PDFUnit-Monitor

User manual

Carsten Siedentop

Table of Contents

Preface

The Current Situation of Testing PDF in Projects

These days telephone bills, insurance policies, official notifications and many types of contract are delivered as PDF documents. They are the result of a process chain consisting of programs in various programming languages using numerous libraries. Depending on the complexity of the documents to be produced, such programming is not easy. The software may have errors and statistically will have errors. So it should be tested in some of the following ways:

- Is the expected text within the expected page region?
- Is the bar code's text the expected text?
- Does the layout fulfill the requirements?
- Do the embedded ZUGFeRD data have the expected values?
- Are the values of the embedded ZUGFeRD data correspond with the visible values?
- Does a PDF comply with DIN 5008?
- Is the PDF signed? When and by whom?

It should scare developers, project managers and CEO's that until now there is almost no way of repeatedly testing PDF documents. And even the options which are available are not used as frequently as they should be. Unfortunately, manual testing is widespread. It is expensive and prone to errors.

With PDFUnit, any document, whether created using a powerful design tool, exported from MS Word or LibreOffice, processed using an API, or dropped out of an XSL-FO workflow, can be tested.

A User Interface for Automated Tests - a Contradiction?

Most automated tests ared started without a Windows-style program. But the PDFUnit-Monitor works interactively and as an interaction-free program. The Monitor watches directories and starts tests immediately when new PDF documents are stored in the directory. So, an IT-team can create documents in any way at any time and the PDFUnit-Monitor shows the validation results whithout any user-interaction.

The PDFUnit-Monitor is easy to use. Test cases will be defined in Excel files, so that non-programmers can write tests. Additionally, the monitor can be used interactively to run individual tests.

Chapter 3: “Features” (p. 10) describes all available functions.

Chapter 1. About this Documentation

Who Should Read it

This documentation is mainly addressed to members of the quality assurance staff.

If there are Problems

If you have problems to test a PDF, please search for a similar problem in the internet. Maybe, you find a solution. Finally, you are invited to write to info[at]pdfunit.com and describe the problem. We'll try to help you.

New Features Wanted?

Do you need other test functions? Please feel free to send your requirements to info@pdfunit.com. You are invited to influence the further development of PDFUnit.

Production of this Documentation

This documentation was created with DocBook-XML and both PDF and HTML are generated from one text source. It is well known that the layout can be improved in both formats, e.g. the pagebreaks in PDF format. And improving the layout is already on the to-do list, but there are other tasks with higher priority.

Feedback

Any kind of feedback is welcomed. Please write to info[at]pdfunit.com.

Chapter 2. Quickstart

The PDFUnit-Monitor is an application that shows the result of PDFUnit tests. Tests are written in Excel files, so non-developers can create and run the tests.

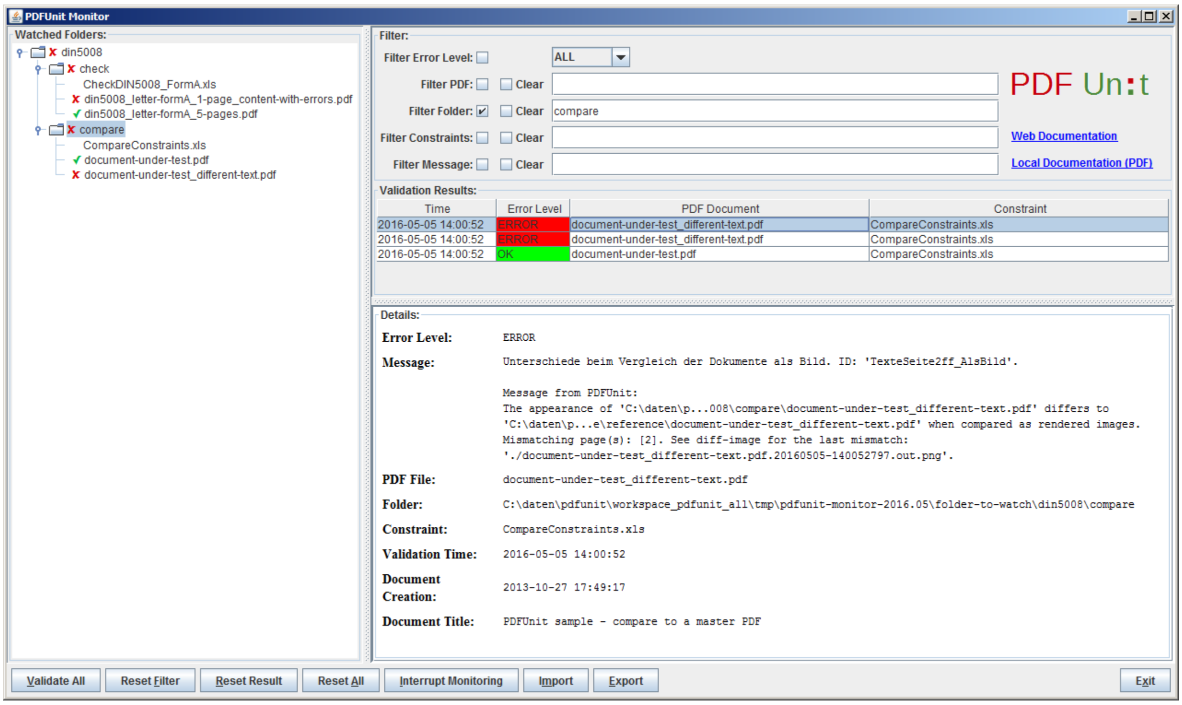
The functional scope of the PDFUnit-Monitor is large. Therefore a detailed description of it exists as a separate file, also a demonstration video is available. Both can be downloaded with this link (download). The separate documentation provides also information about the installation and configuration of the PDFUnit-Monitor. The following sections briefly describe the main features.

Monitored Folders

The PDFUnit-Monitor monitors all PDF documents in a defined directory and its subdirectories. It checks the documents against rules which are read from Excel files. The Excel files have to lay also in the monitored directories. If new PDF documents were copied into the monitored directories, the tests started automatically. A manual start is not necessary but can be done. If a PDF document complies with all rules, its name in the folder tree is decorated with a green checkmark. When a PDF test fails, all violations will be listed. Additionally its name is decorated with a red cross. This status is transferred to the directory name. The name of a folder is decorated with a green checkmark only, when the folder itself and all subdirectories contain valid PDF documents. Otherwise, the folder is decorated with a red cross.

Folder with the name 'reference' are not watched. Such folders contain PDF documents which will be used as references when comparing actual documents under test.

The following picture shows the PDFUnit-Monitor. On the left side the folder structure with PDF and Excel files can be seen. The right side shows the validation results in the upper half and details of a selected message in the lower half.



Watched Folders:

- ✗ din5008
  - ✗ check
    - CheckDIN5008_FormA.xls
    - ✗ din5008_letter-formA_1-page_content-with-errors.pdf
    - ✓ din5008_letter-formA_5-pages.pdf
  - ✗ compare
    - CompareConstraints.xls
    - ✓ document-under-test.pdf
    - ✗ document-under-test_different-text.pdf

Filter:

Filter Error Level: □ ALL
Filter PDF: □ □ Clear
Filter Folder: ☑ □ Clear compare
Filter Constraints: □ □ Clear
Filter Message: □ □ Clear

Validation Results:

Time	Error Level	PDF Document	Constraint
2016-05-05 14:00:52	ERROR	document-under-test_different-text.pdf	CompareConstraints.xls
2016-05-05 14:00:52	ERROR	document-under-test_different-text.pdf	CompareConstraints.xls
2016-05-05 14:00:52	OK	document-under-test.pdf	CompareConstraints.xls

Details:

Error Level: ERROR

Message:

```
Unterschiede beim Vergleich der Dokumente als Bild. ID: 'TexteSeite2ff_AlsBild'.

Message from PDFUnit:
The appearance of 'C:\daten\p...008\compare\document-under-test_different-text.pdf' differs to
'C:\daten\p...e\reference\document-under-test_different-text.pdf' when compared as rendered images.
Mismatching page(s): [2]. See diff-image for the last mismatch:
'./document-under-test_different-text.pdf.20160505-140052797.out.png'.
```

PDF File: document-under-test_different-text.pdf

Folder: C:\daten\pdfunit\workspace_pdfunit_all\tmp\pdfunit-monitor-2016.05\folder-to-watch\din5008\compare

Constraint: CompareConstraints.xls

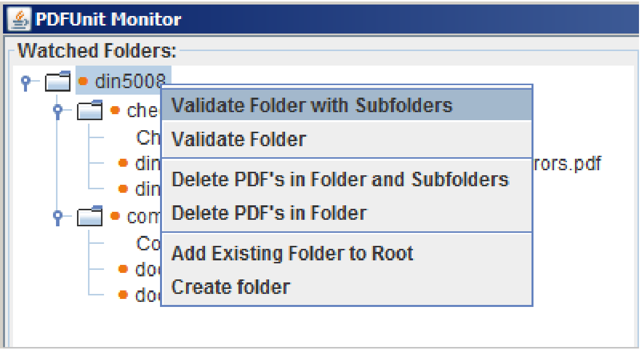
Validation Time: 2016-05-05 14:00:52

Document Creation: 2013-10-27 17:49:17

Document Title: PDFUnit sample - compare to a master PDF

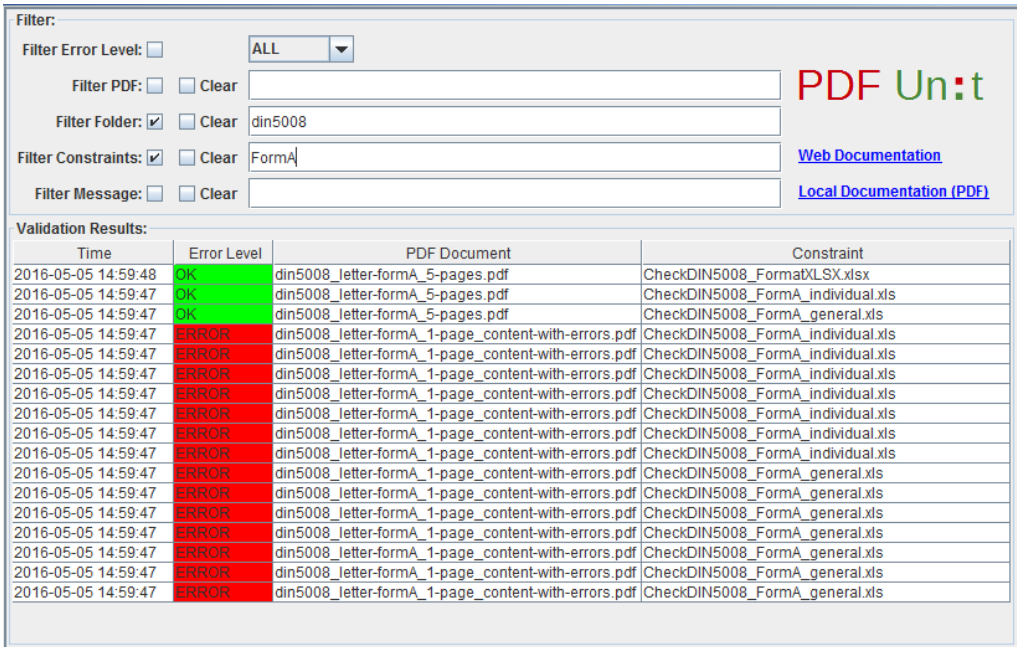
A double click on a PDF or Excel document in the folder structure or in the error list opens the standard application of the operating system for the document.

Each element of the folder structure provides a context menu with various functions. The following figure shows an example:



Overview of Test Results with Filter Options

The Monitor shows all validation results as a list in the upper part of the right-hand side. Each constraint validation is one entry in the list. The details of a validation will be shown in the lower part of the right side when a list entry is selected.



Filter:

Filter Error Level: □ ALL

Filter PDF: □ □ Clear

Filter Folder: ☑ □ Clear din5008

Filter Constraints: ☑ □ Clear FormA

Filter Message: □ □ Clear

Validation Results:

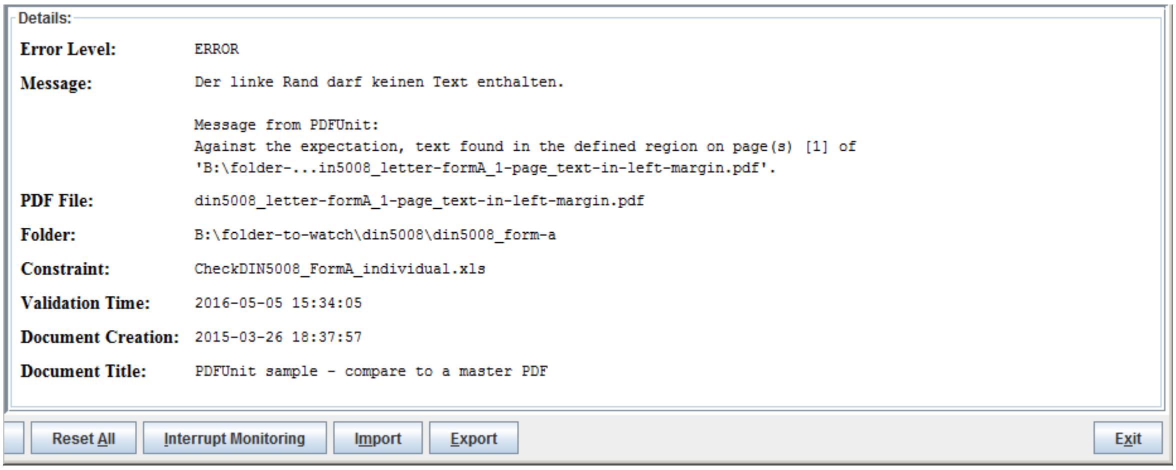
Time	Error Level	PDF Document	Constraint
2016-05-05 14:59:48	OK	din5008_letter-formA_5-pages.pdf	CheckDIN5008_FormatXLSX.xlsx
2016-05-05 14:59:47	OK	din5008_letter-formA_5-pages.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 14:59:47	OK	din5008_letter-formA_5-pages.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 14:59:47	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 14:59:47	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 14:59:47	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 14:59:47	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 14:59:47	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 14:59:47	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 14:59:47	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 14:59:47	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 14:59:47	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 14:59:47	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 14:59:47	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 14:59:47	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 14:59:47	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 14:59:47	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_general.xls

The error list can be filtered by the names of PDF documents, folders, Excel files and regular expressions on error messages. The folder structure on the left side is connected with the filters on the right side. Each time, when a folder or a document is selected in the structure, a corresponding filter is set.

When a cell with a PDF or Excel document is double-clicked, the standard application of the operating system for that document type starts.

Details of an Error

When a row of the error list is selected all details of that entry will be shown in the lower part of the right side of the Monitor.



Details:

Error Level: ERROR

Message: Der linke Rand darf keinen Text enthalten.

```
Message from PDFUnit:
Against the expectation, text found in the defined region on page(s) [1] of
'B:\folder-...in5008_letter-formA_1-page_text-in-left-margin.pdf'.
```

PDF File: din5008_letter-formA_1-page_text-in-left-margin.pdf

Folder: B:\folder-to-watch\din5008\din5008_form-a

Constraint: CheckDIN5008_FormA_individual.xls

Validation Time: 2016-05-05 15:34:05

Document Creation: 2015-03-26 18:37:57

Document Title: PDFUnit sample - compare to a master PDF

The first part of the error message was read from the Excel file. That part is designed by the person who created the tests. The next part of the message comes from PDFUnit. Additionally to the error message itself useful information about the PDF document, the constraint file and the execution time are provided.

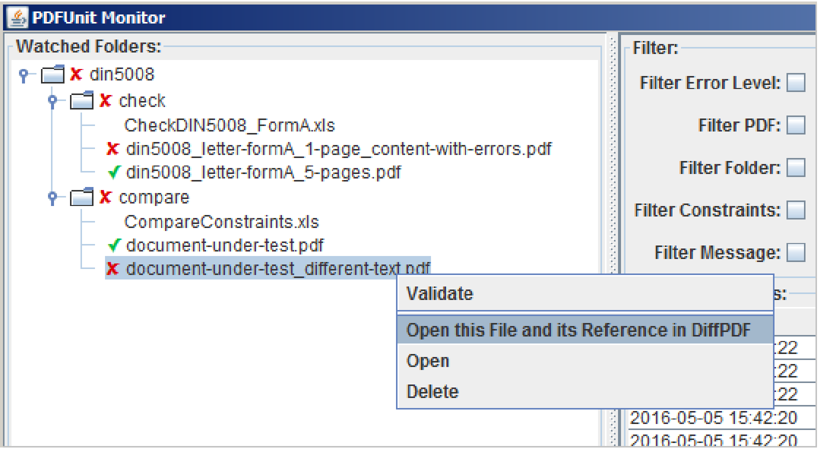
The error messages of PDFUnit are currently in English and in German. Other languages can be provided with a little work, when requested.

Comparing a PDF with a Reference

PDF document can be compared to a reference. The rules for the comparison are read from Excel files. When the PDFUnit-Monitor detects a difference between the test and the referenced PDF the name of the test document will be decorated with a red cross.

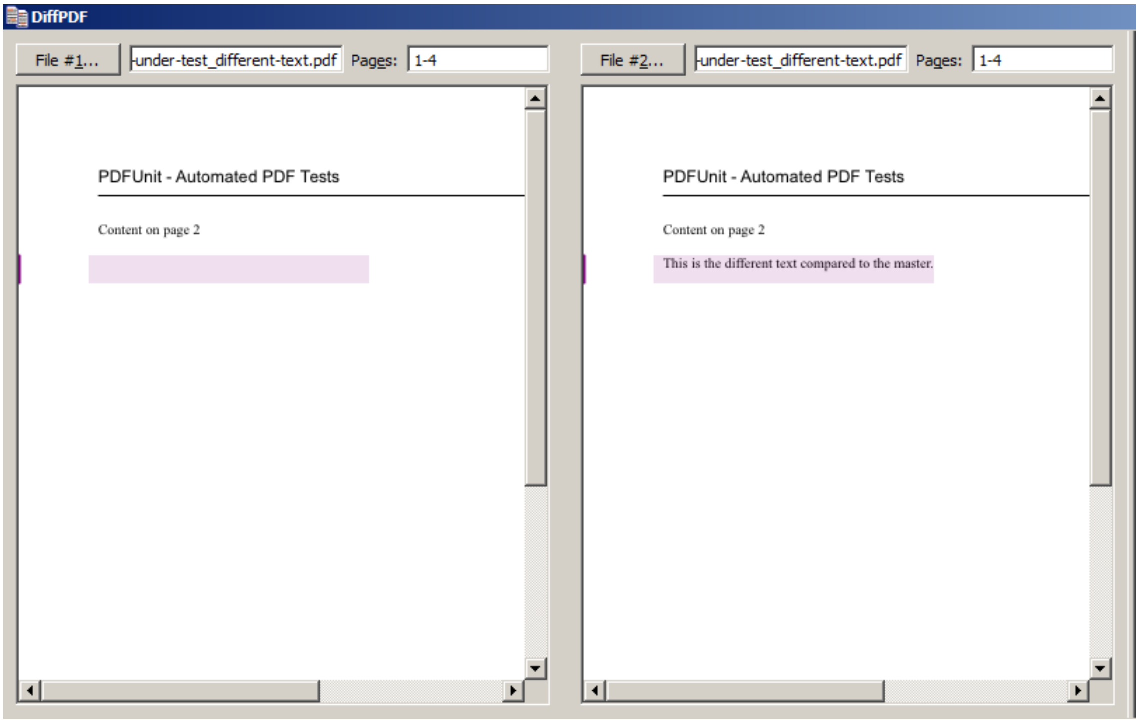
A reference document has to have the same name as the PDF under test and has to be located in a subdirectory named 'reference' below that folder which contains the PDF under test.

If the validation detects a difference, the program 'DiffPDF 1.5.1 portable' can be started by clicking the right mouse button. That program marks the differences very clearly:



The program was created by Mark Summerfield and is available as a 'portable app' from this link (download). DiffPDF can be used in English, German, French and Czech. Many thanks to all who are involved for their work and the great result.

The next image shows the application DiffPDF just after it was started from the PDFUnit-Monitor. The left side shows the referenced PDF and the right side the current test PDF. The application points directly to the first difference, in this case to page 2. The differences are marked with a coloured background. The image does not show the buttons to navigate from one mismatch to the next.



DiffPDF
File #1...
-under-test_different-text.pdf
Pages:
1-4
File #2...
-under-test_different-text.pdf
Pages:
1-4
PDFUnit - Automated PDF Tests
Content on page 2
PDFUnit - Automated PDF Tests
Content on page 2
This is the different text compared to the master.

Export and Import of Test Results

The test results can be exported as XML clicking the button 'Export' With XSLT stylesheets the exported files can be converted into HTML reports. With a click on the button 'Import' the can loaded again into the monitor.

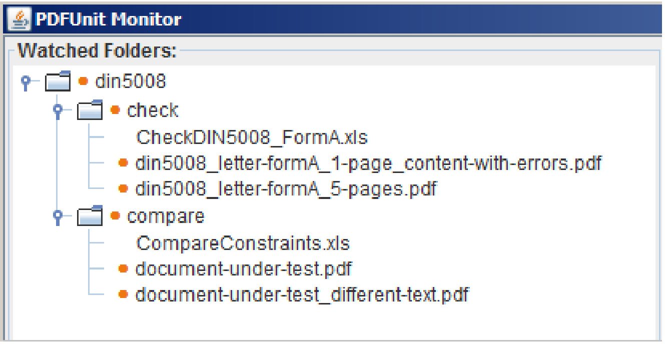
Internationalization

The PDFUnit monitor is currently available for in English and in German. An extension to other languages is structurally prepared and can be realized on demand with little effort.

Chapter 3. Features

3.1. View - Watched Folders

The directory structure in the left part of the monitor shows all files with the extension PDF, XLS and XSLX in all subdirectories below that folder. The topmost folder has to be defined by the Property 'WATCHED_FOLDER' in the configuration file 'pdfunit-monitor.config'. The configuration file is described in chapter 7: “Configuration (pdfunit-monitor.config)” (p. 27).


PDFUnit Monitor
Watched Folders:
din5008
check
CheckDIN5008_FormA.xls
din5008_letter-formA_1-page_content-with-errors.pdf
din5008_letter-formA_5-pages.pdf
compare
CompareConstraints.xls
document-under-test.pdf
document-under-test_different-text.pdf

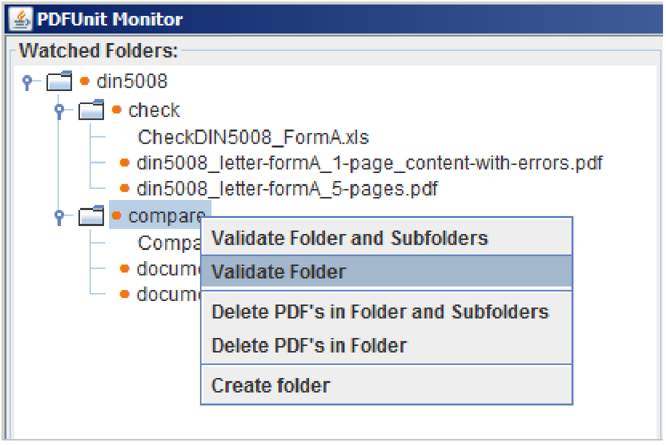
The PDFUnit-Monitor is sensitive to all changes in the watched directories. That means, validation results will be updated whenever files or directories are added, changed or removed.

Each PDF document will be validated by the rules of every Excel file in the same directory. The result of each validation rule will be shown in the result list.

Important hint: PDF documents which are password protected cannot be processed by the PDFUnit-Monitor. But they can be tested with the Java-API of PDFUnit.

Context Menu

When the user starts the monitor he can trigger the validation using the button `Validate All` or the context menu:

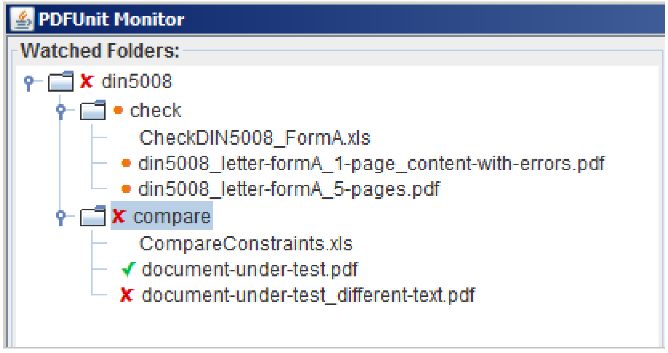

PDFUnit Monitor
Watched Folders:
din5008
check
CheckDIN5008_FormA.xls
din5008_letter-formA_1-page_content-with-errors.pdf
din5008_letter-formA_5-pages.pdf
compare
Compa
docum
docum
Validate Folder and Subfolders
Validate Folder
Delete PDF's in Folder and Subfolders
Delete PDF's in Folder
Create folder

The context menu is different depending on whether the menu relates to a directory, a PDF file or an Excel file.

Colours Indicating the Validation Result

The validation result is shown by a coloured symbol in front of the documents name. A green tick shows that all validation rules passed successfully. A red cross means that a document violates one or more rules.

Only when all tests are successful, a green tick is shown for that folder. A red arrow cross propagates upwards and a yellow circle shows that this file or directory was not yet validated.

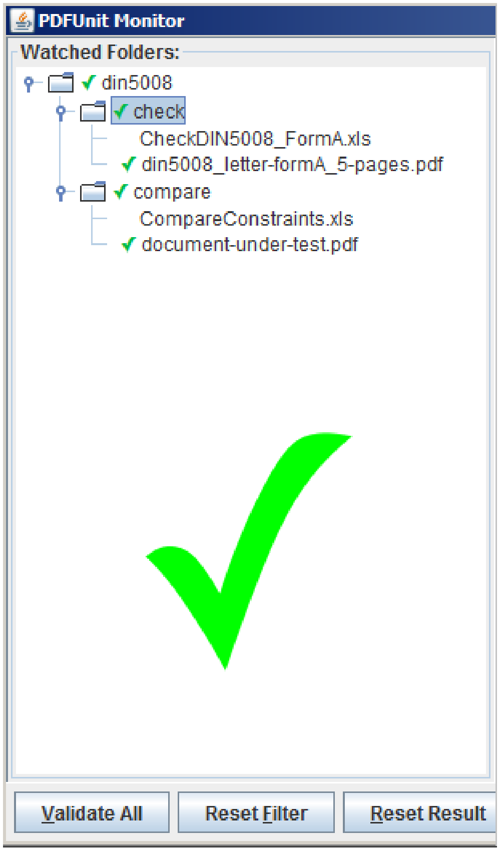


PDFUnit Monitor

Watched Folders:

```
✗ din5008
    ● check
        CheckDIN5008_FormA.xls
        ● din5008_letter-formA_1-page_content-with-errors.pdf
        ● din5008_letter-formA_5-pages.pdf
    ✗ compare
        CompareConstraints.xls
        ✓ document-under-test.pdf
        ✗ document-under-test_different-text.pdf
```

When all PDF documents are error-free, the PDFUnit-Monitor shows a big green tick:



PDFUnit Monitor

Watched Folders:

```
✓ din5008
    ✓ check
        CheckDIN5008_FormA.xls
        ✓ din5008_letter-formA_5-pages.pdf
    ✓ compare
        CompareConstraints.xls
        ✓ document-under-test.pdf
```

Drag and Drop

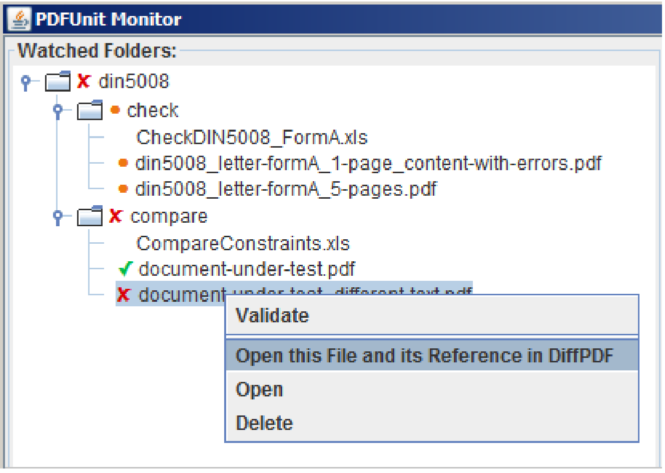
The PDFUnit-Monitor supports the 'Drop' function of 'Drag-and-Drop'. When a PDF document is dropped into a directory, it is copied, not moved! And then it is automatically validated against all Excel files in that folder.

Open Documents

Documents in the directory tree can be opened with the system application which is registered for the file type. Just double-click it.

Showing the Difference between Two PDF Documents

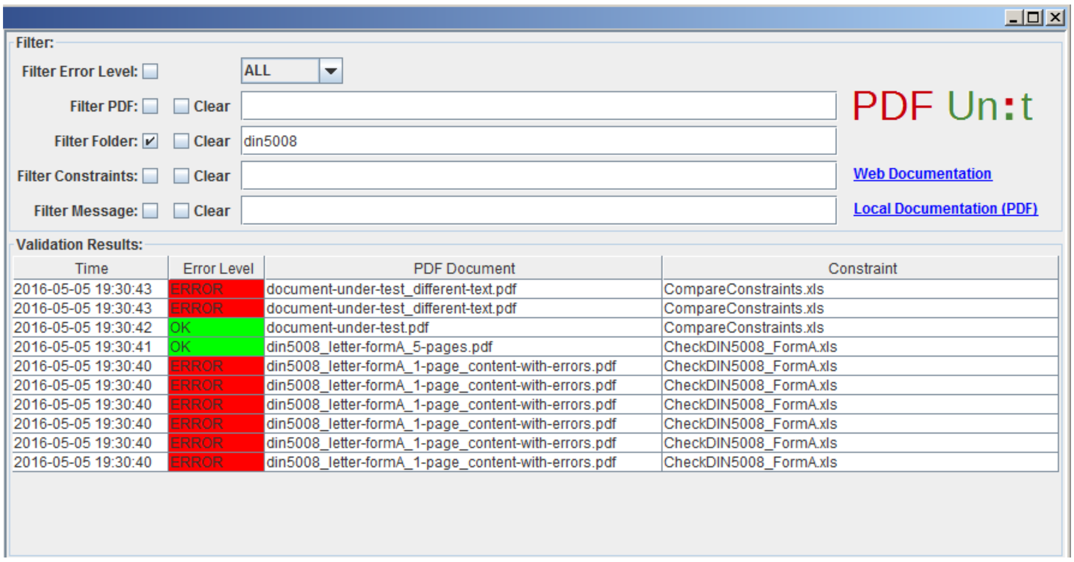
A PDF under test and its related reference document can be opened with the special program 'DiffPDF' to mark the differences. The next image shows the context menu:


PDFUnit Monitor
Watched Folders:
✗ din5008
● check
CheckDIN5008_FormA.xls
● din5008_letter-formA_1-page_content-with-errors.pdf
● din5008_letter-formA_5-pages.pdf
✗ compare
CompareConstraints.xls
✓ document-under-test.pdf
✗ document-under-test_different-text.pdf
Validate
Open this File and its Reference in DiffPDF
Open
Delete

The special program 'DiffPDF' will be described in detail in chapter 3.4: “DiffPDF - Show Mismatches between Two PDF Documents” (p. 15).

3.2. View - Validation Results, Filter

When a validation is completed, all results are shown in the result list:



Filter:

Filter Error Level: □ ALL

Filter PDF: □ □ Clear

Filter Folder: ☑ □ Clear din5008

Filter Constraints: □ □ Clear

Filter Message: □ □ Clear

Validation Results:

Time	Error Level	PDF Document	Constraint
2016-05-05 19:30:43	ERROR	document-under-test_different-text.pdf	CompareConstraints.xls
2016-05-05 19:30:43	ERROR	document-under-test_different-text.pdf	CompareConstraints.xls
2016-05-05 19:30:42	OK	document-under-test.pdf	CompareConstraints.xls
2016-05-05 19:30:41	OK	din5008_letter-formA_5-pages.pdf	CheckDIN5008_FormA.xls
2016-05-05 19:30:40	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA.xls
2016-05-05 19:30:40	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA.xls
2016-05-05 19:30:40	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA.xls
2016-05-05 19:30:40	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA.xls
2016-05-05 19:30:40	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA.xls
2016-05-05 19:30:40	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA.xls

The user can choose which results he wants to see. He can type text into the text field on the right-hand side of the 'Clear' buttons. Alternatively, the filters can be set when he clicks on a file or a directory.

The filters work as follows:

Filter	Behaviour
Filter Error Level	If the level ALL is selected in the drop-down list, all results will be shown. Choosing any value in the drop-down list limits the results to exactly that error level.
Filter PDF	When the user types a text into the text field, results will be shown from those files which contain the text in their name.
Filter Folder	When the user types a text into the text field, results will be shown from those folders which contain the text in their name.
Filter Constraints	When the user types a text into the text field, results will be shown from those Excel files which contain the text in their name.
Filter Message	When the user types a text into the text field, results will be shown when the text of this filter is part of the error message.

The text of all filters are treated case-insensitive.

All filters can be reset using the button 'Reset Filter'

A double-click on the name of a PDF or an Excel file will open the system application for this file type.

Hint: When a user selects a file in the tree, then overtypes the filter and then clicks on the same file again, the filter will not be updated.

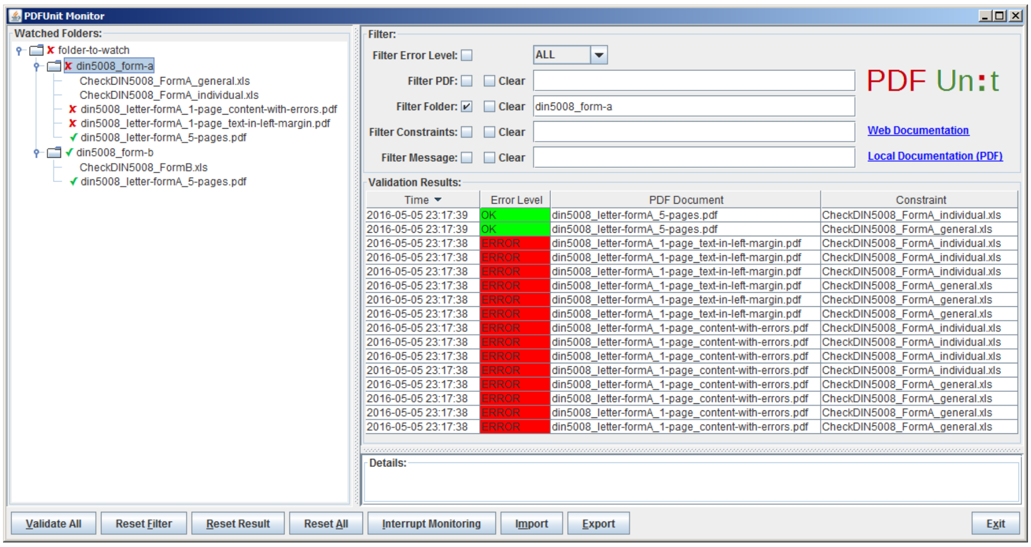
Meaning of the Error Levels

The error levels have the following meanings:

Error Level	Comment
OK	A PDF document passes all rules in all constraint files in the same folder.
ERROR	A PDF document violates one or more constraint definitions in a constraint file in the same folder.
WARNING	This level was set when a validation could not be carried out, e.g. no constraint file exists or an Excel file contains errors itself.
INFO	Constraint files were added or removed from a watched folder.
ALL	When this 'level' is selected, no level filter is active.

Sort the Error List

All entries in the error list can be sorted. Just click on the title of a column. The following two images show the effect.


PDFUnit Monitor
Watched Folders:
✗ folder-to-watch
✗ din5008_form-a
CheckDIN5008_FormA_general.xls
CheckDIN5008_FormA_individual.xls
✗ din5008_letter-formA_1-page_content-with-errors.pdf
✗ din5008_letter-formA_1-page_text-in-left-margin.pdf
✓ din5008_letter-formA_5-pages.pdf
✓ din5008_form-b
CheckDIN5008_FormB.xls
✓ din5008_letter-formA_5-pages.pdf
PDF Un:t
Web Documentation
Local Documentation (PDF)
Validate All
Reset Filter
Reset Result
Reset All
Interrupt Monitoring
Import
Export
Exit

Filter:

Filter Error Level: □ ALL

Filter PDF: □ □ Clear

Filter Folder: ☑ □ Clear din5008_form-a

Filter Constraints: □ □ Clear

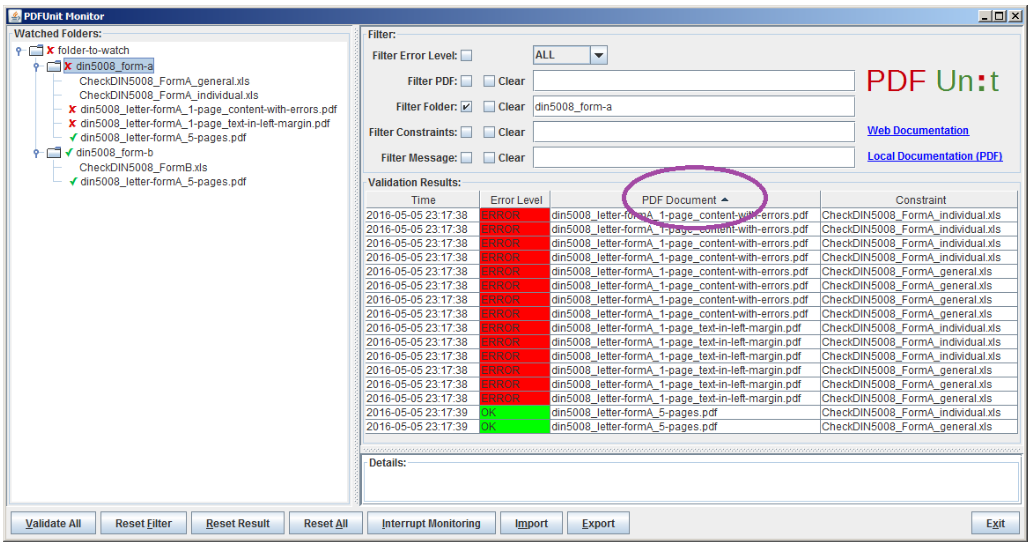
Filter Message: □ □ Clear

Validation Results:

Time ▼	Error Level	PDF Document	Constraint
2016-05-05 23:17:39	OK	din5008_letter-formA_5-pages.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 23:17:39	OK	din5008_letter-formA_5-pages.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_text-in-left-margin.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_text-in-left-margin.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_text-in-left-margin.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_text-in-left-margin.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_text-in-left-margin.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_text-in-left-margin.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_general.xls

Details:

The previous image shows the standard. All events will be sorted by time with the latest at the top. After clicking on the title of the column 'PDF Document', the following image can be seen:


PDFUnit Monitor
Watched Folders:
✗ folder-to-watch
✗ din5008_form-a
CheckDIN5008_FormA_general.xls
CheckDIN5008_FormA_individual.xls
✗ din5008_letter-formA_1-page_content-with-errors.pdf
✗ din5008_letter-formA_1-page_text-in-left-margin.pdf
✓ din5008_letter-formA_5-pages.pdf
✓ din5008_form-b
CheckDIN5008_FormB.xls
✓ din5008_letter-formA_5-pages.pdf
PDF Un:t
Web Documentation
Local Documentation (PDF)
Validate All
Reset Filter
Reset Result
Reset All
Interrupt Monitoring
Import
Export
Exit

Filter:

Filter Error Level: □ ALL

Filter PDF: □ □ Clear

Filter Folder: ☑ □ Clear din5008_form-a

Filter Constraints: □ □ Clear

Filter Message: □ □ Clear

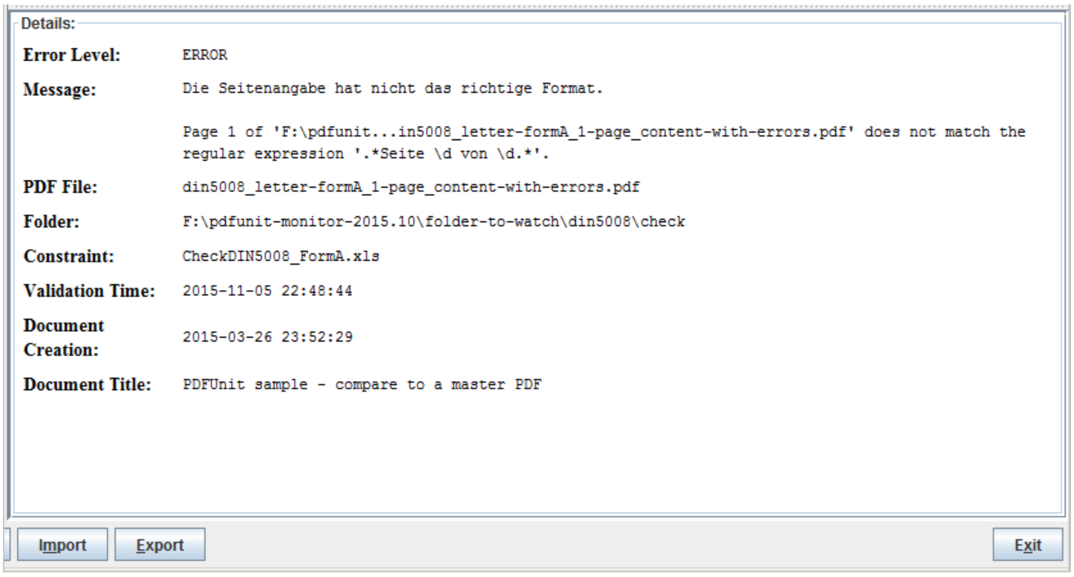
Validation Results:

Time	Error Level	PDF Document ▲	Constraint
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_content-with-errors.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_text-in-left-margin.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_text-in-left-margin.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_text-in-left-margin.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_text-in-left-margin.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_text-in-left-margin.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 23:17:38	ERROR	din5008_letter-formA_1-page_text-in-left-margin.pdf	CheckDIN5008_FormA_general.xls
2016-05-05 23:17:39	OK	din5008_letter-formA_5-pages.pdf	CheckDIN5008_FormA_individual.xls
2016-05-05 23:17:39	OK	din5008_letter-formA_5-pages.pdf	CheckDIN5008_FormA_general.xls

Details:

3.3. View - Validation Details

When a row in the validation list is selected, all detailed information about that result will be shown in the area below the list. The following image gives an example:



Details:

Error Level:	ERROR
Message:	Die Seitenangabe hat nicht das richtige Format. Page 1 of 'F:\pdfunit...in5008_letter-formA_1-page_content-with-errors.pdf' does not match the regular expression '.*Seite \d von \d.*'.
PDF File:	din5008_letter-formA_1-page_content-with-errors.pdf
Folder:	F:\pdfunit-monitor-2015.10\folder-to-watch\din5008\check
Constraint:	CheckDIN5008_FormA.xls
Validation Time:	2015-11-05 22:48:44
Document Creation:	2015-03-26 23:52:29
Document Title:	PDFUnit sample - compare to a master PDF

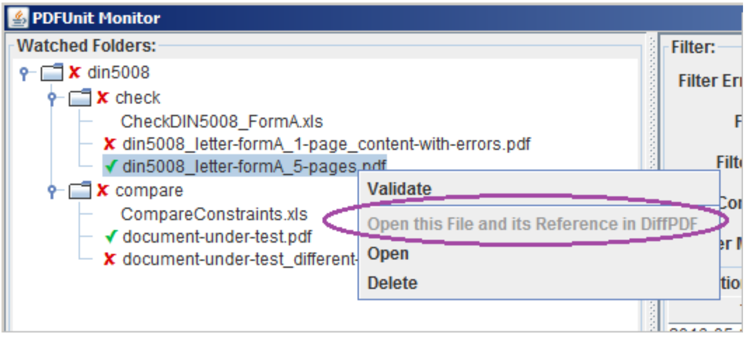
The first line of the 'Message'-section comes from the column 'message' of the Excel sheet. Any text with placeholders is allowed. The next lines in 'Message' represent the validation message from PDFUnit.

Error messages and the labels in the first column of the 'Details' view will be shown in English or German. The Java runtime detects the default language from the operating system. If you want to run PDFUnit in a different language, set appropriate JVM properties. For example, the JVM parameters `-Duser.language=en -Duser.country=UK` force PDFUnit to run in English.

If you need the monitor in another language, please write a short mail to info[at]pdfunit.com.

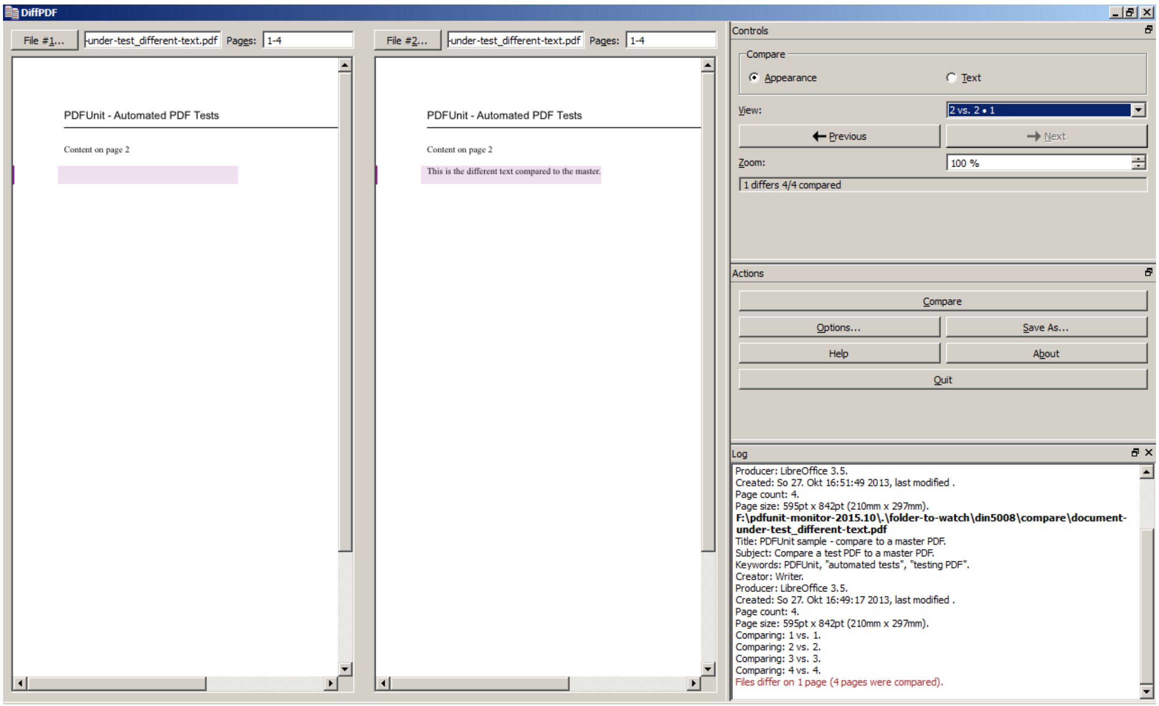
3.4. DiffPDF - Show Mismatches between Two PDF Documents

As described in Chapter 2: “Quickstart” (p. 6), differences between a PDF under test and a reference document can be shown with the program 'DiffPDF'. That program will be started using the context menu of a PDF file. The menu item is only active when the reference file can be found. The reference file is searched for in a subdirectory called 'reference' below the directory of the PDF under test. The referene file has to have the same name as the PDF under test. The following image shows the context menu for the case when no reference PDF was found:



PDFUnit Monitor
Watched Folders:
din5008
check
CheckDIN5008_FormA.xls
din5008_letter-formA_1-page_content-with-errors.pdf
din5008_letter-formA_5-pages.pdf
compare
CompareConstraints.xls
document-under-test.pdf
document-under-test_different
Validate
Open this File and its Reference in DiffPDF
Open
Delete
Filter:
Filter Er

If a reference PDF exists, the application starts with the following image:



File #1...: -under-test_different-text.pdf, Pages: 1-4

PDFUnit - Automated PDF Tests

Content on page 2

File #2...: -under-test_different-text.pdf, Pages: 1-4

PDFUnit - Automated PDF Tests

Content on page 2

This is the different text compared to the master.

Compare: ● Appearance ○ Text

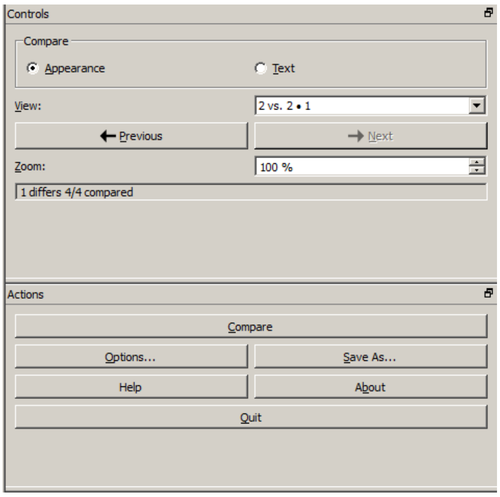
View: 2 vs. 2 • 1

Zoom: 100 %

1 differs 4/4 compared

```
Producer: LibreOffice 3.5.
Created: So 27. Okt 16:51:49 2013, last modified .
Page count: 4.
Page size: 595pt x 842pt (210mm x 297mm).
F:\pdfunit-monitor-2015.10\.\folder-to-watch\din5008\compare\document-
under-test_different-text.pdf
Title: PDFUnit sample - compare to a master PDF.
Subject: Compare a test PDF to a master PDF.
Keywords: PDFUnit, "automated tests", "testing PDF".
Creator: Writer.
Producer: LibreOffice 3.5.
Created: So 27. Okt 16:49:17 2013, last modified .
Page count: 4.
Page size: 595pt x 842pt (210mm x 297mm).
Comparing: 1 vs. 1.
Comparing: 2 vs. 2.
Comparing: 3 vs. 3.
Comparing: 4 vs. 4.
Files differ on 1 page (4 pages were compared).
```

The main part shows both documents page by page. The left-hand side is the reference PDF, the right-hand side is the current PDF under test. The bottom right corner shows some textual information about both documents, the upper right corner provides some buttons for navigating from error to error.



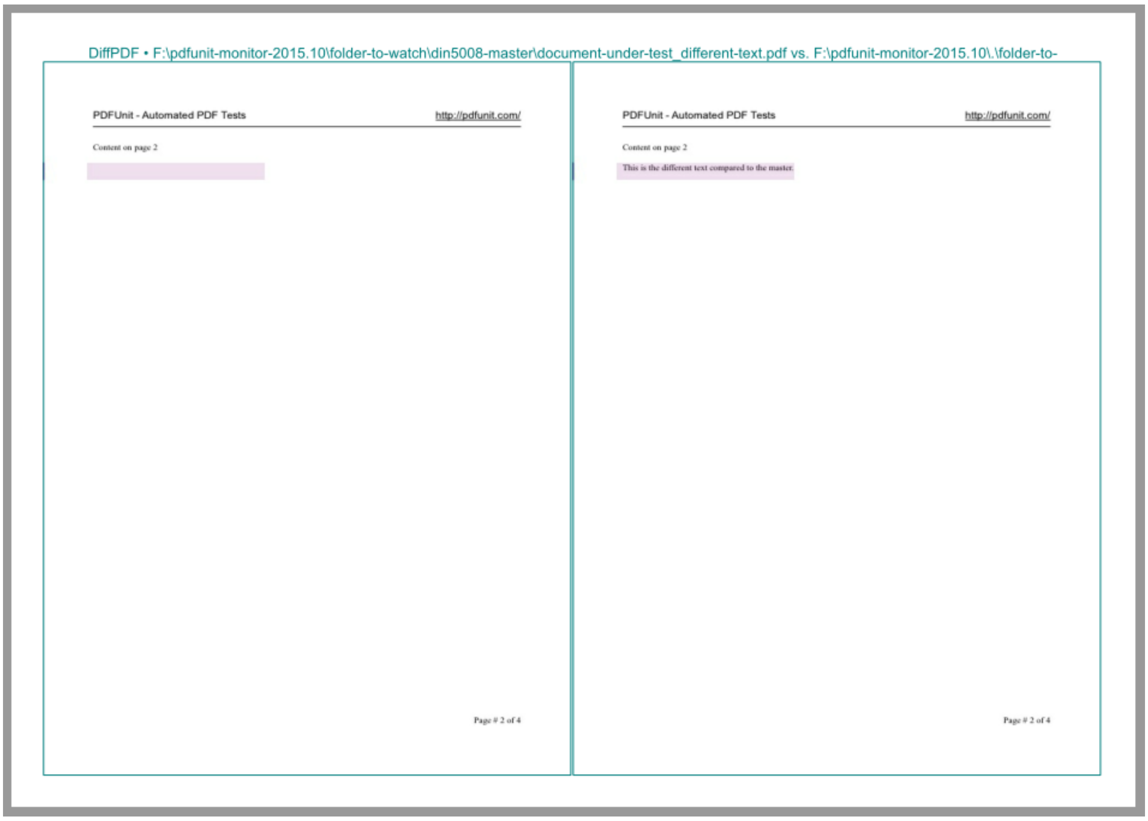
Compare: ● Appearance ○ Text

View: 2 vs. 2 • 1

Zoom: 100 %

1 differs 4/4 compared

With the button 'Save As...' the differences can be exported to a file. The result of such an export looks like this:



DiffPDF • F:\pdfunit-monitor-2015.10\folder-to-watch\din5008-master\document-under-test_different-text.pdf vs. F:\pdfunit-monitor-2015.10\.\folder-to-

PDFUnit - Automated PDF Tests http://pdfunit.com/

Content on page 2

Page # 2 of 4

PDFUnit - Automated PDF Tests http://pdfunit.com/

Content on page 2

This is the different text compared to the master.

Page # 2 of 4

The application 'DiffPDF' can also be used independently from PDFUnit-Monitor.

PDFUnit creates a diff-image for every failed test. The diff-image shows the reference PDF on the left-hand side and the differences in red on the right-hand side. The diff-image files are placed by default in the root folder and have the names `<document-name>.<date-time>.out.png`. The output folder can be set using the property 'output.path.diffimages' in the config file `PDFUNIT_MONITOR_HOME/lib/PDFUnit-<version>/pdfunit.config`.

3.5. Exporting and Importing Test Results

The PDFUnit-Monitor can export the validation result into an XML file. This may be useful for further reporting.

The export is started with the button 'Export'. All visible validation results will be exported. The export ends with a confirmation message.

If a selected destination file exists, a user message is asks whether the export should continue.

The button 'Import' can be used to import a previously exported file.

Hint: Validation results can also be written into a file or they can be sent to other logging monitors. Chapter 9: “Logging with Log4j2” (p. 32) gives some information about the configuratioin of Log4j2.

Chapter 4. Validation Constraints in Excel Files

Rules for validating PDF documents can be written in Excel files. Their structure and the provided validation functions are described in the following chapters.

Structure of an Excel File

An Excel file is searched for sheets with the following names:

Excel Sheet Name	Comment
regions	Definitions of page regions
check	Definition of test cases for individual PDF documents
compare	Definition of test cases comparing the document under test with a reference document

The expected structure of all three sheets is described in the following sections. In all three sheets a star '*' in the first column indicates a commented line.

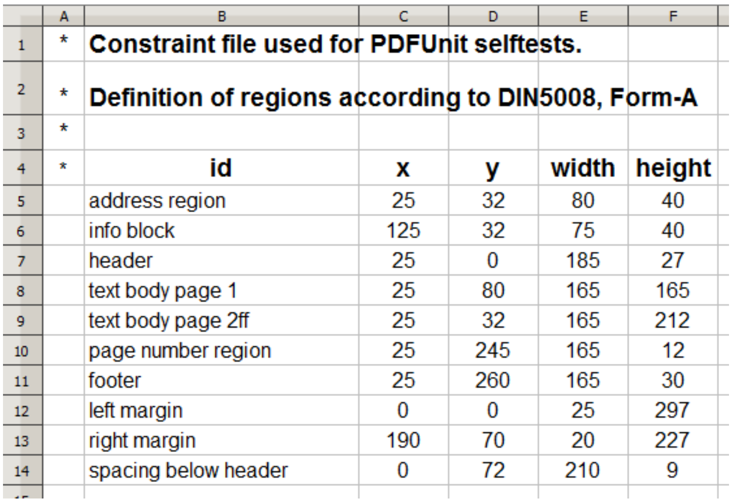
The order of the columns must not be changed. Additional columns after the expected ones are allowed. Additional sheets are also allowed.

A sheet can have empty lines. However, when a sheet has too many of them, the last lines with data may not be read. It's therefore better to use the character '*' in the first column of each empty line.

Definition of Page Regions - Sheet 'regions'

Mostly, you will need to restrict a validation rule to a region of a page, not just to the whole page. For example, it makes no sense to compare the text of full pages between two documents when the text contains a date. So, PDFUnit requires that each test case references a defined page region.

A page region is given by 4 values: the x and the y coordinates of the upper-left corner, as well as the width and the height of the region. All values are interpreted in millimeters. The values may have decimal values, but PDFUnit rounds the decimals to the next integer values. The following image shows examples:



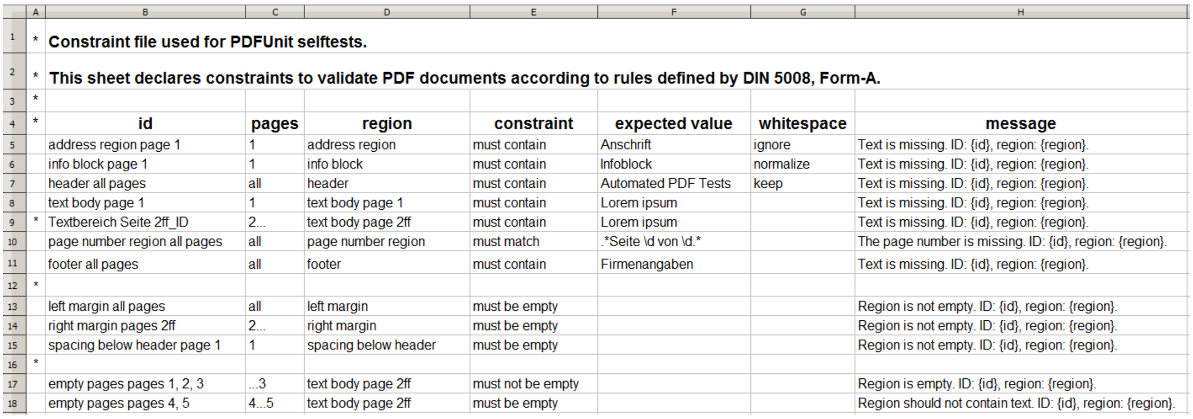
	A	B	C	D	E	F
1	*	**Constraint file used for PDFUnit selftests.**				
2	*	**Definition of regions according to DIN5008, Form-A**				
3	*					
4	*	**id**	**x**	**y**	**width**	**height**
5		address region	25	32	80	40
6		info block	125	32	75	40
7		header	25	0	185	27
8		text body page 1	25	80	165	165
9		text body page 2ff	25	32	165	212
10		page number region	25	245	165	12
11		footer	25	260	165	30
12		left margin	0	0	25	297
13		right margin	190	70	20	227
14		spacing below header	0	72	210	9

You can see that the sheet contains the column `id` in addition to the 4 columns with numeric values. This ID has to be unique and will be referenced by the test case definitions of the Excel sheets 'check' and 'compare'.

Test Cases for Single PDF Documents - Sheet 'check'

The sheet 'check' must be used to define test cases which relate to a single PDF document. It does not cover test cases in which two documents are compared with each other. The existing columns are:

Column Name	Comment
id	Name (ID) of the test case
pages	Pages this test should be used on
region	Name of a page regions which is defined in the Excel sheet 'regions'
constraint	Kind of validation. The allowed values are described below.
expected value	The expected value, if a validation needs one.
whitespace	This column contains a value indicating how whitespace will be handled. The allowed values are described below.
message	In this column an error message with placeholders can be defined. The placeholders are also described below.



	A	B	C	D	E	F	G	H
1	*	**Constraint file used for PDFUnit selftests.**						
2	*	**This sheet declares constraints to validate PDF documents according to rules defined by DIN 5008, Form-A.**						
3	*							
4	*	**id**	**pages**	**region**	**constraint**	**expected value**	**whitespace**	**message**
5		address region page 1	1	address region	must contain	Anschrift	ignore	Text is missing. ID: {id}, region: {region}.
6		info block page 1	1	info block	must contain	Infoblock	normalize	Text is missing. ID: {id}, region: {region}.
7		header all pages	all	header	must contain	Automated PDF Tests	keep	Text is missing. ID: {id}, region: {region}.
8		text body page 1	1	text body page 1	must contain	Lorem ipsum		Text is missing. ID: {id}, region: {region}.
9	*	Textbereich Seite 2ff_ID	2...	text body page 2ff	must contain	Lorem ipsum		Text is missing. ID: {id}, region: {region}.
10		page number region all pages	all	page number region	must match	.*Seite \d von \d.*		The page number is missing. ID: {id}, region: {region}.
11		footer all pages	all	footer	must contain	Firmenangaben		Text is missing. ID: {id}, region: {region}.
12	*							
13		left margin all pages	all	left margin	must be empty			Region is not empty. ID: {id}, region: {region}.
14		right margin pages 2ff	2...	right margin	must be empty			Region is not empty. ID: {id}, region: {region}.
15		spacing below header page 1	1	spacing below header	must be empty			Region is not empty. ID: {id}, region: {region}.
16	*							
17		empty pages pages 1, 2, 3	...3	text body page 2ff	must not be empty			Region is empty. ID: {id}, region: {region}.
18		empty pages pages 4, 5	4...5	text body page 2ff	must be empty			Region should not contain text. ID: {id}, region: {region}.

Pages to which a Test Case is Restricted - Column 'page'

A test case definition is often restricted to individual pages of a document. The following list shows all available syntax elements:

Pages	Syntax in Excel
a single page	1
multiple, individual pages	1, 3, 5
all pages	all
all pages after a given page (inclusive)	2...
all pages before a given page (inclusive)	...5
all pages between (inclusive)	2...5

Two page numbers must be separated by a blank. A comma is optional.

Different Ways of Comparing Text - Column 'constraint'

Values in the column 'constraint' in the sheet 'check' are used for specifying how the actual content of a document and the expected text will be compared. The following list shows the allowed values:

Keyword	Behaviour
'must contain'	The text in the column 'expected value' must be part of the page region. Additionally, this constraint type requires whitespace handling information.

Keyword	Behaviour
'must not contain'	The text in the column 'expected value' must not exist in the page region of the document. Additionally, this constraint type requires whitespace handling information.
'must be empty'	The referenced region must not contain any text.
'must not be empty'	The referenced region must have text.
'must match'	The text in the column 'expected value' will be taken as a regular expression and executed against the text in the referenced region. At least one piece of text must match.
'must not match'	The text in the column 'expected value' will be executed as a regular expression against the text in the referenced region. The test is successful if no match is found.

The column 'constraint' must not be empty. In such a case, PDFUnit throws an error message.

The sheet 'compare' may have other values in the column 'constraint'. Those values are described later below.

Validate Signatures and Images - Column 'constraint'

The column 'constraint' in the sheet 'check' can also contain keywords for validating signatures or images in a document:

Keyword	Behaviour
'is signed'	The PDF document has to be signed.
'is signed by'	A PDF document has to be signed. The expected name of the signatory has to be put into the column 'expected value'.
'has number of images'	The number of all visible images in the referenced page region will be compared with the number of the expected images, which must be put into the column 'expected value'.

Handling whitespaces - Column 'whitespace'

Text comparisons may fail due to differences in the whitespace. For example, text which is rendered in different fonts may have line breaks at different positions.

Keyword	Behaviour
'ignore'	All whitespaces are removed from the text before two strings are compared.
'keep'	Whitespaces will be taken 'as is'.
'normalize'	Whitespace at the beginning or the end of a text are deleted. Multiple whitespaces between words are reduced to one blank.

Wrong values in the column 'whitespace' will cause an error message. If the 'whitespace' column is left blank, the program defaults to 'normalize'.

Some validations, such as comparing bookmarks, are independent of whitespaces. For such validations, any declaration of whitespace handling will be ignored.

Expected Value - Column 'expected value'

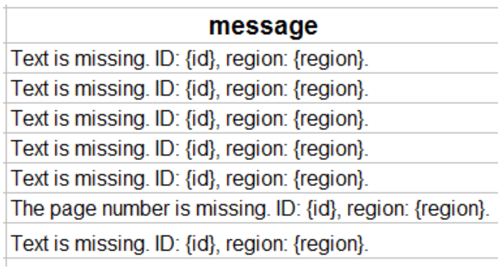
For a test case to verify that a region of a test document contains an expected value, the Excel file must provide a column for that value. This column is named 'expected value'.

If the 'constraint' column has the values 'must match' or 'must not match', the contents of the 'expected value' column are used as a regular expression. More information about regular expressions can be found online, for example on Wikipedia.

If the column 'constraint' has the value 'has number of images', then the content of the 'expected value' column will be parsed to an integer value.

Error Messages with Placeholders - Column 'message'

Individual error messages can be defined in the Excel sheets. These messages are shown in addition to PDFUnit's validation messages. An error message in the Excel sheet may have placeholders for runtime data. The following image shows some examples:



message
Text is missing. ID: {id}, region: {region}.
Text is missing. ID: {id}, region: {region}.
Text is missing. ID: {id}, region: {region}.
Text is missing. ID: {id}, region: {region}.
Text is missing. ID: {id}, region: {region}.
The page number is missing. ID: {id}, region: {region}.
Text is missing. ID: {id}, region: {region}.

The image shows clearly that placeholders in a text are enclosed with curly brackets. The following placeholders can be used:

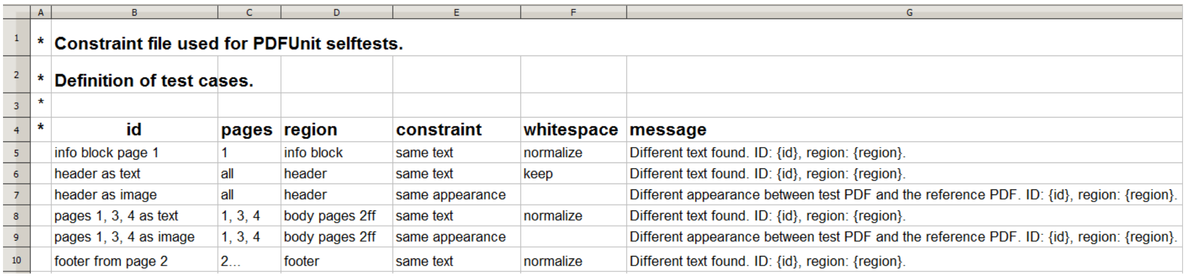
Placeholder	Meaning
{id}	The ID of the current test case
{pages}	The page-number of the page where the error is detected
{region}	The value in the column 'region'
{constraint}	The value in the column 'constraint'

Placeholders can be used anywhere inside a text. The values of the placeholders at runtime are enclosed in single quotation marks, so their is no need to use single quotes in the error messages in the Excel sheet.

Test Cases Comparing Two Documents - Sheet 'compare'

This Excel sheet can be used to declare validation rules for comparing two PDF documents. One document is the 'document under test' and the second document is a reference document.

For comparative testing no information about an expected text need be given in the Excel sheet, so the column 'expected value' is not provided in the sheet 'compare'.

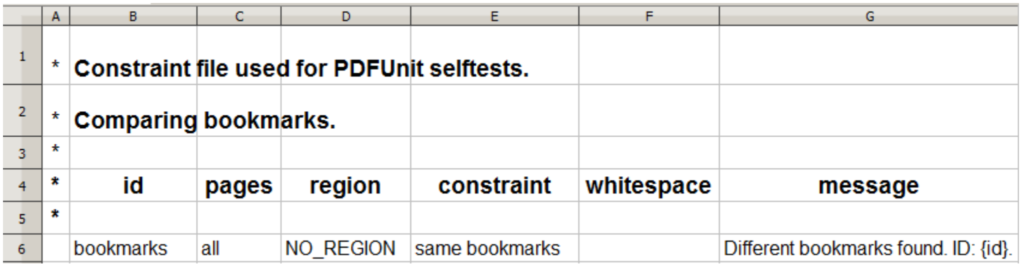


	A	B	C	D	E	F	G
1	*	**Constraint file used for PDFUnit selftests.**					
2	*	**Definition of test cases.**					
3	*						
4	*	**id**	**pages**	**region**	**constraint**	**whitespace**	**message**
5		info block page 1	1	info block	same text	normalize	Different text found. ID: {id}, region: {region}.
6		header as text	all	header	same text	keep	Different text found. ID: {id}, region: {region}.
7		header as image	all	header	same appearance		Different appearance between test PDF and the reference PDF. ID: {id}, region: {region}.
8		pages 1, 3, 4 as text	1, 3, 4	body pages 2ff	same text	normalize	Different text found. ID: {id}, region: {region}.
9		pages 1, 3, 4 as image	1, 3, 4	body pages 2ff	same appearance		Different appearance between test PDF and the reference PDF. ID: {id}, region: {region}.
10		footer from page 2	2...	footer	same text	normalize	Different text found. ID: {id}, region: {region}.

The meaning of columns is the same as described in the above sections for the sheet 'check'. But in the 'compare' sheet, other values are allowed in the 'constraint' column:

Keyword	Behaviour
'same text'	Two PDF documents must have the same text in the referenced region. Additionally, this constraint type requires whitespace handling information.
'same appearance'	The referenced regions of the two PDF documents must be identical when compared as rendered images.
'same bookmarks'	The two PDF documents must have the same bookmarks. Obviously this validation does not require page regions, but for technical reasons the column 'region' must not be empty. Instead, the value 'NO_REGION' should be supplied.

The image below shows a test comparing bookmarks of a 'PDF under test' with a reference PDF:



	A	B	C	D	E	F	G
1	*	Constraint file used for PDFUnit selftests.					
2	*	Comparing bookmarks.					
3	*						
4	*	id	pages	region	constraint	whitespace	message
5	*						
6		bookmarks	all	NO_REGION	same bookmarks		Different bookmarks found. ID: {id}.

PDFUnit searches the reference documents for a given PDF document in a subdirectory of its folder. The subdirectory has to have the name 'reference'. The filename of the reference has to be the same as the PDF under test.

Error Messages at Runtime

The validation of a PDF document does not end with the first detected failure. All defined rules of an Excel file are processed, and then an error message is created for each detected failure.

Chapter 5. Install, Update, Uninstall

5.1. Technical Requirements

PDFUnit needs Java 7 or higher.

Tested Environments

PDFUnit was successfully tested in these environments:

Operating System	Java Version
• Windows-7, 32 + 64 Bit	• Oracle JDK-1.7, 32 + 64 Bit
• Kubuntu Linux 12/04, 32 + 64 Bit	• Oracle JDK-1.8, Windows, 32 + 64 Bit
• Mac OS X, 64 Bit	• IBM J9, R26_Java726_SR4, Windows 7, 64 Bit

More combinations of Java and operating systems will be tested in the future.

If you have any problems with the installation, please contact us at info[at]pdfunit.com.

5.2. Installation

Downloading of PDFUnit-Monitor

Download the file `pdfunit-monitor-<VERSION>.zip` from the project's site: .

Unzip the file into a folder, for example into `PROJECT_HOME/lib/pdfunit-monitor-VERSION`. In the following text that folder is referred to as PDFUNIT_MONITOR_HOME.

Start Script and Classpath

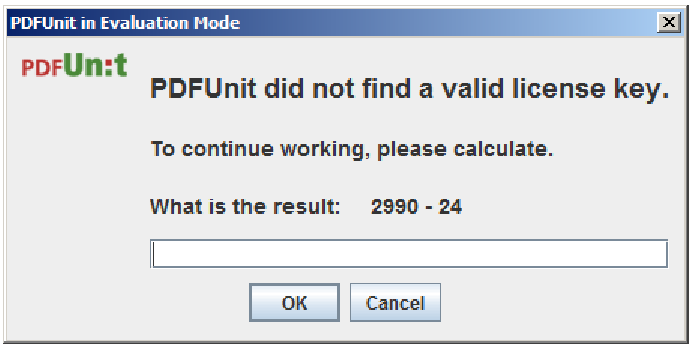
The folder PDFUNIT_MONITOR_HOME contains a script named `startPDFUnitMonitor.bat` to start the PDFUnit-Monitor under Windows and `startPDFUnitMonitor.sh` to start it under Unix.

If you want to store PDFUnit-Java in a directory other than `PDFUNIT_MONITOR_HOME/lib`, you have to change the CLASSPATH. Look for the variable PDFUNIT_JAVA_HOME in the start script.

If you have licensed PDFUnit, add the folder containing the license key file `license-key_pdfunit-java.lic` to the classpath.

Using PDFUnit-Monitor without a License Key

You are welcome to evaluate PDFUnit without a license. In that case the number of tests is limited to 50. And you have to solve a simple math question which is shown in a message box. If you calculate successfully, the test will run, otherwise you have to restart your test and calculate again.



PDFUnit did not find a valid license key.

To continue working, please calculate.

What is the result: 2990 - 24

Sometimes, the message box is covered by other applications and PDFUnit-Monitor is blocked. Minimize all applications to look for the message box.

Order a License Key

The PDFUnit-Monitor and PDFUnit-Java need a license when they are used in a commercial context. Write a mail to info[at]pdfunit.com and ask for a license. You will receive an answer as soon as possible.

The license fee is calculated individually. A small company should not pay as much as a big company, and someone testing only a few PDF documents, of course, pays less. And if you want to get a free license, give us some reasons. It is not impossible.

Using License Key

If you have ordered a license you will receive a ZIP file and a separate file `license-key_pdfunit-java.lic`. Install the contents of the ZIP file as described above and include the license file in your classpath.

Any change to the license file makes it unusable. If this happens, contact PDFUnit.com and ask for a new license file.

Verifying the Installation

If you have a problem with the configuration, start the script `verifyInstallation.bat` or `verifyInstallation.sh`. You will get a detailed problem analysis. See chapter 5.4: "Verifying the Configuration" (p. 24).

5.3. Set Paths Using System Properties

The files `pdfunit-monitor.config` and the license key file need not be mentioned in the classpath. But then their locations have to be declared using JVM arguments. The keys are:

- `-Dpdfunit.configfile`
- `-Dpdfunit.licensekeyfile`

Dependent on your test environment (Eclipse, ANT or Maven) these parameters can be set in many ways. You can set properties in the start script `startPDFUnitMonitor.bat` (*.sh). As a less known option you can use the operating system environment variable `_JAVA_OPTIONS`:

```
set _JAVA_OPTIONS=-Dpdfunit.configfile=..\myfolder\pdfunit-monitor.config
```

If you have questions about this topic, write a mail to info[at]pdfunit.com.

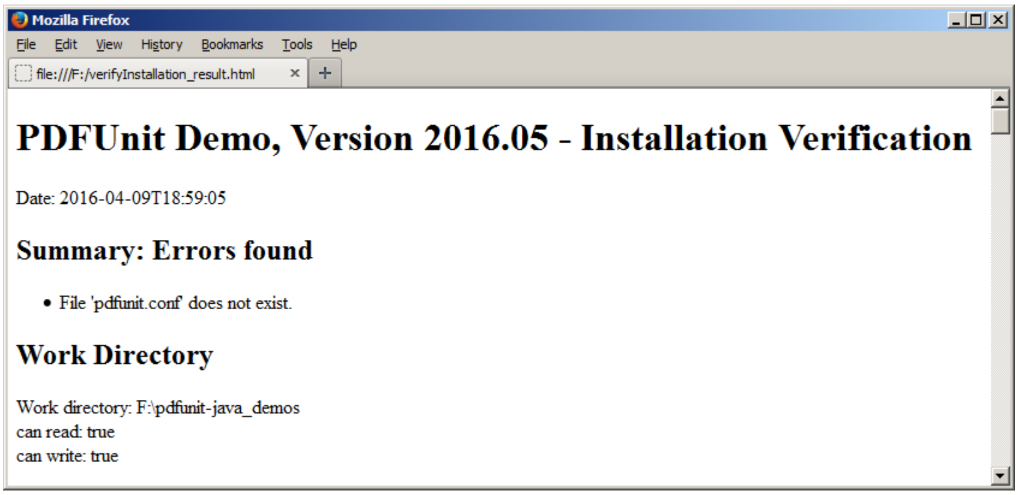
5.4. Verifying the Configuration

The installation of PDFUnit can be checked using a special program, started with the script `verifyInstallation.bat` or `verifyInstallation.sh`:

The script produces the following output on the console:

```
Checking installation ...
... finished. Report created, see 'verifyInstallation_result.html'.
```

The resulting report file shows errors and information about the classpath, environment variables and other runtime related data:



Mozilla Firefox
File Edit View History Bookmarks Tools Help
file:///F:/verifyInstallation_result.html

PDFUnit Demo, Version 2016.05 - Installation Verification

Date: 2016-04-09T18:59:05

Summary: Errors found

- File 'pdfunit.conf' does not exist.

Work Directory

Work directory: F:\pdfunit-java_demos
can read: true
can write: true

5.5. Update to a New Version

The installation of a new release of PDFUnit-Monitor runs just like the initial installation, because PDFUnit is always delivered as a full release, never as an incremental release.

Getting the New Release

If you use PDFUnit without a license, download the new ZIP file from the internet: .

If you use PDFUnit with a license, you will receive a new release by mail with an attached ZIP file and a license file.

Before Update

Before you start installing the new release, run all existing unit tests with the old release. They should be “green”.

Save your project.

Install the Update

Unzip the new release into a new folder and read the file `release-notes.txt`.

Copy the file `pdfunit-monitor.config` from the directory with the old release into the folder with the new release. You will find an entry in the `release-notes.txt` if the structure of the config file has changed in the new release. In such a case, transfer your individual configuration from the old release into the new config file.

If you use PDFUnit with a license, copy the new license file `license-key_pdfunit-java.lic` into the classpath of the new release.

Last Step

Run your existing tests with the new release. If there are no documented incompatibilities between the new and the old release, the tests should run “green” again. Otherwise, read the release information.

5.6. Uninstall

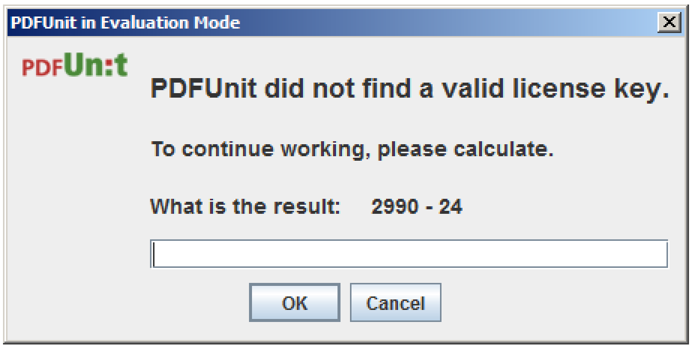
PDFUnit can be uninstalled cleanly by deleting the installation directories. PDFUnit doesn't create any entries in the registry or system directories, so none need to be removed. Don't forget to remove the references to JAR files and PDFUnit directories from your own scripts.

Chapter 6. Starting the Application

The home directory of the PDFUnit-Monitor contains the start scripts `startPDFUnit-Monitor.bat` and `startPDFUnitMonitor.sh`. The monitor will start if all required JAR files can be found.

Running in Evaluation Mode

Starting the monitor with a license is the same as starting it in evaluation mode. But without a license, a small dialog box appears when a validation was started and the user has to solve a simple arithmetic question. This user interaction prevents the monitor from beeing used without a license in scripts.



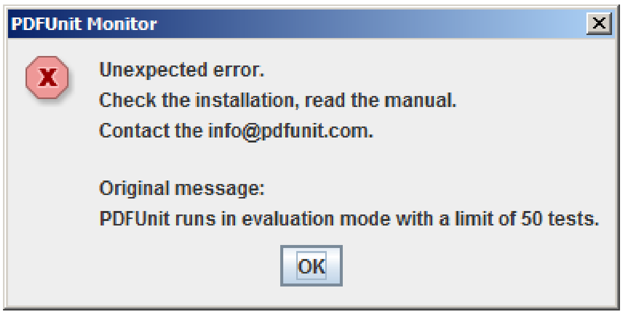
PDFUnit did not find a valid license key.

To continue working, please calculate.

What is the result: 2990 - 24

If the arithmetic challenge was solved, the PDFUnit-Monitor runs up to 50 validations. The validations are counted from the start of the monitor. If the arithmetic question was not answered correctly, the PDFUnit-Monitor needs to be restarted. Please note that the PDFUnit-Monitor is free of this restriction when it is used with a license file.

When the validation limit was reached, the following message appears:



Unexpected error.
Check the installation, read the manual.
Contact the info@pdfunit.com.

Original message:
PDFUnit runs in evaluation mode with a limit of 50 tests.

If the limit was reached, you have to restart the monitor for further tests.

Chapter 7. Configuration (pdfunit-monitor.config)

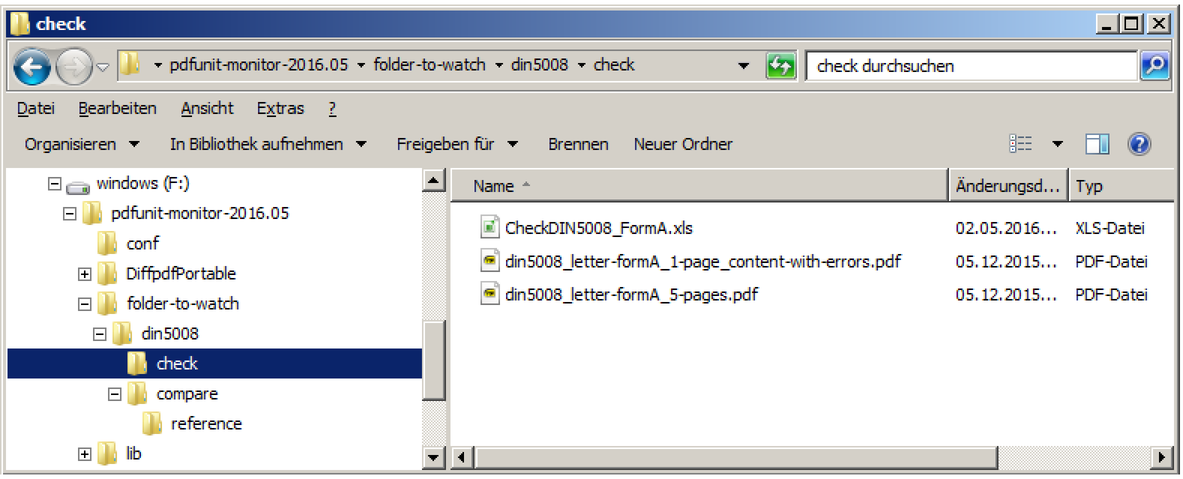
Watched folders

The configuration file `pdfunit-monitor.config` must contain the key `WATCHED_FOLDER`. Its value is the topmost directory to be monitored. All subdirectories of that directory will be watched.

The PDFUnit-Monitor searches for reference documents in a subdirectory `reference/` immediately below the directory which contains the PDF under test. A referenced document has to have the same name as the PDF under test.

The following are examples for the configuration and the corresponding file structure:

```
#
# Base folder for PDF document which are validated.
# All subfolders will be watched too.
#
WATCHED_FOLDER = ./folder-to-watch/din5008
```


check
pdfunit-monitor-2016.05 ▸ folder-to-watch ▸ din5008 ▸ check
check durchsuchen
Datei Bearbeiten Ansicht Extras ?
Organisieren ▾ In Bibliothek aufnehmen ▾ Freigeben für ▾ Brennen Neuer Ordner
windows (F:)
pdfunit-monitor-2016.05
conf
DiffpdfPortable
folder-to-watch
din5008
check
compare
reference
lib

Name	Änderungsd...	Typ
CheckDIN5008_FormA.xls	02.05.2016...	XLS-Datei
din5008_letter-formA_1-page_content-with-errors.pdf	05.12.2015...	PDF-Datei
din5008_letter-formA_5-pages.pdf	05.12.2015...	PDF-Datei

Logging Port

The PDFUnit-Monitor uses Log4j2 internally and therefore by default port 4445. If this port is used by another application, a different port has to be configured in the files `pdfunit-monitor.config` and `log4j2.xml`.

```
#
# The port defined here must be the same as in the configuration file
# for Log4J "log4j2.xml". Look for "SocketAppender" in that file.
#
PORT_TO_LISTEN = 4445
```

The file `log4j2.xml` is used to configure the logging framework. Details are described in chapter 9: “Logging with Log4j2” (p. 32).

Look and Feel

The appearance of the application can be changed. By default the typical Java-Look is used, but with an entry in the file `pdfunit-monitor.config` it can be changed into the typical view of the operating system:

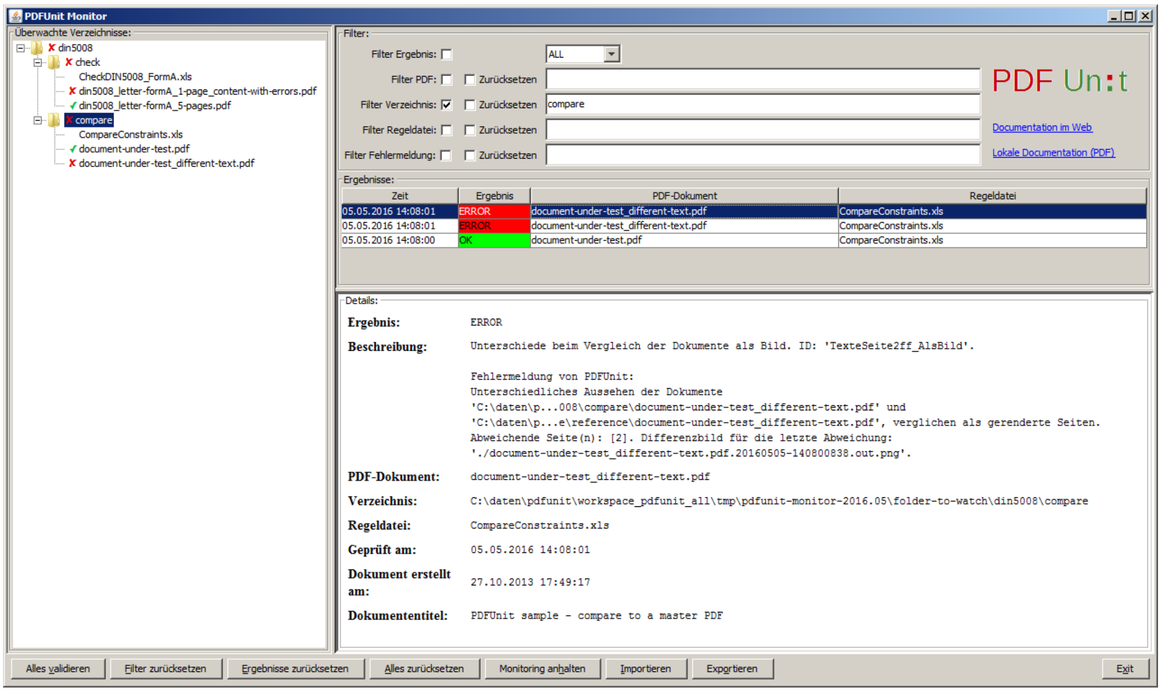
```
#
# Uncomment the next line to set the general look and feel of the operating system
#LOOK_AND_FEEL = System default
```

Language

The PDFUnit-Monitor's graphical user interface and error messages are delivered in English and German. The default language is the language of the operating system. But the monitor can explicitly be set to the other language by using the Java system properties. Those properties must be set before the application is started.

```
-Duser.language=en -Duser.country=UK
```

The following image shows the PDFUnit-Monitor in German and in the Windows-7 Look-and-Feel:



Filter:

Filter Ergebnis: □ ALL

Filter PDF: □ □ Zurücksetzen

Filter Verzeichnis: ☑ □ Zurücksetzen compare

Filter Regeldatei: □ □ Zurücksetzen

Filter Fehlermeldung: □ □ Zurücksetzen

Ergebnisse:

Zeit	Ergebnis	PDF-Dokument	Regeldatei
05.05.2016 14:08:01	ERROR	document-under-test_different-text.pdf	CompareConstraints.xls
05.05.2016 14:08:01	ERROR	document-under-test_different-text.pdf	CompareConstraints.xls
05.05.2016 14:08:00	OK	document-under-test.pdf	CompareConstraints.xls

Details:

Ergebnis:	ERROR
Beschreibung:	Unterschiede beim Vergleich der Dokumente als Bild. ID: 'TexteSeite2ff_AlsBild'. Fehlermeldung von PDFUnit: Unterschiedliches Aussehen der Dokumente 'C:\daten\p...008\compare\document-under-test_different-text.pdf' und 'C:\daten\p...e\reference\document-under-test_different-text.pdf', verglichen als gerenderte Seiten. Abweichende Seite(n): [2]. Differenzbild für die letzte Abweichung: './document-under-test_different-text.pdf.20160505-140800838.out.png'.
PDF-Dokument:	document-under-test_different-text.pdf
Verzeichnis:	C:\daten\pdfunit\workspace_pdfunit_all\tmp\pdfunit-monitor-2016.05\folder-to-watch\din5008\compare
Regeldatei:	CompareConstraints.xls
Geprüft am:	05.05.2016 14:08:01
Dokument erstellt am:	27.10.2013 17:49:17
Dokumententitel:	PDFUnit sample - compare to a master PDF

Installation Folder of DiffPDF

When two PDF documents have different content, it's possible to show the differences with a special diff-program, called 'DiffPDF'. This program, created by Mark Summerfield, is deployed with PDFUnit in a subdirectory of PDFUNIT_MONITOR_HOME. If you want to move it to another folder, you have to change the key `DIFFPDF_FOLDER` in the config file `pdfunit-monitor.config` Hint: The name of the folder must end with a forward slash.

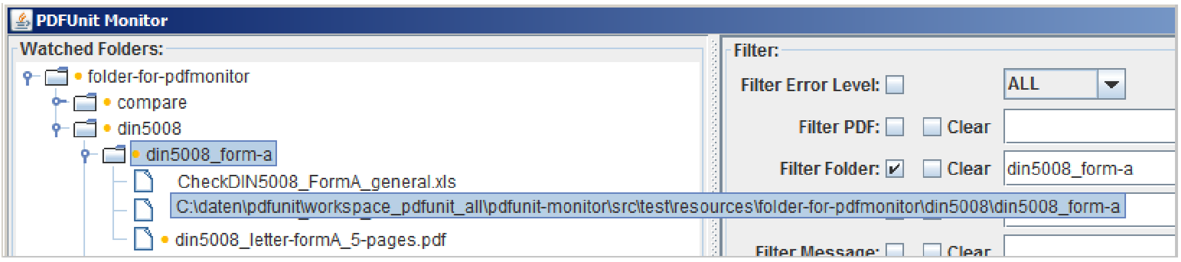
```
#
# Installation definitions for DiffPDF, the free Viewer.
# (Don't forget the trailing slash for the folder.)
# Information about DiffPDF are here: http://www.qtrac.eu/diffpdf.html
#
DIFFPDF_FOLDER     = DiffpdfPortable/
DIFFPDF_EXECUTABLE = DiffpdfPortable.exe
```

Delay for tooltip windows

The path of a file will be shown as a tooltip when the mouse cursor is over the filename. The delay until the tooltip was shown can be configured with the key `TOOLTIP_VISIBILITY_DURATION_MILLISECONDS` in `pdfunit-monitor.config`:

```
#
# The following property sets the wait time,
# before a tooltip is visible.
#
TOOLTIP_VISIBILITY_DURATION_MILLISECONDS = 1500
```

Here an example for the tooltip:



PDFUnit Monitor

Watched Folders:

- folder-for-pdfmonitor
  - compare
  - din5008
    - din5008_form-a
      - CheckDIN5008_FormA_general.xls
      - C:\daten\pdfunit\workspace_pdfunit_all\pdfunit-monitor\src\test\resources\folder-for-pdfmonitor\din5008\din5008_form-a
      - din5008_letter-formA_5-pages.pdf

Filter:

Filter Error Level: □ ALL

Filter PDF: □ □ Clear

Filter Folder: ☑ □ Clear din5008_form-a

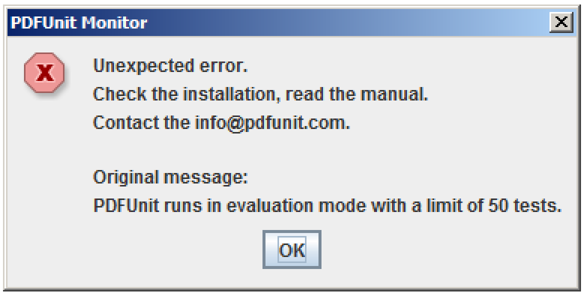
Filter Message: □ □ Clear

Chapter 8. Typical Errors

Errors may happen and good error messages will help to solve a problem. Some typical problems will be described in the following sections.

PDFUnit Evaluation Mode - Limit reached

50 test cases can be executed by the PDFUnit-Monitor when it runs in evaluation mode. Each following test case will fail with the error message:



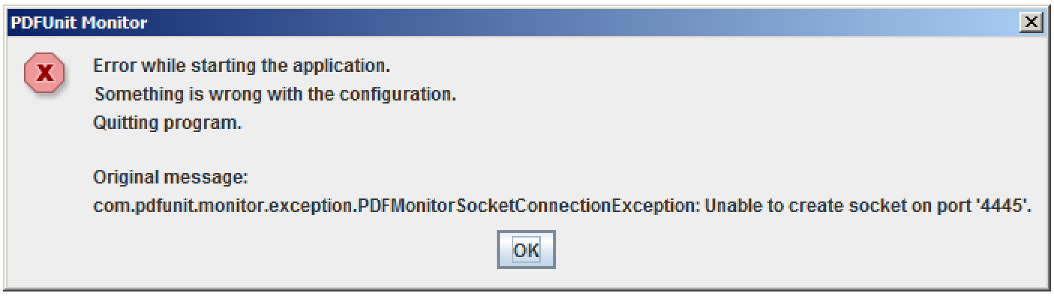
Unexpected error.
Check the installation, read the manual.
Contact the info@pdfunit.com.

Original message:
PDFUnit runs in evaluation mode with a limit of 50 tests.

You have to restart the application. Ask for a test license to evaluate PDFUnit without these messages. Write a mail to info[at]pdfunit.com.

Starting the Application Twice - Logging Port in Use

The internal logging appender 'PDFUnitMonitorAppender' (see log4j2.xml) sends the validation results to the user interface using port 4445. If the application was started a second time, the port is in use by the first started PDFUnit-Monitor and an error message is shown:



Error while starting the application.
Something is wrong with the configuration.
Quitting program.

Original message:
com.pdfunit.monitor.exception.PDFMonitorSocketConnectionException: Unable to create socket on port '4445'.

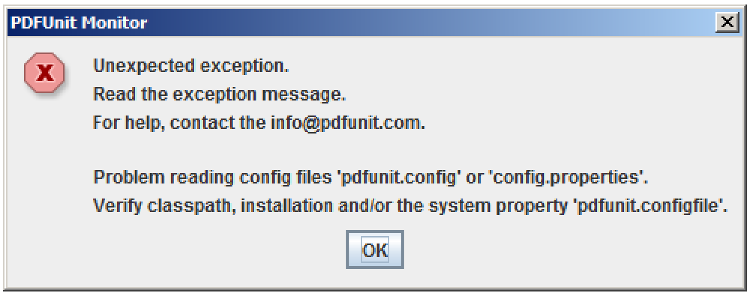
If this port is used by another application, change the port in `pdfunit-monitor.config` and in `log4j2.xml`.

Incomplete Rendering of the Directory Tree

Sometimes, a file name appears incomplete. By pressing the button 'Reset All', all tests run again and the view is refreshed. If you need to preserve the result, export it first.

Configuration `pdfunit.config` not in Classpath

If the config file `pdfunit.config` is not found in installation directory, the PDFUnit-Monitor sends the message:



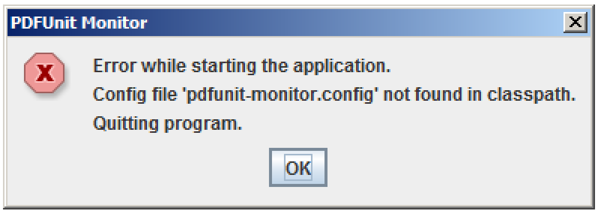
Unexpected exception.
Read the exception message.
For help, contact the info@pdfunit.com.

Problem reading config files 'pdfunit.config' or 'config.properties'.
Verify classpath, installation and/or the system property 'pdfunit.configfile'.

If you want to verify that all required files can be found in the classpath, run the script `verifyInstallation.bat` (oder `verifyInstallation.sh`) and read the report `verifyInstallation_result.html`. You should see 'Summary: No Errors'.

Configuration `pdfunit-monitor.config` not in Classpath

If the file `pdfunit-monitor.config` is not in the classpath, then the following message is shown immediately after you try to start PDFUnit-Monitor:



Error while starting the application.
Config file 'pdfunit-monitor.config' not found in classpath.
Quitting program.

Chapter 9. Logging with Log4j2

All validation results will be reported using the logging system 'Log4j version 2'. This may seem unnecessarily complex, but it allowes the monitoring to be extended by other logging appenders. Other appenders may log to a file or to a log management system like Graylog2 and LogStash.

Current Configuration

The configuration file `log4j2.xml` is located in the directory `PDFUNIT_MONITOR_HOME/conf`.

Log4j2 is configured by PDFUnit to send log events not only to the appender 'PDFUnitMonitorAppender' but also to a file appender which writes to `PDFUNIT_MONITOR_HOME/log/pdfunit-monitor.log`.

```
<Configuration monitorInterval="10">

  <Appenders>
    <File name="XmlFileAppender"
          fileName="./log/pdfunit-monitor.log"
          append="true"
          bufferedIO="false"
          bufferSize="100"
          immediateFlush="true"
    >
      <XMLLayout complete="true" charset="UTF-8" compact="false"/>
    </File>
    <Console name="ConsoleAppender" target="SYSTEM_OUT">
      <PatternLayout pattern="%d %-5p [%t] %C{2} (%F:%L) - %m%n"/>
    </Console>
    <Socket name="PDFUnitMonitorAppender" host="localHost" port="4445">
      <SerializedLayout />
    </Socket>
  </Appenders>

  <Loggers>
    <Logger name="PDFConstraints" level="all" additivity="false">
      <AppenderRef ref="XmlFileAppender"/>
      <AppenderRef ref="PDFUnitMonitorAppender"/>
<!--       <AppenderRef ref="ConsoleAppender"/> -->
    </Logger>
    <Root level="off">
      <AppenderRef ref="ConsoleAppender"/>
    </Root>
  </Loggers>

</Configuration>
```

Log-events can be filtered out by using a log level. By default, the level of the logger `PDFConstraints` is set to `all`. That means, all events will be sent to the result list. They can be filtered out from this list later by using the existing error-level filter.

Configure the Logging Port for Internal Use

The logging appender 'PDFUnitMonitorAppender' sends all validation events to the user interface using a socket. If its default port 4445 is beeing used by another application, change the port number in the files `pdfunit-monitor.config` and `log4j2.xml`.

Information about Log4j2

Further information about Log4j2 can be found in the Internet:

- Project home: http://logging.apache.org/log4j/2.x/index.html.
- Wiki: https://wiki.apache.org/logging/log4j2.
- Apache Log4j 2, User's Guide: log4j-users-guide.pdf.

Index