Bugs & Wish list

copy&paste source text from media wiki into tiki wiki this char "→" breaks "index rebuilding" | Tiki Wiki CMS Groupware :: Development

copy&paste source text from media wiki into tiki wiki this char "→" breaks "index rebuilding"

Status
● Open

Subject
copy&paste source text from media wiki into tiki wiki this char "→" breaks "index rebuilding"

Version
15.x

Category
- Release Blocker

Feature
Import-Export

Resolution status
New

Submitted by
RadoS

Lastmod by
RadoS

Rating
(0)

Description
In an old media wiki with iso-8859-1 charset, somebody supposedly has copy&pasted some M$ Office/Word DOC text into the old wiki.
Now, when migrating from the old media wiki to a new tiki wiki 15.2., somebody else copy&pastes the same text from source to source, but after saving it into the new wiki, "index rebuilding" fails because of the char "→", and with that the whole tiki wiki is stalled, won't produce any usable page anymore.

Uhm, apparently there are multiple versions of "right arrow" in M$ Word, which all seem to be converted to the same UTF char ... now.
Not sure anymore which of those it was... I can't reproduce in old environment, because the pages have been deleted after migration.

Will investigate if I can recover those broken pages/ chars, back later.

Now... funny thing, I didn't notice immediately, but other users later on, and when I "rebuild index" via

the search control panel, I get the following errorlist for simply putting this little example on a page:

"Bla → KK"

Error(s)

Indexing failed while processing "12" (type activity) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "1" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "2" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "3" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "4" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "6" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "7" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "8" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "10" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "12" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "13" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "14" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "15" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "16" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "17" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "18" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "19" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "20" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "21" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "22" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "23" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"

Indexing failed while processing "24" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "25" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "26" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "27" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "28" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "29" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "30" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "32" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "33" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "34" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "35" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "36" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "37" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "38" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "39" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "40" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "41" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "42" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "43" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "44" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "45" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
Indexing failed while processing "46" (type category) with the error "Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1"
The search index could not be rebuilt.
Could not perform index modification: Incorrect string value: '\xE2\x86\x92 KK...' for column 'data' at row 1

Importance
8

Demonstrate Bug
Please demonstrate your bug on show2.tikiwiki.org

Version: trunk ▼ Create show2.tikiwiki.org instance

Ticket ID
6219

Created
Tuesday 27 December, 2016 13:26:24 GMT-0000

LastModif
Friday 10 February, 2017 15:27:35 GMT-0000

Comments

luciash d' being ♂ 27 Dec 16 14:22 GMT-0000

Please add minimal steps how to reproduce. Thanks!

RadoS 27 Dec 16 15:29 GMT-0000

Once I figure it out to reproduce, I will. ;-)

drsassafras 28 Dec 16 02:53 GMT-0000

Could this be related to https://dev.tiki.org/item6189?highlight=4-

RadoS 13 Jan 17 13:54 GMT-0000

I don't know utf8 stuff well enough (i.e. in how many ways a right arrow can be encoded or how many right arrows exist), but sounds possible as there obvioulsy are >2 variants. ;-)
How can I check my DB setup with regard to utf (4byte capable or not)?

drsassafras 14 Jan 17 03:09 GMT-0000

I opened the doc you posted, but the characters were the same. Could you copy them directly into http://unicode.scarfboy.com/, then copy the "URL-encoded UTF8" line and paste them here. That will show exactly what the character is before it goes though any encoding, somewhere along the way.

The code I get with the arrows looks like:URL-encoded UTF8 %E2%86%92

Tiki is not currently able to save 4-bytestring characters. (hence the bug report) It requires upping the database requirements that tiki currently asks for, and also some time applying the changes in tiki. If its done quickly, it could impact performance for every tiki site... The biggest issue however that someone needs to step up and commit to implementing the changes... as is always the case. But lets figure out if his is related first.

RadoS 17 Jan 17 14:43 GMT-0000

Oh, I misunderstood it then that the DB is misconfigured and that
alone were relevant, not tiki.

I'll come back when I found a reproducible case.

RadoS 09 Feb 17 16:38 GMT-0000

I can't find it anymore, the user forgot the source.
I assume it was some leftover from migrating from iso
to utf (db remained iso, WebServer switched, copied
"raw" char from iso made no sense in utf8, pasted into
tiki made it fail because invalid utf8).
Should I hit another such case, I'll re-open this report.

RadoS 10 Feb 17 15:29 GMT-0000

I updated the description with my new findings.
a) I cannot reproduce this "just like that" in show, the same string & actions don't lead to the same results.
b) I have no clue what to inspect/ look for at DB, PHP or any other level, pointer & wild guesses welcome.

Attachments

	filename	created	hits	comment	version	filetype
	kkk.pdf	27 Dec 16 15:28 GMT-0000	0	different "right arrow" entities in M$ word		

The original document is available at http://dev.tiki.org/item6219-copy-paste-source-text-from-media-wiki-into-tiki-wiki-this-char-breaks-index-rebuilding