

Ruby - Bug #6995

Code converter not found (UTF-8 to EUC-TW)

09/09/2012 01:01 AM - blueowl (blue owl)

Status:	Rejected	Backport:
Priority:	Normal	
Assignee:	naruse (Yui NARUSE)	
Target version:		
ruby -v:	ruby 1.9.3p0 (2011-10-30 revision 33570) [i686-linux]" and "ruby 2.0.0dev (2012-09-07 trunk 36920) [i686-linux]	

Description

Hello, recently I was doing some conversion from Unicode into Chinese encodings, and I came across what may be a bug in Ruby. Attempting to transcode a traditional Chinese character from UTF-8 to EUC-TW results in a "code converter not found" error. This character exists in Unicode (U+8B6F), and if it were missing in EUC-TW, then I would expect "Encoding::UndefinedConversionError" rather than "Encoding::ConverterNotFoundError". Relevant code and Ruby versions are shown below for reproducing this issue.

```
$ ruby -v -e "''.encode("EUC-TW")'
ruby 1.9.3p0 (2011-10-30 revision 33570) [i686-linux]
/tmp/test.rb:3:in encode': code converter not found (UTF-8 to EUC-TW) (Encoding::ConverterNotFoundError) from /tmp/test.rb:3:in '
```

```
$ ~/ruby_vm/nightly/bin/ruby -v -e "''.encode("EUC-TW")'
ruby 2.0.0dev (2012-09-07 trunk 36920) [i686-linux]
/tmp/test.rb:3:in encode': code converter not found (UTF-8 to EUC-TW) (Encoding::ConverterNotFoundError) from /tmp/test.rb:3:in '
```

History

- #1 - 09/09/2012 10:33 AM - nobu (Nobuyoshi Nakada)**
- Status changed from Open to Assigned
 - Assignee set to naruse (Yui NARUSE)
 - ruby -v changed from "ruby 1.9.3p0 (2011-10-30 revision 33570) [i686-linux]" and "ruby 2.0.0dev (2012-09-07 trunk 36920) [i686-linux]" to ruby 1.9.3p0 (2011-10-30 revision 33570) [i686-linux]" and "ruby 2.0.0dev (2012-09-07 trunk 36920) [i686-linux]"
- ```
=begin
Some transcoders are missing.

$ ruby -e 'u = Encoding::UTF_8' -e 'puts Encoding.list.find_all{|e|e != u and !Encoding::Converter.search_convpath(e, u) rescue true}'
Emacs-Mule
EUC-TW
IBM864
Windows-1258
GB1988
macCentEuro
macThai
ISO-2022-JP-2
MacJapanese
UTF-7
=end
```
- #2 - 11/03/2012 09:25 AM - naruse (Yui NARUSE)**
- Status changed from Assigned to Feedback
- We don't have UTF-8:EUC-TW converter yet.  
If you want it, make a feature request ticket.
- But as far as I know EUC-TW is not widely used, it is only used in goverment.  
Instead of EUC-TW, many people use Big5.  
Do you really need EUC-TW?  
(for workaround you can use Iconv to convert EUC-TW in 1.9.3)
- #3 - 01/26/2013 07:00 AM - drbrain (Eric Hodel)**
- Status changed from Feedback to Rejected

Marking rejected due to lack of feedback from the submitter.

Files

|         |          |            |                    |
|---------|----------|------------|--------------------|
| test.rb | 61 Bytes | 09/09/2012 | blueowl (blue owl) |
|---------|----------|------------|--------------------|