



# 3-Heights™ PDF OCR Import Shell

## User's Manual

Version 1.91

---

Contact: [pdfsupport@pdf-tools.com](mailto:pdfsupport@pdf-tools.com)

Owner: **PDF Tools AG**  
Geerenstrasse 33  
CH-8185 Winkel  
Switzerland  
[www.pdf-tools.com](http://www.pdf-tools.com)

June 28, 2010

---

## Table of Contents

<b>1</b>	<b>Introduction.....</b>	<b>3</b>
1.1	Descriptions.....	3
1.2	Functions .....	3
	Features.....	3
1.3	About pdfovr.exe .....	3
<b>2</b>	<b>Installation .....</b>	<b>4</b>
<b>3</b>	<b>Getting started and User's Guide .....</b>	<b>5</b>
<b>4</b>	<b>Reference Manual .....</b>	<b>6</b>
4.1	Switches .....	6
-le	List available OCR Engines .....	6
-o	Set Owner Password .....	6
-ocl	Set OCR Language.....	6
-ocp	Set OCR Parameters .....	7
-ocr	Select an OCR Engine.....	7
-p	Set the Permission Flags.....	7
-pw	Password to read encrypted input File .....	8
-u	Set User Password .....	8
-v	Verbose Mode .....	8

# 1 Introduction

## 1.1 Descriptions

---

The 3-Heights™ OCR Enterprise Add-On compliments several 3-Heights™ products with a high performance optical character recognition (OCR) function. There are no page limits.

Even large archives can be quickly and reliably converted into PDF- or PDF/A-Files that can be searched in full text. Multiple languages are supported. Together with the corresponding basic product, the add-on ensures a reliable OCR functionality.

## 1.2 Functions

---

The 3-Heights™ OCR Enterprise Add-On is an OCR module, which is used as an option with several 3-Heights™ products. Based on the ABBYY FineReader Engine it recognizes text contents and embeds these as Unicode Text in the PDF- and PDF/A-File. This makes the PDF files full-text searchable. Numerous options in image manipulation, image pre-processing and text recognition allow a recognition process ideally coordinated to your needs. Almost 200 languages are supported; almost 50 languages are supported by dictionaries and morphologic tools.

### Features

- Recognition of machine generated texts
- Recognition of typewriter scripts and barcodes (1D)
- Image manipulation
- Image pre-processing

## 1.3 About pdfocr.exe

---

The purpose of this tool is to use it in combination with an optical character recognition (OCR) engine to make PDF documents searchable by performing OCR on embedded images.

The PDF OCR Import Shell is part bundled with the product 3-Heights™ Image to PDF Converter Shell.

OCR related features are handled equally as in the Image to PDF Converter.

## **2 Installation**

---

See manual 3-Heights™ Image to PDF Converter Shell:

[www.pdf-tools.com/public/downloads/manuals/i2ps.pdf](http://www.pdf-tools.com/public/downloads/manuals/i2ps.pdf)

### 3 Getting started and User's Guide

---

After the PDF OCR Import Shell and the 3-Heights™ OCR Enterprise Add-On are installed, you can list the available OCR Add-Ons to retrieve the name of the OCR engine using the command "pdfocr -le" as shown below:

```
pdfocr -le
List of available OCR engines:
- abbyy
- service
End of list.
```

The list should contain the two entries above: "abbyy" and "service". The entries in the list indicate the two Add-Ons "pdfocrpluginService.ocr" and "pdfocrpluginAbbyy.ocr" are found. The Add-Ons are required to communicate with the actual OCR-engine or service. Being able to list the Add-Ons does not necessarily mean the OCR-engine is installed and ready. How the OCR-engine is installed is described in the documentation "ocre.pdf".

Once the name (e.g. "abbyy") is known, it is provided as argument to the switch -ocr. The command following example is the basic command to apply OCR to a document. i.e. the input document *input.pdf* is read, OCR is applied, and the resulting, ocr'ed document is saved as *output.pdf*.

**Example:** Set the OCR engine to the "Abbyy FineReader 8.1 OCR Engine":

```
pdfocr -ocr abbyy input.pdf output.pdf
```

Additional OCR engine dependant settings or settings related to encryption are described in the chapter "Reference Manual".

## 4 Reference Manual

### 4.1 Switches

---

#### **-le List available OCR Engines**

OCR engines are accessed through the corresponding OCR interface DLLs. At present there is one OCR engine supported: Abbyy FineReader 8.1 OCR Engine. This engine is accessed by the OCR interface DLL *pdfocrAbbyy.ocr*.

The OCR interface DLL is provided by the *3-Heights™ Image to PDF Converter Shell*.

The OCR engine is provided as a separate product: *3-Heights™ OCR Enterprise Add-On*.

In order to make use of the OCR engine, the OCR interface DLL and the OCR engine must be installed. The switch `-le` lists all available OCR interface DLLs. It does not verify the corresponding OCR engines are installed and can be initialized. The OCR engine is loaded with the switch `-ocr`.

```
pdfocr -le
List of available OCR engines:
- abbyy
- service
End of list.
```

#### **-o Set Owner Password**

Set an owner password (password will be required to modify the PDF document security settings, such as permission flags or passwords).

**Example:** Set the owner password to "owner".

```
pdfocr -o owner input.pdf output.pdf
```

#### **-ocl Set OCR Language**

In order to optimize the performance of the OCR engine, it can be given hints what languages are used. The default language of the Abbyy FineReader 8.1 OCR Engine is English. This switch can only be used if the switch `-ocr` is set.

**Example:** Set the OCR languages to English and German.

```
pdfocr -ocr abbyy -ocl "English, German" input.pdf output.pdf
```

See also documentation for the *3-Heights™ OCR Add-On*.

June 28, 2010

**-ocp Set OCR Parameters**

Using this switch OCR engine specific parameters (key/value pairs) can be set to optimize the performance.

**Example:** Enable the balanced mode to improve the speed and do not detect whether text is bold or not.

```
pdfocr -ocr abbyy -ocp "BalancedMode=TRUE, DetectBold=FALSE" input.pdf
output.pdf
```

See also documentation for the 3-Heights™ OCR Add-On.

**-ocr Select an OCR Engine**

If a PDF document has to be made fully text searchable even if the text is part of a raster image then the images which are contained in the PDF document must be run through an OCR engine. With this switch the user can select an OCR engine, e.g. "Abbyy", and instruct the tool to embed the recognized text as a hidden layer on top of the image. If the add-in is not found or the engine cannot be initialized (because it is not installed or the license key is not valid) then an error message is issued.

The name of the OCR engine can be retrieved using the switch **-le**. If the switch **-ocr** is not used, no OCR is applied.

**Example:** Set the OCR engine to the "Abbyy FineReader 8.1 OCR Engine":

```
pdfocr -ocr abbyy input.pdf output.pdf
```

See also documentation for the 3-Heights™ OCR Add-On.

**-p Set the Permission Flags**

Set the permission flags. It is only usable in combination with encrypted documents. By default no permissions are granted. The permissions that can be granted are listed in the table: Permission Flags.

Table: Permission Flags	
Value	Description
<b>p</b>	low resolution printing
<b>m</b>	modify the document
<b>c</b>	copy objects
<b>o</b>	add or modify annotations
<b>f</b>	form filling
<b>s</b>	support disabilities
<b>a</b>	assembling
<b>d</b>	high quality printing

**Example:** The following command sets the owner password to "owner" and the permission flags to allow printing in low resolution (p) and allow form filling (f).

June 28, 2010

---

```
pdfocr -o owner -p pf input.pdf output.pdf
```

Note that "high quality printing" (*d*) requires the "low resolution printing" (*p*) flag to be set as well:

```
pdfocr -o owner -p pd input.pdf output.pdf
```

For further information about the permission flags, see PDF Reference Manual section 3.5.2.

### **-pw Password to read encrypted input File**

If the input file is encrypted with a user password (password required to open PDF document), then either the user or the owner password must be provided, or the document cannot be processed.

### **-u Set User Password**

Set a user password (password will be required to open the PDF document).

**Example:** Set the user password of the PDF document to "user".

```
pdfocr -u user input.pdf output.pdf
```

### **-v Verbose Mode**

Enable the verbose mode to output more detailed information about the processing steps.