

**Investintech.com Inc.  
PDF2Text Command Line Tool  
User's Guide**

**April 19, 2007**  
**<http://www.investintech.com>**

Copyright 2007 Investintech.com Inc. All rights reserved.

# Contents

<b>Preface .....</b>	<b>1</b>
About This Documentation.....	1
Typographical Conventions Used in This Document.....	1
Getting More Information.....	1
Customer Service and Technical Support.....	1
Fax and Mailing Address .....	2
<b>PDF2Text Command Line .....</b>	<b>3</b>
What is Command Line? .....	4
What is PDF2Text Command Line?.....	4
System Requirements.....	4
Starting the Installation .....	5
Using PDF2Text .....	12
Using PDF2Text from Command Prompt .....	12
Conversion type .....	13
Schedule.....	14
Range .....	14
Input file(s) .....	14
Output file(s).....	15
Examples of Using PDF2Text .....	15
Example #2 .....	15
Example #3 .....	16
Example #4 .....	16
Troubleshooting .....	17
<b>Index .....</b>	<b>19</b>



## Preface

- Welcome to Investintech.com Inc. PDF2Text Command Line. This tool provides technologies for converting PDF files into the text file formats. This user's guide provides the basic information needed to use the command line tool.

## About This Documentation

This documentation describes the purpose, installation and usage of the PDF2Text command-line program that allows users to convert PDF files into various text format

## Typographical Conventions Used in This Document

The following table describes typographical and naming conventions used in this document:

<i>Font</i>	<i>Used for</i>	<i>Examples</i>
monospaced	Expected input from the user	pdf2Word -help
	Paths and filenames	c:\investintech\pdf2Word.exe
	Source code	MessageBox("ABC");
monospaced blue	Source code keywords	If
monospaced green	Source code comments	//this is a comment
monospaced bold	Method name	The <b>AboutBox()</b> method
'Single quoted bold'	Elements of the user interface like buttons, icons etc.	click on ' <b>Start</b> ' button, press ' <b>Enter</b> '
blue underlined	Internet links	<a href="http://www.investintech.com">http://www.investintech.com</a>

## Getting More Information

Investintech.com Inc. conducts research in data conversion technology and develops data conversion products utilizing results of the research. More information about our company is available at the company's website <http://www.investintech.com>.

## Customer Service and Technical Support

Investintech.com Inc. strives to provide the best possible technical support to its customers and prospective customers. If you would like personal assistance, please feel free to call us or send e-mail to Customer Service or Technical Support. We are available by phone during regular business hours and in the vast majority of cases, we will return our e-mail with an answer the same day it is received.

	<i>Customer Service</i>	<i>Technical Support</i>
Telephone:	+1 416 920 5884	+1 416 920 2539
E-mail:	<a href="mailto:cs@investintech.com">cs@investintech.com</a>	<a href="mailto:techsupport@investintech.com">techsupport@investintech.com</a>
Hours:	Our Business Hours are Monday to Friday 9am-6pm (Eastern Time GMT-5:00).	

## Fax and Mailing Address

Mailing Address: Investintech.com Inc.  
410-96 Spadina Avenue  
Toronto, ON  
M5V 2J6 Canada

Fax: +1 416 920 5848

# Investintech.com Inc. Software Development Kit User's Guide

**PDF2Text Command Line**

## ***What is Command Line?***

Microsoft Windows operating system is graphical user interface (GUI) operating system – commands are usually given by using mouse to interact with graphical elements on screen and results are visible through changes on graphics drawn on computer screen. Microsoft DOS is command line interface operating system – commands are typed on keyboard, executed when ‘**Enter**’ key on keyboard is pressed and results are visible in text mode on screen.

While GUI is generally better way of using computers, there are few cases where command line is much better solution for particular need – like when we want to repeat same command over and over again.

In order to get command line interface window in any version of Windows, click on ‘**Start**’ (usually at lower left side of screen), point to ‘**All Programs**’, then point to ‘**Accessories**’ and click on ‘**Command Prompt**’ (or ‘**MS DOS Prompt**’ in older versions of Windows).

## ***What is PDF2Text Command Line?***

PDF2Text Command Line is command line, shell or batch version providing a subset of features found in Able2Doc. It may be used to automate the conversion of multiple documents.

***The demo version of PDF2Text is available, free of charge, for download from our company website, <http://www.investintech.com> Installation Instructions for PDF2Text***

Below you will find step-by-step instructions to install PDF2Text Command Line on your system. You may want to print this instruction sheet for reference before beginning to install PDF2Text.

## **System Requirements**

The minimum computer system resources required to install and run PDF2Text are:

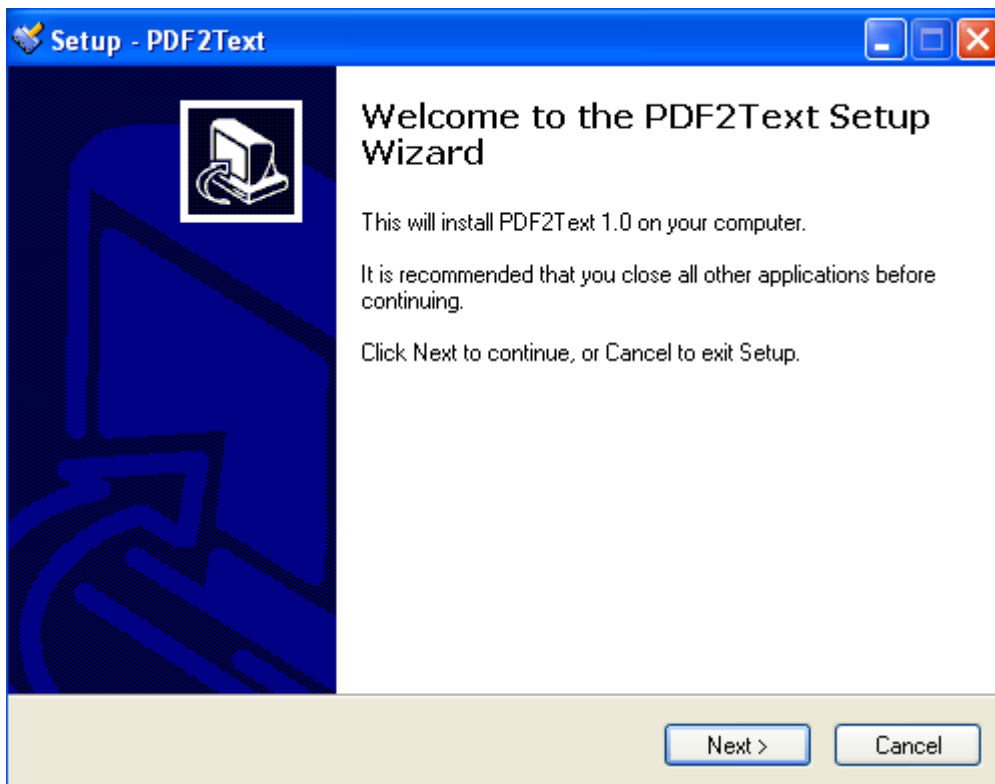
- Microsoft Windows 98/Me/2000/XP/2003
- 64 MB of RAM
- 4 MB of available hard disk space

In order to use PDF2Text comfortably, it is recommended that your computer system meet the following requirements:

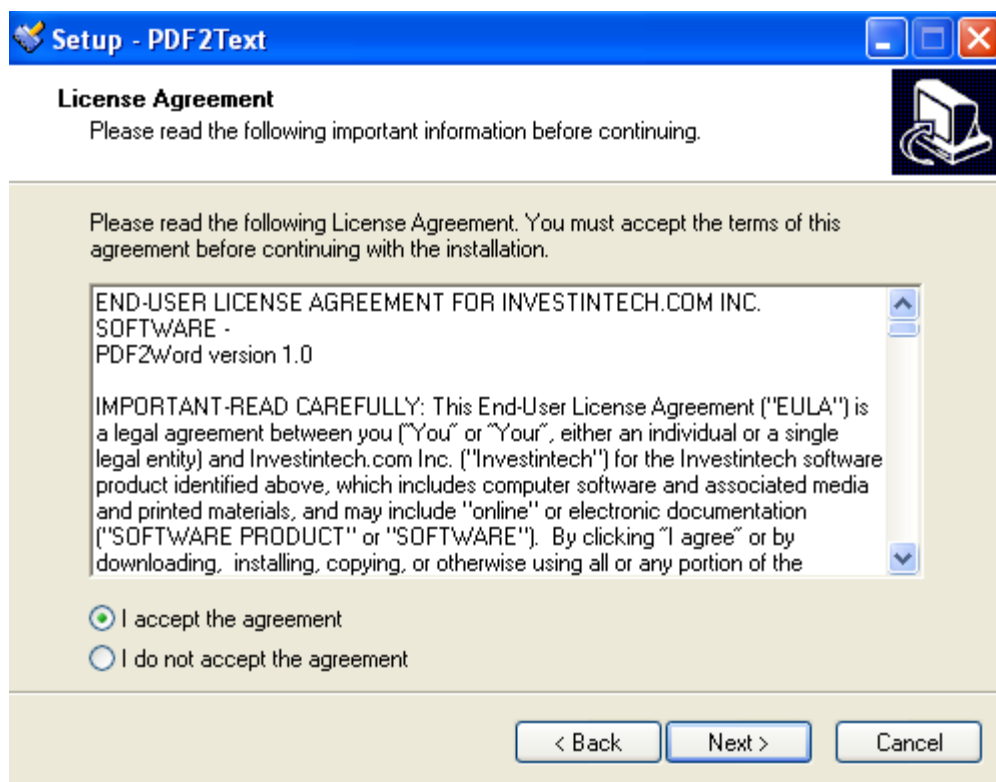
- Microsoft Windows 98/Me/2000/XP/2003
- 512 MB of RAM
- 100 MB of available hard disk space

## Starting the Installation

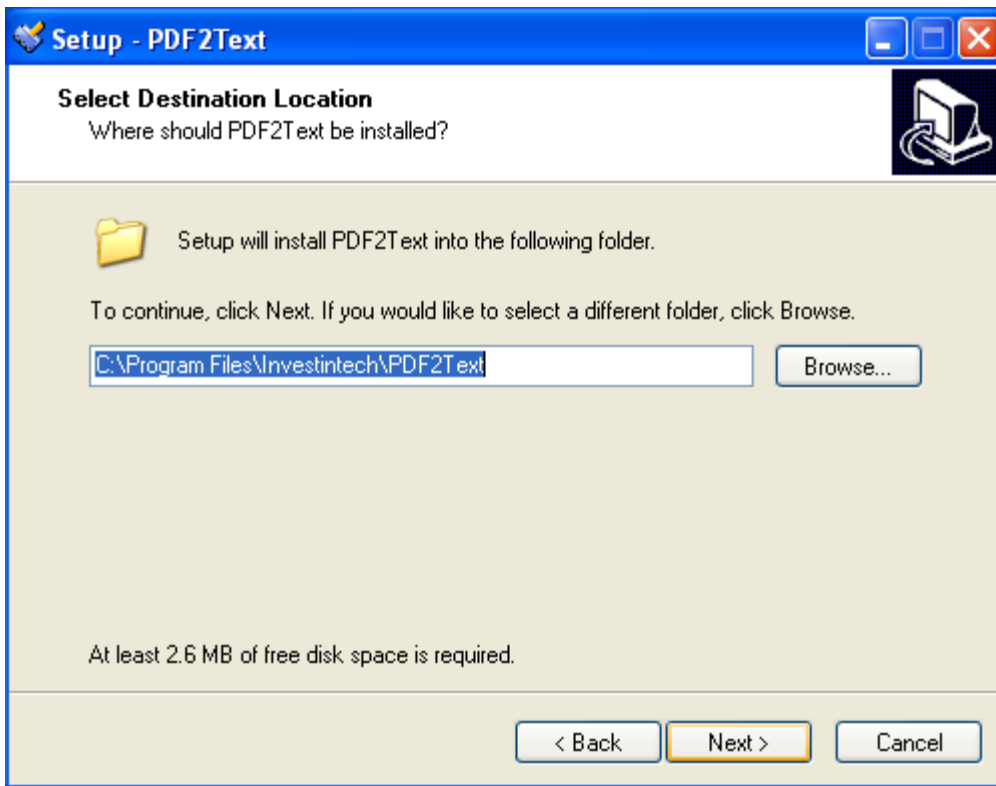
Start the installation by launching the `InstallPDF2Text.exe` file, e.g. by double clicking it in the Windows Explorer. After that you should see a window that looks like this:



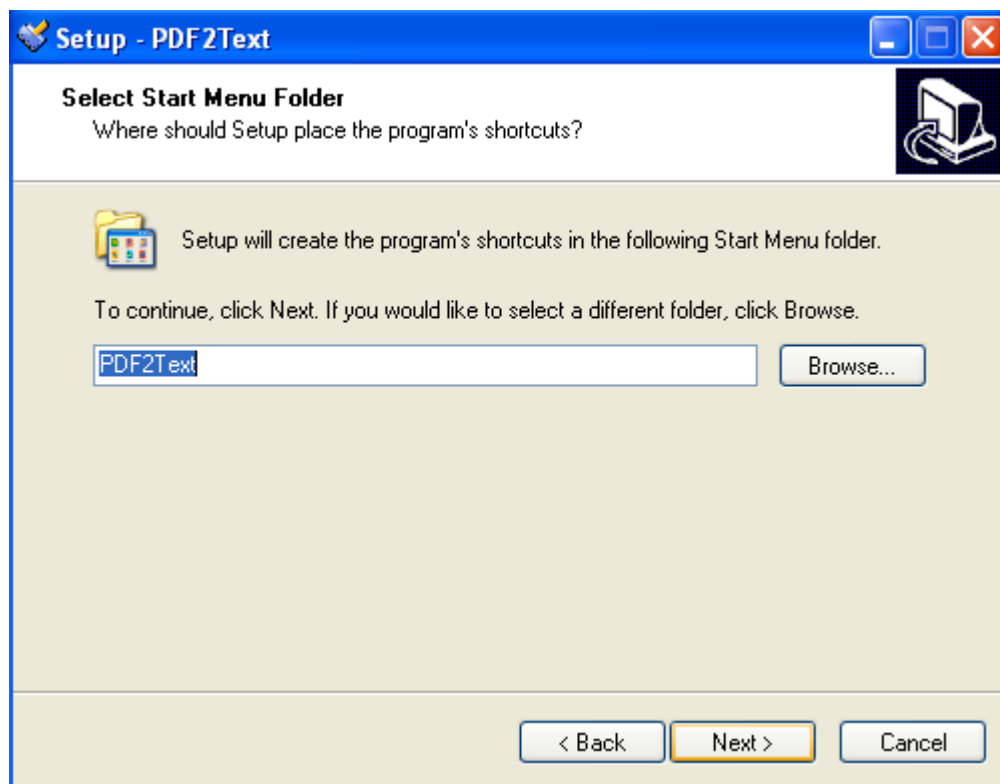
Click the 'Next' button to continue installation. You should see the next window:



Please read the Investintech Software License Agreement carefully and thoroughly. If you wish to continue with installation, choose **'I agree with the above terms and conditions'** and click **'Next'**. Otherwise, click **'Exit'** and installation will terminate. If you clicked **'Next'** you should see the next window:

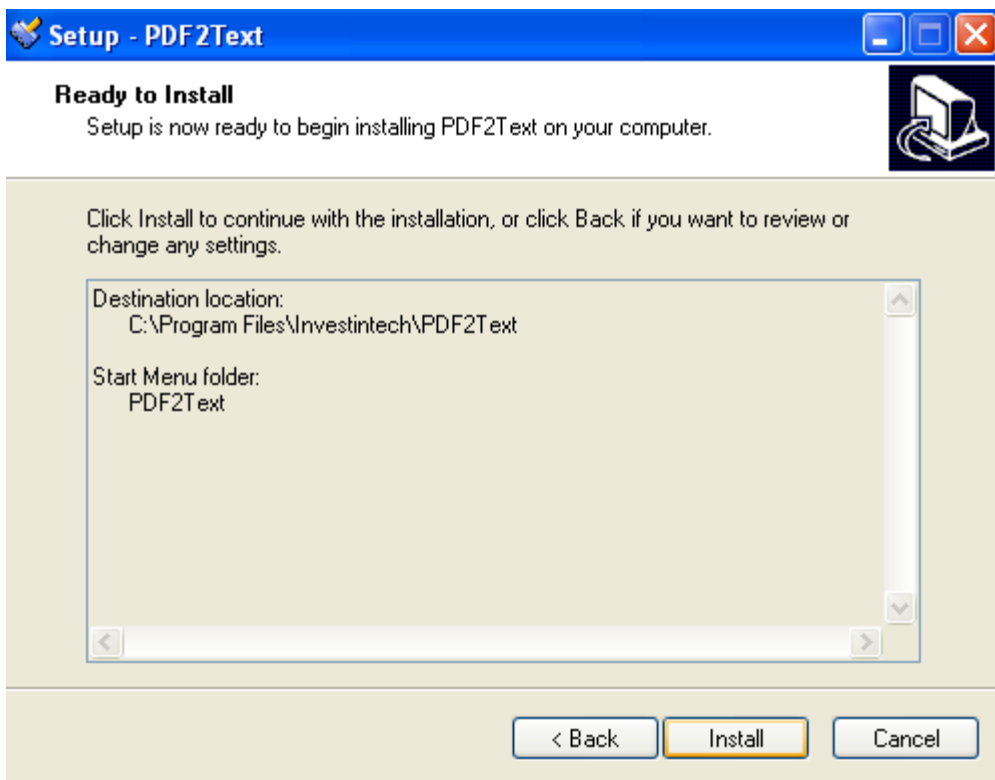


On this screen you can choose the installation directory for the PDF2Text software. The default installation folder will be contained in the `Program Files` folder on your system disk. It is possible to change the location of installation folder either by typing its path directly or by clicking the 'Browse...' button. If the destination folder (directory) you have chosen does not exist, the installation program will create it. Click the '**Next**' button to continue the installation.



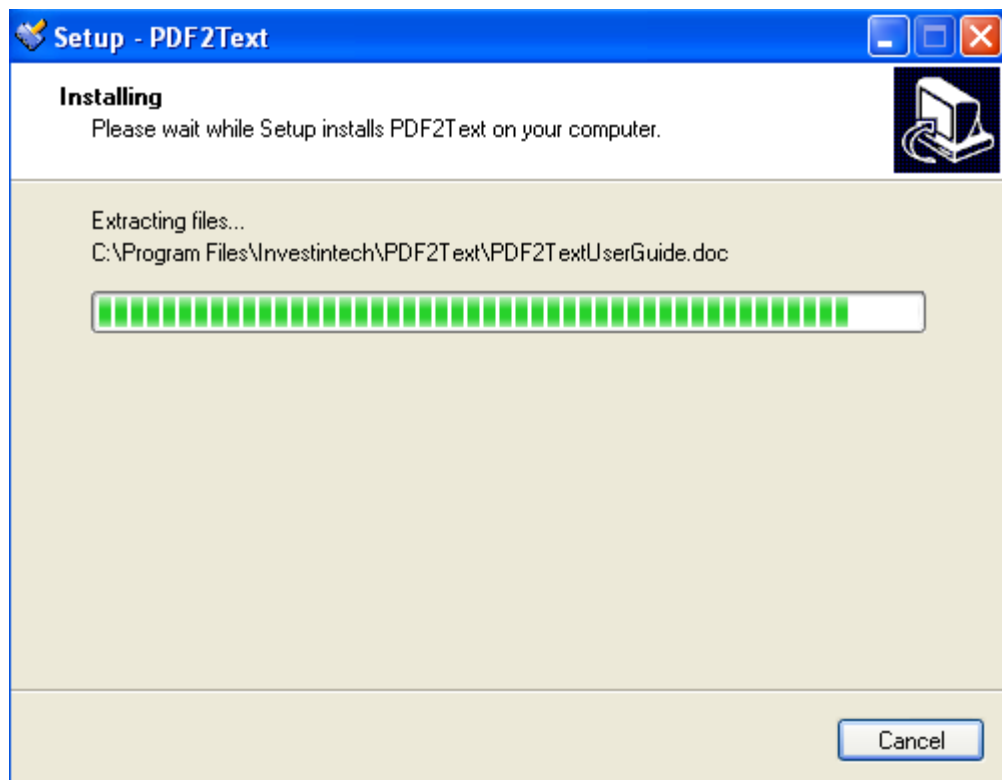
This screen prompts you for the for the start menu folder containing shortcuts to PDF2Text. You may accept the default folder name, type a new folder name, or browse existing folders by clicking the “Browse...” button. Click the ‘Next’ button to continue the installation.

## PDF2Word Command Line

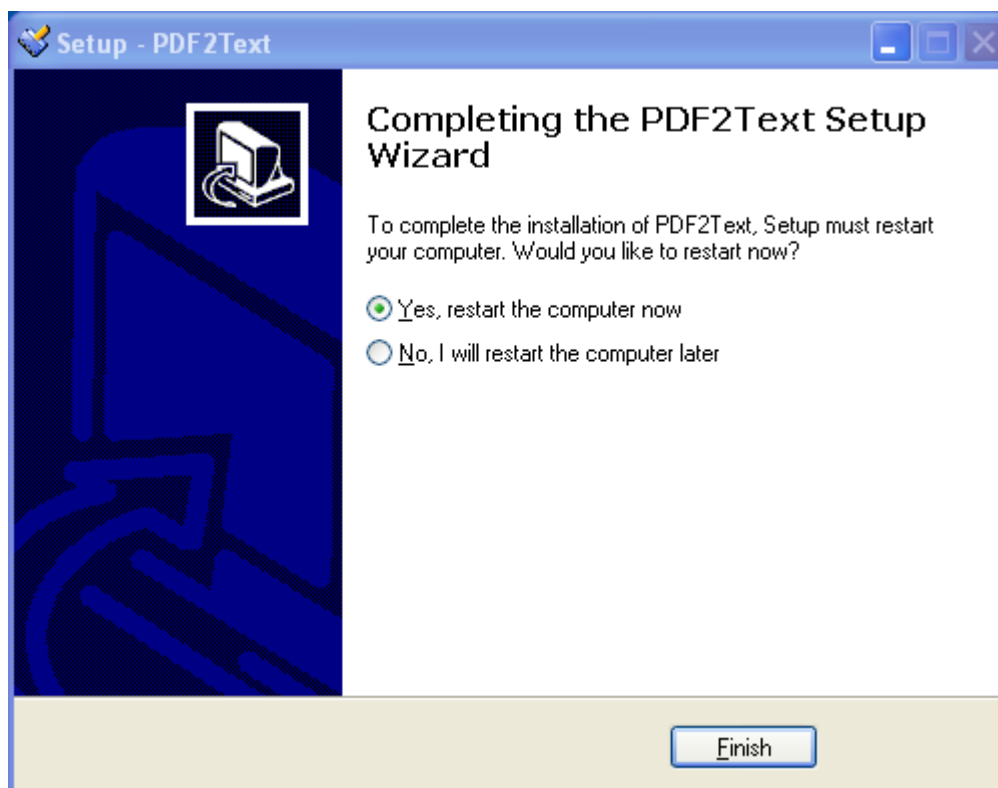


Clicking the **'Install'** button will start the installation.

PDF2Text Command Line



Setup will install the PDF2Text Command Line product on your computer. After a very short period of time, usually about 30 seconds, you should see the next window:



The installation is now finished. Click '**Finish**' to finish the installation.

## ***Using PDF2Text***

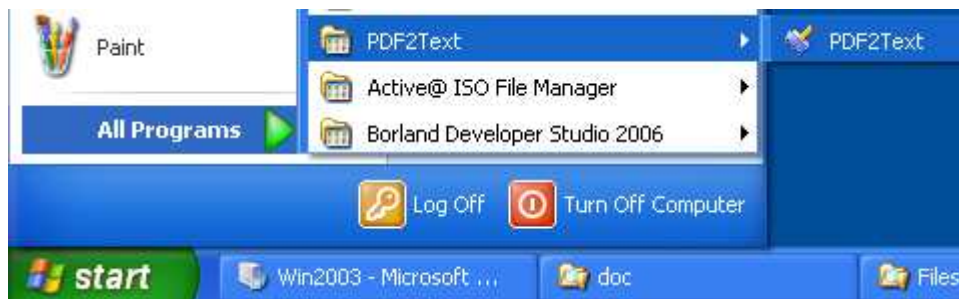
This section provides detailed information about using PDF2Text software. The following topics are covered:

- using PDF2Text from command prompt
- using PDF2Text for batch conversion

### **Using PDF2Text from Command Prompt**

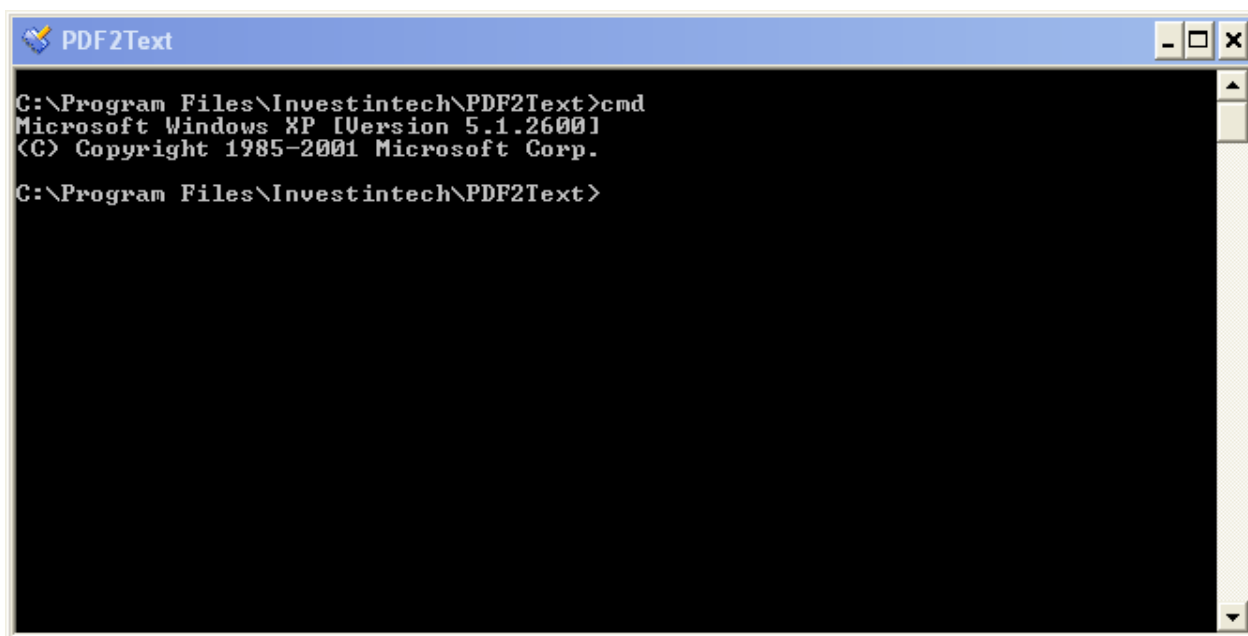
Prior to using PDF2Text, you must open a command prompt window. To open the command prompt window:

Click the **'Start'** button, then point to **'All Programs'**, then point to **'PDF2Text'** and finally point and click on the **'PDF2Text'** menu item.



The window titled “PDF2Word” should open, as shown.

## PDF2Word Command Line



```
C:\Program Files\Investintech\PDF2Text>cmd
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:\Program Files\Investintech\PDF2Text>
```

To display the PDF2Text help message, simply type PDF2Text and press **Enter**.

```

C:\Program Files\Investintech\PDF2Text>pdf2text
Investintech.com Inc. PDF2Text - Command Line Interface
Copyright 2007 Investintech.com Inc. All rights reserved

Unlimited edition

Usage:
  PDF2Text [-s<schedule>] [-t<template>] [-p<range>] -i<input> -o<output>

  where:
  <schedule> is time at which conversion will occur
  <template> is name of template file describing output file format
  <range> indicates page range
  <input> specifies input pdf document (may include wildcards)
  <output> specifies target output file path and extension

  Note: input and output files must contain full type extension
  Wild cards:
  ? Matches any single character in document name
  * Matches 0 or more characters in document name

Example:
  PDF2Text -iSource.pdf -oTarget.txt
  PDF2Text -iSource.pdf -oTarget.txt
  PDF2Text -i*.pdf -o*.txt
  PDF2Text -s3PM -iSource.pdf -oTarget.txt
  PDF2Text -p5-9,15 -iSource.pdf -oTarget.txt
  PDF2Text -p5-9,15o -iSource.pdf -oTarget.txt
  PDF2Text -p4-12,16 -iSource.pdf -oTarget.txt
  PDF2Text -p4-12,16e -iSource.pdf -oTarget.txt
  PDF2Text -tTemplate.t2ae -iSource.pdf -oTarget.txt

-h or -help This help

```

The help message describes command usage, some examples and possible conversion types. The command line format is of the form:

```
PDF2Text [-s<schedule>] [-p<range>] -i<input_file(s)> -o<output_file(s)>
```

### **Schedule**

This optional argument allows you to automate a one-off or recurring conversion operation at some time in the future.

A valid PDF2Text schedule parameter excludes whitespace and is of the form:

```
time[/EVERY:date[,...]/NEXT:date[,...]]
```

where time is 24 hour time to the nearest second e.g. hh:mm:ss

hour: 00 - 23

minute: 00 - 59

second 00 - 59

## PDF2Word Command Line

date either a day of the week e.g. M,T,W,Th,F,S,Su

or a day of the month e.g. 1,2,3,...31

/EVERY clause, /NEXT clause are optional but only one optional clause is permitted

PDF2Text will construct a valid AT command format e.g.

```
AT [\\computername] time [/INTERACTIVE] [/EVERY:date[,...] | /NEXT:date[,...]]  
"command"
```

Note: The Schedule feature depends on the Windows 'Task Scheduler' service.

## **Range**

This optional argument allows use to convert a subset of the PDF document. A continuous range of pages may be specified by the syntax: <starting page>-<ending page> where <starting page> - <ending page> are valid page numbers in the document and the starting page occurs before the ending page.

Example: 10-20

A list of pages may be specified with the syntax: <page i >,<page j>,<page k> where <page i >,<page j>,<page k> are valid page numbers in the document. In addition, you can select even numbered pages or odd numbered pages with the syntax "o" - to select odd numbered pages or "e" - to select even numbered pages.

## **Input file(s)**

This argument determines the source file, i.e. the file you want to convert data from. Please note the following:

- You must include both file name and extension separated with the dot character ('.'). For example, file `business.pdf` consists of file name 'business' and extension 'pdf'.
- You must provide full path to the input file unless it is in the same folder as `PDF2Text.exe`. For example, if input file `business.pdf` is in `MyFiles` folder on C drive, you should type `c:\MyFiles\business.pdf` as an input file argument. If any folder name or file name in the path is longer than 8 characters or contains spaces you must surround its path with double quotation marks, e.g. `"c:\very long path\business.pdf"`

## **Output file(s)**

This argument is a string expression determining the destination file, i.e. the file containing the results of the conversion. The naming rules used by PDF2Text are as follows. PDF2Text will use the specified directory or the current working directory if none is specified. It will use the name corresponding to the name of the source file. It will use the extension. Finally, the same path and file name rules apply for output file as well as for input file.

## Examples of Using PDF2Text

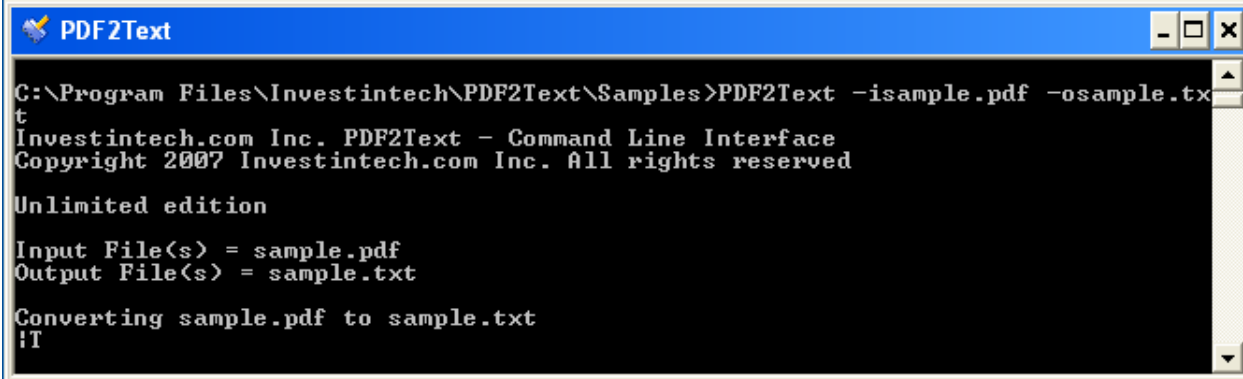
Here are some cases illustrating the use of PDF2Text.

### Example #1

In this example we will assume that both input and output files are in the same folder as PDF2Text. If you type

```
pdf2Text -isample.pdf -osample.csv
```

and press **Enter**, you will convert from PDF to text file. The resulting text file can be opened by any text editor. The picture below shows a successful conversion. For each converted page PDF2Text will print “|” on the screen to let user know that it is active and working.



```
C:\Program Files\Investintech\PDF2Text\Samples>PDF2Text -isample.pdf -osample.txt
Investintech.com Inc. PDF2Text - Command Line Interface
Copyright 2007 Investintech.com Inc. All rights reserved

Unlimited edition

Input File(s) = sample.pdf
Output File(s) = sample.txt

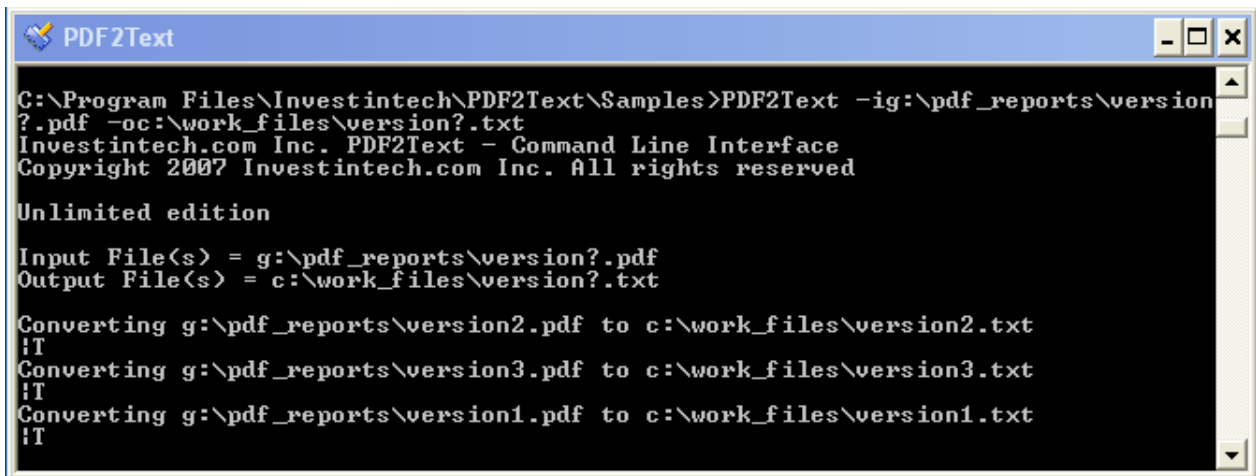
Converting sample.pdf to sample.txt
|
```

### Example #2

In this example we will assume that directory pdf\_reports on the G drive contains a series of PDF documents. Further, we assume input file is in the same folder as PDF2Text, but we want output file to be created in the different folder, e.g. c:\text\_reports. If you type:

```
PDF2Text -ig:\pdf_reports\version?.pdf -
oc:\html_reports\version?.xls
```

and press **Enter**, you will convert the matching documents from PDF to text format. The picture below shows a successful conversion.



```

C:\Program Files\Investintech\PDF2Text\Samples>PDF2Text -ig:\pdf_reports\version
?.pdf -oc:\work_files\version?.txt
Investintech.com Inc. PDF2Text - Command Line Interface
Copyright 2007 Investintech.com Inc. All rights reserved

Unlimited edition

Input File(s) = g:\pdf_reports\version?.pdf
Output File(s) = c:\work_files\version?.txt

Converting g:\pdf_reports\version2.pdf to c:\work_files\version2.txt
!T
Converting g:\pdf_reports\version3.pdf to c:\work_files\version3.txt
!T
Converting g:\pdf_reports\version1.pdf to c:\work_files\version1.txt
!T

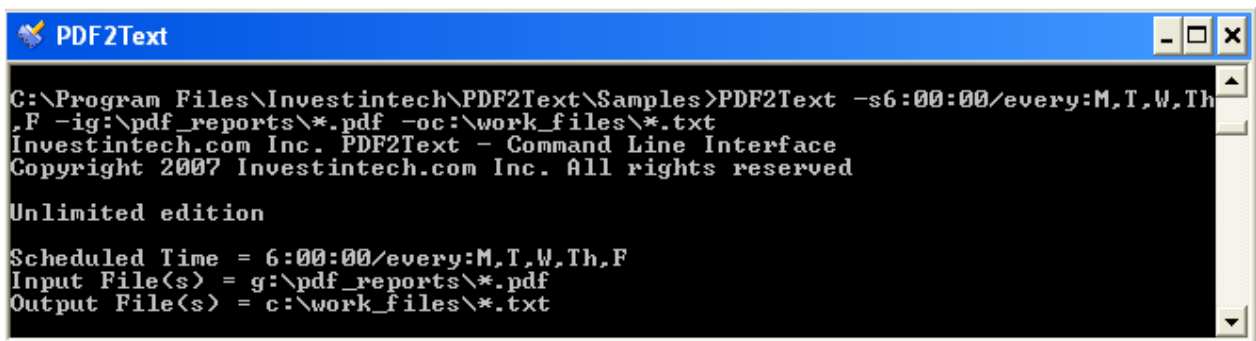
```

### Example #3

In this example we will assume that neither input nor output file is in the same folder as PDF2Text. Input file is in d:\src folder and we want output file to be created in the d:\conv. If you type

```
PDF2Word -cxls -s6:00AM/every:M,T,W,Th,F -ig:\pdf_reports\*.pdf -oc:\text_reports\*.xls
```

and press **‘Enter’**, you will convert from PDF to text file. The picture below shows a successful conversion.



```

C:\Program Files\Investintech\PDF2Text\Samples>PDF2Text -s6:00:00/every:M,T,W,Th
,F -ig:\pdf_reports\*.pdf -oc:\work_files\*.txt
Investintech.com Inc. PDF2Text - Command Line Interface
Copyright 2007 Investintech.com Inc. All rights reserved

Unlimited edition

Scheduled Time = 6:00:00/every:M,T,W,Th,F
Input File(s) = g:\pdf_reports\*.pdf
Output File(s) = c:\work_files\*.txt

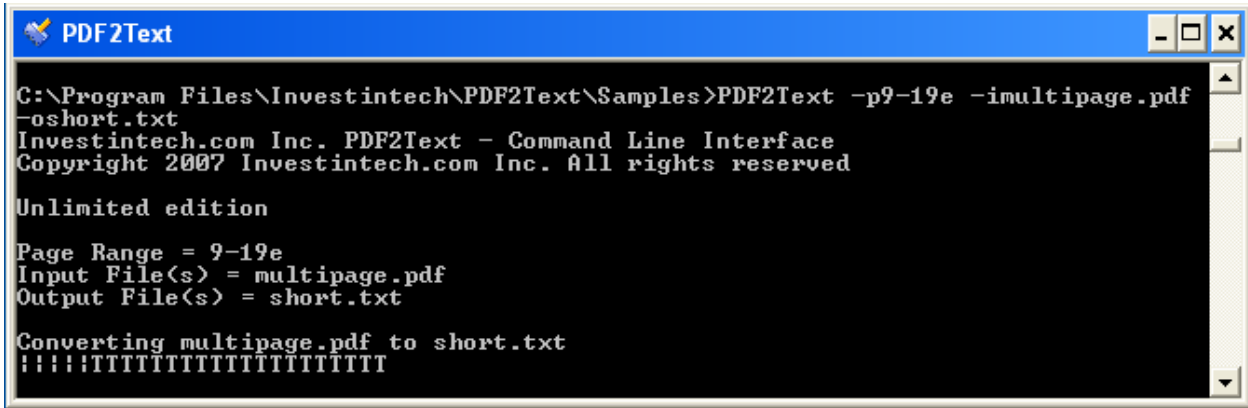
```

### Example #4

In this example we will assume that neither input nor output file is in the same folder as PDF2Text. Input file is in d:\src folder and we want output file to be created in the d:\conversion result. If you type:

```
PDF2Text -p19-41e -imultipage.pdf -short.txt
```

and press **‘Enter’**, you will convert from PDF to text file. The picture below shows a successful conversion.



```
C:\Program Files\Investintech\PDF2Text\Samples>PDF2Text -p9-19e -imultipage.pdf
-oshort.txt
Investintech.com Inc. PDF2Text - Command Line Interface
Copyright 2007 Investintech.com Inc. All rights reserved

Unlimited edition

Page Range = 9-19e
Input File(s) = multipage.pdf
Output File(s) = short.txt

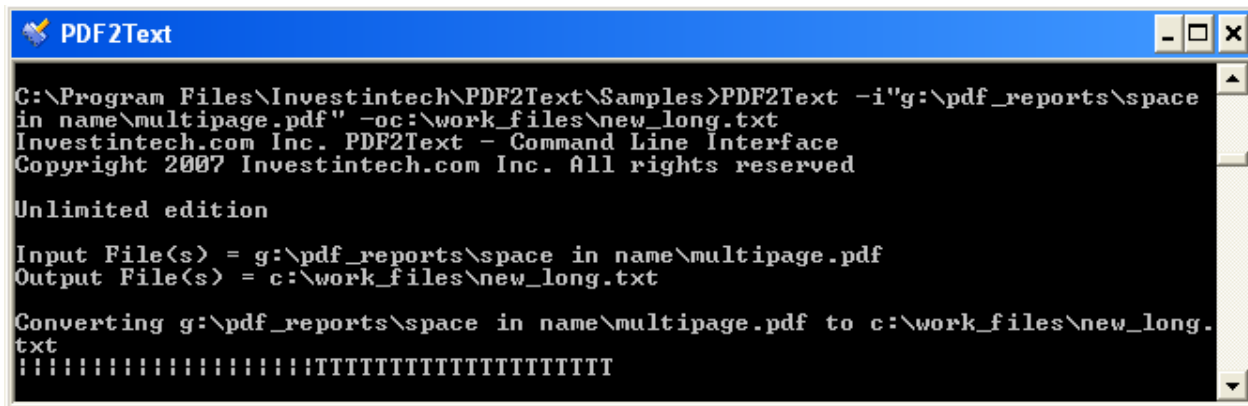
Converting multipage.pdf to short.txt
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
```

### Example #5

In this example we will assume that neither input nor output file is in the same folder as PDF2Word. Input file is in g:\pdf\_reports\space in name folder and we want output file to be created in the oc:\work\_files folder. If you type:

```
PDF2Text -i"g:\pdf_reports\space in name\multipage.pdf" -
oc:\work_files\new_long.txt
```

and press **Enter**, you will convert from PDF to text file. The picture below shows a successful conversion.



```
C:\Program Files\Investintech\PDF2Text\Samples>PDF2Text -i"g:\pdf_reports\space
in name\multipage.pdf" -oc:\work_files\new_long.txt
Investintech.com Inc. PDF2Text - Command Line Interface
Copyright 2007 Investintech.com Inc. All rights reserved

Unlimited edition

Input File(s) = g:\pdf_reports\space in name\multipage.pdf
Output File(s) = c:\work_files\new_long.txt

Converting g:\pdf_reports\space in name\multipage.pdf to c:\work_files\new_long.
txt
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
```

### Troubleshooting

There are a few cases in which PDF2Text is unable to perform the requested file conversion. Please consult the following table if you experience any of errors while working with PDF2Text.

<i>Error</i>	<i>Cause</i>	<i>Solution</i>
Unable to convert - did not find input document file name expression.	Misspelled or omitted input document file parameter. It is also possible you do not have permissions to access input file	use -i<filename.pdf> expression to specify PDF document file(s) to be converted or contact your system administrator to give you permissions to read input file.
Unable to convert - did not find output document file name expression.txt	Misspelled or omitted input document file parameter. It is also possible you do not have permissions to access input file	Use -o<filename.pdf> to specify file(s) that will contain converted information. Try a path for which you are have write permissions.
Empty converted document when using page range	Page range is invalid e.g. -p50-40	Use ascending order in page range e.g. -p40-50
Task not scheduled when using schedule parameter	Schedule time specified incorrectly.	Use a formatted 24 hour time with valid optional recurring sub parameters e.g. -s:30:00 -s06:30:00:/next:1 -23:59:59:/every:M,T,W,Th,F
Input file is damaged or not valid PDF file!	Input file was found but it either is not PDF file or it is damaged or corrupted PDF file	If Acrobat Reader can open the file without any error message please send the file to our technical support e-mail and we will look further into the problem. If Adobe Acrobat (full version) can open the file with error message, save the file back - this way Adobe Acrobat will repair the file. Then try conversion again.

## Index

Command Line	3, 4
Examples of Using	16
Input file	15
Output file	14, 15
System Requirements	4
Using	12
Customer Service	2
Investintech.com Inc.	1
Technical Support	2



ERROR: undefined  
OFFENDING COMMAND: wnd

STACK:

/Ref\_cmd  
/Dest  
-mark-