Bugs & Wish list

Display of many text after upgrade of a database coded UTF8 defined latin1 as UTF8 code | Tiki Wiki CMS Groupware :: Development
Display of many text after upgrade of a database coded UTF8 defined latin1 as UTF8 code

Status
● Open

Subject
Display of many text after upgrade of a database coded UTF8 defined latin1 as UTF8 code

Version
5.x

Category
- Error
- Regression
- Consistency
- Conflict of two features (each works well independently)

Feature
All / Undefined
i18n (Multilingual, l10n, Babelfish)

Resolution status
New

Submitted by
Bernard TREMBLAY

Lastmod by
Bernard TREMBLAY

Rating
★ ★ ★ ★ ★ (0) ❓

Description
Hello,

When you use an existing database (UTF8 coded but charset=latin1 and collate=latin1-swedish-ci) converted by tiki-install 5.0 :

1- the titles and contents are displayed : as pure utf8 is saying the French texts (tables defined charset=latin1) :

A- Titles of any item
"L'échec et ses conséquences" is displayed :
" L'Ã©checs et ses consÃ©quences"

B- Texts of dictionnary (language.php UTF8 encoded) as
"créé à partir d'une structure" (comment the update or creation origin of a document) is too displayed
"crÃ©Ã© Ã partir d'une structure"

C- Contents of Wiki or articles : id

D- Menu options : are well displayed
The text comes from the same file language.php (fr)

E- Tiles of column of tables etc. : well displayed

F- Warning messages ?

Everything behaves like the titles and contents even defined as charset=latin1 are dsiplyed as UTF8 while the others are threated well.
Is it a relation with the fact that "modify" action never displays anything because of the query, I will try to understand.

Everything was displayed well in 4.x versions

trebly

Importance
9 high

Priority
45

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
3484

Created
Saturday 10 July, 2010 13:01:52 GMT-0000
by Bernard TREMBLAY

LastModif
Saturday 10 July, 2010 13:01:52 GMT-0000

Comments

Marc Laporte 18 Jul 10 08:49 GMT-0000

from svn or daily build, and please report back

Bernard TREMBLAY 04 Aug 10 22:46 GMT-0000

The changes made are not OK for the 4.2 wrong-encoded data.
The lonely solution that I found which functions (no anomaly found is :
1- export all database to SQL (with DROP table)
2- cut all the sequence x83C2 and save
3- Import

It seems that a mis UTF8 encode (doubled) has generated this wrong data conversions (nevertheless

well displayed by 4.x).

Luckily, this double encode (result known for latin1 orginal charset) generates only the introduction of the string x83C2. If the string is suppressed everywhere, we get a normal UTF8 encoding.

Luckily, again all strings column in database are concerned and the sequence can't be found in latin1 normally converted, so a full treatment of SQL database image is an easy solution.

Bernard TREMBLAY 09 Sep 10 02:20 GMT-0000

The import procedure has been updated but the problem of the sequences x83C2, present into 4.x data, are like parasite code.
The programming solution should be to suppress these characters which seems to come from a double encoding, by a translation of each field.
They can be suppressed without any problem (solves all problem and don't wakeup other.

As It take a time to develop, I export in sql, make the change with an Hexa editor and reload the database.
Then run the install 5.x with 4.1 (as) datas which make others useful conversion and data treatments.

Bernard TREMBLAY 12 Oct 10 21:29 GMT-0000

Hello Marc,

The manual conversion has been OK. for 5.0

After many personal problems put me out job till now.

At the beginning of august I not succeed normally with 5.1b, 5.1s 5.2 for install, but I could not report.

These versions could not start till 5.3 because of several problems :

1- bug on tra.php see #3481 still remaining in 5.3. When we upgrade with an existing data base, at starting tra.php is called very soon and tries to use the object $tikilib introducing a block (if $tikilib We avoid the crash (joined patched version with comments)

2- For 5.1 to 5.2 Even the install runs completely (detects a imaginary latin1 database) at starting as the install tells SUCCESSFUL at starting we reach the system error screen (this can be looped ad vitam).

I had for 4.x test developed an error handler with error.tpl which gives many details on the context (Only for test, because there no securities), but I had no time to implement it before the 5.3. So I don't know what was the problem.

3-now with 5.3 if tra.php is OK the install succeeds and there is no crash. On my data's I test with the option utf-8 and not, the obtained upgraded and converted database is identical with the one of 5.0 that I create manually and it functions on my local server.

As 5.3 functions normally on my server (windows) I began to make the OVH installation (Linux) and I got functionalities anomalies (main and unique cause ,I believe, is an utf-8 problem).
nota : As I have not yet redact the corresponding tracker you will find his reference into the next

comment.

Bernard TREMBLAY 12 Oct 10 22:40 GMT-0000

The ref is Bugs&Wish list : #3578

Attachments

	filename	created	hits	comment	version	filetype
	tra.php	12 Oct 10 21:33 GMT-0000	80	patched version for 5.x	N/A	

The original document is available at
https://dev.tiki.org/item3484-Display-of-many-text-after-upgrade-of-a-database-coded-UTF8-defined-latin1-as-UTF8-code