

Bugs & Wish list

Sitemap.xml &pagenum=2 -v- &pagenum=2 | Tiki Wiki CMS Groupware :: Development
Sitemap.xml &pagenum=2 -v- &pagenum=2

Status

✖ Closed

Subject

Sitemap.xml &pagenum=2 -v- &pagenum=2

Version

15.x

18.x

Category

- Feature request
- Support request

Feature

Wiki Syntax (text area, parser, external wiki, etc)

Infrastructure

Resolution status

Invalid

Submitted by

EmmArrBee

Lastmod by

EmmArrBee, luciash d' being ☐

Rating

★ ★ ★ ★ ★ (0) ?

Description

This issue is probably external to the wiki ... but is exhibited by your current demo systems and also our Apache based system

With pagination enabled or when calling a wiki page with a user defined variable value, the full url contains an ampersand character,
for example `tiki-index.php?page=Test-Multi-Page&pagenum=2`

However ampersand can not be included in an XML file such as a SiteMap and needs to be escaped as `&`;

TW seems to ignore any parameter prefixed with `&`; which is a valid url
for example `tiki-index.php?page=Test-Multi-Page&pagenum=2`

Does a default installation of TW with Apache need a special Apache rewrite rule to handle this?
Does this perceived issue exist with other web servers?

Just spotted your current docs system has the same issue. See
`https://doc.tiki.org/tiki-index.php?page=Wiki+Argument+Variables&lalala=torololo` which works
but
`https://doc.tiki.org/tiki-index.php?page=Wiki+Argument+Variables&lalala=torololo` does not work

Would appreciate any guidance to resolve this issue.

Thx

M

Priority

25

Demonstrate Bug (Tiki 19+)

Please demonstrate your bug on show2.tiki.org

Version: trunk ▼

Demonstrate Bug (older Tiki versions)

Please demonstrate your bug on show.tikiwiki.org

Version: 18.x ▼

Ticket ID

6620

Created

Friday 13 April, 2018 09:39:58 GMT-0000

by EmmArrBee

LastModif

Sunday 15 April, 2018 09:59:25 GMT-0000

Comments



luciash d' being 13 Apr 18 10:51 GMT-0000

Can you try with CDATA tags in your sitemap XML?



```
<![CDATA[ This is some text with ampersands & other funny characters. >> ]]>
```

Or try to use %26 instead of the ampersand in the URL...



luciash d' being 13 Apr 18 11:00 GMT-0000

Actually, I am not sure what is the issue really? Does the sitemap.xml work with encoded ampersands for Google? I think it is not used by anything else than by that bot. So that bot should handle the conversion to simple & for the links, or... am I missing something?



EmmArrBee 13 Apr 18 11:58 GMT-0000

I may be getting confused here... using XML means I can't directly use an ampersand but using the html

entity & in its place results in an url which isn't understood.

If I encode the ampersand as %26 per your suggestion, it seems to result in a page not found (unless I'm using it incorrectly) see

<https://doc.tiki.org/tiki-index.php?page=Wiki+Argument+Variables%26lalala=torololo>

Or have I got the wrong end of the stick here....



luciash d' being 13 Apr 18 12:17 GMT-0000

If you can paste a snippet of your sitemap.xml here in PluginCode it could help to understand the issue. Anyway still I think these links from the XML are not being used directly as the resulting links (the final links should be using normal & form).



EmmArrBee 13 Apr 18 14:33 GMT-0000

Thanks for your responses. As requested

Snippet from SiteMap.xml



```
<?xml version="1.0" encoding="UTF-8"?> <urlset
xmlns="https://www.sitemaps.org/schemas/sitemap/0.9"
xmlns:xsi="https://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="https://www.sitemaps.org/schemas/sitemap/0.9
https://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd">
<url><loc>https://mydomain.tld/data/tiki-index.php?page=FAQ_Adding_US_State_or_UK_County&
mp;pagenum=2</loc></url>
<url><loc>https://mydomain.tld/data/tiki-index.php?page=ZZY-TRP-SiteMap-Sub&trpsubj=15
</loc></url> </urlset>
```

I can understand and accept your suggestion that the crawling app knows how to handle the contents of a SiteMap.xml file and that I shouldn't expect another app (a web browser) to handle part of a random line cut from it correctly. It just makes checking that the content of the SiteMap file is valid a bit more tricky.

However, I understood that url encoding using % plus two hex digits was perfectly valid.

So <https://doc.tiki.org/Wiki-Argument-Variables> and
<https://doc.tiki.org/Wiki%2dArgument-Variables> are equally valid.

As are https://doc.tiki.org/tiki-index.php?page=Wiki-Argument-Variables&lalala=test_me_now and
https://doc.tiki.org/tiki-index.php?page=Wiki%2dArgument-Variables&lalala=test_me_now and
https://doc.tiki.org/tiki-index.php?page=Wiki%2dArgument-Variables&lalala=%54est_me_now

BUT using your neat suggestion of %26 in place of the ampersand fails...

https://doc.tiki.org/tiki-index.php?page=Wiki-Argument-Variables%26lalala=%54est_me_now
as does using %3F in place of a question mark also fails.

https://doc.tiki.org/tiki-index.php%3fpage=Wiki%2dArgument-Variables&lalala=%54est_me_now

Again, I'm probably missing something fundamental about the way all of this is intended to work....

Thanks again,

M



luciash d' being 13 Apr 18 15:07 GMT-0000

Yeah, thank you for the snippet. As you wrote, the examples are perfectly valid but "?" (and "&" and "=" and "/" plus maybe "+") characters in URLs are special control characters so they should not be escaped in any way if they are meant to serve their purpose. What needs to be escaped are on the other hand such characters if they are meant to be part of some values, like e.g. page named "Look & Feel" - that is the case where you want the "&" URL or HTML escaped for sure.

If it bothers you that you cannot check the URLs from your sitemap.xml manually easily by copypasting them I would suggest do not escape the & characters in the XML file but replace it with the following:



```
<?xml version="1.0" encoding="UTF-8"?> <urlset
xmlns="https://www.sitemaps.org/schemas/sitemap/0.9"
xmlns:xsi="https://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="https://www.sitemaps.org/schemas/sitemap/0.9
https://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd">
<url><loc><![CDATA[https://mydomain.tld/data/tiki-index.php?page=FAQ_Adding_US_State_or_
UK_County&pagenum=2]]></loc></url>
<url><loc><![CDATA[https://mydomain.tld/data/tiki-index.php?page=ZZY-TRP-SiteMap-Sub&trps
ubj=15]]></loc></url> </urlset>
```

Using the CDATA the xml validator should not complain about usage of unescaped ampersands anymore.

Also I would recommend switching to SEFURLs. Anyway, I do not think this is actually a Tiki bug in any way. Can we close this item?



EmmArrBee 15 Apr 18 07:44 GMT-0000

Again, many thanks for your thoughts. Very happy to see this item closed.

Is it possible to get our domain name redacted in the above conversation, just so the Search Engines don't hang on to pages which may go away in the future.

For future reference, is there a better way to raise queries such as this?

Thx,

M



luciash d' being ☐ 15 Apr 18 10:00 GMT-0000

Done. Forums or user / developer mailing list is fine.

luci

Attachments

filename	created	hits	comment	version	filetype
----------	---------	------	---------	---------	----------

No attachments for this item

The original document is available at

<https://dev.tiki.org/item6620-Sitemap-xml-pagenum-2-v-amp-pagenum-2>