

Backport191 - Backport #4028

substring selection and utf8 encoding problem

11/06/2010 02:30 AM - barcala (Fco. Mario Barcala Rodríguez)

Status:	Assigned	
Priority:	Normal	
Assignee:	yugui (Yuki Sonoda)	
Description		
<pre>=begin Substring selection does not work with some utf8 encoded strings. Below is an example. The first substring is well extracted but the second not (extrange characters appear at the end of the substring). It seems it occurs when the string includes letters with umlauts, accents, etc. \$ irb ruby-1.9.1-p378 > word = "Ábaco" => "Ábaco" ruby-1.9.1-p378 > substr = word[word.length-1,word.length] => "o" ruby-1.9.1-p378 > word = "Coordinador de ONG's do País Valenciano" => "Coordinador de ONG's do País Valenciano" ruby-1.9.1-p378 > substr = word[word.length-1,word.length] => "o\x00\x00\x01\x00\x01\x00\x00\x00" =end</pre>		
Related issues:		
Is duplicate of Ruby trunk - Bug #2379: String#[] returns invalid values for ...		Closed 11/18/2009

History

#1 - 11/06/2010 02:54 AM - barcala (Fco. Mario Barcala Rodríguez)

```
=begin
The same error occurs in ruby-1.9.1-p430
=end
```

#2 - 11/06/2010 02:56 AM - barcala (Fco. Mario Barcala Rodríguez)

```
=begin
It seems to be solved in ruby-1.9.2-p0 version. I can't reproduce the error in 1.9.2-p0
=end
```

#3 - 11/12/2010 01:00 AM - barcala (Fco. Mario Barcala Rodríguez)

```
=begin
Showed example uses substring selection in a wrong way. Example should be:
```

```
ruby-1.9.1-p378 > word = "Ábaco"
=> "Ábaco"
ruby-1.9.1-p378 > substr = word[word.length-1,1]
=> "o"
ruby-1.9.1-p378 > word = "Coordinador de ONG's do País Valenciano"
=> "Coordinador de ONG's do País Valenciano"
ruby-1.9.1-p378 > substr = word[word.length-1,1]
=> "o"
```

```
This new example works fine, so the problem arises only when the second value of substring selection exceeds the limits of the string.
=end
```

#4 - 11/12/2010 01:30 AM - naruse (Yui NARUSE)

- Status changed from Open to Assigned

- Assignee set to yugui (Yuki Sonoda)

- Priority changed from 5 to Normal

```
=begin
```

```
Confirmed:
ruby 1.9.1p430 (2010-08-16 revision 28997) [x86_64-freebsd8.1]
ruby-1.9.1-p378 > word = "Coordinador de ONG's do País Valenciano"
=> "Coordinador de ONG's do País Valenciano"
ruby-1.9.1-p378 > substr = word[word.length-1,word.length]
=> "\x00\x00\x01\x00\x01\x00\x00\x00"
=end
```