

Ruby - Bug #9806

URI#encode doesn't encode characters '[' and ']'. They should be encoded as %5B and %5D respectively.

05/06/2014 01:29 AM - charlez (Charles Leu)

Status:	Closed	
Priority:	Normal	
Assignee:		
Target version:		
ruby -v:	2.2.0 and prior versions as well	Backport: 2.0.0: UNKNOWN, 2.1: UNKNOWN
Description The subject says it all. IRB session demonstrating the problem: charlez\$ irb head :001 > RUBY_VERSION => "2.2.0" head :002 > require 'uri' => true head :003 > my_str = '[' futsal club]' => "[" futsal club]" head :004 > URI.encode(my_str) => "[%20futsal%20club%20]" head :005 > Note: Testing using JavaScript function encodeURIComponent('[' futsal club]') produces "%5B%20futsal%20club%20%5D" which is the correct result.		
Related issues: Related to Ruby - Bug #12235: URI.encode issue with square brackets		Closed

History

#1 - 05/06/2014 09:13 PM - charlez (Charles Leu)

Notes:

- Per RFC 2396 section 2.4.3 "Data corresponding to excluded characters **must** be escaped in order to be properly represented within a URI."
- Per RFC 2396 section 2.2 reserved characters are ";", "/", "?", ":", "@", "&", "=", "+", "\$", ",",
- Per URI::REGEXP::PATTERN reserved characters are ";", "/", "?", ":", "@", "&", "=", "+", "\$", ",", "[", "]"
- Thus there appears to be an inconsistency between RFC 2396 2.2 and URI::REGEXP::PATTERN
- After changing URI::REGEXP::PATTERN[:RESERVED] to omit characters '[' and ']', URI.encode('[' futsal club]') produces "%5B%20futsal%20club%20%5D", which I believe is correct.

#2 - 05/10/2014 01:37 PM - mame (Yusuke Endoh)

I'm unfamiliar with URI spec, but I guess RFC 2732 is related.

<http://www.ietf.org/rfc/rfc2732.txt>

This document includes an update to the generic syntax for Uniform Resource Identifiers defined in RFC 2396 [URL]. It defines a syntax for IPv6 addresses and allows the use of "[" and "]" within a URI explicitly for this reserved purpose.

--
Yusuke Endoh mame@tsg.ne.jp

#3 - 05/18/2014 11:23 AM - johnnymugs (Jonathan Mukai)

It looks like URI.encode/escape was deprecated in favor of either CGI.escape or URI.encode_www_form_component per <https://github.com/ruby/ruby/commit/238b979f1789f95262a267d8df6239806f2859cc> and some discussion here: <https://www.ruby-forum.com/topic/207489>

Both options give you the output you want.

However, I'm sure there's plenty of code hanging around that uses `URI.escape`. I wonder what the policy is for updating deprecated methods like this?

Johnny

#4 - 08/05/2014 06:50 PM - charlez (Charles Leu)

Yusuke Endoh wrote:

I'm unfamiliar with URI spec, but I guess RFC 2732 is related.

<http://www.ietf.org/rfc/rfc2732.txt>

This document includes an update to the generic syntax for Uniform Resource Identifiers defined in RFC 2396 [URL]. It defines a syntax for IPv6 addresses and allows the use of "[" and "]" within a URI explicitly for this reserved purpose.

--

Yusuke Endoh mame@tsg.ne.jp

FYI: Refer to the current W3.org BNF for URI syntax http://www.w3.org/Addressing/URL/5_URI_BNF.html

Note the statement 'The "national" and "punctuation" characters do not appear in any productions and therefore may not appear in URIs.'. That statement is at odds with RFC 2732.

It appears that authors of the standards docs aren't always aware of, and/or consistent with, other standards docs. Thus it is not surprising there is confusion regarding what is or isn't a valid URI encoding.

#5 - 03/29/2017 02:01 PM - znz (Kazuhiro NISHIYAMA)

- *Related to Bug #12235: URI.encode issue with square brackets added*

#6 - 03/30/2022 11:04 PM - jeremyevans0 (Jeremy Evans)

- *Status changed from Open to Closed*

`URI.encode` was removed in Ruby 3.0.