

Re: urlencode vs rawurlencode

Source: <http://coding.derkeiler.com/Archive/PHP/comp.lang.php/2004-04/1722.html>

From: John Dunlop (*usenet+2004_at_john.dunlop.name*)

Date: 04/24/04

Date: Sat, 24 Apr 2004 04:00:12 +0100

If I sound confused, that's because I am.

John Dunlop wrote:

> Consider:

>

> 1. <<http://domain.example/?baz=foo+bar>>

> 2. <<http://domain.example/?baz=foo%20bar>>

> 3. <<http://domain.example/?baz=foo%2Bbar>>

[...]

> Reading Björn Höhrmann's explanation of reserved characters in

>

> "Re: Good/Bad – URI encoding in HTML editor",

> <http://lists.w3.org/Archives/Public/uri/2002May/0032.html>

>

> we see that numbers one and two are **not** equivalent.

Actually, I think, numbers one and two are equivalent. Hopefully I've got this straight in my head now. :-)

RFC1630, which I hadn't read before, sums up Tim BL's original intent:

| Within the query string, the plus sign is reserved as shorthand
| notation for a space. Therefore, real plus signs must be encoded.
| This method was used to make query URIs easier to pass in systems
| which did not allow spaces.

According to RFC1738, sec. 3.3, however, plus signs weren't reserved in the query component ("searchpart") of an HTTP URL. That means they had no reserved purpose, so a plus sign meant a plus sign, not a space, and they didn't need encoded.

Then came along RFC2396 and the plus sign became reserved in the query component again. Real plus signs must now be encoded. It doesn't say what the reserved purpose is for plus signs. I guess, then, plus signs are shorthand for spaces.

comp.lang.php: Re: urlencode vs rawurlencode

Previously, I was under the impression that a question mark mustn't appear in query components. It seems I was wrong. A URI may contain more than one question mark, although URI generators are discouraged from generating such URIs. The second "?" should always be treated as data by parsers. See

Roy T. Fielding, 2002-11-17, "Re: Ambiguity of Allowed/Recommended URI Syntax and Escaping",
<http://lists.w3.org/Archives/Public/uri/2002Nov/0015.html>

Refs.:

RFC1630 (informational), 1994-06, "Universal Resource Identifiers in WWW: A Unifying Syntax for the Expression of Names and Addresses of Objects on the Network as used in the World-Wide Web",
<http://www.ietf.org/rfc/rfc1630.txt>

RFC1738 (proposed standard), 1994-12, "Uniform Resource Locators (URL)",
<http://www.ietf.org/rfc/rfc1738.txt>

RFC2396 (draft standard), 1998-08, "Uniform Resource Identifiers (URI): Generic Syntax",
<http://www.ietf.org/rfc/rfc2396.txt>

--
Jock