Ruby master - Feature #6910
Loading syck's broken yaml with psych
08/23/2012 11:04 AM - naruse (Yui NARUSE)

Status: Rejected
Priority: Normal
Assignee: tenderlovemaking (Aaron Patterson)
Target version: 2.0.0

Description
You know, syck outputs wrong yaml.
For example, syck works as following:

```
ruby-1.9.2 > "\u3042",Time.at(0).to_s).to_yaml
=> "---
- "\xE3\x81\x82"
- 1970-01-01 09:00:00 +09:00"
```

It should be

```
ruby-1.9.3 > "\u3042",Time.at(0).to_s).to_yaml
=> "---
- あ
- '1970-01-01 09:00:00 +0900'
```

syck's dump of Unicode string is interpreted as "\u00E3\u0081\u0082".
syck's dump of Time like string is interpreted as Time.
It is hard to migrate old data to new and correct data, so it is useful if psych has a such compatibility option.

History
#1 - 08/23/2012 01:23 PM - Anonymous
- File noname added

On Thu, Aug 23, 2012 at 11:04:22AM +0900, naruse (Yui NARUSE) wrote:

Issue #6910 has been reported by naruse (Yui NARUSE).

Feature #6910: Loading syck's broken yaml with psych
https://bugs.ruby-lang.org/issues/6910

Author: naruse (Yui NARUSE)
Status: Assigned
Priority: Normal
Assignee: tenderlovemaking (Aaron Patterson)
Category: ext
Target version: 2.0.0

You know, syck outputs wrong yaml.
For example, syck works as following:

```
ruby-1.9.2 > "\u3042",Time.at(0).to_s).to_yaml
=> "---
- "\xE3\x81\x82"
- 1970-01-01 09:00:00 +09:00"
```

It should be

```
ruby-1.9.3 > "\u3042",Time.at(0).to_s).to_yaml
=> "---
- あ
- '1970-01-01 09:00:00 +0900'
```

syck's dump of Unicode string is interpreted as "\u00E3\u0081\u0082".
syck's dump of Time like string is interpreted as Time.
It is hard to migrate old data to new and correct data, so it is useful if psych has a such compatibility option.

It's possible to have both syck and psych loaded in 1.9.3 (also 1.9.2 I think):

```ruby
require 'syck'
require 'yaml'
require 'psych'
```
def convert legacy
    Psych.dump YAML.load legacy
end

legacy = YAML.dump ["\u3042",Time.at(0).to_s]

puts convert legacy

I'm pushing syck to a gem so people can do this even further in the future.

--
Aaron Patterson
http://tenderlovemaking.com/

#2 - 08/23/2012 03:49 PM - naruse (Yui NARUSE)
I want to migrate gradually.
The way needs explicit big bang conversion.

#3 - 08/24/2012 01:23 AM - Anonymous
- File noname added

On Thu, Aug 23, 2012 at 03:49:03PM +0900, naruse (Yui NARUSE) wrote:

    Issue #6910 has been updated by naruse (Yui NARUSE).

        I want to migrate gradually.
The way needs explicit big bang conversion.

Then I guess you need a way to differentiate converted yaml from
non-converted yaml. Most people I've helped with this use something out
of band like a database column to mark the conversion. However, it is
possible to use a YAML version. Syck does not support YAML 1.1, so we
can't use the YAML as version 1.1. Unfortunately, syck will not raise an
exception on the version identifier, so we have to test for it ourselves. Here is an example:

    require 'syck'
    require 'yaml'
    require 'psych'
    require 'minitest/autorun'

class Loader < MiniTest::Unit::TestCase
  def converted? text
    text =~ /^\A%YAML 1.1/  
  end

  def load_yaml text
    if converted? text
      Psych.load text
    else
      YAML.load text
    end
  end

  def dump_yaml object
    Psych.dump object, {,:version => [1,1]}  
  end

  def test_convert
    obj = ["\u3042",Time.at(0).to_s]
    legacy_yaml = YAML.dump obj
    obj2 = load_yaml legacy_yaml

    # we can load legacy yaml
    assert_equal obj, obj2

    converted_yaml = dump_yaml obj

    # make sure the yaml is tagged when dumping
    assert converted? converted_yaml
# make sure object loaded from converted yaml is same
assert_equal obj, load_yaml(converted_yaml)
end
end

--
Aaron Patterson
http://tenderlovemaking.com/

#4 - 10/27/2012 02:12 PM - tenderlovemaking (Aaron Patterson)
- Status changed from Assigned to Rejected

I've pushed a gem for people that want to upgrade to Ruby 2.0.0, but still have Syck's YAML:

https://rubygems.org/gems/syck

Since libyaml does the actual YAML parsing, we would either have to include syck with psych, or ship libyaml with ruby and modify it's parser. I think the gem option is the best route. People can still use it to parse old YAML.

Files

<table>
<thead>
<tr>
<th>File</th>
<th>Size</th>
<th>Date</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>noname</td>
<td>500 Bytes</td>
<td>08/23/2012</td>
<td>Anonymous</td>
</tr>
<tr>
<td>noname</td>
<td>500 Bytes</td>
<td>08/24/2012</td>
<td>Anonymous</td>
</tr>
</tbody>
</table>