



PDF Text and Image Extraction

Xtractor is a versatile server-based tool for finding and extracting text and images from PDF files. A whole variety of PDF content suddenly becomes available such as addresses, invoice and report data. The library extracts text or images to match specified criteria, and then the results are used to automate document workflow, enhance usability and much more. The straight-forward .NET interface eliminates the complexities involved with coding. Xtractor may be deployed as a standalone application or used as a companion to ActivePDF Toolkit to maximize flexibility and further extend your document processes.



[TRY IT FOR FREE](#)

[CONTACT US](#)

Xtractor Features:

- **High Volume Batch Extraction**
Quickly locate and extract PDF-based text or images for use in other applications.
- **Automate document processing based on Content**
Analyze and evaluate content of PDF documents based on contextual information such as keywords, location and more.
- **Extract information for document classification and indexing**
Automated document inspection, extracting key information that can be used in categorizing or indexing documents for archiving, classification and more.
- **Streamline cross-functional document processing**
Process categorized information such as metadata, barcodes, or keywords used to control printing sorting or packaging.

- **Find and return text location**
- **Specify location and return text**
- **Search word or phrase**
- **Return preceding or trailing text**
- **Extract metadata for PDF indexing**
- **Extract Document or by page**
- **Extract TIFF, JPEG, BMP, or PNG**
- **Save results to memory or file**
- **Locate and extract invisible text**
- **Multithreaded**
- **.NET API**

Requirements:

- Operating System**
- Windows Server® 2012
 - Windows Server® 2008

Additional Software Requirements

- Microsoft .NET Framework

Code Snippet: Get Text by Keyword

```
//Open the PDF and extract 20 characters after
the string "INVOICE"
exResult = ext.OpenPDF(inputfile);
textResult = ext.GetTextRight("INVOICE", 1, 1, 20,
APXtractor.Xtractor.UnitType.Character);

//Write the results to today's invoice log file i
invoicelog.WriteLine(textResult.Text);
invoicelog.Close();

// Close file and exit clean
exResult = ext.ClosePDF();
ext = null;
textResult = null;
```