QGIS Application - Bug report #18601
QGIS3: texts with leading zero get integers

2018-03-31 06:47 PM - Tobias Wendorff

Status:	Closed		
Priority:	Normal		
Assignee:			
Category:	Data Provider/Delimited Text		
Affected QGIS version:	3.0.0	**Regression?:**	No
Operating System:	Microsoft Windows 7, 64-bit	**Easy fix?:**	No
Pull Request or Patch supplied:		**Resolution:**	
Crashes QGIS or corrupts data:	No	**Copied to github as #:**	26489

Description

When loading delimited text without CSVT, fields with leading zeros (like German ZIP codes (f.e. 01099) or phone numbers) get smashed into integers.
I've written a pretty complex regex to identify numbers and numbers-alike strings. I think, for QGIS it's better to have an Excel or Calc like behavior to let the user chose, which format the columns are in?

Same behavior in commit:6ddd9ab0c7 and below.

Related issues:

Related to QGIS Application - Bug report # 18789: Import of CSV-Data	**Closed**	**2018-04-22**

Associated revisions

Revision b4f2069c - 2018-07-19 10:39 AM - Chris Crook

Delimited text detectTypes flag - fixes #18601

[FEATURE] Adds a detectTypes flag to the delimited text provider url. If
set to "no" then type detection is not done and all attributes are
treated as text fields. Otherwise the original behaviour of
detecting field types is preserved.

[needs-docs] Adds a "Detect field types" check box to the record and
field options section of the delimited text provider GUI.

This addresses (at least partially) issue #18601. A more complete
solution would be to allow users to set field types.

History

#1 - 2018-04-04 10:18 PM - Tobias Wendorff

Demo dataset (text file), which get parsed as numeric instead of string:

01234
+1555487218
001555487218
0123456789
01234567890123456789
012345678901234567890123456789
0123456789012345678901234567890123456789

#2 - 2018-04-14 06:29 AM - Tobias Wendorff

Could a mod please edit my first post? Actually, it is destroying data. I missed to add this.

#3 - 2018-04-21 02:09 PM - Nyall Dawson

- Status changed from Open to Feedback

Is this using the delimited text provider? Or ogr? You've marked it as the delimited text provider yet only ogr supports csvt.

#4 - 2018-04-21 02:16 PM - Tobias Wendorff

Sorry, it's often very hard to find the correct category. I don't have have a clue, which is the right one this time. You can find the dialog this way:

Menu: "Layer > Add Layer > Add Delimited Text Layer"
Data Source Manager: "Delimited Text"

I just released that double-quoting the values also cast the strings them to integer. As of RFC 4180, text-qualified strings should be enclosed in quotes.

#5 - 2018-04-22 03:22 PM - Jürgen Fischer

- Related to Bug report #18789: Import of CSV-Data added

#6 - 2018-04-22 03:24 PM - Jürgen Fischer

- Description updated

#7 - 2018-04-22 03:25 PM - Jürgen Fischer

- Status changed from Feedback to Open

#8 - 2018-06-25 01:00 AM - Tobias Wendorff

This still has not been fixed. I think, we should rework this feature ... it should work like XLSX import, where you can set the field type of every column. Also, a "scan 25/50/100% of file" feature is needed to correctly identify the fields.

Right now, this function is broken. It's highly destructive. An easy work-around would be to cast every field to text in import.

#9 - 2018-06-25 09:26 AM - Chris Crook

Hi Tobias. I hear your frustration.

Setting every field type would be good but is not a quick option, and harder to do well in a GUI that is already too busy. At the moment the parser identifies field type by trying to read each field as integer and float (and potentially scanning the whole file). This is further complicated as QGIS now accepts many more field types than it did, so to do this properly would require handling dates for example, which means providing some means of specifying the date format.

One option that could be added quickly though is to add an option to set every column type to text as you suggest. The expression builder could still be used to convert the text fields to other types if required. And the X and Y columns could then be converted to floats to generate geometry and return null geometries where they cannot be read as numbers. I think this needs to be an option rather than a default, as there is much value in having numeric fields available by default - most users find the "highly destructive" delimited text function very useful, and get a lot of value from having numeric data that can be

used directly for rendering, statistics, etc without having to use the expression builder. However the delimited text GUI is good at remembering options you select.

This would be one extra checkbox in the GUI, and one extra parameter in the layer URL, which seems a good solution.

How does that sound to you?

#10 - 2018-07-19 10:39 AM - Chris Crook

- % Done changed from 0 to 100

- Status changed from Open to Closed

Applied in changeset commit:qgis|b4f2069c6fd1fc6a583918327a7fbc9eccabe2ff.