

Manual

# PDF Extractor

---

Revised

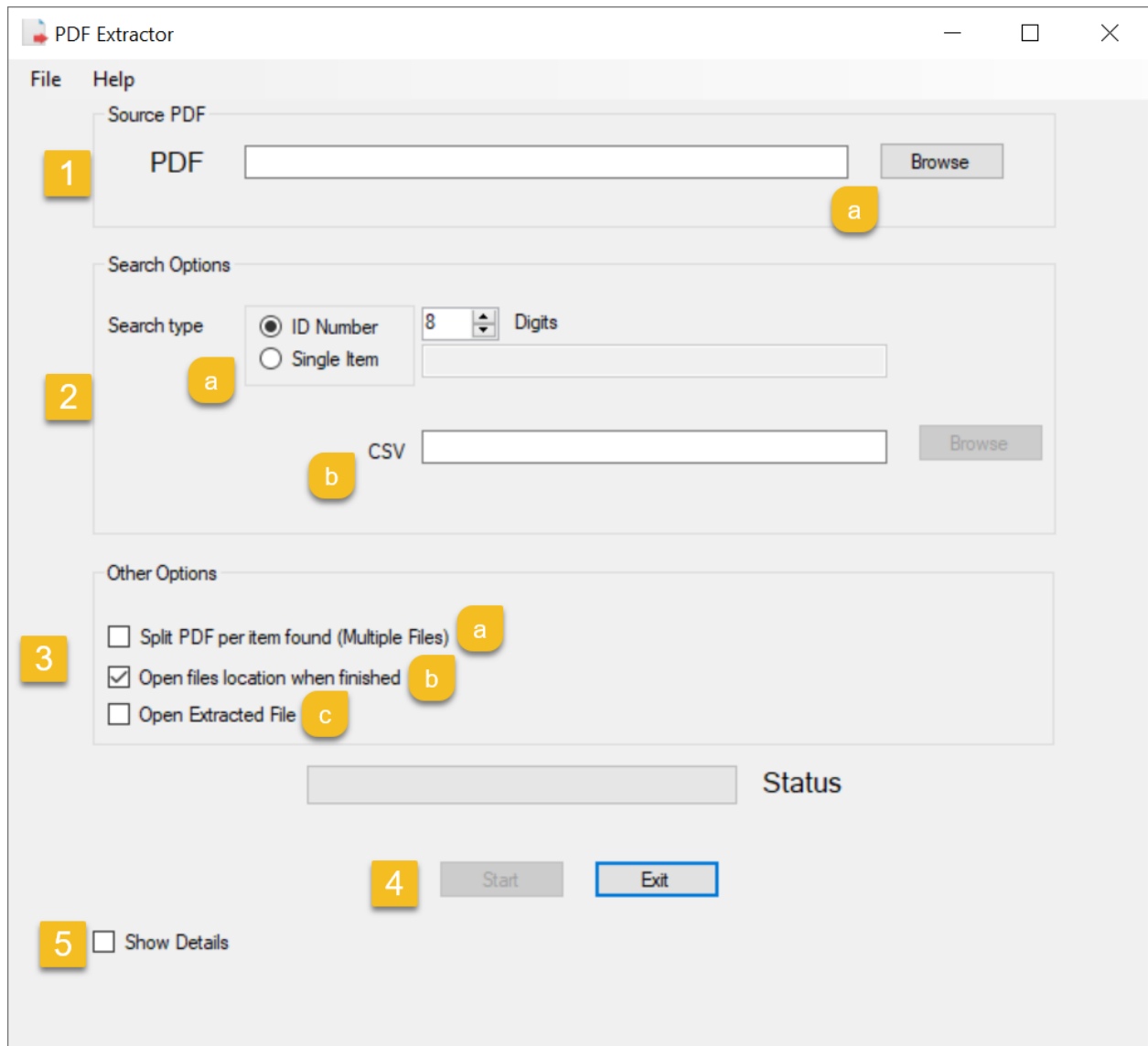
04/22/2022




## Introduction

The PDF extractor app was created for the extraction of pages from a single large file. The extraction is based on a CSV file containing all the 8 digits that need to be found or on a single phrase or test that may be typed manually. The application will extract the files matching a set number of digits that are unique per page or a specified word or phrase.

## Interface



- 
1. Source PDF File that contains multiple pages
    - a. Browse button for the main PDF file.
  2. Search options for the pages that are needed.
    - a. The application indexes the PDF file based on this criteria. By default the application will search for an 8 digit number and index that page. You can change the number of digits to index, by either typing the number of digits needed or by using the up or down arrows. If you would like to search by a specific word and phrase. You may do so by clicking on “Single Item”. This will allow you to search for a single phrase, text or number. Note that the application will match this exactly as it is typed.
    - b. If you are searching for a large number of digits, you must use a csv file that contains all the numbers you are looking for. The application will extract them in the order that they are in the file. Note: you do not need a heading for the column.
  3. Other Options
    - a. Split PDF per Item found will split the PDF into multiple files based on the item that was searched for. The application will automatically add the searched item to the file name.
    - b. When the application finishes processing and extracting the pages, this will automatically open the destination folder. (The same folder as the PDF)
    - c. Only when extracting a single item. This will automatically open the extracted PDF file with the system’s default PDF viewer.
  4. Start and exit. The start button will become active if all the correct settings are chosen. When you click “Start”, the application will begin processing the PDF.
  5. Clicking this box will expand a window where you can see the details on what the application is processing.

## Extracting a single file from a CSV list

- Browse for the PDF file (1a)
- Browse for the csv file (2b)

Once the files have been selected click on “Start” and the files will begin to be processed.

Sample CSV

	A	B	C	D
1	11111111			
2	22222222			
3	33333333			
4	44445444			
5				
6				

Settings:

The screenshot shows the 'PDF Extractor' application window. It has a menu bar with 'File' and 'Help'. The main area is divided into three sections: 'Source PDF', 'Search Options', and 'Other Options'. In the 'Source PDF' section, the 'PDF' field contains 'C:\Temp\testFile.pdf' and there is a 'Browse' button. In the 'Search Options' section, the 'Search type' is set to 'ID Number' with a value of '8' and a 'Digits' label. There is also a 'CSV' field containing 'C:\Temp\testlist.csv' and a 'Browse' button. In the 'Other Options' section, there are three checkboxes: 'Split PDF per item found (Multiple Files)' (unchecked), 'Open files location when finished' (checked), and 'Open Extracted File' (unchecked). At the bottom, there is a 'Status' label, 'Start' and 'Exit' buttons, and a 'Show Details' checkbox.

PDF Extractor

File Help

Source PDF

PDF C:\Temp\testFile.pdf Browse

Search Options

Search type

☒ ID Number 8 Digits

☐ Single Item

CSV C:\Temp\testlist.csv Browse

Other Options

☐ Split PDF per item found (Multiple Files)

☒ Open files location when finished

☐ Open Extracted File

Status

Start Exit

☐ Show Details

## Extract based on a phrase or text

- Browse for the PDF file (1a)
- Click on single item and type your text (2a)

Once the info is entered “Start” and the file will begin to be processed.

### Settings

The screenshot shows the 'PDF Extractor' application window. It has a menu bar with 'File' and 'Help'. The main interface is divided into several sections:

- Source PDF:** A text field containing 'C:\Temp\testFile.pdf' and a 'Browse' button.
- Search Options:**
  - Search type:** Two radio buttons, 'ID Number' (unselected) and 'Single Item' (selected).
  - Search value:** A text field containing 'Test phrase'.
  - CSV:** A text field and a 'Browse' button.
- Other Options:** Three checkboxes:
  - ☐ Split PDF per item found (Multiple Files)
  - ☒ Open files location when finished
  - ☐ Open Extracted File
- Status:** A progress bar and a 'Status' label.
- Buttons:** 'Start' and 'Exit' buttons.
- Show Details:** A checkbox at the bottom left.



Licensed under GNU Affero General Public License v3.0

This application uses iText AGPL open source code.

