QGIS Application - Bug report #6013
Delimited text provider/plugin doesn't work as expected with quoted strings containing the delimiter

2012-07-11 02:17 AM - Giuseppe Sucameli

Status:	Closed		
Priority:	Normal		
Assignee:	Giuseppe Sucameli		
Category:	Data Provider		
Affected QGIS version:	master	**Regression?:**	No
Operating System:		**Easy fix?:**	No
Pull Request or Patch supplied:	No	**Resolution:**	
Crashes QGIS or corrupts data:	No	**Copied to github as #:**	15401

Description

In a CSV file created from "Save as..." some string values are escaped in quotes (e.g. those which contain the delimiter), but the Delimited text provider/plugin doesn't understand a row like 1.0,"City, Country"
contains 2 values only (and not 3).

I guess this problem should be handle by the plugin/provider.

As workaround I'm using a regexp delimiter like "?,(?!\\s)"?
but it can fail in some cases.

Associated revisions

Revision beb70d31 - 2012-07-16 12:51 AM - Giuseppe Sucameli

move delimitedtext plugin functionality to the provider (fix #6013):

allow GUI and provider to use the same splitLine method.

History

#1 - 2012-07-15 04:13 AM - Giuseppe Sucameli

The provider already trims quotes from beginning and ending of the string, but it has to skip the delimiter within the string.

#2 - 2012-07-15 04:43 AM - Jürgen Fischer

How are quotes inside quoted text escaped?

#3 - 2012-07-15 05:13 AM - Giuseppe Sucameli

Jürgen Fischer wrote:

> *How are quotes inside quoted text escaped?*

I don't know, but if a CSV file follows the RFC 4180 then double quotes must be duplicated (like SQL):

- DOS-style lines that end with (CRLF) characters
- An optional header record (there is no sure way to detect whether it is present, so care is required when importing).
- Each record "should" contain the same number of comma-separated fields.
- Any field may be quoted (with double quotes).
- Fields containing a line-break, double-quote, and/or commas should be quoted. (If they are not, the file will likely be impossible to process correctly, so this should is better taken as must).
- A (double) quote character in a field must be represented by two double quote characters.

#4 - 2012-07-15 05:18 AM - Giuseppe Sucameli

- Category set to C++ Plugins

Found it!

Using a "plain" separator the provider works as expected, so it skips separators within quoted text.
The "Add Delimited Text Layer" dialog instead displays a wrong result in the "sample text" table although the layer created is ok.

#5 - 2012-07-15 05:48 AM - Jürgen Fischer

- Category changed from C++ Plugins to Data Provider

Giuseppe Sucameli wrote:

> *The "Add Delimited Text Layer" dialog instead displays a wrong result in the "sample text" table although the layer created is ok.*

I suggest to move the plugin functionality to the provider (selectWidget()...) - that would also allow GUI and provider to use the same splitLine method.

#6 - 2012-07-15 03:26 PM - Giuseppe Sucameli

- Status changed from Open to In Progress
- Assignee set to Giuseppe Sucameli

Jürgen Fischer wrote:

> *I suggest to move the plugin functionality to the provider (selectWidget()...) - that would also allow GUI and provider to use the same splitLine method.*

Right, I agree. This would also remove duplicated code.

#7 - 2012-07-15 03:52 PM - Giuseppe Sucameli

- Status changed from In Progress to Closed

Fixed in changeset commit:"beb70d3175d9995804ffe178e57368224cc9576e".

#8 - 2013-04-15 05:37 PM - Chris Crook

The handling of delimiters has been reworked in 2.0. This should now reliably handle CSV formats, including quotes and new line fields within quotes.

Committed at fab2c57478f67be01a9ac91f0ce27a1f739d0501