

QGIS Application - Bug report #19973

QGIS 3.3 master delete duplicate geometries does not work

2018-09-27 09:11 PM - salvatore fiandaca

<b>Status:</b>	Closed	
<b>Priority:</b>	High	
<b>Assignee:</b>		
<b>Category:</b>	Processing/Core	
<b>Affected QGIS version:</b>	3.3(master)	<b>Regression?:</b> No
<b>Operating System:</b>	win 10 - osgeo4w	<b>Easy fix?:</b> No
<b>Pull Request or Patch supplied:</b>		<b>Resolution:</b>
<b>Crashes QGIS or corrupts data:</b>		<b>Copied to github as #:</b> 27795
<b>Description</b>		
<p>I extracted the vertices of the polygonal vector that I attach, subsequently I launch the geo-algorithm 'delete duplicate geometries', the process is fast but I stay still at 99% and does not create anything</p> <p>it also does not work in 3.2.3</p>		

Associated revisions

Revision 9698444f - 2018-09-28 11:37 PM - Nyal Dawson

[processing] Fix inefficiencies in Delete Duplicate Geometries algorithm

..and make progress bar more accurate.

Use a spatial index to avoid comparing every feature to every other feature, and only compare against features with intersecting bounding boxes instead. Also optimise feature requests and loop logic.

Benchmarks:

Point layer, 6000k features

Before: 30 seconds

After: 0.15 seconds

Point layer, 45k features

Before: > 10 minutes

After: 7 seconds

Fixes #19973

History

#1 - 2018-09-27 09:24 PM - salvatore fiandaca

- File delete2.png added

EDIT:

for a dataset of <6k points it takes 243 seconds

## **#2 - 2018-09-28 03:50 AM - Andrea Giudiceandrea**

On my system, core i5-460M, 8 GB RAM, Windows 7 64 bit:

vertices extracted: 5629 points

running "delete duplicate geometries" on 5629 points takes

~50 seconds with QGIS 2.18.23 64 bit

50 seconds with QGIS 3.2.3 64 bit

88 seconds with QGIS 3.3.0 (80723e89fd) 64 bit (slower probably due to the debug build slowness)

resulting in a 3948 points layer

All the three versions however are affected by the "stuck at 99%" misleading strange behaviour you reported.

## **#3 - 2018-09-28 04:58 AM - Nyal Dawson**

Not a regression - the algorithm is just extremely inefficient and doesn't scale for large layers (it compares EVERY feature with EVERY other). The solution here is probably to add a spatial index so that only features with intersecting bounding boxes are tested for equality

## **#4 - 2018-09-28 05:02 AM - Nyal Dawson**

(For reference -- there's a dedicated, optimised, 'remove duplicate vertices' algorithm which may be of use here)

## **#5 - 2018-09-28 05:34 AM - Nyal Dawson**

How's "Execution completed in 0.18 seconds" sound?

## **#6 - 2018-09-28 05:49 AM - Nyal Dawson**

- Status changed from Open to In Progress

<https://github.com/qgis/QGIS/pull/8047/files>

## **#7 - 2018-09-28 02:40 PM - salvatore fiandaca**

Nyal Dawson wrote:

| How's "Execution completed in 0.18 seconds" sound?

wow, it sounds great

thank you

#8 - 2018-09-28 11:37 PM - Nyal Dawson

- % Done changed from 0 to 100
- Status changed from In Progress to Closed

Applied in changeset commit:qgis|9698444f4af4712cb4a6508c839c014fa0b335e1.

Files			
delete.png	33.6 KB	2018-09-27	salvatore fiandaca
duplicate.7z	46.3 KB	2018-09-27	salvatore fiandaca
delete2.png	32.7 KB	2018-09-27	salvatore fiandaca