> **Note: This product is distributed on a 'try-before-you-buy' basis. All features described in this documentation are enabled. The registered version does not insert a watermark in your generated pdf documents.**

About Boxoft PDF Content Split

Boxoft PDF Content Split is a desktop utility program that lets you split Acrobat files into smaller pdf files base on location and text information within the files. It can be used to split composite PDF documents (such as invoices, records or salaries) to pieces by invoice number, account number or employees name which statement in a pdf file. Boxoft PDF Content Split will locate the position, extract text and compare text, then generate small file. It provides a power split rule editor to allow you define how the file will be split and how the split output files are uniquely named. You even can define and save multiple rules to apply different type files.

Boxoft PDF Content Split provides fast and flexible script to custom output filename, Boxoft PDF Content Split provides also a command line function, which can be called by script. It is useful when you want to split file automatically and/or periodically.

Boxoft PDF Content Split does NOT require Adobe Acrobat, and produces documents compatible with Adobe Acrobat Reader Version 5 and above.

Working Example for Boxoft PDF Content Split

Boxoft PDF Content Split   Feature List

- ➢ Generates multiple pdf's from a single pdf by splitting on textual content
- ➢ Splits on unique text by using word position searches
- ➢ Support defining text location information by sampling visualization
- ➢ Support Batch Pdf file processing. can split on text information within many PDF's
- ➢ Support defined splits rules that can be used again. The rules can be imported and exported for use in other computers.
- ➢ Support define output files name pattern.
- ➢ Support custom script for output files name (such as Splittag, Basename)
- ➢ Support password-protected pdf files
- ➢ Support command line.
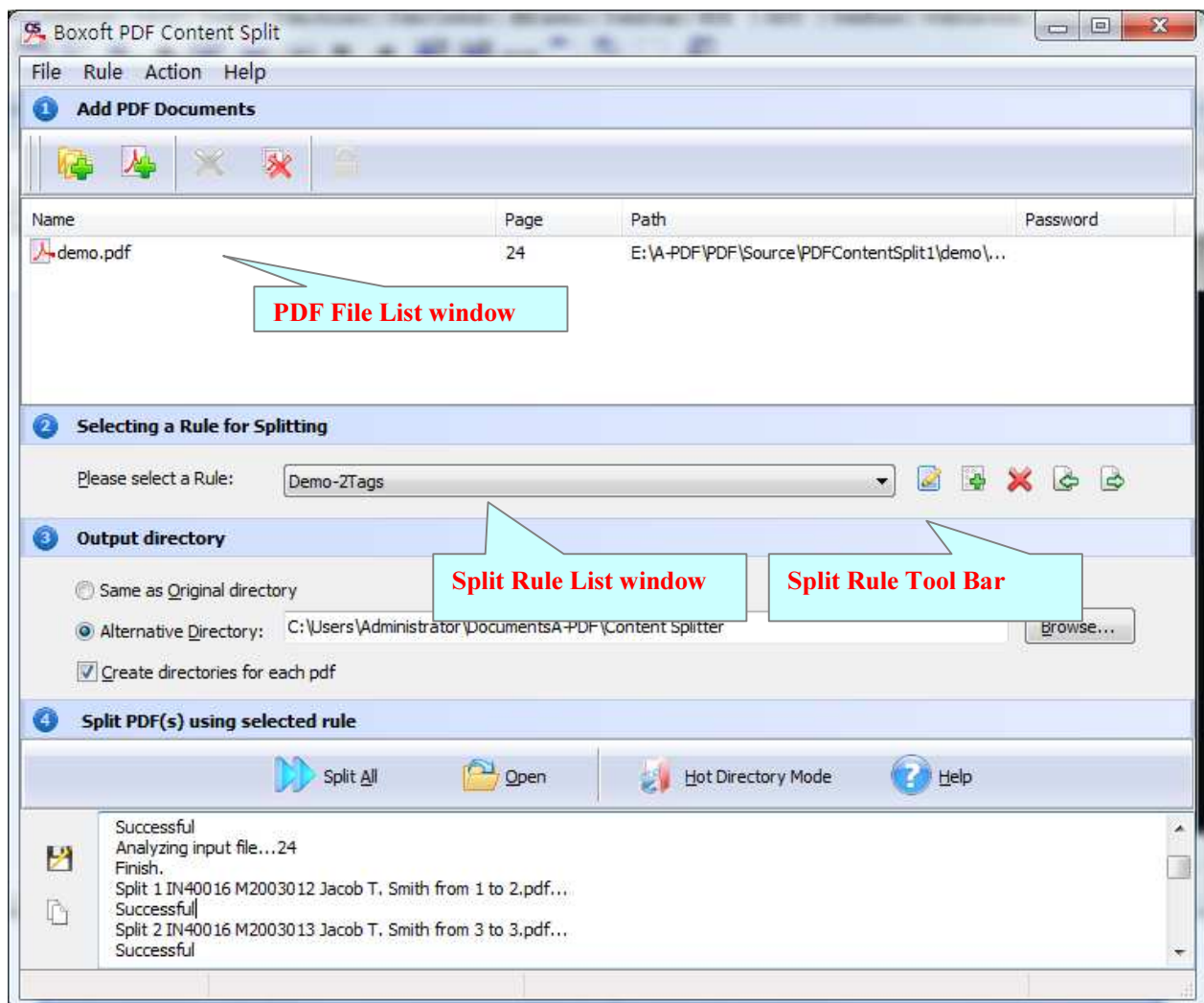- ➢ Allow splitting part of pages, e.g. page range etc

Using Boxoft PDF Content Split

Boxoft PDF Content Split can be started either by launching the application via a desktop shortcut, or directly from within Windows/Windows Explorer.

When you start the program, you will be presented with the primary screen from which all actions are taken.

Selecting PDF File

Clicking the **Add PDF Document** icon, or selecting **File** then choosing **Add PDF Document** will open the standard file browse and select window. From that window you can navigate to the folder and find the files you want to add to the list of files to split.

Alternatively, if you are using Windows Explorer to locate files, you have the convenience of being able to 'drag and drop' files into the document window.

Boxoft PDF Content Split even allows you to select all files in a particular directory/folder by using the **Add a Directory** option from the **File** menu! Add a directory also supports sub-folders.
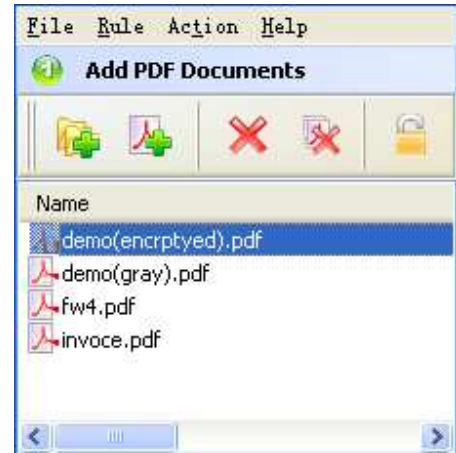
Each added file is displayed in the PDF File List window, as seen in the example below. You can view any of the files in the window by 'double-clicking' them. That will open pdf files in your default pdf reader program (usually Adobe Reader).

If any of the files you want to split are password-protected (security may be set within Adobe Acrobat for Read and other permissions), then before Adding Watermark to PDF files you need to provide the file password(s).

Files that contain passwords show a lock icon in left. Click the file name and the **Password** icon becomes active. Then clicking the **Password** icon produces an input box for each password required.
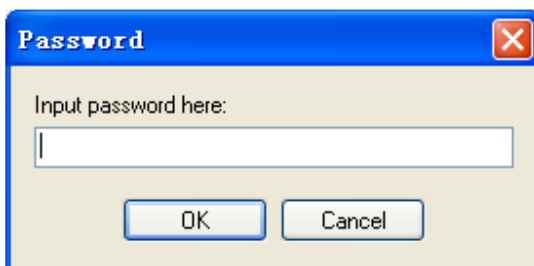
Select Split Rule

What is Split Rule? A Split Rule defines text **position** information within the sample pdf. Boxoft PDF Content Split according to the rule to search for text, compare text and then split the document up for you automatically. You can find an example to create a split rule below sections.

The Split Rule for applying to PDF(s) are listed in the **Split Rule List window**, where you can find rule items list in here, the current **Rule** you selected will to be applied to PDF(s)

Split Rule Tool Bar Function
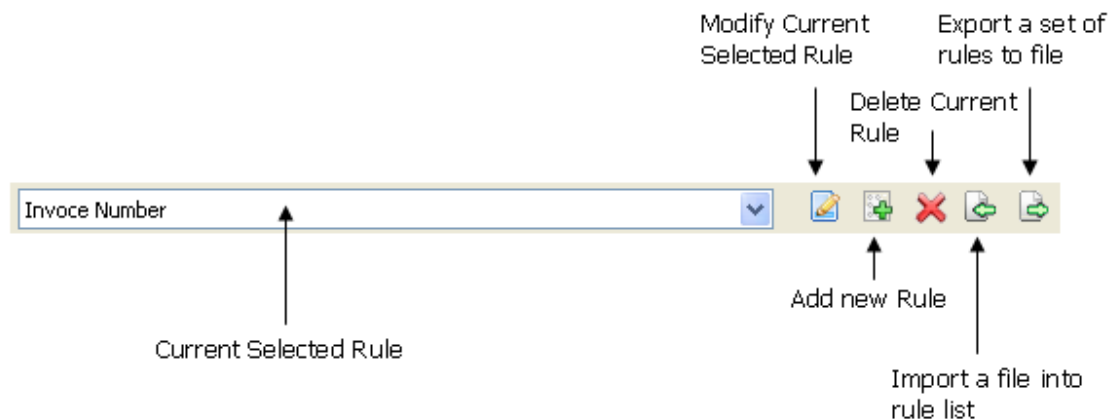
Set Output Directory

Same as Original directory: The spitted PDFs will be saved to the same as Original directory.

Alternative Directory: You can browse an alternative folder to save output

Create directories for each pdf: The option checked means Boxoft PDF Content Split will create folders for each source PDF under the folder specified above, then save the output files to proper folders.

Splitting PDF

Click the **Split All** button to split all of PDF files in the PDFs document windows. Boxoft PDF Content Split will to split PDF files base on the selected rule.

The log splitting will be shown in log list.

A simple example of how Boxoft PDF Content Split is used will demonstrate how easy and convenient this utility is for splitting pdf files into parts.

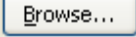## How to Create a Split Rule (Example)?

I have a PDF document generated by my financial management system (for example, "invoices.pdf"). The file contains hundreds of invoices. I want to split it on invoice number information within the PDF.

Example for creating and applying an "Invoice Number" rule

Step 1, In **Split Rules list window**, click the icon to open the **New Split Rule** configuration window. You can configure the split rule properties here. Once you set the desired options for the rule, you can click Save Button to save it.

Step 2, Click the **Browse** button in configuration window. You can select a pdf file (for example: "invoices.pdf") as sampling document. Normally the sampling document can be the pdf document that you want to split.

Step 3, In the opened **sample PDF view**. You can find the Invoice Number words (for example the sample words is: **90001155**) in it. Click left mouse button to select it. Then you will find have a Red dashed border on it.
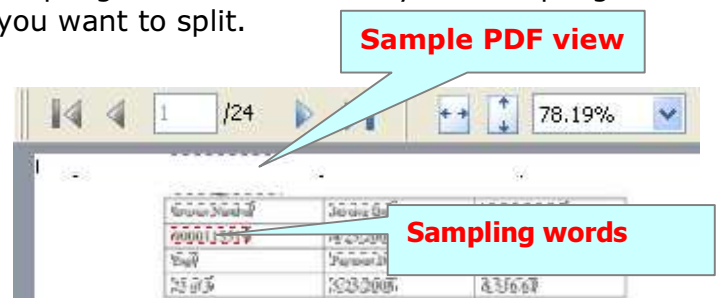
Step 4, Click the **Add to Split Tag List ->** button to add the **Position** of selected sample words to Split Tag list. Then open the **Add Samples to Split Tag List** configuration window. You can adjust the **Position deviation (X,Y)** to mitigate the impact of position deviation in sample PDF.

If the text align of sampling is left. The **Samples Align** value should be **"Left"** If the text align of sampling is middle. The **Samples Align** value should be **"middle"** and if the text align of sampling is right, the **Samples Align** value should be **"right"**.
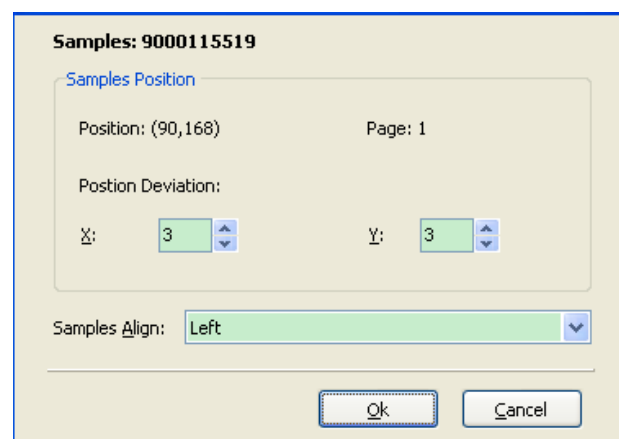
Next Click the **Ok** button to add the Samples to tag list (That means the split condition (in this case): **Finding unique "Invoice Number text" using text position searches (in same position (90,168) and in each page) until "Invoice Number text" changed)**

Position	Split Tag in position
(90,168)	Sample text : 9000115519, from page : 1)

Step5. Entering a name for the Rule (note the text that will appear in the Split Rules list, must be a unique name). (For example: Invoice Number)

Rule Name: Invoce Number

Step6. Boxoft PDF Content Split allows you to choose how the split output files will be named. The '**Filename pattern**' dropdown lets you choose from several 'patterns' of nomenclature so that each of the split files is logically named and can be easily identified later. **If you need to custom output filename flexible, please refer to "Script for Output"**

These name patterns are macros that automatically increment as necessary for the Final split files.

Ouput Filename

Filename pattern:

{basename}_{splittag}_({pagebegin}/{pageend})

Example: **invoice_9000115519_(1/24).pdf**

For example, choosing the {basename}.{part:0000} macro will name your split files from the unsplit filename and add sequential part numbers for each split file. 0000 is a place-holder and can be substituted with your own choice of size and starting part number. An example of this pattern and macro could name split files like example.part 0002.pdf. Choosing {basename}.{pagebegin:0000}-{pageend:0000} will generate split files which include the start and end page of the original unsplit file that each split file contains.

For even more flexibility, you can enter your own macro selection to give split files any rational name and sequence you want.

Patterns and their associated macros are a great convenience for naming your split files conveniently and logically with almost no effort on your part.

MACRO	DESCRIPTION
{basename}	The unsplit file name without extension name.
{splittag}	The text searched. For example, 900001234.
{part:0…0}	A sequential number from 1. 0...0 is place-holder. For example, you are using {part:0000}, will get 0001, 0002 etc. You can use other place-holder: A..A : the number will be uppercase letter, instead of digit, such as: A, B, C B..B: The number will be lowercase letter. Such as: a,b,c R..R: The number will be uppercase roman letter. Such as: I, II, III M..M: The number will be lowercase roman letter. Such as: i, ii, iii
{filecount:0..0}	The count of resultant files. The place-holder is same as above.
{pagebegin:0…0}	The output file begin page number in original file. The place-holder is same as above.
{pageend:0…0}	The output file end page number in original file. The place-holder is same as above.
{pagecount:0..0}	The page count in the file. The place-holder is same as above.

Step 7, A range of pages can be selected – either the entire contents of your original file, or only part of the file. This is a particularly useful option when you really only want a smaller set of split files produced from part of your original.

Step 8, Next click save button [Save] to save the '**Invoice Number**' Rule.

Step 9, Return to the main window. In the PDF File List window, click the **Add PDF Document**

icon to add PDF file which you what to split. (For example: "invoice.pdf")
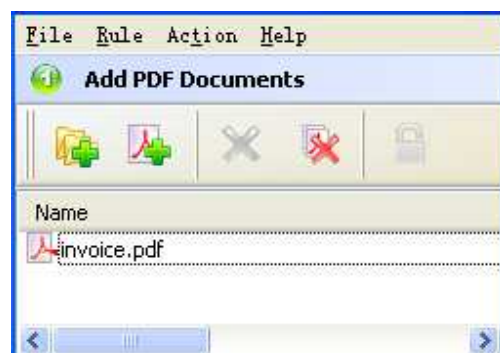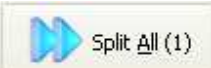
Step 10, In the **Split Rule List window,** select
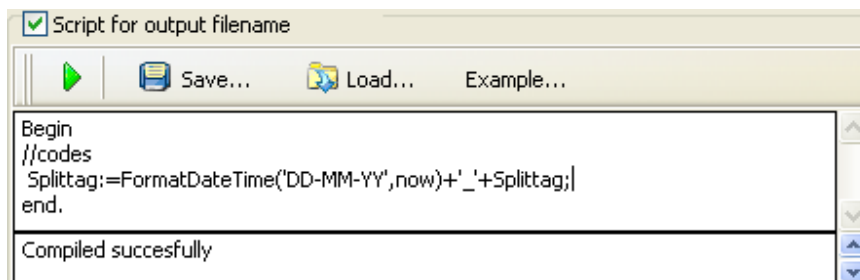
---

a rule (for example: Invoice Number)

② Selecting a Rule for Splitting

Please select a Rule: | Invoce Number

Step11. Click the **Split All** button Split All (1) to split, the output file will split on "Invoice Number" and the result as blew (splitting base on invoice number)

invoice_9000115519_(1_13).pdf
invoice_9000115520_(2_13).pdf
invoice_9000115521_(3_13).pdf
invoice_9000115522_(4_13).pdf
invoice_9000115523_(5_13).pdf
invoice_9000115524_(7_13).pdf
invoice_9000115525_(8_13).pdf
invoice_9000115526_(9_13).pdf
invoice_9000115527_(10_13).pdf
invoice_9000115528_(11_13).pdf
invoice_9000115529_(12_13).pdf
invoice_9000115530_(13_13).pdf

Script for Output

Boxoft PDF Content Split allows you to custom how the split output files will be named by using script. to enable script function, you need to check 'Script for Output Filename' option, then a script editor will be opened, which you can script 7 variables ('basename', 'part', 'pagebegin', 'pageend', 'pagerange', 'FileCount', 'SplitTag') in here.

☑ Script for output filename

▶   📄 Save...   📂 Load...   Example...

```
Begin
//codes
 Splittag:=FormatDateTime('DD-MM-YY',now)+'_'+Splittag;
end.
```

Compiled succesfully

Variable type table:

Variable name	Variable type
basename	string
SplitTag	string
pagerange	string
part	integer
pagebegin	integer
pageend	integer
FileCount	integer

A script for variable 'Basename' example here
// example #1 - Trim Basename
// Script code:
begin
 Basename:=trim(Basename);
end;
// example #1 end;

More examples please see: **CS-Script-Example.txt**

Ending Your Session
When you have completed your work, choose the **File** option and select **Exit** from the dropdown list.

More information please visit: http://www.Boxoft.com

Boxoft PDF Content Split Command Line Help

Boxoft PDF Content Split Command line (PCSCMD) is a Windows console utility that extracts pieces of PDF files based on the content position user defined. PCSCMD is a standalone program. It does not need Adobe Acrobat. You can call the command line at your script or batch file.

Boxoft PDF Content Split Command Line is included in the package installed.

USAGE

```
PCSCMD <input file> [-S<password>] <rule file> [-O<output dir>]


Parameters:
<input file>    : The PDF file to be split. Such as "C:\invoice.pdf"
<output dir>    : The output directory. If it is blank, will be the
                  "output" under current directory.
<rule file>     : Content Splitter Split Rule file name, PCSCMD will
                  split <input File> base on the rule, the rule can be
                  defined by  Boxoft PDF Content Split  GUI
<password>      : A password to open <input file> (if the pdf have open
                  Password Security). if the <input file> no Password
                  Security, it is not need.


Examples:

PCSCMD.exe "C:\invoice.pdf" "InvoiceNumber.rul"

PCSCMD.exe "C:\invoice.pdf" –Spwd "IN.rul" -o"C:\split" >c:\log.txt


Return Code:
 0: Split successful
 1: Show help message only. Parameters error.
 2: Input file does not exist.
 3: Rule file does not exist.
 4: Load input file(PDF file) error.
 5: Split file error.
 6: Create output directory failed.
 99: Unknown exception error.
```

More information please visit: http://www.Boxoft.com