobid:C:\Jobs\job001.xml
```

Run the job specified in the job file C:\Program Files\PDF Junction\jobs\job001.xml :

```
cscript pdfjunction.wsf /jobid:job001
```

3.2.2 Example Split Commands

Split a PDF file, c:\demo\in\demo.pdf into single page PDF files placed in the directory c:\demo\out

```
cscript pdfjunction.wsf /operation:split /joboptions:-s  
/target:c:\demo\out c:\demo\in\demo.pdf
```

Split a PDF file, c:\demo\in\demo.pdf into single pages, each named demo_split999.pdf where demo is the page number. The files will be placed in the directory c:\demo\out

```
cscript pdfjunction.wsf /operation:split "/joboptions:-s -f  
%F%_split%U3%" /target:c:\demo\out c:\demo\in\demo.pdf
```

Split a PDF file, c:\demo\in\demo.pdf by bookmarks. The files will be placed in the directory c:\demo\out and named according to the bookmark.

```
cscript pdfjunction.wsf /operation:split "/joboptions:-k -b -f %B%"  
/target:c:\demo\out c:\demo\in\demo.pdf
```

Split all PDF files in c:\demo\in into single page PDF files placed in the directory c:\demo\out\name where name is the name of the original pdf file, creating the directories where they don't exist.

```
cscript pdfjunction.wsf /operation:split /sourcetype:folder  
"/joboptions:-c -s -f %F%/U%" /target:c:\demo\out c:\demo\in
```

Split all PDF files in c:\demo\in and each subdirectory into single page PDF files placed in the directory c:\demo\out\sub\name where sub is the subdirectory name, and name is the name of the original pdf file, creating the directories where they don't exist.

```
cscript pdfjunction.wsf /operation:split /sourcetype:tree  
"/joboptions:-c -s -f %F%/U%" /target:c:\demo\out c:\demo\in
```

3.2.3 Example Merge Commands

Merge PDF files demo1.pdf, demo2.pdf, demo3.pdf in c:\demo\in\ into a single PDF file named demomerge.pdf in the directory c:\demo\out

```
cscript pdfjunction.wsf /operation:merge /target:c:\demo\out
"/joboptions:-f demomerge
c:\demo\in\demo1.pdf*c:\demo\in\demo2.pdf*c:\demo\in\demo3.pdf
```

Merge all the PDF files in c:\demo\in\ into a single PDF file named demomerge.pdf in the directory c:\demo\out, overwriting the file if it already exists.

```
cscript pdfjunction.wsf /operation:merge /sourcetype:folder
/target:c:\demo\out "/joboptions:-o -f demomerge" c:\demo\in
```

Merge all the PDF files in c:\demo\in\ and each subdirectory into a single PDF file per source directory. The result file will be named dir.pdf where dir is the source directory name, and placed in the corresponding directory in c:\demo\out\

```
cscript pdfjunction.wsf /operation:merge /sourcetype:tree
/target:c:\demo\out "/joboptions:-c -o -f %D%" c:\demo\in
```

3.2.4 Example Set Commands

Note that the set parameters (/metadata ..) can be used with the split and merge commands to set characteristics of the result files.

Creates a copy of c:\demo\in\demo3.pdf with the metadata settings shown, in the directory c:\demo\out

```
cscript pdfjunction.wsf /operation:set /target:c:\demo\out
c:\demo\in\demo3.pdf "/metadata:Author=Tony Bush;Title=A Special
Document"
```

Creates a copy of each PDF file in c:\demo\in\ with a user password of "secret", with printing and modification of annotations allowed. The result PDF files are placed in in the directory c:\demo\out.

```
cscript pdfjunction.wsf /operation:set /sourcetype:folder /joboptions:-o
/target:c:\demo\out c:\demo\in "/security:userpassword=Secret;
permissions=allowprinting,allowmodifyannotations"
```

Creates a copy of each PDF file in c:\demo\in\ with viewer preferences set to hide toolbar and to open the document showing bookmarks. The result PDF files are placed in in the directory c:\demo\out.

```
cscript pdfjunction.wsf /operation:set /sourcetype:folder /joboptions:-o
/target:c:\demo\out c:\demo\in "/docoptions:hidetoolbar,pagemodeuseoutlines"
```

3.2.5 Example Get Command

Generates an XML file with metadata from c:\demo\in\demo3.pdf with the metadata settings shown, in the directory c:\demo\out

```
cscript pdfjunction.wsf /operation:get /target:c:\demo\out
c:\demo\in\demo3.pdf "/joboptions:-f %F%_metadata"
```


3.3 /joboption specifications

3.3.1 Job Options for Split Operations

| Parameter | Notes | | | | | | | | | | | | |
|----------------------|--|-----------|-------------------------|-----------|---|-----------|---------------------------------------|-------------|--|-------------|---|---------------|--------------|
| -b | Include bookmarks | | | | | | | | | | | | |
| -p | Password for opening document | | | | | | | | | | | | |
| -m | Retain metadata | | | | | | | | | | | | |
| -c | Create result directories if required | | | | | | | | | | | | |
| -o | Overwrite result files if they already exist | | | | | | | | | | | | |
| -r range{,range} | Set of page ranges separated by commas that defines which pages from the original should be extracted. The following types of page ranges are allowed : <table border="1" data-bbox="501 712 1353 909"> <tbody> <tr> <td>1</td> <td>Specifies a single page</td> </tr> <tr> <td>1-3</td> <td>Specifies a range of pages</td> </tr> <tr> <td>10-20even</td> <td>Specifies even pages within the range</td> </tr> <tr> <td>10-20odd</td> <td>Specifies odd pages within the range</td> </tr> <tr> <td>20-lastpage</td> <td>Note that <i>lastpage</i> can be used to represent the last page in the document.</td> </tr> </tbody> </table> | 1 | Specifies a single page | 1-3 | Specifies a range of pages | 10-20even | Specifies even pages within the range | 10-20odd | Specifies odd pages within the range | 20-lastpage | Note that <i>lastpage</i> can be used to represent the last page in the document. | | |
| 1 | Specifies a single page | | | | | | | | | | | | |
| 1-3 | Specifies a range of pages | | | | | | | | | | | | |
| 10-20even | Specifies even pages within the range | | | | | | | | | | | | |
| 10-20odd | Specifies odd pages within the range | | | | | | | | | | | | |
| 20-lastpage | Note that <i>lastpage</i> can be used to represent the last page in the document. | | | | | | | | | | | | |
| -g <i>grouppages</i> | Apply the page range to each set of <i>grouppages</i> pages within the document. For example if -r 2-4 and -e 5 is specified, then the range is re-applied every 4 pages. Hence the file is split such that the first output file contains pages 2-4 from the original document, the second contains pages 6-8 and so on. | | | | | | | | | | | | |
| -k | Split by top-level bookmark | | | | | | | | | | | | |
| -s | Split by single pages (The Default) | | | | | | | | | | | | |
| -f string | Output file name specifier. If -f is not specified, the default parameter setting is %FILENAME%UNIQUE6% which would give (for example) <i>filename.pdf</i> split into <i>filename000001.pdf filename000002.pdf</i> etc. <table border="1" data-bbox="501 1290 1353 1832"> <tbody> <tr> <td>%BOOKMARK</td> <td>Bookmark title</td> </tr> <tr> <td>%PAGEn</td> <td>The first page number from source file, zero padded to n digits. If n is not supplied or is zero, then zero padding is not used</td> </tr> <tr> <td>%FILENAME</td> <td>Ssource file name without .pdf</td> </tr> <tr> <td>%UNIQUEn</td> <td>Unique number starting at 1. If n is supplied, then zero padding to n digits is used. If n is not supplied or is zero, then zero padding is not used</td> </tr> <tr> <td>%DIRNAME</td> <td>Source directory name of the directory currently being processed.</td> </tr> <tr> <td><i>String</i></td> <td>(Any string)</td> </tr> </tbody> </table> <p>The .pdf extension is automatically added if not specified.</p> | %BOOKMARK | Bookmark title | %PAGE n | The first page number from source file, zero padded to n digits. If n is not supplied or is zero, then zero padding is not used | %FILENAME | Ssource file name without .pdf | %UNIQUE n | Unique number starting at 1. If n is supplied, then zero padding to n digits is used. If n is not supplied or is zero, then zero padding is not used | %DIRNAME | Source directory name of the directory currently being processed. | <i>String</i> | (Any string) |
| %BOOKMARK | Bookmark title | | | | | | | | | | | | |
| %PAGE n | The first page number from source file, zero padded to n digits. If n is not supplied or is zero, then zero padding is not used | | | | | | | | | | | | |
| %FILENAME | Ssource file name without .pdf | | | | | | | | | | | | |
| %UNIQUE n | Unique number starting at 1. If n is supplied, then zero padding to n digits is used. If n is not supplied or is zero, then zero padding is not used | | | | | | | | | | | | |
| %DIRNAME | Source directory name of the directory currently being processed. | | | | | | | | | | | | |
| <i>String</i> | (Any string) | | | | | | | | | | | | |

3.3.2 Job Options for Merge Operations

| Parameter | Notes | | | | |
|---------------|---|----------|---|---------------|--------------|
| -b | Include bookmarks | | | | |
| -p | Password for document reading | | | | |
| -m | Retain metadata (taken from the first file in the merge list). | | | | |
| -c | Create result directories if required | | | | |
| -o | Overwrite result files if they already exist | | | | |
| -f string | Output file name specifier. If -f is not specified, the merged file will be called mergeoutput.pdf. <table border="1" data-bbox="501 712 1353 875"> <tbody> <tr> <td>%DIRNAME</td> <td>Source directory name of the directory currently being processed.</td> </tr> <tr> <td><i>String</i></td> <td>(Any string)</td> </tr> </tbody> </table> <p>The .pdf extension is automatically added if not specified.</p> | %DIRNAME | Source directory name of the directory currently being processed. | <i>String</i> | (Any string) |
| %DIRNAME | Source directory name of the directory currently being processed. | | | | |
| <i>String</i> | (Any string) | | | | |

3.3.3 Job Options for Set Operations

| Parameter | Notes | | | | | | |
|---------------|--|-----------|-------------------------------|----------|---|---------------|--------------|
| -c | Create result directories if required | | | | | | |
| -o | Overwrite result files if they already exist | | | | | | |
| -f string | Output file name specifier. <table border="1" data-bbox="501 1357 1353 1585"> <tbody> <tr> <td>%FILENAME</td> <td>Source file name without .pdf</td> </tr> <tr> <td>%DIRNAME</td> <td>Source directory name of the directory currently being processed.</td> </tr> <tr> <td><i>String</i></td> <td>(Any string)</td> </tr> </tbody> </table> <p>The .pdf extension is automatically added if not specified.</p> | %FILENAME | Source file name without .pdf | %DIRNAME | Source directory name of the directory currently being processed. | <i>String</i> | (Any string) |
| %FILENAME | Source file name without .pdf | | | | | | |
| %DIRNAME | Source directory name of the directory currently being processed. | | | | | | |
| <i>String</i> | (Any string) | | | | | | |
| -p | Password for document reading | | | | | | |

3.3.4 Job Options for Get Operations

| Parameter | Notes | | | | | | | | | | | | |
|---------------|---|-----------|----------------|-----------|--|-----------|--------------------------------|-------------|--|----------|---|---------------|--------------|
| -c | Create result directories if required | | | | | | | | | | | | |
| -o | Overwrite result files if they already exist | | | | | | | | | | | | |
| -f string | Output file name specifier. <table border="1" data-bbox="501 580 1351 1122"> <tr> <td>%BOOKMARK</td> <td>Bookmark title</td> </tr> <tr> <td>%PAGEn</td> <td>The first page number from source file, zero padded to n digits. If n is not supplied or is zero, then zero padding is not used.</td> </tr> <tr> <td>%FILENAME</td> <td>Ssource file name without .pdf</td> </tr> <tr> <td>%UNIQUEn</td> <td>Unique number starting at 1. If n is supplied, then zero padding to n digits is used. If n is not supplied or is zero, then zero padding is not used</td> </tr> <tr> <td>%DIRNAME</td> <td>Source directory name of the directory currently being processed.</td> </tr> <tr> <td><i>String</i></td> <td>(Any string)</td> </tr> </table> <p>The .xml extension is automatically added</p> | %BOOKMARK | Bookmark title | %PAGE n | The first page number from source file, zero padded to n digits. If n is not supplied or is zero, then zero padding is not used. | %FILENAME | Ssource file name without .pdf | %UNIQUE n | Unique number starting at 1. If n is supplied, then zero padding to n digits is used. If n is not supplied or is zero, then zero padding is not used | %DIRNAME | Source directory name of the directory currently being processed. | <i>String</i> | (Any string) |
| %BOOKMARK | Bookmark title | | | | | | | | | | | | |
| %PAGE n | The first page number from source file, zero padded to n digits. If n is not supplied or is zero, then zero padding is not used. | | | | | | | | | | | | |
| %FILENAME | Ssource file name without .pdf | | | | | | | | | | | | |
| %UNIQUE n | Unique number starting at 1. If n is supplied, then zero padding to n digits is used. If n is not supplied or is zero, then zero padding is not used | | | | | | | | | | | | |
| %DIRNAME | Source directory name of the directory currently being processed. | | | | | | | | | | | | |
| <i>String</i> | (Any string) | | | | | | | | | | | | |
| -p | Password for document reading | | | | | | | | | | | | |

3.4 /metadata specification

This may be used to specify metadata properties as in the following example :

```
cscript pdfjunction.wsf /operation:set /target:c:\demo\out
c:\demo\in\demo3.pdf "/metadata:Author=Tony Bush;Title=A Special
Document"
```

The attributes that may be set are :

- Author
- Title
- Subject
- Keywords
- Creator

3.5 /security specification

This may be used to specify security properties as in the following example :

```
cscript pdfjunction.wsf /operation:set /sourcetype:folder /joboptions:-o
/target:c:\demo\out c:\demo\in "/security:userpassword=Secret
;permissions=allowprinting,allowmodifyannotations"
```

| Parameter | Description | | | | | | | | | | | | | | | | | | | | |
|------------------------|--|------------|-------------|---------------|-----------------------------|---------------------|---|-----------|---|------------------------|-----------------------------------|-------------|------------------------------|--------------------|--|---------------|---|-----------------------|----------------------------|----------|---------------------------|
| Userpassword | A password that will be required to open the document. | | | | | | | | | | | | | | | | | | | | |
| Ownerpassword | A password that will be required to change the document permissions. | | | | | | | | | | | | | | | | | | | | |
| Permissions | A comma separated list of document permissions <table border="1" data-bbox="501 450 1353 837"> <thead> <tr> <th>Permission</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>allowprinting</td> <td>Allow high-quality printing</td> </tr> <tr> <td>allowmodifycontents</td> <td>Allow assembly (see below) and other document medications</td> </tr> <tr> <td>allowcopy</td> <td>Allow text and graphic copying and extraction</td> </tr> <tr> <td>allowmodifyannotations</td> <td>Allow modification of annotations</td> </tr> <tr> <td>allowfillin</td> <td>Allow filling of form fields</td> </tr> <tr> <td>allowscreenreaders</td> <td>Allow extraction of text and graphics in support of accessibility.</td> </tr> <tr> <td>allowassembly</td> <td>Allow rotation, insertion or deletion of pages.</td> </tr> <tr> <td>allowdegradedprinting</td> <td>Allow low-quality printing</td> </tr> <tr> <td>allowall</td> <td>Allow all the permissions</td> </tr> </tbody> </table> | Permission | Description | allowprinting | Allow high-quality printing | allowmodifycontents | Allow assembly (see below) and other document medications | allowcopy | Allow text and graphic copying and extraction | allowmodifyannotations | Allow modification of annotations | allowfillin | Allow filling of form fields | allowscreenreaders | Allow extraction of text and graphics in support of accessibility. | allowassembly | Allow rotation, insertion or deletion of pages. | allowdegradedprinting | Allow low-quality printing | allowall | Allow all the permissions |
| Permission | Description | | | | | | | | | | | | | | | | | | | | |
| allowprinting | Allow high-quality printing | | | | | | | | | | | | | | | | | | | | |
| allowmodifycontents | Allow assembly (see below) and other document medications | | | | | | | | | | | | | | | | | | | | |
| allowcopy | Allow text and graphic copying and extraction | | | | | | | | | | | | | | | | | | | | |
| allowmodifyannotations | Allow modification of annotations | | | | | | | | | | | | | | | | | | | | |
| allowfillin | Allow filling of form fields | | | | | | | | | | | | | | | | | | | | |
| allowscreenreaders | Allow extraction of text and graphics in support of accessibility. | | | | | | | | | | | | | | | | | | | | |
| allowassembly | Allow rotation, insertion or deletion of pages. | | | | | | | | | | | | | | | | | | | | |
| allowdegradedprinting | Allow low-quality printing | | | | | | | | | | | | | | | | | | | | |
| allowall | Allow all the permissions | | | | | | | | | | | | | | | | | | | | |
| strength | Unless this is set to some other value (such as 40), the default of 128 is used. | | | | | | | | | | | | | | | | | | | | |

3.6 /docoptions specification

This may be used to specify document viewing preferences as in the following example :

```
cscript pdfjunction.wsf /operation:set /sourcetype:folder /joboptions:-o /target:c:\demo\out c:\demo\in "/docoptions:hidetoolbar,pagemodeuseoutlines"
```

| Property | Description |
|----------------------------------|---|
| pagelayoutsinglepage | Single Page (SinglePage) |
| pagelayoutonecolumn | Continuous(OneColumn) |
| pagelayouttwocolumnleft | ContinuousFacing [odd pages left] (TwoColumnLeft) |
| pagelayouttwocolumnright | ContinuousFacing [odd pages right] (TwoColumnRight) |
| pagemodeusenone | No Thumbnails or Bookmarks Visible (UseNone) |
| pagemodeuseoutlines | Bookmarks Visible (Use Outlines) |
| pagemodeusethumbs | Thumbnails Visible (UseThumbs) |
| pagemodef fullscreen | Full Screen Mode (FullScreen) |
| hidetoolbar | The viewer's tool bar will be hidden |
| hidemenubar | The viewer's menu bar will be hidden |
| hidewindowui | The viewer's UI elements (scrollbars etc) will be hidden |
| fitwindow | The viewer will resize the document's window to fit the size of the first displayed page. |
| centerwindow | The document window will be positioned in the center of the screen. |
| nonfullscreenpagemodeusenone | On exiting full screen mode, neither outline or thumbnail images are shown. |
| nonfullscreenpagemodeusenoutline | On exiting full screen mode outlines are visible. |
| nonfullscreenpagemodeusethumbs | On exiting full screen mode thumbnails are visible. |

4 CONVERSION TO PDF – GUI AND COMMAND LINES INTERFACES (PROFESSIONAL EDITION)

The Professional Edition of PDF Junction includes facilities to convert Microsoft Office and other document types to PDF via GUI and command line interfaces.

4.1 GUI Interface

The standard PDF Junction interface is used, and one of the following Tasks selected

Convert a File to PDF

Convert a Folder of Files to PDF

Convert a Folder Tree of Files to PDF

The conversion process can be tailored according to the PDF Conversion Settings selected. This screen can be accessed via the “More Options” tab.

The screenshot shows the 'Document Options' dialog box with the 'PDF Conversion Settings' tab selected. The dialog is divided into several sections:

- Conversion Control:** Includes a text input for 'Conversion Timeout (Seconds)', a checkbox for 'Continue On Error', and a text input for 'Conversion Error Folder' with a folder selection icon.
- Microsoft Office Conversion Settings:** Includes checkboxes for 'Convert Bookmarks' and 'Convert Hyperlinks'. The 'Convert Bookmarks' section has a 'Depth' dropdown menu set to 'Heading 1'. Other options include 'Print All Sheets (Excel)', 'Print BG Color (IE)', and 'Print Scale % (Visio)'. There are also text inputs for 'Header (IE)' and 'Footer (IE)'.
- Image and Font Conversion Settings:** Includes a dropdown for 'Image Compression', a checkbox for 'Image Downsizing', a text input for 'Image JPEG Quality (5-100)', and a text input for 'Downsize Resolution (DPI)'. It also has a dropdown for 'Font Embedding', a checkbox for 'Embed Fonts as Type 0', and a dropdown for 'Font Substitution'.
- Page Settings and Margins:** Includes text inputs for 'Top', 'Bottom', 'Left', 'Right', 'Height', and 'Width' margins, and a dropdown for 'Paper Orientation'.

At the bottom right of the dialog are 'OK' and 'Cancel' buttons.

The conversion settings available are as follows :

| Option | Command Line / Config File Name | Description |
|-------------------------|---------------------------------|--|
| Conversion Timeout | FileConversionTimeout | Limits the amount of time in seconds that can be spent on a conversion. A value of zero means wait indefinitely. |
| Continue on Error | ContinueOnError | For multiple file processing, continue if an error occurs with one file. |
| Conversion Error Folder | ErrorFolder | Define the folder where documents are placed if their conversion raised an error. |
| | | |
| Convert Bookmarks | Convertbookmarks | For MS Word, convert bookmarks |
| Depth | BookmarkDepth | <p>This property will take effect only when the <u>ConvertBookmarks</u> property is set to True. Numbers defining bookmark levels must be equal to or larger than one. Word style names must not repeat in the string. The string must not start or end with the delimiter. When this property is empty, the default style mapping (Heading one through nine will be mapped to level one through nine) will be used. Therefore, empty string is functionally equivalent to:</p> <p>Heading 1 1 Heading 2 2 Heading 3 3 Heading 4 4 Heading 5 5 Heading 6 6 Heading 7 7 Heading 8 8 Heading 9 9</p> <p>Note: If you use non-English version of Microsoft Word, then you may need to replace the word "Heading" with its localized version.</p> |
| Convert Hyperlinks | ConvertHyperlinks | Sets the flag to indicate whether to convert Word hyperlinks to PDF hyperlinks. |
| Print All Sheets | PrintAllSheets | The flag that indicates whether to print all Excel worksheets or not. |
| Print BG Color | PrintBGColor | For files printed via IE Sets the flag that indicates whether to print background color or not when printing. |
| Print Scale | PrintScale | For Visio files, sets the print scale |
| Header | Header | This property modifies Internet Explorer's header setting. |
| Footer | Footer | This property modifies Internet Explorer's footer setting. |
| | | |
| Image Compression | ImageCompression | If you want a lossless image compression, use PRN_IMAGE_COMPRESS_ZIP (ZIP compression). |
| Image Downsizing | ImageDownsizing | If the this property is set to True, then the resolution of images are reduced to the DPI value specified in the <u>ImageDownsizeResolution</u> property. |
| Image JPEG Quality | ImageJPEGQuality | The allowed value range is from 5 to 100 with 100 being the highest quality. |
| Downsize Resolution | ImageDownsizeResolution | If the <u>ImageDownsizing</u> property is set to True, then the resolution of images are reduced to the DPI value specified in this property. |

| Option | Command Line / Config File Name | Description |
|-----------------------|--|--|
| Font Embedding | FontEmbedding | The option PRN_FONT_EMBED_FULLSET (embedding fullset of font) will cause significant increase in PDF file size, especially for CJK font, and therefore not recommended. If you need to embed font, PRN_FONT_EMBED_SUBSET (embed subset of font) will be a better choice. |
| Font Substitution | FontSubstitution | For the PRN_FONT_SUBST_TABLE (use font substitution table) option, you need to configure the substitution table. The table is stored under the "Device Setting" section of the printer driver properties (can be accessed from Control Panel). |
| Embed Fonts as Type 0 | FontEmbedAsType0 | This option is recommended if you have non-standard fonts like barcode font. |
| | | |
| Top Margin | MarginTop | Sets top margin. |
| Bottom | MarginBottom | Sets bottom margin. |
| Left | MarginLeft | Sets left margin. |
| Right | MarginRight | Sets right margin. |
| Width | PageWidth | Sets a custom page width. |
| Height | PageHeight | Sets a custom page height. |
| Paper Orientation | PaperOrientation | Sets paper orientation. Eg PRN_PAPER_ORIENT_LANDSCAPE |

4.2 Command Line Interface

A command line interface to PDF generation is provided topdf.exe in the PDF Junction bin directory

```
topdf.exe /source=sourcefiledefinition {/option=value}
```

Options can be any of the options listed in section 4.1 in addition to the following :

| Command Line / Config File Name | Description |
|---------------------------------|--|
| /Source | Source file definition. Can include wildcards. |
| /TargetFolder | Target folder for converted file. |
| /Output | |
| /CreateFolders | Create output and error folders where required. |
| /PaperSize | Word does not use the paper size setting from printer driver. If you need to change the paper size, use this property instead. If you need to use a custom paper size, set this property to PRN_MSO_PAPER_CUSTOM, and set the <u>PageWidth</u> and <u>PageHeight</u> properties. |
| /PrintColorType | With this property, you can set the PowerPoint to print with either color, gray scale, or black and white. |
| /QueueWaitTimeOut | Since PowerPoint cannot perform multiple conversion at the same time, each conversion request will be put into queue, and they are converted one at a time. A waiting print job will timeout if it waits for more than the value specified in QueueWaitTimeout property. Timeout value is in milliseconds. Set to zero if you want to wait indefinitely. |
| /DisableScriptDebugger | Enables/Disables Internet Explorer's script debugger. It is highly recommended that Internet Explorer's script debugger is turned off during printing so that PDF conversion process is not interrupted. |
| /DisplayErrorDialogOnEveryError | Enables/Disables Internet Explorer's error dialog displayed on every error. It is highly recommended that this feature is turned off from Internet Explorer so that PDF conversion process is not interrupted. |
| /pdfa | Specifies that the output PDF should be PDF/A-1b compliant |
| /pdfx1a | Specifies that the output PDF should be PDF/X-1a compliant |
| /pdfx3 | Specifies that the output PDF should be PDF/X-3 compliant |
| /weboptimize | Specifies that the output PDF should be Linearized (web optimized) |
| /convertattachments | When converting from .msg files, the output PDF will contain converted attachment pages as well as the converted email. |
| /IncludeDocumentMarkups | For word documents, setting this option to true will include comments and markups in the generated PDF. |

Example Commands :

Convert the Excel spreadsheet sales.xls to PDF. The result file will be in the same directory as the source file (as targetfolder was not specified).. All the sheets in the spreadsheet will be printed, using landscape orientation.

```
topdf.exe /source=c:\infolder\sales.xls  
/paperorientation=PRN_PAPER_ORIENT_LANSCAPE /printallsheets=true
```


Convert all the files in c:\infolder to PDF and place the output files in c:\outfolder. The output file name will be converted_FILENAME.pdf where FILENAME was the original file name prefix. Eg sales.xls is converted to converted_sales.xls.

```
topdf.exe /source=c:\infolder\*.* /targetfolder=c:\outfolder
/output=converted_%FILENAME%.pdf
```

4.3 Configuration File

The configuration file topdf.exe.config (in the PDF Junction bin directory) is used to define a default set of parameters for PDF conversion.

Additional Configuration variables that can only be set via the configuration file are as follows :

| Command Line / Config File Name | Description |
|---------------------------------|--|
| WordExtension | A list of file extensions to be printed to PDF via MS Word. |
| ExcelExtension | A list of file extensions to be printed to PDF via Excel |
| PowerPointExtension | A list of file extensions to be printed to PDF via PowerPoint |
| VisioExtension | A list of file extensions to be printed to PDF via Visio |
| IEExtension | A list of file extensions to be printed to PDF via IE. |
| OutlookExtension | A list of file extensions to be printed to PDF via Outlook |
| OpenOfficeExtension | A list of file extensions to be printed to PDF via Open Office |
| AutoExtension | A list of file extensions to be printed according to automatic determination. |
| GenericExtension | A list of file extensions to be printed via the Generic interface which does not support any of the application-specific parameters. |
| ExcludedExtensions | A list of file extensions to be excluded from processing. |

```
<?xml version="1.0" encoding="utf-8" ?>
<configuration>
<appSettings>

<!-- Word Settings -->
<add key="Convertbookmarks" value="false" />
<add key="Bookmarkdepth" value="Heading 1|1|Heading 2|2|Heading 3|3" />
<add key="Converthyperlinks" value="false" />
<add key="Documentpassword" value="" />

<!-- Extension Mapping -->
<add key="AutoExtension" value="" />
<add key="GenericExtension" value="" />
<add key="WordExtension" value=".rtf.doc.docx.txt" />
<add key="ExcelExtension" value=".xls.xlsx" />
<add key="PowerPointExtension" value=".ppt.pptx" />
<add key="VisioExtension" value=".vsd" />
<add key="IEExtension" value=".htm.xml" />
<add key="OutlookExtension" value=".msg" />
<add key="OpenOfficeExtension"
value=".odt.swx.wpd.ods.sxc.odp.sxi.odg.sxd" />
<add key="ExcludedExtensions" value=".pdf.zip.exe.pps.ps.chm" />

<!-- Create Folders -->
<add key="CreateFolders" value="true" />
```

```
<!-- Page Settings -->
<add key="PaperOrientation" value="PRN_PAPER_ORIENT_PORTRAIT" />

<!-- Timeout -->
<add key="FileConversionTimeout" value="0" />

</appSettings>
</configuration> </configuration>
```

The configuration file can be amended to change the product behaviour – for example, to use Office Office rather than Microsoft Word to convert .doc .docx .rtf and .txt files you can make the following two changes :

```
<add key="WordExtension" value="" />
<add key="OpenOfficeExtension"
value=".odt.swx.wpd.ods.sxc.odp.sxi.odg.sxd.rtf.doc.docx.txt" />
```

5 PDF TO TIFF CONVERSION COMMAND LINE INTERFACE

5.1 Command Line Interface

A command line interface to PDF to TIFF conversion is provided topdf.exe in the PDF Junction bin directory

```
pdf2tiff.exe /source=sourcefiledefinition {/option=value}
```

| Command Line / Config File Name | Description |
|---------------------------------|--|
| /Source | Source file definition. |
| /Target | Output folder. If not specified the result file will be placed in the same folder as the source file. |
| /Output | Output file name. If not specified the output file name will be the same as the source file with a .tif extension instead of a .pdf one. |
| /TIFFCompression | GROUP4 or LZW. If not specified the default of GROUP4 will be used. |
| /TIFFResolution | The DPI of the output file. If not specified the default of 200 DPI will be used. |

6 PDF TEXT EXTRACTION COMMAND LINE INTERFACE

6.1 Command Line Interface

A command line interface to PDF text extraction is provided to `pdf2txt.exe` in the PDF Junction bin directory

```
pdf2txt.exe /source=sourcefiledefinition {/option=value}
```

| Command Line / Config File Name | Description |
|---------------------------------|--|
| /Source | Source file definition. |
| /Target | Output folder. If not specified the result file will be placed in the same folder as the source file. |
| /Output | Output file name. If not specified the output file name will be the same as the source file with a .txt extension instead of a .pdf one. |
| /PageFrom | First page to be processed. If not specified processing will start at page 1. |
| /PageTo | Last page to be processed. If not specified processing will end at the last page. |
| /PageSep | Page separator string which will be inserted in the text file between pages. |

7 WEB OPTIMIZATION (LINEARIZATION) COMMAND LINE INTERFACE

7.1 Command Line Interface

A command line interface to web optimization is provided to `pdfo.exe` in the PDF Junction bin directory

```
pdfo.exe /source=sourcefiledefinition {/option=value}
```

| Command Line / Config File Name | Description |
|---------------------------------|---|
| /Source | Source file definition. |
| /Target | Output folder. If not specified the result file will be placed in the same folder as the source file. |
| /Output | Output file name. |

PDF Junction allows the setting of XMP Metadata and custom DocInfo metadata. XMP (Extensible Metadata Platform) facilitates the embedding of metadata in PDF files using a subset of RDF. For further XMP background information see <http://www.adobe.com/products/xmp/>.

Note : Any existing XMP metadata in the source file will not be copied into the new file, so the directive XML file needs to include all required XMP metadata. However, existing DocInfo metadata can be automatically mapped to XMP – see below for further details.

8.1 Setting Metadata

To set metadata, an XML directive file is required. By default the file is expected to be named *filename.xml* where *filename.pdf* is the name of the source PDF file.

Directive files should follow the model shown below. A sample PDF and XML directive file can be found in the PDF Junction xmp folder (usually C:\Program Files\Aquaforest\PDF Junction 2.5\xmp).

```
<?xml version='1.0' encoding='utf-8'?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
<setxmp type='MAPINFO'>
  <custom name='RiskID' value='12345'></custom>
  <custom name='Department' value='AS12'></custom>
  <rdf:Description rdf:about=''
xmlns:pdfaid='http://www.aiim.org/pdfa/ns/id/'>
    <pdfaid:part>1</pdfaid:part>
    <pdfaid:conformance>B</pdfaid:conformance>
  </rdf:Description>
  <rdf:Description about=''
xmlns:stEvt='http://ns.adobe.com/xap/1.0/sType/ResourceEvent#'
xmlns:xmpMM='http://ns.adobe.com/xap/1.0/mm/'>
    <xmpMM:DocumentID>adobe:docid:indd:9705C6F1-8db6-11d5-a7dd-
cdf3d5439594</xmpMM:DocumentID>
    <xmpMM:History>
      <rdf:Seq>
        <rdf:li rdf:parseType='Resource'>
          <stEvt:action>rendition created</stEvt:action>
          <stEvt:parameters>some additonal info</stEvt:parameters>
          <stEvt:softwareAgent>Adobe Photoshop 6.0 for
Windows</stEvt:softwareAgent>
          <stEvt:when>2001-07-25T09:45:06-07:00</stEvt:when>
        </rdf:li>
        <rdf:li rdf:parseType='Resource'>
          <stEvt:action>version 5 created</stEvt:action>
          <stEvt:parameters>helpful info</stEvt:parameters>
          <stEvt:softwareAgent>Adobe InDesign 2.0 for
Windows</stEvt:softwareAgent>
          <stEvt:when>2001-08-10T10:45:39Z</stEvt:when>
        </rdf:li>
      </rdf:Seq>
    </xmpMM:History>
  </rdf:Description>
</setxmp>
</rdf:RDF>
```

| Tag | Notes | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-----------------|---|-----------------|-------------|-------------|-----|------|-------|-------------|----------|-------------|--|--------|-------------|------------|------------|-------------|---------|-------------|------------|-------------|--|----------|-------------|--------------|-------------|--|---------|-------------|-----------------|-------------|--|----------|-------------|--------------|-------------|--|--------------|-------------|----------------|-------------|--|---------|-------------|----------------|-------------|--|
| Setxmp | <p>Required. Type should be either “MAPINFO” or “NOMAPINFO”. If set to “MAPINFO” then the standard PDF Info Dictionary items will be mapped to their XMP equivalents as described in the ISO 19005 (PDF/A) crosswalk as shown below.</p> <table border="1"> <thead> <tr> <th>Entry</th> <th>PDF</th> <th>Property</th> <th>XMP</th> <th>type</th> </tr> </thead> <tbody> <tr> <td>Title</td> <td><i>text</i></td> <td>dc:title</td> <td><i>Text</i></td> <td></td> </tr> <tr> <td>Author</td> <td><i>text</i></td> <td>dc:creator</td> <td><i>seq</i></td> <td><i>Text</i></td> </tr> <tr> <td>Subject</td> <td><i>text</i></td> <td>dc:subject</td> <td><i>Text</i></td> <td></td> </tr> <tr> <td>Keywords</td> <td><i>text</i></td> <td>pdf:Keywords</td> <td><i>Text</i></td> <td></td> </tr> <tr> <td>Creator</td> <td><i>text</i></td> <td>xmp:CreatorTool</td> <td><i>Text</i></td> <td></td> </tr> <tr> <td>Producer</td> <td><i>text</i></td> <td>pdf:Producer</td> <td><i>Text</i></td> <td></td> </tr> <tr> <td>CreationDate</td> <td><i>date</i></td> <td>xmp:CreateDate</td> <td><i>Date</i></td> <td></td> </tr> <tr> <td>ModDate</td> <td><i>date</i></td> <td>xmp:ModifyDate</td> <td><i>Date</i></td> <td></td> </tr> </tbody> </table> | Entry | PDF | Property | XMP | type | Title | <i>text</i> | dc:title | <i>Text</i> | | Author | <i>text</i> | dc:creator | <i>seq</i> | <i>Text</i> | Subject | <i>text</i> | dc:subject | <i>Text</i> | | Keywords | <i>text</i> | pdf:Keywords | <i>Text</i> | | Creator | <i>text</i> | xmp:CreatorTool | <i>Text</i> | | Producer | <i>text</i> | pdf:Producer | <i>Text</i> | | CreationDate | <i>date</i> | xmp:CreateDate | <i>Date</i> | | ModDate | <i>date</i> | xmp:ModifyDate | <i>Date</i> | |
| Entry | PDF | Property | XMP | type | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Title | <i>text</i> | dc:title | <i>Text</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Author | <i>text</i> | dc:creator | <i>seq</i> | <i>Text</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Subject | <i>text</i> | dc:subject | <i>Text</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Keywords | <i>text</i> | pdf:Keywords | <i>Text</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Creator | <i>text</i> | xmp:CreatorTool | <i>Text</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Producer | <i>text</i> | pdf:Producer | <i>Text</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CreationDate | <i>date</i> | xmp:CreateDate | <i>Date</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ModDate | <i>date</i> | xmp:ModifyDate | <i>Date</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Custom | Optional. This can be used to specify name / value pairs for custom entries in the PDF Info Dictionary. The custom DocInfo metadata is visible in Adobe Acrobat via File Document Properties Custom. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Rdf:Description | <p>Optional. This can be used to define the RDF for a particular XMP schema. XMP Metadata is visible in Adobe Acrobat via File Document Properties Description Additional Metadata Advanced.</p> <p>The Adobe XMP reference fully describes XMP syntax.: http://www.adobe.com/devnet/xmp/pdfs/xmp_specification.pdf</p> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

8.2 Command Line Interface

A command line interface to PDF to TIFF conversion is provided topdf.exe in the PDF Junction bin directory

```
setxmp.exe /source=sourcefile /target=targetfile /output=outfilefile [/debug=true] [/xmlfile=xmlfile]
```

For example, the following command processes the file test.pdf using the directive file test.xmp in the same folder (this is the default assumption unless /xmlfile is specified). The output PDF is written to the d:\dev\out folder and is called test_xmp.pdf.

```
setxmp.exe" /output="%FILENAME_xmp.pdf" /Source="D:\dev\test.pdf" /Target="D:\dev\out"
```

| Command Line / Config File Name | Description |
|---------------------------------|---|
| /Source | Source file definition. |
| /Target | Output folder. If not specified the result file will be placed in the same folder as the source file. |
| /Output | Output file name mask. May include %FILENAME If not specified the output file name will be based on the source file name. |
| /xmlfile | This specifies the XML directive file. If not specified, the XML directive file is assumed to be %FILENAME.xml where the source file is %FILENAME..pdf. |

9 XML JOB TICKET FILES

PDF Junction uses XML Job Ticket files to store details of a particular job specification before this is run using the product either from the GUI or using the /jobid option from the command line – see 3.2.1. Note that some functionality, in particular custom scripts and logfiles can only be used in conjunction with XML Job Ticket files. Other functionality can also be used via the equivalent command line flags.

Developers may wish to create or generate XML job files using the information below.

A sample job file is shown here :

```
<pdfjunction_job>
<operation>split</operation>
<sourcetype>folder</sourcetype>
<sourcefiles>C:\qatest\pj\in\set001</sourcefiles>
<target>c:\qatest\pj\out\qa001</target>
<joboptions>-o -c -f %F%\%U%</joboptions>
<metadata></metadata>
<security></security>
<docoptions></docoptions>
<custom_script>custom.wsf</custom_script>
<custom_script_pre_job>>true</custom_script_pre_job>
<custom_script_post_job></custom_script_post_job>
<custom_script_pre_folder></custom_script_pre_folder>
<custom_script_post_folder></custom_script_post_folder>
<custom_script_pre_file>>true</custom_script_pre_file>
<custom_script_post_file></custom_script_post_file>
<custom_script_on_error></custom_script_on_error>
<logfile>qa001.txt</logfile>
</pdfjunction_job>
```

| Attribute | Description |
|---------------------------|--|
| operation | Must be split, merge, set, get or pdf (for conversion to PDF) |
| sourcetype | Source type (file, folder, tree). The default value is file. |
| target | Target directory for result files. If not specified, the source directory will be used as the target. |
| Joboptions | Set of options for the specified operation. Described in detail in section 5.3. |
| Metadata | Metadata settings for result PDF files. By default metadata fields are left blank. Described in detail in section 5.4. |
| Security | Security settings for result PDF files. By default files are not secured. Described in detail in section 5.5. |
| docoptions | Document option settings for result PDF files. Described in detail in section 5.6. |
| Sourcefiles | Source PDF file, folder or tree. May be multiple files for merge operations. |
| custom_script | If not left blank, specifies the name of a custom windows script file in <i>PDFJUNCTIONDIR</i> custom that will be called in accordance with the custom_script_* settings below. |
| custom_script_pre_job | If=Y, the custom script will be called at the start of the job |
| custom_script_post_job | If=Y, the custom script will be called at the end of the job |
| custom_script_pre_folder | If=Y, the custom script will be called at the start of processing each folder |
| custom_script_post_folder | If=Y, the custom script will be called at the end of processing each folder |
| custom_script_pre_file | If=Y, the custom script will be called at the start of processing each file |
| custom_script_post_file | If=Y, the custom script will be called at the end of processing each file |
| custom_script_on_error | If=Y, the custom script will be called when a processing error occurs. |
| logfile | If specified, output will be logged to afile with this name in <i>PDFJUNCTIONDIR</i> \logs |

10 CUSTOM SCRIPTS

The functionality of PDF Junction can be extended by using custom scripts which may be called at various points during the processing of a particular job. The custom scripts are Windows Script Files which allow scripting in either VBScript or Jscript.

Custom scripts reside in *PDFJUNCTIONDIR\custom*. In order to implement custom scripts, a developer should take a copy of the script template (custom.wsf, shown below) and implement the script according to their requirements. The custom script name should be specified and required exists should be flagged as being called, either using the GUI or directly in the Job Ticket File (See section 4).

Parameters are passed as follows :

| Parameter | Description |
|---------------|---|
| jobid | The current job id |
| currentfolder | The path of the folder currently being processed, or “null” for non-applicable exits (eg job_start) |
| currentfile | The path of the file currently being processed, or “null” for non-applicable exits (eg job_start) |
| exitpoint | One of the following string values, according to the exit point that was called. pre_job post_job pre_folder post_folder pre_file post_file on_error |

```
<job>
<runtime>
  <named name="jobid" helpstring="Job ID" type="string" required="false"/>
  <named name="currentfolder" helpstring="Current Folder" type="string" required="false" />
  <named name="currentfile" helpstring="Current File" type="string" required="false" />
  <named name="exitpoint" helpstring="Script Exit Point" type="string" required="false" />
</runtime>
<script language="VBScript">
'
' Use this section for VBScript
'
'jobID=WScript.Arguments.Named("jobid")
'currentFolder=WScript.Arguments.Named("currentfolder")
'currentFile=WScript.Arguments.Named("currentfile")
'exitPoint=WScript.Arguments.Named("exitpoint")
'WScript.StdOut.WriteLine("Parameter Values - "+jobID+"."+currentFolder+"."+currentFile+"."+exitPoint)
</script>
<script language="JScript">
//
// Use this section for JScript
//
var jobID=WScript.Arguments.Named("jobid");
var currentFolder=WScript.Arguments.Named("currentfolder");
var currentFile=WScript.Arguments.Named("currentfile");
var exitPoint=WScript.Arguments.Named("exitpoint");
WScript.StdOut.WriteLine("Parameter Values - "+jobID+"."+currentFolder+"."+currentFile+"."+exitPoint);
</script>
</job>
```


11 XML “INFORMATION” FILES

PDF Junction can be used to generate XML files containing metadata relating to source PDF files, as in the following example :

```
cscript pdfjunction.wsf /operation:get /target:c:\demo\out
c:\demo\in\demo3.pdf "/joboptions:-f %F%_metadata"
```

| Attribute | Description |
|---|--|
| Path | Full path name of the PDF file |
| Pages | The number of pages in the document |
| Title Author Subject Keywords Creator | PDF Document Metadata |
| Pdfversion | Version of PDF that the document complies with |
| Filesize | The size in bytes of the file |
| Permissions | Permissions bit mask. If set to 0, there is no security set on the file. |
| Isencrypted | True if the file is encrypted |
| Xmlmetadata | The XML PDF metadata as extracted from the document. See XXX for more details. |

```
<?xml version="1.0"?>
<pdfdocument>
<path>c:\demo\in\demo3.pdf</path>
<pages>3</pages>
<title>1</title>
<author>Neil Pitman</author>
<subject></subject>
<keywords></keywords>
<creator>Acrobat PDFMaker 6.0 for Word</creator>
<pdfversion>5</pdfversion>
<filesize>517811</filesize>
<permissions>0</permissions>
<isencrypted>>false</isencrypted>
<xmlmetadata>
<?xpacket begin='ï¿½' id='W5M0MpCehiHzreSzNTczkc9d'?>
<?adobe-xap-filters esc="CRLF"?>
<x:xmpmeta xmlns:x='adobe:ns:meta/' x:xmptk='XMP toolkit 2.9.1-13, framework 1.6'>
<rdf:RDF xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'
xmlns:iX='http://ns.adobe.com/iX/1.0/'>
....
<rdf:Description rdf:about='uuid:3a543dda-858c-4fb6-88f9-3b7358591e42'
xmlns:dc='http://purl.org/dc/elements/1.1/'
dc:format='application/pdf'><dc:title><rdf:Alt><rdf:li xml:lang='x-
default'>1</rdf:li></rdf:Alt></dc:title><dc:creator><rdf:Seq><rdf:li>Neil
Pitman</rdf:li></rdf:Seq></dc:creator><dc:subject><rdf:Seq><rdf:li></rdf:Seq>
</dc:subject></rdf:Description>
</rdf:RDF>
</x:xmpmeta>
<?xpacket end='w'?>
</xmlmetadata>
</pdfdocument>
```

12 PDF JUNCTION FILES AND DIRECTORIES

After installation, the following subdirectories are created in the PDF Junction install directory :

| Directory | Contents |
|-----------|--|
| bin | Executables and Scripts. |
| docs | Reference Guide |
| jobs | Standard directory for XML Job Ticket files. |
| logs | Log files |
| temp | Temporary storage for XML Job Ticket files. |
| License | License File. |

13 SUPPORT

Any problems running this application should give evident error messages, but if you are unable to resolve a problem please contact support@aquaforest.com

14 ACKNOWLEDGEMENTS

PDF Junction makes use of the iText Java library under the Mozilla Public License. See <http://www.lowagie.com/iText/> for more information.

PDF Junction makes use of the JSAP argument parser under the LGPL, with clarification of usage as dynamic linking. See <http://www.martiansoftware.com/jsap/index.html> for more information.