QGIS Application - Bug report #1310
Saving/Loading Unicode data in shapefiles

2008-09-11 08:45 AM - zachariahyoder -

Status:	Closed			
Priority:	Low			
Assignee:	nobody -			
Category:				
Affected QGIS version:		**Regression?:**	No	
Operating System:	Windows	**Easy fix?:**	No	
Pull Request or Patch supplied:		**Resolution:**	fixed	
Crashes QGIS or corrupts data:		**Copied to github as #:**	11370	

Description

[You will note this is my first Ticket. Please have patience if I have made any mistakes. As far as I can tell this is not a duplicate ticket...]

Saving: Add unicode data to a shapefile through the user interface either by
 1) using "Open table" and entering the data in the table, or by
 2) using "Capture Point", adding a point and entering the data

Enter data that is non-Ascii (e.g. ŋɪʃ)
Set the label to be data-defined.

If the point has not yet been saved to the database the Unicode characters will appear correctly in the labels. After saving all the non-Ascii characters will change to question marks (?).

Loading:
1) Create a shapefile and close QGIS
2) Change a database file to be encoded in Unicode and enter a few non-ASCII characters (I do this with [[OpenOffice]] Calc by a) copying to a new file, b) saving as dbf in UTF-8 encoding, c) cosing [[OpenOffice]] and d) renaming the new file with the original name).
3) Open QGIS. The shapefile will load and show the ASCII characters (which are a subset of Unicode), but the Unicode only characters will be interpreted as if ASCII.

I'm not familiar with the Shapfile specifications, but I'm surprised that I can't save Unicode data.

I'm using Windows XP, but I assume this is true for all platforms and have thus selected "ALL".
I'm testing on version 0.11.0 but assuming it affects the HEAD as well.

History

#1 - 2008-11-29 02:58 AM - Anne Ghisla

- Resolution set to worksforme

- Status changed from Open to Closed

This problem does not affect QGIS 1.0 preview1 for Windows. I modified a field in a shapefile adding non-ASCII characters (àòè§ç), saved edits and reopened the table. They are all correctly displayed.

Anyway, if I open the dbf file with [[OpenOffice]] or Excel, the characters are correctly displayed if I choose Western Europe (ISO-8859-15/EURO) or (Windows-1252/WinLatin 1). It doesn't work with UTF-8 nor Wester Europe (DOS/OS2-850/International) encoding, the latter is [[OpenOffice]] default choice.

I work on Windows XP home with an Italian keyboard layout.

#2 - 2009-01-19 12:03 AM - zachariahyoder -

- Status changed from Closed to Feedback

- Resolution deleted (worksforme)

Please try it again with a Unicode encoding, such as UTF-8.

Here are some examples you can cut and paste to try:

□□□□□□□□(Japanese)

ŋʃtooŋɱmɱ (IPA)

Thanks!

#3 - 2009-01-19 06:51 AM - cdavilam -

If I select utf-8 encoding when I load a shapefile into a qgis project, non ASCII characters are correctly displayed, but if I close and reopen qgis, next time I open the project, non ASCII characters are wrong and I have to remove the layer and reload it so that they are displayed properly.

Shouldn't QGIS save the encoding of the layer as it was loaded first time?

#4 - 2009-01-20 02:22 AM - zachariahyoder -

While implementing saving the encoding of a file...

Please add an option in New Vector Layer dialog to choose the encoding when first making a new vector/shapefile. (Layer -> New Vector Layer -> New Vector Layer dialog)

It seems the user is currently not given a choice.

P.S. A big thank you to the people who spend their time to fix bugs like this!!!

#5 - 2009-01-20 03:02 AM - zachariahyoder -

- Status changed from Feedback to Closed

- Resolution set to worksforme

I have repeated the steps described by cdavilam and it works for me in both 0.11.0 and 1.0.0 preview II (which is called .preview1 in the about window)

See #1496 "setting the encoding..."

#6 - 2009-03-25 02:32 AM - harrikoo -

- Status changed from Closed to Feedback

- Resolution deleted (worksforme)

Replying to [comment:5 zachariahyoder]:

I have repeated the steps described by cdavilam and it works for me in both 0.11.0 and 1.0.0 preview II (which is called .preview1 in the about window)

I'm running QGis 1.0.1-Kore (reported by QGis), installed on Windows using the OSGeo4W installer, and this bug seems to exist still:

When adding an ESRI Shapefile containing UTF-8 data as a new layer to a project, the labels are shown correctly. Saving the project, closing the program, restarting it and reopening the project, non-ascii characters in the layer labels are not shown correctly.

I did look through the saved project file, and the string "utf" does not appear anywhere there, definitively not below the layer definition, so it seems, that at least in this version, the layer encoding is NOT saved in the project files.

Harri K.

#7 - 2009-04-12 12:36 PM - Marco Hugentobler

- Status changed from Feedback to Closed

- Resolution set to fixed

Provider encoding is now saved to and read from project file in (backport to 1.0 follows soon). I don't have any japanese characters to test, for my german special characters it works now.

#8 - 2009-08-22 01:01 AM - Anonymous

Milestone Version 1.0.1 deleted