

Importing PDF & HTM Files

ABOUT

This document covers how to import PDF and HTM files into RoboHelp without unwanted characters.

These characters may be seen if for example:

- The source was a PDF converted to an HTM file.
- In an HTM file created outside of RoboHelp.

You may see characters such as à or squares where the character cannot read.

PDF DOCUMENTS

Convert the Document to HTM

If you want to bring a PDF into RoboHelp 2020 you need to convert it to an HTML file.

That can be done with Acrobat DC using the Export to HTML option.

The imported HTML file will likely show some unwanted characters in either Authoring view or the Output. If it does, then some cleanup is required.

How to Clean the HTM File

Open the HTM file in Word.

Click on Save As.

Select the Web Page Filtered Option.

01_2021-04572		
Word Document (*.docx) 👻		
Word Document (*.docx)		
Word Macro-Enabled Document (*.docm)		
Word 97-2003 Document (*.doc)		
Word Template (*.dotx)		
Word Macro-Enabled Template (*.dotm)		
Word 97-2003 Template (*.dot)		
PDF (*.pdf)		
XPS Document (*.xps)		
Single File Web Page (*.mht, *.mhtml)		
Web Page (*.htm, *.html)		
Web Page, Filtered (*.htm, *.html)		
Rich Text Format (*.rtf)		
Plain Text (*.txt)		
Word XML Document (*.xml)		
Word 2003 XML Document (*.xml)		
Strict Open XML Document (*.docx)		
OpenDocument Text (*.odt)		

Click the More Options link.

↑ 🗁 D: > ViaData > Import Tests > Converted Docs > 02 Second Run		
01_2021-04572		
Web Page, Filtered (*.htm, *.html)		
More options		

Select Web Options > Encoding and set it to UTF 8 as shown.

Web Options	?	×					
Browsers Files Pictures Encoding Fonts							
Encoding							
Reload the current document as:		~					
Save this document as:							
Unicode (UTF-8)							
Always save Web pages in the default encoding							
ОК	С	ancel					

Click OK and Save. Import the file again and it should now be clean. If it is not, then see the next section.

You can change the file extension from HTML to HTM without any impact if preferred.

HTM FILES WITH CHARACTER ISSUES

If an imported HTM file has any characters issues, changing the encoding may help.

Create a copy of the file first in case this does not help, then open the file in Notepad++ and go to the Encoding menu.

Whatever it shows, click the highlighted option. Then import the file into RoboHelp again.

E <u>n</u> c	oding	<u>L</u> anguage	Se <u>t</u> tings	T <u>o</u> ols		
	ANSI					
•	UTF-8	UTF-8				
	UTF-8-BOM					
	UCS-2 BE BOM					
	UCS-2	2 LE BOM				
	Chara	acter sets		>		
	Conv	ert to ANSI				
	Conv	ert to UTF-8				
	Conv	ert to UTF-8-	BOM			
	Conv	ert to UCS-2	BE BOM			
	Conv	ert to UCS-2	LE BOM			