

QGIS Application - Bug report #6883

curved placement for labels breaks Indic scripts (Khmer, Lao, Nepale, Bangladeshe, etc.) strings

2012-12-16 08:17 PM - Mathieu Pellerin - nIRV

Status:	Closed	
Priority:	Normal	
Assignee:	Larry Shaffer	
Category:	Labelling	
Affected QGIS version:	master	Regression?: No
Operating System:		Easy fix?: No
Pull Request or Patch supplied:		Resolution:
Crashes QGIS or corrupts data:		Copied to github as #: 16011
Description		
QGIS currently breaks Indic scripts when labelling lines with the curved placement. I believe this is due to the curved placement logic assuming a "latin" behavior to all strings. While in a latin-based language each char follows the previous one, it is not the case with Indic scripts. Indic scripts are broken down into clusters. Unicode char data for a basic cluster looks like this: [consonant],[leg(s)],[vowel]. The consonant is used as the center glyph, while leg(s) and vowel can be placed <i>all around</i> the center glyph / consonant. You can spot the broken rendering when you see the dotted-circle (which is used by rendering engines such as harfbuzz and uniscribe to indicate that legs and vowels are not attached to a needed consonant to form a cluster. This is happening in QGIS because the curved placement breaks down strings into single chars and place it.		

Associated revisions

Revision 2dc5d95f - 2015-05-30 01:36 PM - Nyall Dawson

Fix broken rendering of curved labels for scripts which use >1 char graphemes (fix #6883)

Revision 71b5a993 - 2015-06-24 01:45 PM - Nyall Dawson

Fix broken rendering of curved labels for scripts which use >1 char graphemes (fix #6883)

Cherry-picked from 2dc5d95f00770c602497f633612a6dabb8be4962

History

#1 - 2012-12-16 08:21 PM - Mathieu Pellerin - nIRV

Note: Hermann Kraus (<https://github.com/herm>) has recently implemented curved placement labels into mapnik relying on harfbuzz-ng. He might be a resourceful person to rely on. Behdad Esfahbod (<https://github.com/behdad>), the creator of harfbuzz, is also a great person to talk to.

#2 - 2012-12-16 08:54 PM - Mathieu Pellerin - nIRV

This commit, applied to mapnik, insures clusters are respected and using same rotation angle:
<https://github.com/mapnik/mapnik/commit/f10d5b107f5fd62a2592cc1b0315fb9fcca38990> -- this might be useful in figuring out which functions are needed.

#3 - 2013-07-12 10:50 AM - Paolo Cavallini

- Category set to Labelling

#4 - 2014-06-21 12:49 AM - Mathieu Pellerin - nIRV

Good news, everyone!

It seems there's actually no need to talk to harfbuzz directly, a QT function (QTextLayout::isValidCursorPosition) will actually validate whether the cursor in a text string is valid or not (i.e., "In a Unicode context some positions in the text are not valid cursor positions, because the position is inside a Unicode surrogate or a grapheme cluster.").

Currently, in qgspallabeling.cpp, the curved text placement simply breaks a string into its individual chars (line 147: for (int i = 0; i < mText.count(); i++)) which breaks Indic-based scripts (and most probably other languages) which rely on clusters of characters that can't be dissociated.

The code would have to be reworked to call isVlaidCursorPosition and accumulate clusters of chars to be compatible with non-Latin strings.

#5 - 2014-06-21 12:50 AM - Mathieu Pellerin - nIRV

- Assignee set to Larry Shaffer
- Target version set to Future Release - Lower Priority

#6 - 2015-05-30 04:41 AM - Nyal Dawson

- Status changed from Open to Closed

Fixed in changeset commit:"2dc5d95f00770c602497f633612a6dabb8be4962".

Files			
label-correct-paralle_placement.jpg	45 KB	2012-12-16	Mathieu Pellerin - nIRV
label-broken-curved_placement.jpg	49.5 KB	2012-12-16	Mathieu Pellerin - nIRV