

## Ruby master - Bug #11706

### Clean up files etc/unicode/name2ctype.{h,blt,kwd,src}

11/18/2015 05:04 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>	<b>Backport:</b> 2.0.0: UNKNOWN, 2.1: UNKNOWN, 2.2: UNKNOWN
<b>Description</b>	
<p>The files name2ctype.{h,blt,kwd,src} in etc/unicode are intermediate products that are not needed in the repository, and haven't been committed consistently. I propose to remove them.</p> <p>[I'm not sure this is a bug or a feature, but it doesn't provide any new functionality, so feature doesn't seem right.]</p> <p>[I've assigned this to Nobu for feedback; I can execute it once we agree on a way forward.]</p> <p>On 2015/11/17 15:39, Nobuyoshi Nakada wrote:</p> <p style="padding-left: 40px;">Please update name2ctype.{h,blt,kwd,src} files too.</p> <p>Thanks for the reminder. I had a look at these files. Maybe before further commits, we can try to simplify things a bit, and/or to ignore irrelevant stuff.</p> <p>Sorry this message is long. Looking at the three files you mentioned, I noticed the following:</p> <p>enc/unicode/name2ctype.h.kwd was produced on the Onigmo side, when I worked on the update (see also <a href="https://github.com/k-takata/Onigmo/pull/58">https://github.com/k-takata/Onigmo/pull/58</a>), too. However, it is not part of the Onigmo distribution. It was last committed by Yui Naruse at r36070, on 2012/06/14. This is way before the update to Unicode 7.0.0 with r46831.</p> <p>On 2011/11/20, K. Takata introduced <a href="https://github.com/k-takata/Onigmo/blob/master/tool/convert-name2ctype.sh">https://github.com/k-takata/Onigmo/blob/master/tool/convert-name2ctype.sh</a>, which is used as: <code>convert-name2ctype.sh name2ctype.kwd &gt; name2ctype.h</code> to directly convert from name2ctype.kwd to name2ctype.h (although it produces a few numbered intermediary files which are removed in the last step).</p> <p>enc/unicode/name2ctype.h.blk was last committed by yourself in r49292 on 2015/01/17. Your log message mentions r46831, but it is unclear why you updated .h.blk and not .kwd and .src. The last commit before this was r36070, same as for name2ctype.h.kwd.</p> <p>enc/unicode/name2ctype.src also was last committed in r36070.</p> <p>Looking at Makefile.in, it contains instructions to create enc/unicode/name2ctype.h from enc/unicode/name2ctype.kwd at <a href="http://svn.ruby-lang.org/cgi-bin/viewvc.cgi/trunk/Makefile.in?view=markup#l340">http://svn.ruby-lang.org/cgi-bin/viewvc.cgi/trunk/Makefile.in?view=markup#l340</a>. There, .h.blk and .src are mentioned, but my knowledge of shell syntax isn't good enough to understand what's exactly supposed to go on.</p> <p>My conclusions so far would be:</p> <ul style="list-style-type: none"><li>• name2ctype.{h,blt,kwd,src} are all intermediary files that are not actually used directly for building Ruby.</li><li>• In the last few years, these three files have been committed only rarely and accidentally, not in any visible sync with actual bug fixes or feature additions.</li><li>• Onigmo no longer uses name2ctype.h.blk and .src, and does not commit .kwd.</li><li>• The build process on the Onigmo side, although I did it manually, was well documented and painless; on the Ruby side, it may be possible to build enc/unicode/name2ctype.h (the file that's finally used for compilation), but I haven't found how to do so.</li><li>• For a process that needs to be done about once a year, this amount of manual work seems perfectly fine (at least for me, and I volunteer to do it again next year).</li><li>• Therefore, I suggest that we don't care about committing name2ctype.{h,blt,kwd,src}. If you want me to commit enc/unicode/name2ctype.h.kwd, I can do it (because I have the new version). Indeed, it might be better to remove these three files; they only make checkouts heavier.</li><li>• If we want to simplify the production process, my preference would be to update Makefile.in based on convert-name2ctype.sh, or to directly integrate convert-name2ctype.sh into tool/enc-unicode.rb (why would one want to use sed and friends if we already use ruby?)</li></ul>	

**Related issues:**

Related to Ruby master - Feature #11563: Update Onigmo regular expression eng...

**Closed****History****#1 - 11/19/2015 09:10 AM - duerst (Martin Dürst)**

- Related to Feature #11563: Update Onigmo regular expression engine to Unicode Version 8.0.0 added

**#2 - 01/03/2016 05:49 PM - chriseaton (Chris Seaton)**

I've been dealing with an issue related to this. When Ruby updated to MRI 7.0 the name2ctype.h was updated but not the name2ctype.src, so they're now inconsistent (look at CR\_Blank for example).

I found this problem when I tried to update JCodings (part of JRuby) which generated its tables from these files. It uses the name2ctype.src, so got the wrong values.

I'll update JCodings to read from name2ctype.h instead.

You've listed name2ctype.h as an intermediate that should be deleted. I'm not sure that's right - it's actually the original source now isn't it? It's the only file in <https://github.com/k-takata/Onigmo/tree/master/enc/unicode>. I don't think that one can be deleted.

<https://github.com/jruby/jcodings/issues/13>

**#3 - 01/05/2016 06:52 AM - duerst (Martin Dürst)**

Chris Seaton wrote:

I've been dealing with an issue related to this. When Ruby updated to MRI 7.0

Do you mean Unicode 7.0?

the name2ctype.h was updated but not the name2ctype.src, so they're now inconsistent (look at CR\_Blank for example).

What do you mean by "now"? What's your current revision/Ruby version? As for inconsistencies, I indeed mentioned that.

I found this problem when I tried to update JCodings (part of JRuby)

Can you tell me where in the JRuby source tree these files are?

which generated its tables from these files. It uses the name2ctype.src, so got the wrong values.

I'll update JCodings to read from name2ctype.h instead.

You've listed name2ctype.h as an intermediate that should be deleted. I'm not sure that's right - it's actually the original source now isn't it?

But I haven't listed it as an intermediary; I only listed name2ctype.h.blk, which isn't the same file.

It's the only file in <https://github.com/k-takata/Onigmo/tree/master/enc/unicode>. I don't think that one can be deleted.

I didn't propose to delete it, but it could be deleted because it's an intermediate file in the sense that the original source of the data is the Unicode database itself.

<https://github.com/jruby/jcodings/issues/13>

I'll add a pointer to here to that issue.

**#4 - 01/05/2016 12:26 PM - chriseaton (Chris Seaton)**

Yes sorry I mean Unicode 7.0.

The JRuby code is at <https://github.com/jruby/jcodings/tree/master/scripts>.

Ah sorry I misread name2ctype.{h.blk,kwd,src} as name2ctype.{h,blk,kwd,src}, so I see you aren't proposing removing the .h.