

## Ruby trunk - Feature #15182

### Update extended grapheme cluster implementation for Unicode 11

10/01/2018 02:23 AM - duerst (Martin Dürst)

<b>Status:</b>	Closed
<b>Priority:</b>	Normal
<b>Assignee:</b>	duerst (Martin Dürst)
<b>Target version:</b>	2.6
<b>Description</b>	
Reported by naruse (Yui NARUSE) at <a href="https://bugs.ruby-lang.org/issues/14802#change-74213">https://bugs.ruby-lang.org/issues/14802#change-74213</a> :	
The definition of extended grapheme cluster is changed in Unicode 11 (Unicode® Standard Annex #29 UNICODE TEXT SEGMENTATION revision 33: <a href="https://www.unicode.org/reports/tr29/tr29-33.html">https://www.unicode.org/reports/tr29/tr29-33.html</a> ) This affects Regexp <code>\X/</code> which is hardcoded in <code>node_extended_grapheme_cluster()</code> in <code>regparse.c</code> .	
<pre>( CRLF   Prepend* ( RI-sequence   Hangul-Syllable   !Control ) ( Grapheme_Extend   SpacingMark ) *   . )  crlf   Control   precore* core postcore*</pre>	
<b>Related issues:</b>	
Blocks Ruby trunk - Feature #14802: Update Unicode data to Unicode Version 11...	<b>Closed</b>
Blocked by Ruby trunk - Feature #15341: Provide emoji version as RbConfig::CO...	<b>Closed</b>
Blocked by Ruby trunk - Bug #15343: String#each_grapheme_cluster wrongly spli...	<b>Closed</b>

#### History

##### #1 - 10/01/2018 02:24 AM - duerst (Martin Dürst)

- Blocks Feature #14802: Update Unicode data to Unicode Version 11.0.0 added

##### #2 - 11/26/2018 04:56 AM - duerst (Martin Dürst)

- Blocked by Feature #15341: Provide emoji version as RbConfig::CONFIG['UNICODE\_EMOJI\_VERSION'] added

##### #3 - 11/26/2018 09:03 AM - duerst (Martin Dürst)

- Blocked by Bug #15343: String#each\_grapheme\_cluster wrongly splits some emoji (genie, zombie, wrestling) added

##### #4 - 12/07/2018 09:57 AM - duerst (Martin Dürst)

- Assignee set to duerst (Martin Dürst)

- Status changed from Open to Closed

Implemented though a long series of patches, centered on `regparse.c`.

Related patches start at [r65085](#) and end at [r66269](#). The main patch is [r66213](#).

New tests are at `test/ruby/enc/test_grapheme_breaks.rb` and `test/ruby/enc/test_emoji_breaks.c`.  
`enc/unicode.c` is also modified.