

Some characters are not displayed correctly when I open CSV output in Excel

It is possible to open CSV files directly in Excel by just double clicking on them in an explorer window. But doing this does not allow Unicode characters to display correctly. CSV files with Unicode characters should be imported into Excel and have the encoding set to UTF-8.

To import a file into Excel and change the encoding follow these steps:

1. Open Excel
2. Create New Blank Workbook
3. Click Data and select From Text/CSV on the upper left side of the screen
4. Browse to and select your CSV file
5. Change File Origin to 65001: Unicode (UTF-8)

faqs_270357_export_20171013141325.csv

Question	Answer Html	Section
Introduction to Normalizer	<p class="MsoNormal">Adobe Systems introduced the...	Adobe Normalizer
Requirements and recommended tools	Datalogics PDF Java Toolkit requires the Java D...	Datalogics PDF Java Toolkit

faqs_270357_export_20171013141325.csv

File Origin	Delimiter
1252: Western European (Windows)	Comma
874: Thai (Windows)	
857: Turkish (DOS)	
28599: Turkish (ISO)	
10081: Turkish (Mac)	
1254: Turkish (Windows)	
10017: Ukrainian (Mac)	
1200: Unicode	
1201: Unicode (Big-Endian)	
12001: Unicode (UTF-32 Big-Endian)	
12000: Unicode (UTF-32)	
65000: Unicode (UTF-7)	
65001: Unicode (UTF-8)	
20127: US-ASCII	
1258: Vietnamese (Windows)	
20005: Wang Taiwan	
850: Western European (DOS)	
20105: Western European (IA5)	
28591: Western European (ISO)	
10000: Western European (Mac)	
1252: Western European (Windows)	

faq5_270357_export_20171013141325.csv

File Origin	Delimiter	Data Type Detection
65001: Unicode (UTF-8)	Comma	Based on first 200 rows

Question	Answer Html	Section
Introduction to Normalizer	<p class="MsoNormal">Adobe Systems introduced the...	Adobe Normalizer
Requirements and recommended tools	Datalogics PDF Java Toolkit requires the Java D...	Datalogics PDF Java Toolkit

If you open the CSV file directly, the cells will appear like this:

î	Alpha	101	â†”	arrowboth	253	
---	-------	-----	-----	-----------	-----	--

This is what you will see if you import as UTF-8:

Alpha	101		arrowboth	253
-------	-----	--	-----------	-----