

## Ruby master - Bug #10416

### Create mechanism for updating of Unicode data files downstreams when we want

10/22/2014 11:27 AM - duerst (Martin Dürst)

<b>Status:</b> Open	
<b>Priority:</b> Normal	
<b>Assignee:</b> nobu (Nobuyoshi Nakada)	
<b>Target version:</b>	
<b>ruby -v:</b> ruby 2.2.0dev (2014-10-22 trunk 48092) [x86_64-cygwin]	<b>Backport:</b> 2.0.0: UNKNOWN, 2.1: UNKNOWN
<b>Description</b> <p>The current mechanism for updating Unicode data files will create the following problem: Downstream compilers/packagegers will download Unicode data files ONE time (they may already have done so).</p> <p>However, if they don't activate ALWAYS_UPDATE_UNICODE = yes, these files will never get updated, and they will stay on Unicode version 7.0 even if in five years Unicode is e.g. on version 12.0.</p> <p>On the other hand, if they activate ALWAYS_UPDATE_UNICODE = yes (and assuming issue <a href="#">#10415</a> gets fixed), they constantly update to the latest version of Unicode. That's good for those who actually want this, but now what our current policy is. What's missing is that we (Ruby core) can make sure downstream checkouts update to a new Unicode version when we want then to do so (as we e.g. can do for other parts that are based on Unicode data, see e.g. <a href="https://bugs.ruby-lang.org/issues/9092">https://bugs.ruby-lang.org/issues/9092</a>), without sending an email to everybody and hoping they read and follow it.</p> <p>[Currently, the only solution I know will work is the one pointed out by Yui Naruse in <a href="https://bugs.ruby-lang.org/issues/10084#note-17">https://bugs.ruby-lang.org/issues/10084#note-17</a>, but I'm okay with any other solution.]</p>	
<b>Related issues:</b>	
Related to Ruby master - Bug #10458: After r48196, make cannot complete becau...	<b>Closed</b>

#### History

##### #1 - 10/22/2014 05:47 PM - nobu (Nobuyoshi Nakada)

It affects only developers who build from the repository.  
Released packages should have the latest (and fixed) version at the release time.

##### #2 - 10/27/2014 01:05 PM - naruse (Yui NARUSE)

For years, file structures of Unicode Data was changed some times.  
Therefore there's no guarantee that Unicode 12 can work with the current script.

##### #3 - 10/28/2014 03:30 AM - duerst (Martin Dürst)

Yui NARUSE wrote:

For years, file structures of Unicode Data was changed some times.  
Therefore there's no guarantee that Unicode 12 can work with the current script.

I agree (but see last paragraph of this comment). But that's not what this issue is about.

What I'm talking about is that next year, at some point in time, we decide that ruby trunk is upgraded to Unicode 8.0 (and so on probably every year). This was the case this year for Unicode 7.0, see issue [#9092](#).

We do this after checking that the new Unicode data files work with the current script (first the beta files and then the final releases), and if they don't work, then we upgrade the script. Then we commit, and everybody on trunk gets the changes when they update. But currently, this is not the case for the Unicode data files, and people on trunk will have to use a special effort to upgrade.

Besides committing lib/unicode\_normalize/tables.rb (nobu reverted it but didn't give any reason why), there's another way to achieve this goal:

Note in a file the versions or timesteps of the 'official' version of the Ruby trunk Unicode data files. This could be part of a .mk file, or a new file. Of the three files we currently download, two have a header (first two lines) like this:

## NormalizationTest-7.0.0.txt

**Date: 2013-11-27, 09:54:41 GMT [MD]**

So we could note the version and/or date we want people on trunk to use, and check against it. But one file, UnicodeData.txt, doesn't contain the information in the file, so we have to rely on the date of the Last-Modified http header (which we already use to avoid repeated downloads of the same file).

The reason why UnicodeData.txt doesn't contain is these header lines is that this is a very old file and the Unicode Consortium is actually quite careful to not make any changes that could affect the users of a file. If data of a different type is needed, then it is provided in a separate file.

**#4 - 10/30/2014 04:30 AM - duerst (Martin Dürst)**

I committed r48194, switching the download location to <http://www.unicode.org/Public/7.0.0/ucd/> (i.e. Unicode Version 7.0.0), as discussed at the meeting yesterday. This does not yet address this bug, because when we change this to <http://www.unicode.org/Public/8.0.0/ucd/> next year, the new files won't automatically be downloaded.

**#5 - 10/30/2014 08:31 AM - duerst (Martin Dürst)**

- *Related to Bug #10458: After r48196, make cannot complete because of Unicode file download problem added*

**#6 - 01/05/2018 09:01 PM - naruse (Yui NARUSE)**

- *Target version deleted (2.2.0)*