

Ruby master - Bug #9539

REXML XPath UTF8 encoding problem

02/20/2014 09:10 AM - mbarcala (Mario Barcala)

Status:	Closed	
Priority:	Normal	
Assignee:	kou (Kouhei Sutou)	
Target version:		
ruby -v:	ruby 2.1.0	Backport: 1.9.3: UNKNOWN, 2.0.0: UNKNOWN, 2.1: UNKNOWN
Description I found some problems in REXML when processing XPath expressions with Unicode not ascii characters. I attached a sample script and a sample document. If you see the script output, you will see two different problems: 1) text() XPath function does not work properly when there is an accent or tilde character. 2) two different XPath paths, one with an accent and the other without it, are considered the same. Thank you, Mario Barcala		

Associated revisions

Revision ba5ed845 - 02/23/2014 09:01 AM - kou (Kouhei Sutou)

- lib/rexml/xmltokens.rb: Add missing non ASCII valid characters to element name characters. Now, REXML name tokens exactly match "[5] Name" in the XML spec and "[4] NCName" in the Namespaces in XML spec. See comment about the details. [Bug #9539] [ruby-core:60901] Reported by Mario Barcala. Thanks!!!
- test/rexml/xpath/test_node.rb: Add tests for the above case.

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@45153 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 45153 - 02/23/2014 09:01 AM - kou (Kouhei Sutou)

- lib/rexml/xmltokens.rb: Add missing non ASCII valid characters to element name characters. Now, REXML name tokens exactly match "[5] Name" in the XML spec and "[4] NCName" in the Namespaces in XML spec. See comment about the details. [Bug #9539] [ruby-core:60901] Reported by Mario Barcala. Thanks!!!
- test/rexml/xpath/test_node.rb: Add tests for the above case.

Revision 45153 - 02/23/2014 09:01 AM - kou (Kouhei Sutou)

- lib/rexml/xmltokens.rb: Add missing non ASCII valid characters to element name characters. Now, REXML name tokens exactly match "[5] Name" in the XML spec and "[4] NCName" in the Namespaces in XML spec. See comment about the details. [Bug #9539] [ruby-core:60901]

Reported by Mario Barcala. Thanks!!!

- test/rexml/xpath/test_node.rb: Add tests for the above case.

Revision 45153 - 02/23/2014 09:01 AM - kou (Kouhei Sutou)

- lib/rexml/xmltokens.rb: Add missing non ASCII valid characters to element name characters. Now, REXML name tokens exactly match "[5] Name" in the XML spec and "[4] NCName" in the Namespaces in XML spec. See comment about the details. [Bug #9539] [ruby-core:60901]
Reported by Mario Barcala. Thanks!!!
- test/rexml/xpath/test_node.rb: Add tests for the above case.

Revision 45153 - 02/23/2014 09:01 AM - kou (Kouhei Sutou)

- lib/rexml/xmltokens.rb: Add missing non ASCII valid characters to element name characters. Now, REXML name tokens exactly match "[5] Name" in the XML spec and "[4] NCName" in the Namespaces in XML spec. See comment about the details. [Bug #9539] [ruby-core:60901]
Reported by Mario Barcala. Thanks!!!
- test/rexml/xpath/test_node.rb: Add tests for the above case.

Revision 45153 - 02/23/2014 09:01 AM - kou (Kouhei Sutou)

- lib/rexml/xmltokens.rb: Add missing non ASCII valid characters to element name characters. Now, REXML name tokens exactly match "[5] Name" in the XML spec and "[4] NCName" in the Namespaces in XML spec. See comment about the details. [Bug #9539] [ruby-core:60901]
Reported by Mario Barcala. Thanks!!!
- test/rexml/xpath/test_node.rb: Add tests for the above case.

Revision 45153 - 02/23/2014 09:01 AM - kou (Kouhei Sutou)

- lib/rexml/xmltokens.rb: Add missing non ASCII valid characters to element name characters. Now, REXML name tokens exactly match "[5] Name" in the XML spec and "[4] NCName" in the Namespaces in XML spec. See comment about the details. [Bug #9539] [ruby-core:60901]
Reported by Mario Barcala. Thanks!!!
- test/rexml/xpath/test_node.rb: Add tests for the above case.

History

#1 - 02/23/2014 09:01 AM - kou (Kouhei Sutou)

- % Done changed from 0 to 100

- Status changed from Open to Closed

Applied in changeset r45153.

- lib/rexml/xmltokens.rb: Add missing non ASCII valid characters to element name characters. Now, REXML name tokens exactly match "[5] Name" in the XML spec and "[4] NCName" in the Namespaces in XML spec. See comment about the details. [Bug #9539] [ruby-core:60901] Reported by Mario Barcala. Thanks!!!
- test/rexml/xpath/test_node.rb: Add tests for the above case.

#2 - 02/23/2014 09:03 AM - kou (Kouhei Sutou)

- Assignee set to kou (Kouhei Sutou)

Thanks for your report!
I've fixed it in trunk.

It was very helpful that you attach a sample script and sample XML to reproduce the problem. :-)

Files

sample.rb	366 Bytes	02/20/2014	mbarcala (Mario Barcala)
sample.xml	224 Bytes	02/20/2014	mbarcala (Mario Barcala)