JSON.dump can generate invalid UTF-8 sequence

06/10/2012 02:48 AM - shyouhei (Shyouhei Urabe)

| Status:    | Closed                          |
| Priority:  | Normal                          |
| Assignee:  | naruse (Yui NARUSE)             |
| Target version: | ruby 2.0.0dev (2012-06-09) [x86_64-linux] | Backport: | 2.3: UNKNOWN, 2.4: UNKNOWN, 2.5: UNKNOWN |

Description

Look, in the following code JSON.dump outputs a sequence invalid as UTF-8.

```ruby
# encoding: utf-8
require 'json'
IO.popen('hexdump -C', 'w') do |fp|
  JSON.dump(['\xea'], fp)
end
```

RFC4627 says that to encode JSON as a Unicode is a "SHALL". So this is an RFC violation.

History

#1 - 06/10/2012 07:19 AM - nobu (Nobuyoshi Nakada)
- File bug-6566.diff added

=begin
A bit simpler, it seems wrong that
JSON.generate(['\xea']).valid_encoding?
returns (({false})).
I think this would be a bug in json generator, but what should happen
in this case? Seems (((convert_UTF8_to_JSON_ASCII()))) wants to reject
invalid sequence.
=end

#2 - 06/11/2012 11:11 AM - naruse (Yui NARUSE)
json is not only for 1.9, so nobu's patch is not acceptable.
I made https://github.com/flori/json/pull/139

#3 - 02/17/2013 07:53 PM - naruse (Yui NARUSE)
- Target version changed from 2.0.0 to 2.6

#4 - 12/25/2017 06:15 PM - naruse (Yui NARUSE)
- Target version deleted (2.6)

#5 - 08/03/2019 12:01 AM - jeremyevans0 (Jeremy Evans)
- Status changed from Assigned to Closed

I think this was fixed upstream last year: https://github.com/flori/json/commit/91a24ecac3d722bc3e545fdd0657ed225e91c2df

Files

<table>
<thead>
<tr>
<th>File</th>
<th>Size</th>
<th>Date</th>
<th>User</th>
</tr>
</thead>
<tbody>
<tr>
<td>bug-6566.diff</td>
<td>1.62 KB</td>
<td>06/10/2012</td>
<td>nobu (Nobuyoshi Nakada)</td>
</tr>
</tbody>
</table>