

Bugs & Wish list

Inserting a greek chi in an article leads to immediate loss of data | Tiki Wiki CMS Groupware :: Development

Inserting a greek chi in an article leads to immediate loss of data

Status

● Open

Subject

Inserting a greek chi in an article leads to immediate loss of data

Version

18.x Regression

Category

- Release Blocker
- Bug

Feature

Article

Resolution status

New

Submitted by

hman

Lastmod by

hman

Rating

★ ★ 1/1 ★ ★ ★ ★ 1/1 ★ ★ ★ ★ 1/1 ★ ★ ★ ★ 1/1 ★ ★ ★ ★ (0) ?

Description

When you enter text into an article, or when you edit an article's content, there might come a situation where you want to use a character that is not readily on your keyboard.

Unfortunately, in standard setting, you cannot enter it's unicode address directly with ampersand hash.

That is the time to use the edit feature 'special characters', represented by a keyboard symbol. Clicking this opens a pop-up window containing several special chars, among them are some greek characters.

Btw, strangely there is a greek uppercase chi, but I did not see a lowercase one. The font (currently I use EB Garamond) not only carries all greek chars, but even accented greek and diacritic greek...

This is what happened:

But I just needed an uppercase chi, and it was there, so I clicked it to insert that into the text. It was inserted, the text looked okay, so I clicked on Save. Result: ALL text after the chi was gone, and instead of a chi there was just a]6.

I was lucky, there were just two sentences after the chi, which I could reproduce. I tried to replay the

scenario, and it could be replicated. Each and every time the text gets truncated at the chi. And because articles do not have a history (I'll have to open a feature request to have that added, it's dearly needed) this leads to immediate loss of data (once again, that is not available as a bug classification).

Importance

10 high

Easy to solve?

5

Priority

50

Demonstrate Bug (Tiki 19+)

This bug has been demonstrated on show2.tikiwiki.org

SVN update

Ticket ID

7829

Created

Sunday 19 September, 2021 11:27:59 GMT-0000

by hman

LastModif

Sunday 19 September, 2021 11:27:59 GMT-0000

Comments



hman 19 Sep 21 12:17 GMT-0000

Recreating the error in the show2 instance (on Tiki 24):

No lower case chi, also.

The text does not get truncated.

But an error is thrown. See screenshots. The chi makes it into the article, and the subsequent text is also there, even when editing the article.



Marc Laporte 19 Sep 21 13:19 GMT-0000

Thank you hman. Good catch!

We'll assign a junior dev to pinpoint the issue, and hopefully solve.



hman 19 Sep 21 15:36 GMT-0000

Thanks. Looks like there is a fix already, as there is no truncation in Tiki 24, that will have to be backported. And the error should be checked (but I must admit I did not check whether the show2 instance cried for an index rebuilding right after it's creation, and now that cannot be checked again...).

My first guess would be that it could be an "unsafe" unicode insertion that could, interpreted not as such, but as ASCII, produce an EOT or EOF within the string. Or just closing quote signs 😊

The missing lower case greek letters should be easily added, that seems to be version-independent.



Jonny Bradley 20 Sep 21 10:00
GMT-0000

This was fixed by changing the database storage from utf8 (the MySQL default using only 3 bytes per char) to utf8-mb4 and it was a colossal task, eventually completed and released in 19.x.

It would not be feasible to backport it to 18.x i'm afraid.



Marc Laporte 22 Sep 21
14:24 GMT-0000

Right

Tiki 18.x is still supported until 2023-01 but last 3 years are security-fixes only. We accept also low-risk bug fixes, which is not the case here.

<https://tiki.org/Versions>



hman 22 Sep 21
23:29 GMT-0000

Which spells out to factually only fixing bugs for 2 years instead of the claimed "5 years support period". In reality, even showstopper bugs like the non-functioning menu of 18.8 do not get fixed (reported February 21) or the calendar crashing altogether (reported the same month).

And the visual chart of the support periods on the above linked page doesn't match the declared support periods. By far.

According to that chart, 18.8 went into "S: Security fixes only" in November 2018, which is just ten months after it's release in January 2018. Ten months is significantly less than two years...



Jonny Bradley

23 Sep 21 16:20
GMT-0000

We're always looking for volunteer developers! 🌍



Marc Laporte

23 Sep 21 16:24
GMT-0000

hman:About Visual Chart: I have tasked a junior developer to review and fix. He had

worked on this before
and no one had
noticed the error.



**Marc
Laporte** 30
Sep 21 00:28
GMT-0000

hman
<https://tiki.org/Version>
has been
fixed



Marc Laporte
23 Sep 21 16:28
GMT-0000

hman: About "5 years
support period" and
period of bug fixing vs
security: What do you
propose? We can
improve the policy.



hman 24 Sep 21 08:08
GMT-0000

"This was fixed by changing the
database storage from utf8 (the
MySQL default using only 3 bytes
per char) to utf8-mb4 and it was a
colossal task, eventually
completed and released in 19.x."

My 18.8. already runs on
utf8mb4_unicode_ci. And I didn't
change that manually, so Tiki must
have adopted that earlier...



Marc Laporte 27

Sep 21 13:36

GMT-0000

I added context: `utf8mb4`

`utf8mb4_unicode_ci` could be the default when a new db is created on your system?



Marc Laporte

27 Sep 21 14:08

GMT-0000

Fixing `utf8mb4` should ideally have been done sooner, but it wasn't.

I wonder if we could add a check for emojis and other problematic characters via JavaScript and warn the user not to use...



hman 27 Sep 21

15:02 GMT-0000

The DB was created a couple of years ago...



Marc Laporte 22 Sep 21 14:29 GMT-0000

hman

On an internal task, we are tracking: "At the latest step of a fresh install: rebuild index to avoid the error message: "Click here to rebuild index" (Unified Index) or Contact the site administrator. The index needs rebuilding."

But we have yet to reliably reproduce, so we haven't yet solved. If the presence of utf8-mb4 characters provokes the error, this is valuable info.

Thanks!



Marc Laporte 27 Sep 21 13:59 GMT-0000

Merge request in discussion:

https://gitlab.com/tikiwiki/tiki/-/merge_requests/903



hman 27 Sep 21 15:23 GMT-0000

Tiki just swallowed my reply. It was GONE after clicking Post...



hman 27 Sep 21 15:24 GMT-0000

Reply was: It is not a missing uppercase chi, lowercase chi is missing!



hman 27 Sep 21 15:25 GMT-0000

Again, reply was simply discarded...



hman 27 Sep 21 15:26 GMT-0000

It's not an uppercase chi that is missing, it's lower case chi. Look at the attached screenshot.



hman 27 Sep 21 15:32 GMT-0000

Also missing:

Uppercase alpha, beta, epsilon, zeta,
lowercase theta, uppercase iota, kappa,

my, ny, both cases of omikron, then the order breaks down: The list carries another uppercase theta?!, both cases of rho missing, uppercase tau and ypsilon missing.



hman 27 Sep 21 15:33
GMT-0000

These may look similar to some latin glyphs. But abusing latin glyphs to represent greek glyphs leads into trouble with sorting and comparing...



Marc Laporte 30
Sep 21 00:30
GMT-0000

hman: describing the glyphs is making the things more complicated. Could you just copy-paste what you want added?

Thanks!



hman 30 Sep 21
19:21 GMT-0000

copy-pasting things is an invitation to disaster! Code pages in browsers... No, copy-paste is error prone. There is a good reason why things have names that can be written in ASCII. And actually HTML works that way. If you want an uppercase alpha, you can write Α and if you

want it lower case
make it α

Unicode would be
error-free, too, but
only written as code:

```
&#913  
&#945
```

but they are clearly
MUCH harder to
understand 😊

If you are unsure
about greek letters,
this might help:

https://en.wikipedia.org/wiki/Greek_alphabet



Marc Laporte 30 Sep 21 00:37 GMT-0000

About Tiki swallowing the reply: Do you think this is because of utf8mb4? I will try some emojis now...

U+1F600 🤪 — — — grinning face

U+1F603 🤪 grinning face with big eyes



hman 30 Sep 21 19:24 GMT-0000




Unicode won't work on d.t.o... Also, "grinning face" emojis did not render...



Marc Laporte 30 Sep 21 00:39 GMT-0000

Testing a Greek word:

Attachments

	filename	created	hits	comment	version	filetype
	Bildschirmfoto vom 2021-09-19 14-14-21.png	19 Sep 21 12:17 GMT-0000	0	Creating an article containing an uppercase chi		
	Bildschirmfoto vom 2021-09-19 14-14-58.png	19 Sep 21 12:18 GMT-0000	0	Resulting error		
	Bildschirmfoto vom 2021-09-27 17-11-00.png	27 Sep 21 15:12 GMT-0000	0	Special characters offered		

The original document is available at

<https://dev.tiki.org/item7829-Inserting-a-greek-chi-in-an-article-leads-to-immediate-loss-of-data>

Show PHP error messages