Ruby master - Feature #4066
Encoding GBK needs update
11/18/2010 03:54 AM - oCameLo (oCameLo oTnTh)

Status: Rejected
Priority: Normal
Assignee: naruse (Yui NARUSE)
Target version: 1.9.2

Description

=begin
When GBK was released in 1995, it included 95 characters were not included in Unicode 1.1. Until now (Windows 7), these characters were still assigned Unicode PUA code points in CP936.

GBK isn't an official standard, so I think it won't be updated anymore. But GB18030 is official, and the subset consisting of one-byte and two-byte characters is sometimes also referred to as GBK. In GB18030-2005, 81 characters were assigned to PUA, are now defined in Unicode.


Actually, the remaining 14 characters are now defined in Unicode, too. Please take a look at gbk_fe05.gif, light grey and light yellow ones.

These 95 characters are all defined in Unicode now (see gbk_mod.htm), so I think we should add these characters to gbk-tbl.rb. It won't cause any compatibility issue, at least in Ruby side.
=end

History

#1 - 11/18/2010 12:29 PM - shyouhei (Shyouhei Urabe)
- Category changed from test to M17N
- Status changed from Open to Assigned
- Assignee set to naruse (Yui NARUSE)

=begin
=end

#2 - 11/18/2010 01:03 PM - naruse (Yui NARUSE)

=begin
I understand your point and it seems reasonable. Anyway, is there any other implementation whose conversion table includes such characters?
=end

#3 - 11/19/2010 08:12 PM - oCameLo (oCameLo oTnTh)

=begin
I just can find out only one mailing list thread about this problem here: [http://sources.redhat.com/ml/libc-alpha/2000-09/msg00394.html](http://sources.redhat.com/ml/libc-alpha/2000-09/msg00394.html)

For compatibility, we should accept this patch. But from the angle of standard, let it go.

Both ways are acceptable.
=end

#4 - 11/20/2010 03:16 PM - naruse (Yui NARUSE)

=begin
(2010/11/19 20:12), oCameLo oTnTh wrote:

I just can find out only one mailing list thread about this problem here: [http://sources.redhat.com/ml/libc-alpha/2000-09/msg00394.html](http://sources.redhat.com/ml/libc-alpha/2000-09/msg00394.html)

For compatibility, we should accept this patch. But from the angle of standard, let it go.
Ruby's mapping table should follow de facto or de jure standards. In current situation, I should say the expectation for compatibility is wrong. (you may know, Euro sign is also incompatible).

So until other implementation like converters, editors, or web browsers supports such table, ruby won't support them.

--
NARUSE, Yui naruse@airemix.jp
=end

#5 - 01/14/2011 02:52 PM - naruse (Yui NARUSE)
- Status changed from Assigned to Rejected

=begin
=begin

Files

<table>
<thead>
<tr>
<th>File</th>
<th>Size</th>
<th>Date</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>gbk-tbl.95_chars.diff</td>
<td>4.27 KB</td>
<td>11/18/2010</td>
<td>oCameLo (oCameLo oTnTh)</td>
</tr>
<tr>
<td>gbk_fe05.gif</td>
<td>33.9 KB</td>
<td>11/18/2010</td>
<td>oCameLo (oCameLo oTnTh)</td>
</tr>
<tr>
<td>gbk_mod.htm</td>
<td>6.36 KB</td>
<td>11/18/2010</td>
<td>oCameLo (oCameLo oTnTh)</td>
</tr>
</tbody>
</table>