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 Ã 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.
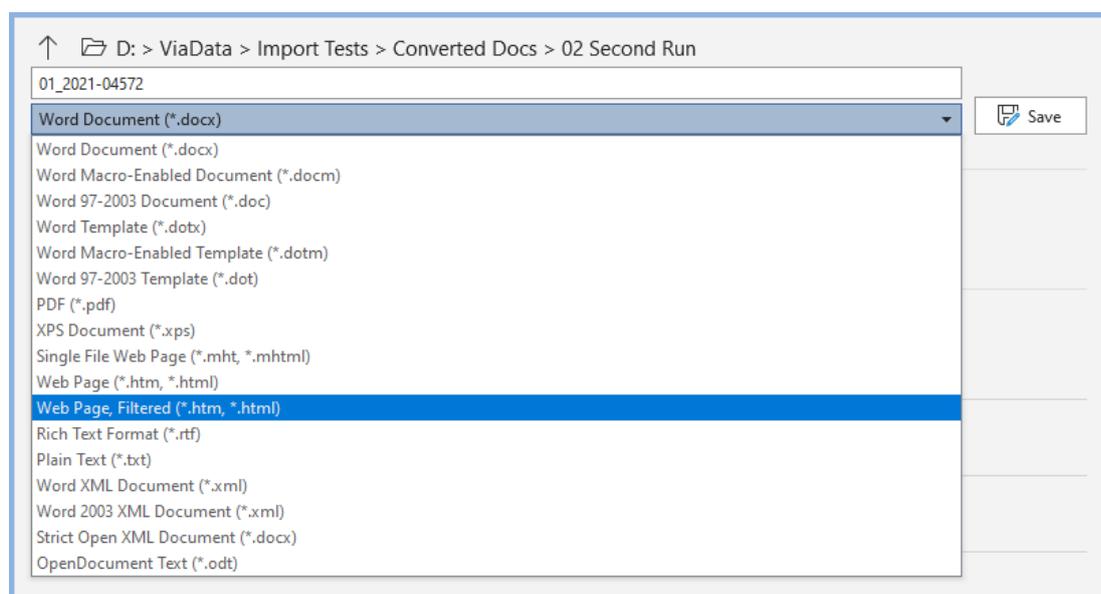
Click the More Options link.
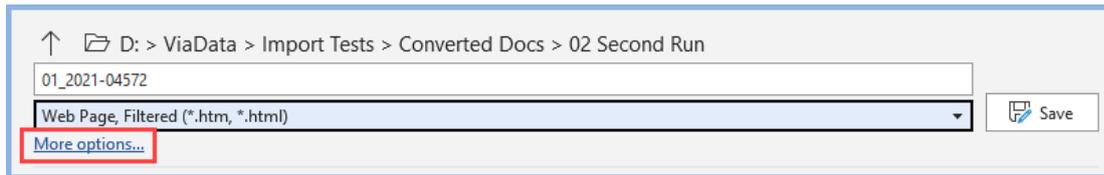
Select Web Options > Encoding and set it to UTF 8 as shown.
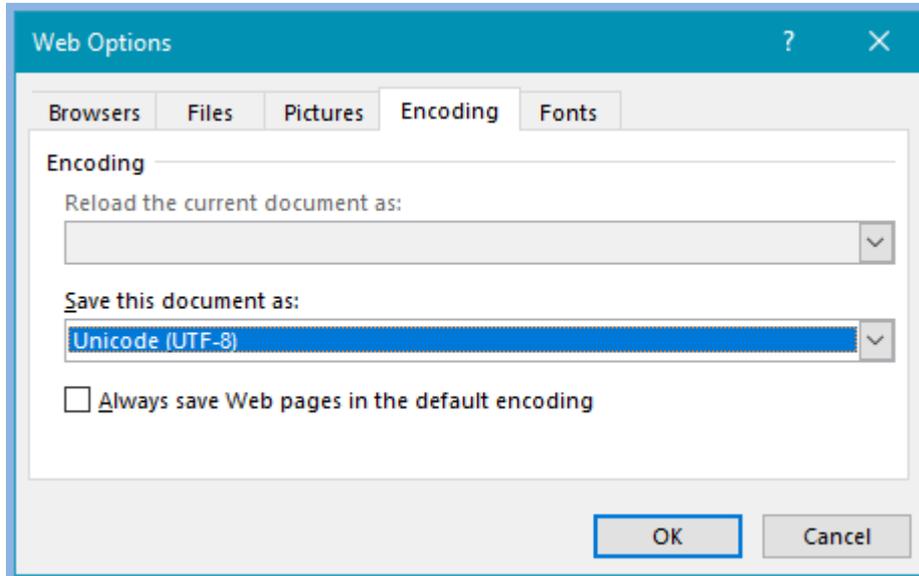
Click OK and Save. Import the file again and it should now be clean. If it is not, then see the next section.

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.