

TYPO3 Core - Bug #87295

Chinese Language url not working with TYPO3 9.5

2018-12-26 06:55 - bharat parmar

Status:	Closed	Start date:	2018-12-26
Priority:	Must have	Due date:	
Assignee:		% Done:	100%
Category:	Link Handling, Site Handling & Routing	Estimated time:	0.00 hour
Target version:	Candidate for patchlevel	Complexity:	
TYPO3 Version:	9	Is Regression:	
PHP Version:	7.2	Sprint Focus:	
Tags:	slug		

Description

I have site with Chinese language and TYPO3 9.5 but Chinese page title in url are replaced with some string like e5b7a5e4bd9ce4b88ee8818ce4b89a

In realurl version its /cn/[]/ and in new(TYPO3 9 slug) its /cn/e5ba94e794a8

So is Chinese supported in TYPO3 9 slug?

According to doc <https://docs.typo3.org/typo3cms/extensions/core/Changelog/9.4/Feature-84729-NewTCATypeSlug.html> unicode is supported.

Its seems

```
line 126 $slug = rawurlencode($slug); \\ \TYPO3\CMS\Core\DataHandling\SlugHelper->sanitize() encoded it.
```

Associated revisions

Revision 0299afec - 2019-03-01 12:37 - Guido Schmechel

[BUGFIX] Support non ASCII url slugs

Resolves: #87295

Releases: master, 9.5

Change-Id: Ib4fb1a8283c79a02cbc8cb52d91e2448ad9292ec

Reviewed-on: <https://review.typo3.org/c/59796>

Tested-by: Benni Mack <benni@typo3.org>

Tested-by: TYPO3com <noreply@typo3.com>

Tested-by: Jürgen Venne <venne@schaffrath-digital.de>

Tested-by: Anja Leichsenring <leichsenring@ab-softlab.de>

Reviewed-by: Benni Mack <benni@typo3.org>

Reviewed-by: Jürgen Venne <venne@schaffrath-digital.de>

Reviewed-by: Anja Leichsenring <leichsenring@ab-softlab.de>

Revision 394decab - 2019-03-01 14:25 - Guido Schmechel

[BUGFIX] Support non ASCII url slugs

Resolves: #87295

Releases: master, 9.5

Change-Id: Ib4fb1a8283c79a02cbc8cb52d91e2448ad9292ec

Reviewed-on: <https://review.typo3.org/c/59822>

Tested-by: TYPO3com <noreply@typo3.com>

Tested-by: Anja Leichsenring <leichsenring@ab-softlab.de>

Reviewed-by: Anja Leichsenring <leichsenring@ab-softlab.de>

History

#1 - 2019-01-27 11:34 - Riccardo De Contardi

- Category set to Link Handling, Site Handling & Routing

#2 - 2019-02-14 10:54 - Ricky Mathew

- Priority changed from Should have to Must have

- Target version set to Candidate for patchlevel

Any updates?

#3 - 2019-02-14 12:39 - Lars Peter Søndergaard

The rawurlencode line can probably be removed.

A few lines earlier there is the regular expression:

```
$slug = preg_replace('/[^\p{L}0-9\/' . preg_quote($fallbackCharacter) . ']/u', '', $slug);
```

That line removes anything that is not a Letter (using unicode properties), not a digit (0-9), not a forward-slash and not the fallback character.

Letters within ASCII are only a-z and A-Z. Anything else has a codepoint beyond U+80, outside of ASCII and those characters should not cause trouble for URLs. None of them are used as URL or HTML delimiters, as far as I know.

Letters, digits and slashes are safe in the path segments. Only the fallback-character might be unsafe if defined unusual.

The only characters that could cause trouble in the path segment of a URL, is the '?', '&', '' and '' (for HTML). Those characters however, are properly encoded by the symfony UrlGenerator class:

```
Symfony\Component\Routing\Generator\UrlGenerator::doGenerate
```

It uses rawurlencode when building the url, but decodes a set of characters predefined as protected \$decodedChars.

I only followed the case where a Site configuration is available, so I wouldn't know what happens in "traditional" setups, or if those slugs are even relevant in that case.

Greetings.

#4 - 2019-02-15 06:30 - Ricky Mathew

I also think there is no need for `rawurlencode()` as the `preg_replace()` before handles everything perfectly. If `rawurlencode()` can't be avoided then I have a workaround of using `rawurldecode()` at the end of `sanitize()` function if `sanitize()` is referenced only on purpose of url generation. What all other thoughts? I think this must be considered in the next patch level as typo3 url system isn't supporting Asian languages at all !!

#5 - 2019-02-18 17:25 - Sven Juergens

here same problem with Arabic language

#6 - 2019-02-24 22:50 - Gerrit Code Review

- Status changed from New to Under Review

Patch set 1 for branch **master** of project **Packages/TYPO3.CMS** has been pushed to the review server. It is available at <https://review.typo3.org/c/Packages/TYPO3.CMS/+59796>

#7 - 2019-03-01 08:29 - Gerrit Code Review

Patch set 2 for branch **master** of project **Packages/TYPO3.CMS** has been pushed to the review server. It is available at <https://review.typo3.org/c/Packages/TYPO3.CMS/+59796>

#8 - 2019-03-01 12:37 - Gerrit Code Review

Patch set 1 for branch **9.5** of project **Packages/TYPO3.CMS** has been pushed to the review server. It is available at <https://review.typo3.org/c/Packages/TYPO3.CMS/+59822>

#9 - 2019-03-01 13:00 - Guido Schmechel

- Status changed from Under Review to Resolved

- % Done changed from 0 to 100

Applied in changeset [0299afec77ad32449b2a929f3c66d630af50cf68](https://review.typo3.org/c/Packages/TYPO3.CMS/+59822).

#10 - 2019-05-07 12:15 - Benni Mack

- Status changed from Resolved to Closed

Files

slug.png	182 KB	2018-12-26	bharat parmar
----------	--------	------------	---------------