

**Investintech.com Inc.
Software Development Kit:
PDF-to-HTML Function Library
User's Guide**

July 13, 2007
<http://www.investintech.com>

Copyright 2007 Investintech.com, Inc. All rights reserved

Adobe® is registered by Adobe Systems Incorporated
Acrobat® is registered by Adobe Systems Incorporated
Postscript® is registered by Adobe Systems Incorporated
Access™ is registered by Microsoft Corporation
Excel® is registered by Microsoft Corporation
Visual Basic® is registered by Microsoft Corporation
Visual C++® is registered by Microsoft Corporation
Visual C#® is registered by Microsoft Corporation
Visual J#® is registered by Microsoft Corporation
Visual Studio® is registered by Microsoft Corporation
Win32® is registered by Microsoft Corporation
Windows® is registered by Microsoft Corporation
Windows NT® is registered by Microsoft Corporation
Windows Vista™ is registered by Microsoft Corporation

Contents

Preface	1
About This Documentation	1
Typographical Conventions Used in This Document	1
Getting More Information	2
Customer Service and Technical Support	2
Fax and Mailing Address.....	2
Investintech PDF-To-HTML Conversion DLL.....	3
What is the Investintech PDF-To-HTML Conversion DLL?	4
Installation Instructions for PDF-To-HTMLDLL	4
System Requirements	4
Starting the Installation.....	4
Using the Investintech PDF-To-HTML Conversion DLL in Microsoft Visual C++ .NET 2003. 11	
Implicit Linking.....	11
Linking .lib file with project	12
Using methods from DLL file.....	13
Investintech PDF-To-HTML Conversion DLL Methods	13
Interface	14
Parameter Type.....	14
File Names	14
Error Handling.....	15
Common Sample Source Code.....	15
<i>Conversion from PDF document to HTML Workbook.....</i>	<i>15</i>
PDF_to_HTML_CBR	16
Prototype	16
Description	16
Calling Convention	16
Parameters	16
Returns	16
Example	16
PDF_to_HTML_CBT.....	16
Prototype	16
Description	16
Calling Convention	17
Parameters.....	17
Returns	17
Example	17
PDF_to_HTML_CLR.....	17
Prototype	17
Description	17
Calling Convention	17
Parameters.....	18
Returns	18
Example	18
PDF_to_HTML_CLT	18
Prototype	18
Description	18

Calling Convention	18
Parameters.....	18
Returns	18
Example	18
PDF_to_HTML_SBR	19
Prototype.....	19
Description.....	19
Calling Convention	19
Parameters.....	19
Returns	19
Example	19
PDF_to_HTML_SBT.....	20
Prototype.....	20
Description.....	20
Calling Convention	20
Parameters.....	20
Returns	20
Example	20
PDF_to_HTML_SLR.....	20
Prototype.....	20
Description.....	21
Calling Convention	21
Parameters.....	21
Returns	21
Example	21
PDF_to_HTML_SLT	21
<i>Prototype</i>	21
Description.....	21
Calling Convention	21
Parameters.....	22
Returns	22
Example	22
VB6 callable code.....	22
PDF_to_HTML_VB6	22
Prototype.....	22
Description.....	23
Calling Convention	23
Parameters.....	23
Returns	23
Example	23
InvestintechConversionDLL_IDString.....	23
Description.....	23
Parameters.....	23
Returns	23
Investintech PDF-To-HTML Conversion COM Server Methods	24
PDF2HTML	24
Prototype.....	24
Description.....	24
Parameters.....	24
Returns	24
Example	24
Index	26

Preface

Welcome to Investintech.com Inc. Software Development Kit (SDK). This SDK provides technologies for converting PDF files into other file. This user's guide provides the basic information needed to use all parts of the SDK. This SDK consists of:

- Investintech PDF To HTML Conversion DLL, a Dynamic-Link Library for converting PDF files

This document provides a comprehensive description of the DLL.

About This Documentation

This documentation consists of:

- Investintech PDF To HTML Conversion DLL, describing usage and methods of DLL designed to convert PDF files into Hypertext Markup Language format (HTML).

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	a2ecl -pdf2text abc.pdf abc.txt
	Paths and filenames	c:\a2ecl\a2ecl.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 AboutBox() method
'Single quoted bold'	Elements of the user interface like buttons, icons etc.	click on 'Start' button, press 'Enter'
blue underlined	Internet links	http://www.investintech.com

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:	cs@investintech.com	techsupport@investintech.com
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

Investintech PDF-To-HTML Conversion DLL

What is the Investintech PDF-To-HTML Conversion DLL?

The Investintech PDF-To-HTML Conversion DLL is a collection of methods compiled, linked and stored in a dynamic-link library (DLL) file. These methods provide a set of methods for converting files from PDF to Hypertext Markup Language (HTML) format.

Installation Instructions for PDF-To-HTMLDLL

Below you will find step-by-step instructions to install PDF-To-HTML DLL Conversion Library on your system. You may want to print this instruction sheet for reference before beginning to install PDF-To-HTML DLL.

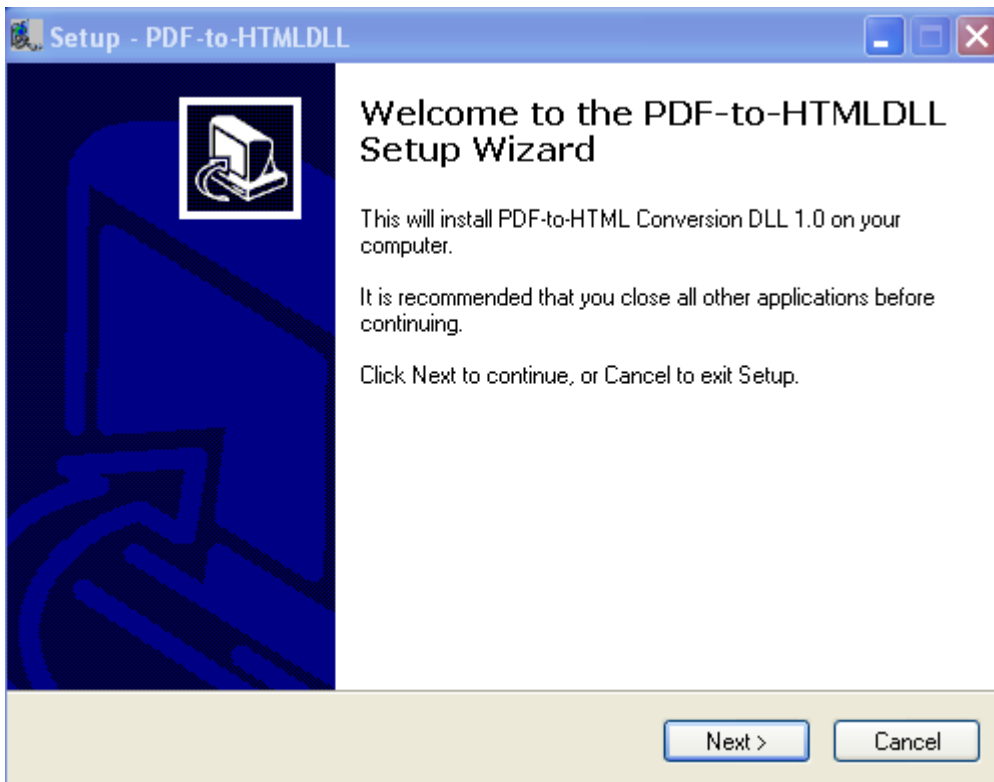
System Requirements

The minimum computer system resources required to install and use PDF-To-HTML DLL are:

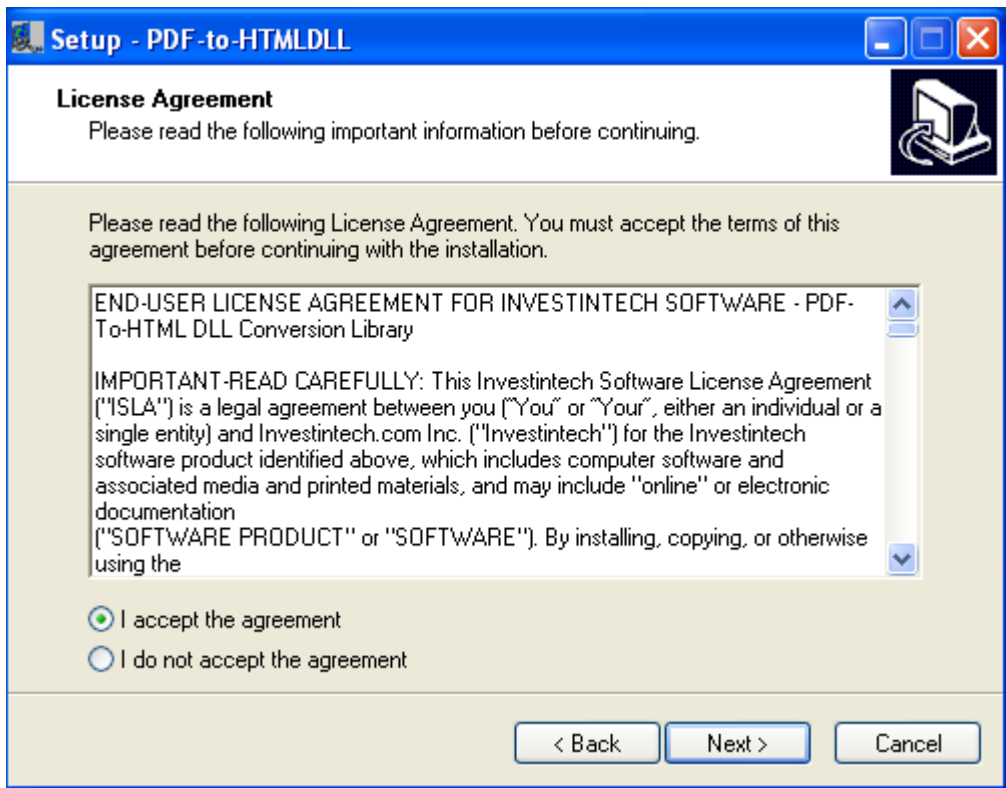
- Microsoft Windows XP
- 64 MB of RAM
- 10 MB of available hard disk space
- A software development environment, such as Microsoft Visual Studio or Borland Delphi

Starting the Installation

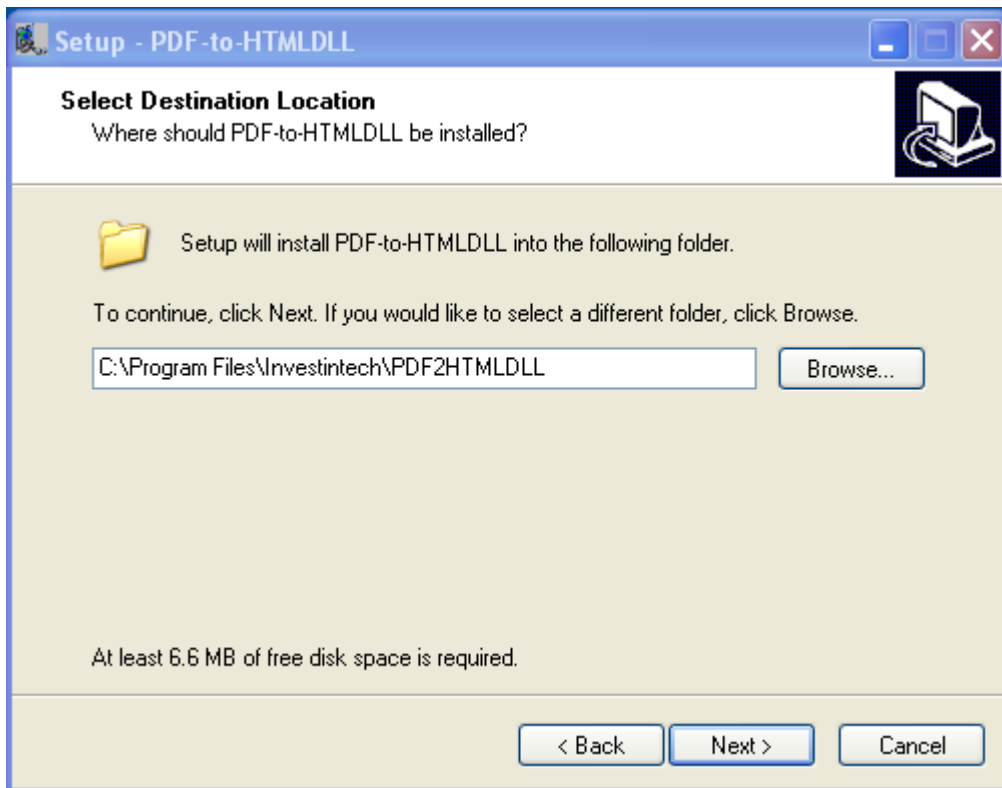
Start the installation by launching the `InstallPDF2HTMLDLL.exe` file, e.g. by double clicking it in the Windows Explorer. After that you should see a window that looks like this:



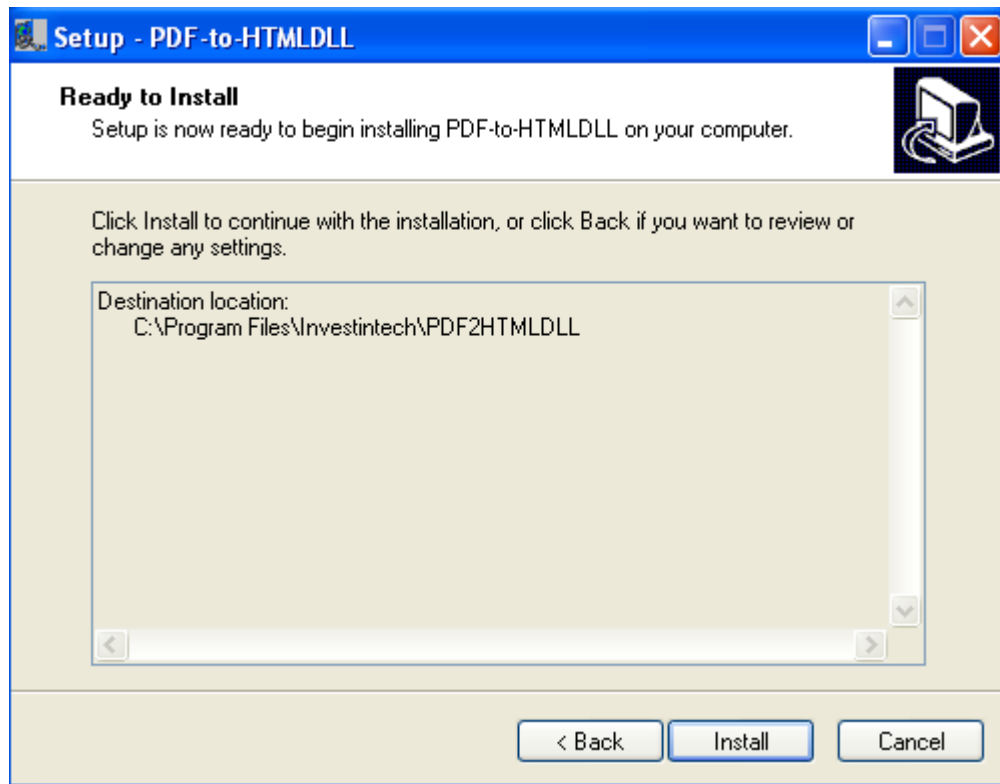
Click the 'Next' button to continue installation. You should see the next window:



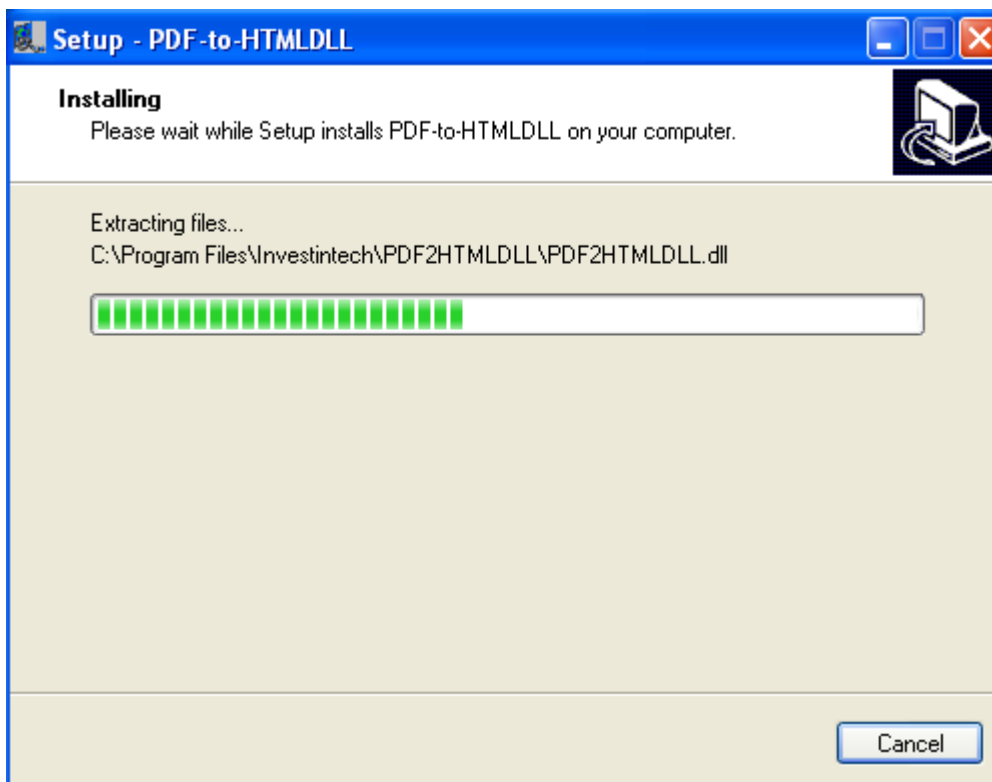
Please read the Investintech Software License Agreement carefully and thoroughly. If you wish to continue with installation, choose ‘**I agree with the above terms and conditions**’ and click ‘**Next**’. Otherwise, click ‘**Exit**’ and installation will terminate. If you clicked ‘**Next**’ you should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



On this screen you can choose the installation directory for the PDF-To-HTMLDLL 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.

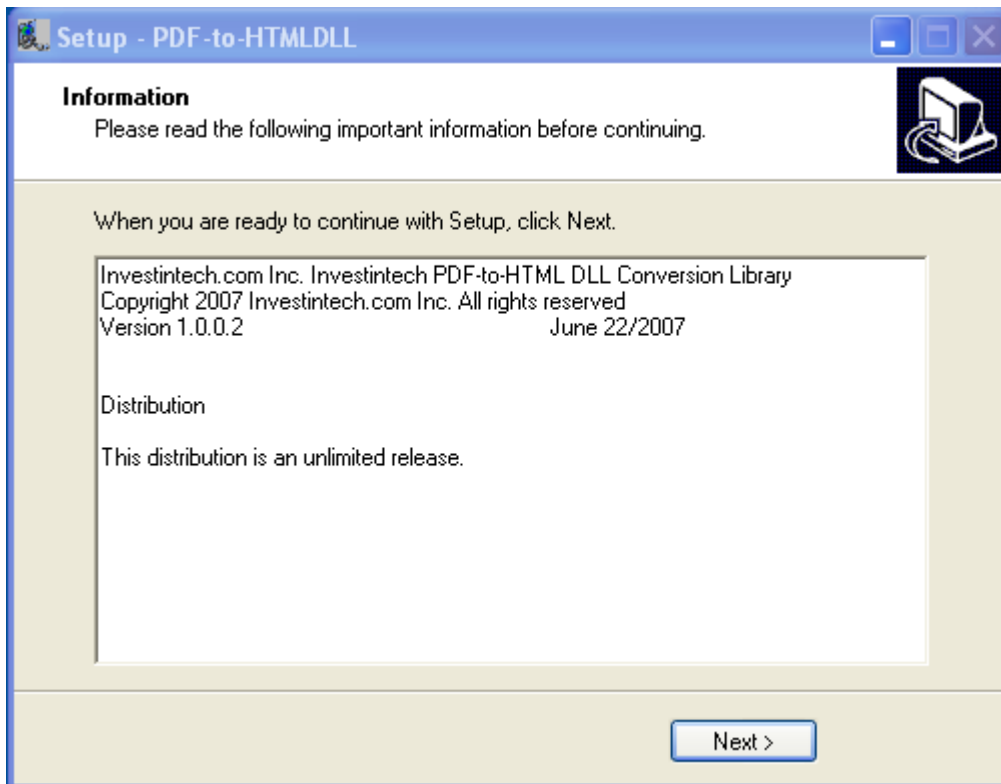


Clicking the **Install** button will start the installation.

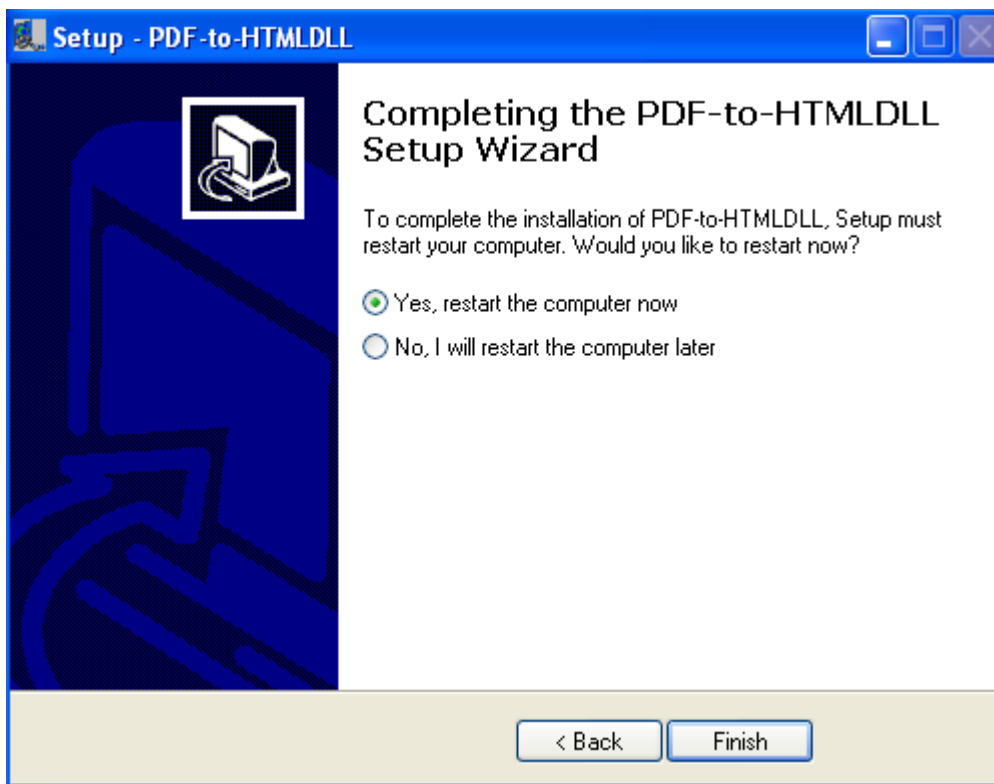


The installation will copy the files that comprise PDF-To-HTMLDLL to the directory you specified earlier. After a very short period of time, usually about 30 seconds, you should see the next window:

PDF2HTMLDLL – Investintech Conversion DLL



This windows describes the distribution of PDF-To-HTMLDLL being installed. Click ‘**Next**’ to continue.



When this window appears, the installation process is complete. Click **Finish** to end the installation program.

Using the Investintech PDF-To-HTML Conversion DLL in Microsoft Visual C++ .NET 2003

In this section you will learn how to use the Investintech PDF-To-HTML Conversion DLL. All pictures and examples from now on will be based on Microsoft Visual C++ .NET 2003.

In order to use methods contained in DLL you have to link an executable file to DLL. Executable file can be either EXE or DLL file. There are two possible ways of linking: implicit and explicit linking. In this documentation implicit linking will be shown.

Implicit Linking

Implicit linking is load-time linking. The following is needed for executables to implicitly link to Investintech Conversion DLL:

- PDF2HTMLDLL.h, the header file which contains declarations of exported methods
- PDF2HTMLDLL.lib, an import library used by the linker

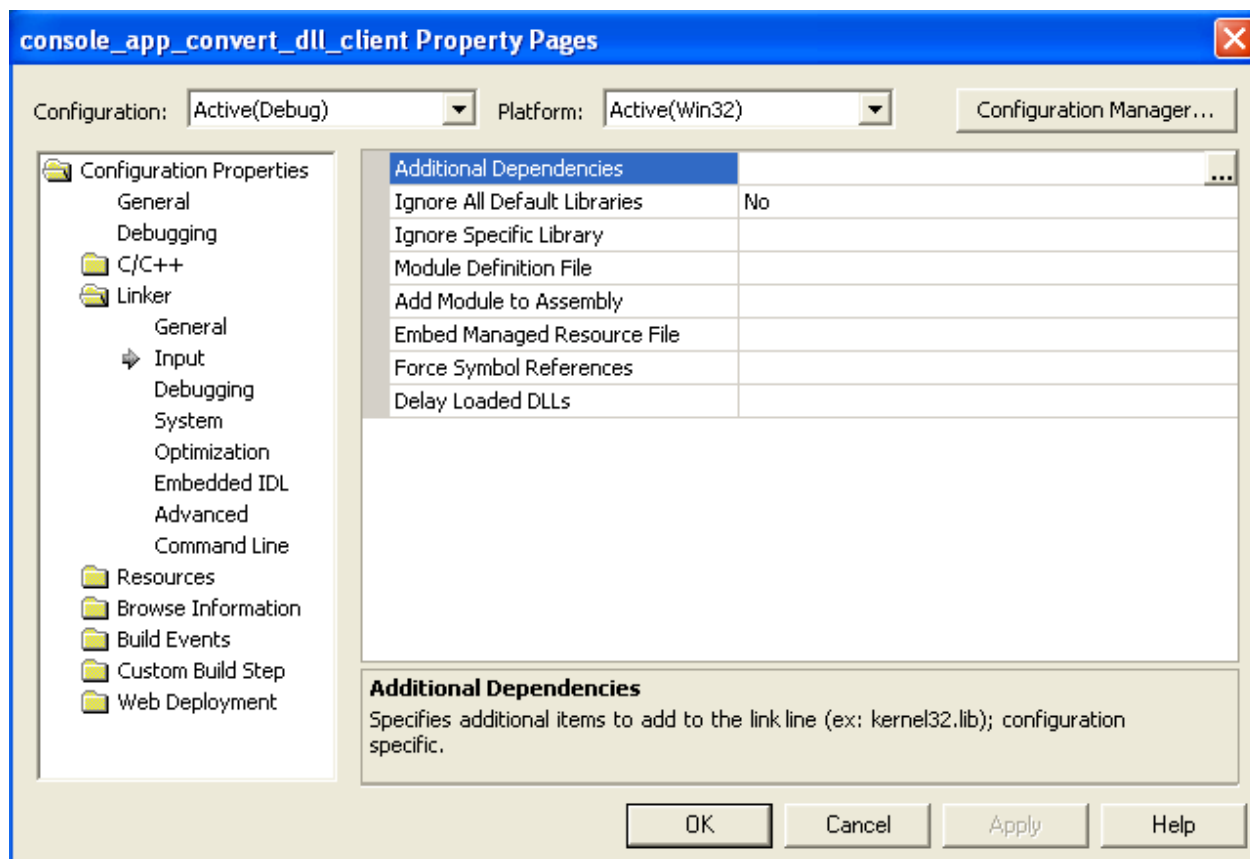
- PDF2HTMLDLL.dll, the Dynamic Link Library file

All these files are provided by Investintech.com Inc.

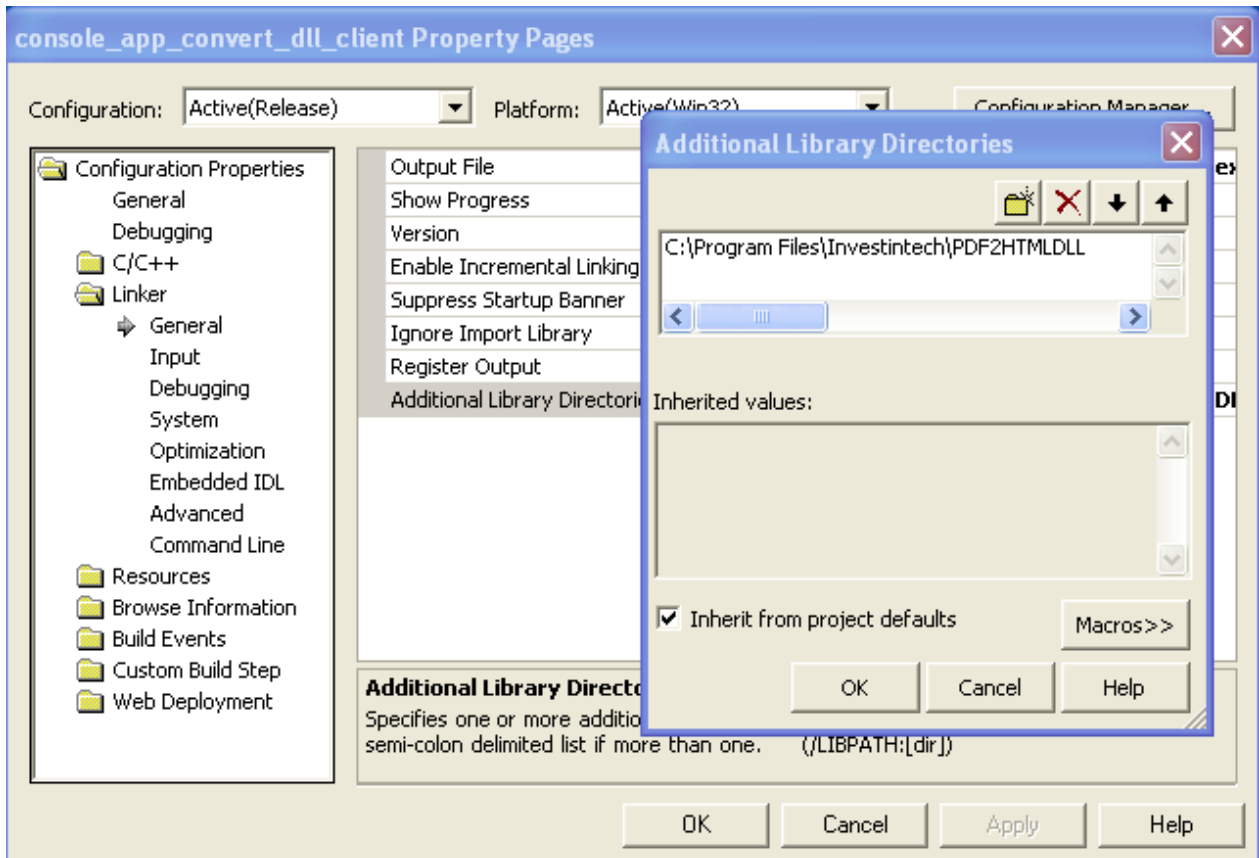
Linking .lib file with project

Below you will find step-by-step instructions on how to link PDF2HTMLDLL.lib file with your project.

Start Visual Studio and select 'File > Open > Project'. Select the project in which you want to use PDF2HTMLDLL.dll and click 'OK'. Select 'Project > Properties'. You should see the window like next one:



Click the 'Linker' folder and open 'Input' property page. Modify the 'Additional Dependencies' property by clicking the '...' button.



The window titled “Additional Dependencies” should open: type PDF2HTMLDLL.lib and click the ‘OK’ button. Click ‘OK’ to close “Property Pages” window. You have successfully linked PDF2HTMLDLL.lib with your project.

Using methods from DLL file

After linking with PDF2HTMLDLL.lib file you can use methods contained in InvestintechConversionDLL.dll file. Simply include the PDF2HTMLDLL.h file in all source files that need to use those methods.

Investintech PDF-To-HTML Conversion DLL Methods

In this section you will find details about each method such as: prototype; description; calling conventions; parameters; return value; and sample use of the methods.

Interface

The interface of the Investintech PDF-To-HTML Conversion DLL consists of its exposed methods. The interface contains an exposed method for each combination of: calling convention; parameter type; and error handling.

The calling convention is either the “cdecl” calling convention or the “stdcall” calling convention.

The “cdecl” calling convention is normally used for calling C/C++ functions. It is characterized by the following points:

- arguments are passed in order from right to left
- the calling function is responsible for popping the arguments from the stack at the conclusion of function execution
- an underscore character (`_`) is prefixed to function names
- no case translation is performed on the function name

The “stdcall” calling convention is typically used to call Windows operating system API functions. It is characterized by the following points:

- arguments are passed in order from right to left
- arguments are passed by value, unless a pointer or reference type is passed.
- The called function is responsible for popping its own arguments from the stack.
- an underscore (`_`) is prefixed to the name. The name is followed by the at sign (`@`) followed by the number of bytes (in decimal) in the argument list. For instance, the function declared as `void foo(int bar)` is decorated as follows: `_foo@4`
- no case-translation is performed on the function name

Parameter Type

The parameter type is either ANSI (`C char *`) or UNICODE (`BSTR`).

File Names

Files may be specified by supplying relative or absolute file names. A relative file name is relative to the application default directory containing the application executable file (e.g. “`..\in_parent_dir.pdf`”, “`in_application_dir.pdf`”, “`subdir\in_sub_directory.pdf`”). An absolute file name includes the directory path to the file beginning at the top of the directory tree (e.g. “`c:\pdfs\absolute_file.pdf`”).

It is also possible to use UNC notation (e.g. “`\\myserver\public\file.pdf`”) provided your user account has sufficient access permissions.

Error Handling

The error handling strategy is either to throw an exception (part of the C++ throw-catch error handling construct) or return in integer/boolean type result code (the C style error handling strategy).

The Investintech Conversion DLL identifies the characteristics of a function by an encoded suffix. The suffix is defined by the rules:

- C or S - cdecl or stdcall calling convention
- B or L - BSTR or LPST parameter type
- R or T - return error code or throw exception

Common Sample Source Code

All sample code snippets share the following initialization code snippet:

```
//.h file must be included if we want to use conversion methods
#include <afx.h>
#include <afxwin.h>           // MFC core and standard components
#include "PDF2HTMLDLL.h"
```

```
CString str_inputFile = _T("c:\pdfs\in.pdf"); //input file name
CString str_outputFile = _T("out.html"); //output file name
```

```
//conversion from CString to BSTR
BSTR bstr_inputFile = str_inputFile.AllocSysString();
BSTR bstr_outputFile = str_outputFile.AllocSysString();
```

Conversion from PDF document to HTML Workbook

The following methods convert a PDF document file to an HTML workbook file.

PDF_to_HTML_CBR

Prototype

```
bool PDF_to_HTML_CBR(BSTR inFile, BSTR outFile)
```

Description

The PDF_to_HTML_CBR() method converts the user specified file from PDF to HTML format. Any web browser can open the created output file. The original PDF file is not modified.

Calling Convention

The function uses the “cdecl” calling convention.

Parameters

inFile	Unicode string containing the absolute or relative filename of the PDF file that will be converted to HTML format.
outFile	Unicode string containing the absolute or relative filename of the HTML file that will contain the result of the conversion.

Returns

true if PDF file is successfully converted to HTML, false otherwise.

Example

```
bool ret = PDF_to_HTML_CBR(bstr_inputFile, bstr_outputFile);
```

PDF_to_HTML_CBT

Prototype

```
bool PDF_to_HTML_CBT(BSTR inFile, BSTR outFile)
```

Description

The PDF_to_HTML_CBT() method converts the user specified file from PDF to HTML format. Any version of Microsoft HTML can open the created output file. The original PDF file is not modified.

Calling Convention

The function uses the “cdecl” calling convention.

Parameters

inFile	Unicode string containing the absolute or relative filename of PDF file that will be converted to HTML format.
outFile	Unicode string containing the absolute or relative filename of HTML file containing the result of conversion.

Returns

true if PDF file is successfully converted to HTML, otherwise throws an exception that should be handled by the calling application.

Example

```
try
{
    bool ret = PDF_to_HTML_CBT(bstr_inputFile, bstr_outputFile);
}
catch(...)
{
    // TODO: Handle exception
}
```

PDF_to_HTML_CLR

Prototype

```
bool PDF_to_HTML_CLR(LPSTR inFile, LPSTR outFile)
```

Description

The PDF_to_HTML_CLR() method converts the user specified file from PDF to HTML format. Any web browser can open the created output file. The original PDF file is not modified.

Calling Convention

The function uses the “cdecl” calling convention.

Parameters

- inFile** ANSI string containing the absolute or relative filename of the PDF file that will be converted to HTML format.
- outFile** ANSI string containing the absolute or relative filename of the HTML file containing the result of conversion.

Returns

true if PDF file is successfully converted to HTML, false otherwise.

Example

```
bool ret = PDF_to_HTML_CLR(str_inputFile.GetBuffer(), str_outputFile.GetBuffer());
```

PDF_to_HTML_CLT

Prototype

```
bool PDF_to_HTML_CLT(LPSTR inFile, LPSTR outFile)
```

Description

The PDF_to_HTML_CLT() method converts the user specified file from PDF to HTML format. Any web browser can open the created output file. The original PDF file is not modified.

Calling Convention

The function uses the “cdecl” calling convention.

Parameters

- inFile** ANSI string containing the absolute or relative filename of the PDF file that will be converted to HTML format.
- outFile** ANSI string containing the absolute or relative filename of the HTML file containing the result of conversion.

Returns

true if PDF file is successfully converted to HTML, otherwise throws an exception that the calling application should handle.

Example

```
try
```

```

{
    bool ret = PDF_to_HTML_CLT(str_inputFile.GetBuffer(), str_outputFile.GetBuffer());
}
catch(...)
{
    // TODO: Handle exception
}

```

PDF_to_HTML_SBR

Prototype

```
UINT PDF_to_HTML_SBR(BSTR inFile, BSTR outFile)
```

Description

The PDF_to_HTML_SBR() method converts the user specified file from PDF to HTML format. Any web browser can open the created output file. The original PDF file is not modified.

Calling Convention

The function uses the “stdcall” calling convention.

Parameters

inFile	Unicode string containing the absolute or relative filename of the PDF file that will be converted to HTML format.
outFile	Unicode string containing the absolute or relative filename of the HTML file containing the result of conversion.

Returns

Non-zero integer value if PDF file is successfully converted to HTML, zero integer value otherwise.

Example

```
UINT ret = PDF_to_HTML_SBR(bstr_inputFile, bstr_outputFile);
```

PDF_to_HTML_SBT

Prototype

```
UINT PDF_to_HTML_SBT(BSTR inFile, BSTR outFile)
```

Description

The PDF_to_HTML_SBT() method converts the user specified file from PDF to HTML format. Any web browser may be used to open the created output file. The original PDF file is not modified.

Calling Convention

The function uses the “stdcall” calling convention.

Parameters

- | | |
|----------------|--|
| inFile | Unicode string containing the absolute or relative filename of the PDF file that will be converted to HTML format. |
| outFile | Unicode string containing the absolute or relative filename of the HTML file containing the result of conversion. |

Returns

Non-zero integer value if PDF file is successfully converted to HTML, otherwise throws an exception that the calling application should handle.

Example

```
try
{
    UINT ret = PDF_to_HTML_SBT(bstr_inputFile, bstr_outputFile);
}
catch(...)
{
    // TODO: Handle exception
}
```

PDF_to_HTML_SLR

Prototype

```
UINT PDF_to_HTML_SLR(LPSTR inFile, LPSTR outFile)
```

Description

The PDF_to_HTML_SLR() method converts the user specified file from PDF to HTML format. Any web browser may be used to open the created output file. The original PDF file is not modified.

Calling Convention

The function uses the “stdcall” calling convention.

Parameters

inFile	ANSI string containing the absolute or relative filename of the PDF file that will be converted to HTML format.
outFile	ANSI string containing the absolute or relative filename of the HTML file containing the result of conversion.

Returns

Non-zero integer value if PDF file is successfully converted to HTML, zero integer value otherwise.

Example

```
UINT retCode = PDF_to_HTML_SLR(str_inputFile, str_outputFile);
```

PDF_to_HTML_SLT**Prototype**

```
UINT PDF_to_HTML_SLT(LPSTR inFile, LPSTR outFile)
```

Description

The PDF_to_HTML_SLT() method converts the user specified file from PDF to HTML format. Any web browser may be used to open the created output file. The original PDF file is not modified.

Calling Convention

The function uses the “stdcall” calling convention.

Parameters

- inFile** ANSI string containing absolute or relative filename of the PDF file that will be converted to HTML format.
- outFile** ANSI string containing the absolute or relative filename of the HTML file containing the result of conversion.

Returns

Non-zero integer value if PDF file is successfully converted to HTML, otherwise throws an exception that should be handled by the calling application.

Example

```
try
{
    UINT retCode = PDF_to_HTML_SLT(str_inputFile, str_outputFile);
}
catch(...)
{
    //TODO: handle error
}
```

VB6 callable code

VB6 client applications must use these functions in order to avoid runtime errors related to stack and string references.

PDF_to_HTML_VB6

Prototype

```
UINT PDF_to_HTML_VB6(BSTR inFile, BSTR outFile)
```

Description

The `PDF_to_HTML_VB6()` method converts the user specified file from PDF to HTML format. Any web browser may be used to open the created output file. The original PDF file is not modified.

Calling Convention

The function uses the “stdcall” calling convention.

Parameters

inFile	Unicode string containing the absolute or relative filename of the PDF file that will be converted to HTML format.
outFile	Unicode string containing the absolute or relative filename of the HTML file containing the result of conversion.

Returns

Non-zero integer value if PDF file is successfully converted to HTML, zero integer value otherwise.

Example

```
UINT ret = PDF_to_HTML_VB6(bstr_inputFile, bstr_outputFile);
```

InvestintechConversionDLL_IDString

```
char* InvestintechConversionDLL_IDString()
```

Description

The `InvestintechConversionDLL_IDString()` method returns vendor's ID string.

Parameters

None.

Returns

"Investintech Conversion DLL Version 1.00 by Investintech.com Inc." string.

Investintech PDF-To-HTML Conversion COM Server Methods

This section provides a description of the conversion methods exposed by the PDF-To-HTML COM Server. The COM Server defines the CoClass CPDF2HTML that implements the interface IPDF2HTML. The interface consists of the following methods.

PDF2HTML

Prototype

```
HRESULT PDF2HTML([in] BSTR sourceFile, [in] BSTR destFile,  
[out,retval] VARIANT_BOOL *successFlag)
```

Description

The PDF2HTML() method converts the user specified file from PDF to HTML format. The output file may be opened with any web browser. The original PDF file is not modified.

Parameters

- | | |
|--------------------|--|
| sourceFile | Unicode string containing the absolute or relative filename of the PDF file that will be converted to HTML format. |
| destFile | Unicode string containing the absolute or relative filename of the HTML file that contains the result of conversion. |
| successFlag | Pointer to boolean variable that will contain the result of the operation: true is the conversion operation is successful; else false. |

Returns

HRESULT enumerated value indicating the result of the COM operation.

Example

```
HRESULT result = PDF2HTML(BSTR_sourceFile BSTR_destinationFile, &boolFlag);
```


Index

A

Able2Extract Command Line	
Installation Instructions	4
System Requirements	4

C

Conversion DLL	3
Customer Service	2

I

Investintech.com Inc.	2
InvestintechSDK DLL	4
Implicit Linking.....	11
InvestintechSDK_IDString	23
InvestintechSDK_PDF_to_HTML.....	16, 17, 18
Methods	13, 24
InvestintechSDK.dll	12
InvestintechSDK.h	11
InvestintechSDK.lib.....	11

T

Technical Support	2
-------------------------	---

ERROR: undefined
OFFENDING COMMAND: ceo

STACK:

/Ref_part3
/Dest
-mark-