

## QGIS Application - Bug report #4343

### Shapefile, created in Qgis, encoding not recognized by Esri ArcGIS 10

2011-10-03 01:41 AM - vvj -

<b>Status:</b>	Closed		
<b>Priority:</b>	Low		
<b>Assignee:</b>			
<b>Category:</b>	Data Provider		
<b>Affected QGIS version:</b>	master	<b>Regression?:</b>	No
<b>Operating System:</b>		<b>Easy fix?:</b>	No
<b>Pull Request or Patch supplied:</b>		<b>Resolution:</b>	fixed
<b>Crashes QGIS or corrupts data:</b>		<b>Copied to github as #:</b>	14280
<b>Description</b>			
<p>Hello,</p> <p>I have created shapefile with UTF-8 attribute values. In QuantumGIS 1.7.1 they are displayed correctly, but in Esri ArcGIS 10 encoding is lost.</p> <p>Registry was modified according to <a href="http://support.esri.com/en/knowledgebase/techarticles/detail/21106">http://support.esri.com/en/knowledgebase/techarticles/detail/21106</a> , correct .cpg file was created - UTF-8 encoding still not recognized.</p> <p>Temporary workaround: open .dbf with LibreOffice Calc (encoding should be selected "Unicode UTF8), save it. Then create .cpg file with correct encoding name ("UTF-8") in it. After these steps, shapefile correctly can be opened in ArcGIS.</p> <p>Attached are two sample shapefiles:</p> <ul style="list-style-type: none"><li>- test_utf - original QuantumGIS shapefile with manually created .cpg</li><li>- test_utf_lo - shapefile .dbf was modified with LibreOffice Calc</li></ul> <p>Also there is screenshot with actual and detected encoding sample.</p>			
<b>Related issues:</b>			
Related to QGIS Application - Bug report # 5911: Language Driver ID in dbf fi...		<b>Closed</b>	<b>2012-06-30</b>

#### History

##### #1 - 2011-10-06 01:01 AM - Giovanni Manghi

- Category set to Data Provider
- Priority changed from Normal to Low

Arcgis does not allow the user to choose the encoding when opening a vector, as QGIS does?

##### #2 - 2011-10-06 03:03 AM - Giovanni Manghi

I'm not really sure this is a QGIS issue...

##### #3 - 2011-10-11 11:19 PM - vvj -

ArcGIS does not allow to choose encoding. For UTF-8 support in shapefiles, several requirements must be met:

1. Registry key with correct encoding must be set (otherwise, default system encoding will be used).
2. "When opening a shapefile and dBASE file in ArcGIS Desktop, the Desktop programs look at the Language Driver ID (LDID) in the header of a dBASE file, or an associated \*.CPG file, which are both used to define the code page, and to determine the code page of the file that is read".

In my case, registry key and .cpg files are created, but encoding is not recognized, until .dbf file is opened with LibreOffice Calc (upon opening, encoding can be choosed) and saved with the same name. dbf, saved with Calc, is slightly different than QGIS created.

I guess, that QGIS incorrectly sets encoding in .dbf header, and yes, it IS QGIS issue.

I have attached "bad" and "good" shapefile samples, maybe you can figure out, what is different.

**#4 - 2011-12-16 02:10 PM - Giovanni Manghi**

- Target version set to Version 1.7.4

**#5 - 2012-04-16 06:29 AM - Paolo Cavallini**

- Target version changed from Version 1.7.4 to Version 1.8.0
- Crashes QGIS or corrupts data set to No
- Affected QGIS version set to master

**#6 - 2012-09-04 11:57 AM - Paolo Cavallini**

- Target version changed from Version 1.8.0 to Version 2.0.0

**#7 - 2013-04-22 01:57 PM - Borys Jurgiel**

- Resolution set to fixed
- Status changed from Open to Closed

Fixed in commit:c0551a68c250489955c9831f5714f187df087d83

The difference between the files is the test\_utf contains LDID=0x57, what is recognized by various software either as "current codepage" or "ISO-8859-1". Seems this value overrides CPG declaration for ArcGIS. LibreOffice Calc resets the LDID value to 0, so the CPG file can act. Now QGIS will reset LDID and create CPG file when creating new layer.

For existing layers you can use Shapefile Encoding Fixer plugin.

Files			
test_utf8.zip	90.8 KB	2011-10-02	wvj -