QGIS Application - Bug report #17150
Processing: output for standard NxT distance matrix now wrong

2017-09-15 03:54 PM - Giovanni Manghi

Status:	Closed			
Priority:	High			
Assignee:	Victor Olaya			
Category:	Processing/QGIS			
Affected QGIS version:	master	**Regression?:**	Yes	
Operating System:		**Easy fix?:**	No	
Pull Request or Patch supplied:		**Resolution:**		
Crashes QGIS or corrupts data:		**Copied to github as #:**	25049	

Description

Up 2.14 (with the ftools tool) the result used to be correct (despite non handling well the accented letters in "unique ID fields"):
The first column had the "unique ID field" for the first input and the first row the "unique ID field" for the second input.

In 2.18 (Processing) the first column seems ok, but the first row as only DIST_1, DIST_2, DIST_3... that make the result useless.

In master the result is as in 2.18 but as additional issue a vector is outputted instead of a table.

Associated revisions

Revision 54031ea6 - 2017-10-06 12:45 PM - Alexander Bruy

[processing] fix Standard (N x T) distance matrix (ref #17150)

Revision e8628fb5 - 2017-10-06 01:18 PM - Alexander Bruy

Merge pull request #5312 from alexbruy/distance-matrix

[processing] fix Standard (N x T) distance matrix (ref #17150)

Revision c64f03d0 - 2017-10-24 02:27 PM - Alexander Bruy

[processing] keep column names in standard distance matrix (fix #17150)

Revision d34c09b3 - 2017-10-24 04:17 PM - Alexander Bruy

Merge pull request #5427 from alexbruy/processing-distance

[processing] keep column names in standard distance matrix (fix #17150)

History

#1 - 2017-09-16 02:05 AM - Nyall Dawson

I agree that the loss of information in the column names is a regression.

> *In master the result is as in 2.18 but as additional issue a vector is outputted instead of a table.*

Why is this an issue? This change at least was intentional - since the geometry can always be ignored/discarded/output saved as csv or other non-geometric format now. But if we don't keep the original geometries, then you are FORCED to rejoin the table back to the original to get these back again. To me it's a big feature that it now maintains these and just adds extra columns to the data.

#2 - 2017-09-16 09:08 AM - Giovanni Manghi

Nyall Dawson wrote:

> *I agree that the loss of information in the column names is a regression.*
>
> > *In master the result is as in 2.18 but as additional issue a vector is outputted instead of a table.*
>
> *Why is this an issue? This change at least was intentional - since the geometry can always be ignored/discarded/output saved as csv or other non-geometric format now. But if we don't keep the original geometries, then you are FORCED to rejoin the table back to the original to get these back again. To me it's a big feature that it now maintains these and just adds extra columns to the data.*

Hi Nyall,
I see your point, thanks.

The problem I see in this approach is the following:

as column names we are placing DIST_1, DIST_2, DIST_3... instead of the "unique ID field" of one of the inputs. I suspect that if we fix it and output a shape (as Processing does) then because of the 10 char limit a number of those "unique ID field" will be truncated with the very real possibility of having a few columns named the same way. If you output as CSV there is no such problem to take into account.

#3 - 2017-09-16 10:53 PM - Nyall Dawson

Processing in 3.0 uses a memory layer as intermediate steps in models, add running individual algs allows choice of output format. So I don't think this will be a restriction.

Otherwise we just flip the name to 1_dist, 2_dist, etc

#4 - 2017-09-22 03:52 PM - Håvard Tveite

I also noticed this regression a few days ago. The result of the "Standard (N x T)" version not only has new (useless) column names, **each row is also ordered** / **sorted on increasing distance**, meaning that it is impossible to find out from the result matrix which pair of points a value corresponds to!

#5 - 2017-09-22 03:54 PM - Giovanni Manghi

Håvard Tveite wrote:

> *I also noticed this regression a few days ago. The result of the "Standard (N x T)" version not only has new (useless) column names, **each row is also ordered** / **sorted on increasing distance**, meaning that it is impossible to find out from the result matrix which pair of points a value corresponds to!*

agree.

#6 - 2017-09-24 12:41 AM - Nyall Dawson

Giovanni - I've been trying to track down when this change was made to see what the rationale behind it was. I **think** it may have been an issue with the processing distance matrix since forever, and it's only since the other python distance matrix tool was replaced with the processing version that this issue became known. Does that sound correct to you?

#7 - 2017-09-25 12:36 PM - Giovanni Manghi

Nyall Dawson wrote:

> *Giovanni - I've been trying to track down when this change was made to see what the rationale behind it was. I **think** it may have been an issue with the processing distance matrix since forever, and it's only since the other python distance matrix tool was replaced with the processing version that this issue became known. Does that sound correct to you?*

Hi Nyall,
I think you are right: the tool in "ftools" worked as expected (despite very slow for large datasets). Cheers!

#8 - 2017-10-06 01:20 PM - Alexander Bruy

Fixed in 2.18

#9 - 2017-10-06 01:22 PM - Giovanni Manghi

- Affected QGIS version changed from 2.18.12 to master

#10 - 2017-10-24 04:16 PM - Alexander Bruy

- % Done changed from 0 to 100

- Status changed from Open to Closed

Applied in changeset commit:qgis|c64f03d0a43822c091e3f60bd4b8b3538d000c1e.