QGIS Application - Bug report #11992
DXF with UTF-8 layer names imported with ascii layer attributes

2015-01-15 11:29 PM - Adam Szieberth

Status:	Closed		
Priority:	Normal		
Assignee:			
Category:	Vectors		
Affected QGIS version:	2.6.1	**Regression?:**	No
Operating System:		**Easy fix?:**	No
Pull Request or Patch supplied:	No	**Resolution:**	not reproducable
Crashes QGIS or corrupts data:	No	**Copied to github as #:**	20198

Description

Compare the values of column "Layer" of the DXF imported layer "entities LineString" with the UTF-8 layer name values of the raw content of the attached DXF file (Open it with UTF-8 encoding).

Esample:

Új-Watt_Építendő KÖF szab. vez.|
˘
Ášj-Watt_Ă‰pĂ-tendĹ‘ KĂ–F szab. vez.

History

#1 - 2015-01-18 02:40 AM - Giovanni Manghi

- Category set to Vectors

- Status changed from Open to Feedback

ogrinfo returns

giovanni@sibirica:~/Downloads > ogrinfo Kaloz_Belmajor20141210.dxf
Warning 1: One or several characters couldn't be converted correctly from CP1250 to UTF-8.
This warning will not be emitted anymore
INFO: Open of `Kaloz_Belmajor20141210.dxf'
using driver `DXF' successful.
1: entities

so if I re-save your dxf with the cp1250 encoding then ogrinfo does not returns that problem anymore and in QGIS the letters are ok.

#2 - 2015-01-19 12:02 AM - Adam Szieberth

Thanks for the workaround.

The attached DXF file was produced by Teigha File Converter which converted it from a DWG. By searching "dwg qgis" on Google, the first match links to gis.stackexchange.com where the answer leads to that software. Would be nice to see the chain of the softwares doing the task nicely.

#3 - 2015-01-19 12:06 AM - Adam Szieberth

Failed to encode the character '²' (U+B2) at column 8 in line 144520 with the encoding "windows-1250".

#4 - 2015-01-19 12:11 AM - Giovanni Manghi

Adam Szieberth wrote:

> *Failed to encode the character '²' (U+B2) at column 8 in line 144520 with the encoding "windows-1250".*

I used "gedit" one of the many text editors on Linux. On Windows try notepad++. Anyway this does not seems a qgis issue, so I suggest to close this ticket.

#5 - 2015-01-20 03:38 AM - Adam Szieberth

Well, no matter what you use for encoding, that won't make the upper indexed "2" into a valid cp1250 character. I am quite sure it got ignored somehow and got replaced with a question mark or whatever.

Still, the issue remains: an UTF-8 DXF is imported incorrectly into QGIS. Maybe that is an OGR issue and should be reported there.

#6 - 2015-01-20 06:37 AM - Giovanni Manghi

Adam Szieberth wrote:

> *Well, no matter what you use for encoding, that won't make the upper indexed "2" into a valid cp1250 character. I am quite sure it got ignored somehow and got replaced with a question mark or whatever.*

Actually I can, saving the file with iso-8859-whatever

> *Still, the issue remains: an UTF-8 DXF is imported incorrectly into QGIS. Maybe that is an OGR issue and should be reported there.*

My gut feeling is that the encoding of your dxf was wrongly defined/saved at the source.

#7 - 2015-01-23 12:40 AM - Adam Szieberth

I was not inteded to install Notepad++ but I did it anyway to do the conversion. Well, that paricular line (144520) changed from:

3x50 mm² ald.

to

3x50 mm2 ald.

Notepad++ encodes in replace mode without warnings. The exception above was raised by jEdit and I am glad it was raised.

```
Python 3.4.2 (v3.4.2:ab2c023a9432, Oct  6 2014, 22:16:31) [MSC v.1600 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> with open("Kaloz_Belmajor20141210.dxf", encoding='utf-8') as f:
...     s = f.read()
...
>>> s_cp1250 = s.encode('cp1250')
```

```
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "c:\\Python34\\lib\\encodings\\cp1250.py", line 12, in encode
    return codecs.charmap_encode(input,errors,encoding_table)
UnicodeEncodeError: 'charmap' codec can't encode character '\\xb2' in position 1207566: character maps to <undefined>
>>> s_cp1250 = s.encode('cp1250', 'ignore')
>>> s_cp1250 = s.encode('cp1250', 'replace')
```

The exception was raised because of the upper indexed "2" again. Nothing is wrong with the source file, it is a valid UTF-8 encoded DXF. However, it can't get encoded to cp1250 losslessly.

#8 - 2015-01-23 04:00 AM - Giovanni Manghi

Adam Szieberth wrote:

> *I was not inteded to install Notepad++ but I did it anyway to do the conversion. Well, that paricular line (144520) changed from:*
>
> *3x50 mm² ald.*
>
> *to*
>
> *3x50 mm2 ald.*

using Kate under KDE works as expected, that line remains as

3x50 mm² ald.

I still do no understand why this should be a qgis issue. cheers!

#9 - 2015-05-10 02:36 AM - Giovanni Manghi

- Resolution set to not reproducable

- Status changed from Feedback to Closed

closing for lack of feedback.

Files

Kaloz_Belmajor20141210.dxf	2.75 MB	2015-01-15	Adam Szieberth