QGIS Application - Bug report #6299
Sextante output do not have the same Encoding as the inputs

2012-09-03 02:36 AM - Filipe Dias

Status:	Closed			
Priority:	High			
Assignee:	Victor Olaya			
Category:	Processing/Core			
Affected QGIS version:	2.4.0	**Regression?:**	No	
Operating System:		**Easy fix?:**	No	
Pull Request or Patch supplied:		**Resolution:**	fixed/implemented	
Crashes QGIS or corrupts data:	No	**Copied to github as #:**	15595	

Description

If I select as an input a layer with special characters such as "´", "~", "`" etc the output of any Sextante tool replaces these characters with weird symbols

Related issues:

Duplicated by QGIS Application - Bug report # 11174: Qgis processing toolbox ...	**Rejected**	2014-09-09

History

#1 - 2012-10-19 09:22 AM - Filipe Dias

- Priority changed from Normal to High

Im changing the priority because this "corrupts" the output's table.

#2 - 2012-10-30 05:26 AM - Alexander Bruy

- Status changed from Open to Feedback

Not sure if I understand problem correctly. Your input file has some non-ASCII characters in attributes and in output file, produced by SEXTANTE this characters are corrupted. Right?

In this case this is not SEXTANTE issue but QGIS and GDAL one, there are number of tickets about this problem (see #5255, #6500 and related tickets). If you use master branch then check "Ignore shapefile encoding" checkbox under Settings-Options-General, restart QGIS and try again.

#3 - 2012-11-01 12:02 PM - Filipe Dias

- Status changed from Feedback to Closed

Yes that was it.

I tried with today's QGIS Master and this no longer happened. It wasn't even necessary to check "ignore shapefile encoding".

Sinc this is/was a QGIS issue Im closing this.

#4 - 2012-11-04 09:09 AM - Pedro Venâncio

- Status changed from Closed to Reopened

Hi,

I'm reopening this ticket because I think that there is even a problem.

I leave here [1] a table with the behavior I get with encoding ISO-8859-1 or UTF-8, in different situations and with the same tools from QGIS Vector menu, from Sextante -> fTools or from Sextante -> SAGA.

The data I am testing is this one (originally in ISO-8859-1) [2].

Could someone confirm? I'm using Xubuntu and QGIS master.

Thanks!

[1] https://dl.dropbox.com/u/5772257/qgis/qgis_sextante_encoding_problem.xls
[2] http://www.igeo.pt/produtos/cadastro/caop/download/CAOP20121_Shapes/Cont_AAd_CAOP20121.zip

#5 - 2012-11-05 12:58 AM - Filipe Dias

Pedro, to make sure I understand correctly (and to make sure others do to):

You made a table in which you record your findings regarding encoding problems. You used different tools and different encoding systems to test this issue. With some combinations of "tool x encoding" you found problems. But we others you didnt, is that it?

I only tested with ftools dissolve and got no error. But after reading your findings, Im lead to believe that there is a "larger issue" behind this.

Did you try Alexander Bruy's recomendation above?

#6 - 2012-11-05 02:00 AM - Pedro Venâncio

Hi Filipe,

Exactly, and in addition, following Alexander's indication, I did the tests with the option "Ignore shapefile encoding" selected (first 4 columns in the spreadsheet attached) and not selected (last 4 columns).

The column "Result encoding" shows the result "Correct" when execution of the operation leads to a shapefile with the correct encoding, and "Incorrect" when the encoding of the resulting shapefile is wrong and different from the input layer.

#7 - 2013-04-30 02:54 AM - Filipe Dias

I understand there have been some changes regarding Encoding in QGis Master. Is this still happening?

#8 - 2013-06-04 08:19 AM - Filipe Dias
- Priority changed from High to Severe/Regression

This is still happening and it's probably the biggest issue on Sextante right now.

#9 - 2013-08-23 10:40 AM - Borys Jurgiel

Although QGIS works properly with and without "Ignore Shapefile encoding" now, please remember there is one unresolvable problem as long as it's turned off (so GDAL conversion is not deactivated):

Since GDAL converts the source encoding to UTF-8 (more or less properly, depending on the declaration in the SHP), it doesn't pass any information about the source encoding to QGIS, so QGIS is unable to apply it to the output file. It always gets UTF8 from GDAL provider and it's a real pain for any SHP processing...

#10 - 2013-08-28 04:41 AM - Giovanni Manghi

- Priority changed from Severe/Regression to High

This seems to be platform dependent, in fact I made a few tests now based on #6299-4 and on Windows I can always see the expected results (no garbled attributes) while on Linux I confirm the issue.

Unfortunately for us Linux users I don't think that at this stage this can be tagged as blocker.

#11 - 2013-08-28 05:12 AM - Giovanni Manghi

Giovanni Manghi wrote:

> *This seems to be platform dependent, in fact I made a few tests now based on #6299-4 and on Windows I can always see the expected results (no garbled attributes) while on Linux I confirm the issue.*
>
> *Unfortunately for us Linux users I don't think that at this stage this can be tagged as blocker.*

it does not seems to depend on the gdal/ogr version. I initially tested qgis 32bit on Windows that uses gdal 1.9.2, then I tested qgis 64bit/Windows and the results are the same (ok) and it uses gdal 1.10.

#12 - 2013-08-28 06:30 AM - Alexander Bruy

I can't download sample file, mentioned in previous comments, but this issue not reproducible here (Linux, GDAL 1.10.0) with my own test data (contains Cyrillic and some other non-ASCII characters).

Seems the only possible way to get broken attributes is using "System" encoding when saving file.

#13 - 2013-08-28 10:22 AM - Giovanni Manghi

It seems to be also a problem not related to QGIS/SEXTANTE, in fact I can replicate issues with the enconding just using SAGA (for example) from the command line.

#14 - 2013-11-19 04:37 AM - Filipe Dias

Alexander Bruy wrote:

> *Seems the only possible way to get broken attributes is using "System" encoding when saving file.*

I still get messed up characters after merging (Merge shapes - SAGA 2.0.8) correctly created .shp with ISO 8859-2

#15 - 2014-02-03 02:28 AM - Filipe Dias

- Priority changed from High to Severe/Regression

#16 - 2014-02-16 01:59 PM - Giovanni Manghi

- Status changed from Reopened to Feedback

- Priority changed from Severe/Regression to High

Filipe Dias wrote:

> Alexander Bruy wrote:
>
>> Seems the only possible way to get broken attributes is using "System" encoding when saving file.
>
>> I still get messed up characters after merging (Merge shapes - SAGA 2.0.8) correctly created .shp with ISO 8859-2

For how painful this can be, I don't think that any Sextante ticket should be a blocker, because Sextante/processing is upgreadable at any moment. This will obviously be more evident when (maybe during next dev meeting) merge the processing tickets into main qgis ones.

On the other hand I still doubt that this in just a processing issue. Did you tested the same operation in native SAGA? I personally get messed attributes there too.

#17 - 2014-02-17 02:10 AM - Filipe Dias

On SAGA 2.10, Merge Shapes doesn't corrupt the encoding.

I also tried v.overlay from QGIS and it doesn't corrupt the encoding.

Can you confirm that on your system only SAGA algs corrupt the encoding?

#18 - 2014-10-04 12:08 PM - Giovanni Manghi

- Project changed from 78 to QGIS Application

- Category deleted (63)

- Affected QGIS version set to 2.4.0

- Crashes QGIS or corrupts data set to No

#19 - 2014-10-04 12:09 PM - Giovanni Manghi

- Category set to Processing/Core

#20 - 2015-12-19 12:10 PM - Giovanni Manghi

- Status changed from Feedback to Closed

- Resolution set to fixed/implemented

it seems to me that this is now fixed (at least in master), please reopen of necessary.