QGIS Application - Bug report #5911
Language Driver ID in dbf file of new shapefile

2012-06-30 04:30 AM - Minoru Akagi

Status:	Closed		
Priority:	High		
Assignee:			
Category:	Data Provider		
Affected QGIS version:	1.8.0	**Regression?:**	No
Operating System:		**Easy fix?:**	No
Pull Request or Patch supplied:	No	**Resolution:**	
Crashes QGIS or corrupts data:	Yes	**Copied to github as #:**	15355

Description

Shapefile created with QGIS has 0x57 value in the LDID field of dbf file regardless of what encoding has been selected in the dialog. The LDID/87 (0x57) value means ISO-8859-1, which is a default. See OGR driver: ESRI Shapefile. This issue causes character corruption in the attribute table.

In detail, though the createEmptyDataSource() receives encoding as one of the parameters, it is not used to create shapefile.

Although the LDID might be set to the codepage specified by the user, the generated dataset that had zero in the LDID field and included .cpg file might be easier to handle as a user. This point is desirable to be discussed.

Best regards.

Related issues:

Related to QGIS Application - Bug report # 5900: QGIS 1.8.0 windows standalon...	**Rejected**	**2012-06-29**
Related to QGIS Application - Bug report # 5255: Wrong codepage of shapefile	**Closed**	**2012-03-29**
Related to QGIS Application - Bug report # 4343: Shapefile, created in Qgis, ...	**Closed**	**2011-10-03**
Related to QGIS Application - Bug report # 5622: layer properties, general, p...	**Closed**	**2012-05-20**
Related to QGIS Application - Bug report # 5508: DBF encoding and cyrillic va...	**Closed**	**2012-04-26**
Related to QGIS Application - Bug report # 5340: QGIS loses non-latin letters...	**Closed**	**2012-04-11**
Related to QGIS Application - Bug report # 5927: ESRI shapefile encoding problem	**Closed**	**2012-07-02**
Related to QGIS Application - Bug report # 5982: vector layer encoding defaul...	**Closed**	**2012-07-09**
Related to QGIS Application - Bug report # 6057: QGIS 1.8 Encoding problem wi...	**Closed**	**2012-07-17**
Related to QGIS Application - Bug report # 13203: When opening Shapefile the ...	**Closed**	**2015-08-10**
Duplicated by QGIS Application - Bug report # 6500: Language Encoding very br...	**Closed**	**2012-10-11**

Associated revisions

Revision 75dc85b4 - 2012-08-16 10:33 PM - Jürgen Fischer

allow to ignore (OGR's interpretation of) shape file encoding (might fix #5911)

Revision 7fb46498 - 2013-02-16 11:27 AM - Jürgen Fischer

also optionally apply SHAPE_ENCODING to layer creation (fixes #5911)

History

#1 - 2012-06-30 04:50 AM - Jürgen Fischer

Mapping between LDID values and codepages (LDID/87 means ISO-8859-1):

http://trac.osgeo.org/gdal/browser/branches/1.9/gdal/ogr/ogrsf_frmts/shape/ogrshapelayer.cpp#L170

#2 - 2012-08-01 05:34 AM - Minoru Akagi

I show three solutions:
1. Create a mapping which converts the MIBenum to text of cpg file(or LDID value). QTextCodec::codecForName(encoding_name)->mibEnum() gives MIBenum.

- The encodings in the listbox of QgsEncodingFileDialog are those that QTextCodec supports.
- Mapping between MIBenum and character set name is at http://www.iana.org/assignments/character-sets
- Some encodings supported by QTextCodec are not supported by Shapefile.

2. Change encodings in the listbox of QgsEncodingFileDialog for supported encodings of Shapefile.
- Shapefile supports code page values in http://resources.arcgis.com/fr/content/kbase?fa=articleShow&d=21106
- There are a few plug-ins which use QgsEncodingFileDialog.
- Some other dialogs (such as "Save Vector Layer As" dialog) also have the encoding listbox.

3. Generate a Shapefile dataset that has zero in LDID field and no cpg file regardless of the selected encoding. Then QGIS opens the dataset with the encoding specified.

I guess number 3 is easiest one.

#3 - 2012-08-01 11:04 PM - Minoru Akagi

- File qgsogrprovider3.patch added

I attach a patch for solution number 3.

#4 - 2012-08-02 09:59 PM - Minoru Akagi

See also http://trac.osgeo.org/gdal/ticket/4739

#5 - 2012-08-04 02:34 AM - Minoru Akagi

If default LDID of OGR Shapefile Driver dataset creation was changed to zero, the encoding problem of shapefiles generated via the "Save vector layer as" dialog would be solved as well. Also that of shapefiles generated by some plug-ins(e.g. fTools).

#6 - 2012-08-16 07:39 AM - Marco Lechner

- Priority changed from Normal to High
- File encodingtest.zip added

I guess this should be priority high, because all shapes not having LDID set or not having an cpg-file (which surely are most of Shapes out there) are forced latin1. Users choice to select the encoding when loading a layer, should always overwrite the default. Otherwise it can not be understood why a Shapes attributetable is always displayed wrong, wether the user tries to define encoding or not. This brakes the behavior of QGIS as known by the user.

I add some Shapes and qgs-files for testing.

btw it depends on gdal-Version 1.9.x

#7 - 2012-08-16 01:34 PM - Jürgen Fischer

- Status changed from Open to Closed

Fixed in changeset commit:"75dc85b4d652116814873bb7674cab15ce6cde66".

#8 - 2012-09-08 01:13 AM - Minoru Akagi

Jef's fix is good at reading shapefiles and creating new shapefiles, so the issue of this ticket has been fixed. However I've found an encoding problem of the shapefile generated via the "Save vector layer as" dialog or fTools is still existing. OGR Shapefile driver converts character encoding from UTF-8 to ISO-8859-1 and rarely garbles attribute strings.

See also GDAL #4808

#9 - 2012-09-10 09:45 PM - Minoru Akagi

Sorry,

Testing it again today, I don't experience any character corruption of shapefiles generated via both "Save vector layer as" and fTools. Maybe I had forgotten to check the option. The fix is very nice!

#10 - 2013-02-16 01:52 AM - Minoru Akagi

- Status changed from Closed to Reopened

I've noticed that the garbling occurs when saving Spatialite/PostGIS layer to Shapefile. The above fix means that if the option "Ignore shapefile encoding" is checked, OGR Shapefile's encoding conversion will be disabled when a OGR layer is loaded. Saving after editing newly created layer has no problem because in the layer creation QGIS generates an empty layer and then loads it. However, in the particular case I encountered, if no OGR layer has been loaded, output will be garbled.

#11 - 2013-02-16 02:28 AM - Jürgen Fischer

- Status changed from Reopened to Closed

Fixed in changeset commit:"7fb46498c9fb3c14a2d0b0fcc8e634dba2f1cade".

#12 - 2013-02-16 03:28 AM - Minoru Akagi

Thank you very much.

#13 - 2013-04-12 04:19 AM - Minoru Akagi

- File japan_poly.zip added

#14 - 2013-04-18 02:02 AM - Ivan Mincik

- File shp-encoding-problem-cp1250.zip added

I am attaching test Shapefile in cp1250 and the same in utf-8 for comparison. Both where made by QGIS 1.8 compiled with GDAL 1.7 (in Debian Squeeze).

#15 - 2013-04-18 02:40 AM - Jürgen Fischer

See also http://ssrebelious.wordpress.com/2012/03/11/qgis-and-gdal1-9-encoding-issue-a-workaround and http://plugins.qgis.org/plugins/shapefile_encoding_fixer/

#16 - 2013-04-29 06:55 AM - Minoru Akagi

In master, LDID is set to zero and .cpg file is appended except for "System" on creating shapefile. Thank you Borys!

#17 - 2018-07-13 04:26 PM - Jürgen Fischer

- Related to Bug report #13203: When opening Shapefile the .cpg file is ignored in Windows 8.1 added

Files

qgsogrprovider3.patch	1.02 KB	2012-08-01	Minoru Akagi
encodingtest.zip	6.16 KB	2012-08-16	Marco Lechner
japan_poly.zip	30.5 KB	2013-04-12	Minoru Akagi
shp-encoding-problem-cp1250.zip	238 KB	2013-04-18	Ivan Mincik