

QGIS Application - Bug report #4091

Create layer from delimited text (csv) does not work properly for quoted strings

2011-07-18 05:26 PM - springmeyer -

Status:	Closed	
Priority:	Low	
Assignee:		
Category:	C++ Plugins	
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 #: 14074
Description		
<p>If you import a csv with values with commas, using 'comma' as the delimiter, only commas that are unquoted should be used to split the columns.</p> <p>Right now (QGIS 1.7.0) the result of a row like:</p> <p>1, "John,Doe", "Mary, Jane"</p> <p>is to split on the , between John and Doe, which is not the right behavior.</p> <p>Assigning to ccrook as i see he's done some recent work on the plugin and can hopefully give feedback on this.</p> <p>The reason I think getting this behavior right is critical is that most csv export software (in my case I'm using LibreOffice) is going to default to quoting strings with commas and using commas as delimiters.</p>		

Associated revisions

Revision 230bbfb4 - 2012-10-29 04:22 PM - Giuseppe Sucameli

use plain delimiter if one delimiter only was selected (partially fix #4091)

History

#1 - 2011-07-18 05:41 PM - springmeyer -

I also meant to mention that when "a value" is imported the quotes are not stripped, as they should be. It is my understanding that quoted strings should be representing string literals so keeping the quotes after import is wrong.

#2 - 2011-12-09 08:36 AM - Paolo Cavallini

- Category set to C++ Plugins
- Pull Request or Patch supplied set to No

#3 - 2011-12-16 02:11 PM - Giovanni Manghi

- Target version set to Version 1.7.4

#4 - 2012-01-16 05:01 PM - Chris Crook

- Affected QGIS version set to master
- Crashes QGIS or corrupts data set to No

Definitely an issue with CSV import! The workaround for the moment is to OGR CSV format (with a VRT file) which works just fine. Will have a look at fixing this in delimited text plugin.

#5 - 2012-01-16 08:39 PM - springmeyer -

Chris Crook wrote:

Definitely an issue with CSV import! The workaround for the moment is to OGR CSV format (with a VRT file) which works just fine. Will have a look at fixing this in delimited text plugin.

Hey, thanks for commenting. I've used the VRT method and was looking for a one-step approach for novice users. I ended up solving things (for my purposes) in Mapnik by writing my own CSV plugin. So, +1 to improving this feature, but at least my original usecase is not longer critical.

#6 - 2012-04-16 06:28 AM - Paolo Cavallini

- Target version changed from Version 1.7.4 to Version 1.8.0

#7 - 2012-09-04 11:59 AM - Paolo Cavallini

- Target version changed from Version 1.8.0 to Version 2.0.0

#8 - 2012-10-29 08:22 AM - Giuseppe Sucameli

- Status changed from Open to Closed

Fixed in changeset commit:"230bbfb459f807a645fa3edbbc44b1012177bdfb".

#9 - 2012-10-29 08:33 AM - Giuseppe Sucameli

- % Done changed from 0 to 50
- Status changed from Closed to Reopened
- Assignee deleted (Chris Crook)
- Priority changed from Normal to Low

Whether you choose only one delimiter from the "selected delimiter" list it is internally converted to "plain delimiter", so now it works also quoted strings (see #6013).

If more delimiters are choosen from the "selected delimiters" list it still uses the "regexp delimiter" and it doesn't parse qouted strings.

The newline problem (quoted strings on more lines are not parsed) is still there, whatever delimiter you're using.

#10 - 2012-11-02 11:12 AM - Chris Crook

I have an update for the delimiter plugin which fixes the newline and comma issues, but it also requires an update to the plugin dialogue which I haven't had time to complete yet. Basically the approach I am considering is to use a couple of alternative parsers - one for regexp, one for plain whitespace, and one for fixed delimiters such as CSV. I'm thinking the dialog could then be a bit simpler (for the user), with an initial selection of parser type (which could include preset types, such as Excel CSV, tab delimited), and then options displayed according to the type of delimiter set.

One development issue that makes this difficult is that both the data provider plugin and the options need to access the same parsing code, but they are

different compilation modules, so I haven't figured where to put the common code, or whether to just replicate it.

#11 - 2013-04-15 05:44 PM - Chris Crook

- *Status changed from Reopened to Closed*

Fixed for 2.0 at commit [fab2c57478f67be01a9ac91f0ce27a1f739d0501](#)