

Bugs & Wish list

Installer encoding issue | Tiki Wiki CMS Groupware :: Development

Installer encoding issue

Status

✖ Closed

Subject

Installer encoding issue

Version

21.x

Category

- Regression
- Upgrade Blocker

Feature

Installer (profiles, upgrades and server-related issues)

Resolution status

Won't Fix

Submitted by

Jonny Bradley

Keep informed

rjsmelo

Lastmod by

Jonny Bradley

Rating

★ ★ ★ ★ ★ (0) ?

Description

Doing an upgrade of an 18.x tiki (using the console command didn't work at all) i am presented with this big red error box:

This doesn't make sense because `utf8_unicode_ci` is a UTF-8 collation.

Solution

Sorry about this everyone, it was a local effect when testing a database which i have previously upgraded. Even though i deleted all the table in it, the database collation was still set to `utf8mb4_unicode_ci`, so the warning was saying that the tables' collations were different from this - maybe it could be better worded but nothing to fix here i think, and probably won't happen in "normal" operation.

Apologies for the noise (but it was a big scary warning! 🙄😅)

Importance

10 high

Easy to solve?

9

Priority

90

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

7270

Created

Thursday 20 February, 2020 18:47:00 GMT-0000

by Jonny Bradley

LastModif

Wednesday 04 March, 2020 17:05:32 GMT-0000

Comments



hman 28 Feb 20 16:47 GMT-0000

Maybe it's just not precise enough for an error message. According to <https://dev.mysql.com/doc/refman/8.0/en/charset-mysql.html> there are many unicode collations, but "utf8_unicode_ci" is not among them, but utf8_general_ci and utf8mb4_0900_ai_ci (although I do not know what the latter does in particular) and a VAST number of localized versions to choose from.



Jean-Marc Libs 02 Mar 20 21:04 GMT-0000

"utf8_unicode_ci" has existed for ages, and it still exists:
<https://dev.mysql.com/doc/refman/8.0/en/charset-unicode-sets.html>. It is not the same as *utf8_general_ci*.



_general_ci Versus *_unicode_ci* Collations For any Unicode character set, operations performed using the *xxx_general_ci* collation are faster than those for the *xxx_unicode_ci* collation. For example, comparisons for the *utf8_general_ci* collation are faster, but slightly less correct, than comparisons for *utf8_unicode_ci*. The reason is that *utf8_unicode_ci* supports mappings such as expansions; that is, when one character compares as equal to combinations of other characters. For example, ß is equal to ss in German and some other languages. *utf8_unicode_ci* also supports contractions and ignorable characters. *utf8_general_ci* is a legacy collation that does not support expansions, contractions, or ignorable characters. It can make only one-to-one comparisons between characters.

The message makes no sense but what is the situation on the database level?

I wonder which tables use *utf8_unicode_ci* instead of an *utf8mb4* encoding after an upgrade. The *index_** tables?



hman 03 Mar 20 09:23 GMT-0000

Since collation is highly relevant to the operation of a Wiki, this a matter that should concern many. Of course all the inner workings of Tiki Wiki must work properly, being the highest priority. But then sort orders might be something admins might want to customize to their needs, so to some degree I think this could (or should?) be placed in the hands of the admin, making it a configurable setting. With some guidance, of course, because this is prone to produce side effects, I believe.



Jonny Bradley 03 Mar 20 11:40 GMT-0000

As i understand it *collation* is not the same as *encoding*, and it's encoding that needs to be right, which mine is (it's UTF-8) so no conversion needed, therefore the warning is spurious imho.



hman 03 Mar 20 14:44 GMT-0000

My understanding ist that collation is something like the order how things get sorted, that 'a' is less than 'b'. That might seem trivial for ASCII, but it certainly isn't for umlauts or accented characters and gets tricky with punctuation or operators. Should 'á' be less than or greather than 'a' or 'ä'? And what about 'ß' that many americans call "the b looking character" or confuse it with the greek beta (it is a ligature of 's' and 'z', and it doesn't exist in Switzerland, although they do have German as official language there). That is also why there are so many collations... So collation does depend on encoding, but not the other way round. cf <https://en.wikipedia.org/wiki/Collation>



Jean-Marc Libs 03 Mar 20 18:56 GMT-0000

Yes, exactly. This message we are discussing is part of the « guidance » we provide. That's why it needs to be accurate, which it is not currently the case.



hman 04 Mar 20 09:26 GMT-0000

Good link for anyone interested in learning more about how complex collation really works: <http://www.unicode.org/reports/tr10/>
Just read the introduction and see how different languages sort completely different, sometimes not even similar, and that there are even discrepancies between sorting in dictionaries and sorting in phone books...



rjsmelo 28 Feb 20 18:07 GMT-0000

Hi @Jonny, can you share the result of running this SQL command:
(make sure you replace replace-with-your-tiki-db with the name of that tiki database)



```
SELECT TABLE_NAME, TABLE_COLLATION FROM INFORMATION_SCHEMA.TABLES WHERE  
TABLE_SCHEMA = "replace-with-your-tiki-db" AND TABLE_COLLATION NOT LIKE "utf8mb4%"  
AND NOT (TABLE_NAME LIKE "index_%" OR TABLE_NAME LIKE "zzz_unused_%");
```



rjsmelo 03 Mar 20 22:46 GMT-0000

Hi Jonny, any luck running the query above?

And also, please run the following:



```
SELECT SCHEMA_NAME, DEFAULT_CHARACTER_SET_NAME, DEFAULT_COLLATION_NAME  
FROM INFORMATION_SCHEMA.SCHEMATA WHERE SCHEMA_NAME = "replace-with-your-tiki-  
db"
```



Jonny Bradley 04 Mar 20 10:31 GMT-0000

Sorry for the delay, just short of time (and it's a 1.5GB database so takes ages to roll back, re-import and test...) - doing it next 🇺🇲



Jonny Bradley 04 Mar 20 11:35 GMT-0000

Oops, just realised you want this running after the update... will try and see why the console database update command failed first...



Jonny Bradley 04 Mar 20 14:25 GMT-0000

Stuck on item7306-Console-commands-no-longer-respect-multitiki-site-param so far...



Hi @rjsmelo, ok, think i've worked it out in the debugger - will close and add the solution and explanation above, thanks for your input and sorry for the extra noise 🙄

Attachments

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

No attachments for this item

The original document is available at <https://dev.tiki.org/item7270-Installer-encoding-issue>