Bugs & Wish list

database charset issue | Tiki Wiki CMS Groupware :: Development
database charset issue

Status
🚫 Closed

Subject
database charset issue

Category
- Error

Feature
Database MySQL (MyISAM)
All / Undefined

Resolution status
Fixed or Solved

Lastmod by
pkdille

Rating
★ ★ ★ ★ ★ (0) ❓

Description
we're having a charset issue when displaying tikiwiki pages.
The pages themselves are correctly sent as UTF-8: data coming from the localization files is fine.
However, data coming from the tikiwiki database is latin1, and those characters are shown as ? in a lozenge.

The database is encoded as UTF-8, but the default encoding for MySQL, set at build-time, is Latin1.
So it appears that there is an implicit conversion done when getting/putting data to and from the DB.
Data received is converted from UTF-8 to Latin1, and data sent is considered to be Latin1 (though it is UTF-8), and converted again: the end result in the database is corrupted again.

It seems that this happens because tikiwiki uses the default charset encoding from the MySQL engine. On another system, with the same MySQL version, but built to use UTF-8 as default (--with-charset=utf-8), there is no problem.

It should probably use the "SET NAMES utf8" command to set the charset encoding, as described there:

http://dev.mysql.com/doc/refman/5.0/en/set-option.html

Software used is:
Solaris 9
Apache 2.2.3
PHP 4.4.4
MySQL 5.0.24a

Priority
25

Demonstrate Bug (Tiki 19+)

Demonstrate Bug (older Tiki versions)

Ticket ID
    929

Created
    Wednesday 08 November, 2006 16:30:00 GMT-0000
    by Unknown

LastModif
    Monday 06 October, 2008 19:25:51 GMT-0000

Comments

Marc Laporte 12 Jun 07 13:59 GMT-0000

"It should probably use the..."

Will this work for everyone? This is the only similar report of a problem. We really need to be sure before making a change like this. I have no problem and I manage dozens of Tiki sites on all kinds of different servers.

Anyone else have this problem?

Thanks!!

Koichi 25 Jul 07 09:12 GMT-0000

Yeah, I'm experiencing the same problem.
Local WindowsXP machine: Apache 2.2.4 + PHP 5.2.3 + mySQL 5.0.45
the mySQL was installed from MSI installer, and uses default charset. That seems to be latin1.
I've found a lot tips about "set names utf8", and they seem to be right.
When I open tables of tikiwiki in phpmyadmin, it shows wrong symbols instead of non-latin symbols.
Here's what I've done to make it work right:

tiki-install.php:
function has_admin() {
// Try to see if we have an admin account
global $dbTiki;
global $admin_acc;
$query = "select hash from users_users where login='admin'";
$dbTiki->Execute("set names utf8");
@$result = $dbTiki->Execute($query);
...

db/tiki-db.php:
...
if ($db_tiki == 'sybase') {
$dbTiki->Execute("set quoted_identifier on");
}

```
$dbTiki->Execute("set names utf8");
// set db version
//$dbTiki->dbversion=$dbversion_tiki;
```
...

Where should I post it or discuss it, to make some change in tikiwiki sources? I mean, they will be OK for mySQL installations with default charset other than utf8.

Koichi 26 Jul 07 10:57 GMT-0000

The solution is on Tikiwiki, PHP, UTF-8 character encoding and MySQL page by nyloth !!
It's mostly installation issue so far, and not a bug anymore. A review is advised.

Simon Pucher 31 Oct 07 14:35 GMT-0000

Simon Pucher 31 Oct 07 14:44 GMT-0000

Like Koichi I have also encountered this utf8 issue.
The problem was, that Tikiwiki and the mysql server communicated not in the same connection context (utf8). So the writing and reading between Tiki and the database was okay but if I was reading the data (or writing)from another application to the tikiwiki tables some letters (e.g. in German Ã¤,Ã¼,Ã¶,and so on) were not interpreted correctly.

The solution was to implement the following line to the db/tiki-db.php:
```
if ($db_tiki == 'mysql') {
$dbTiki->Execute("set names utf8");
}
```

fogsnow 10 Apr 08 05:11 GMT-0000

My host server's mysql default charset is GBK, so I add "set names utf8" just as koichi did. It works when I use Chinese, but when I upload pictures, it blurs, sometimes even black. How should fix it?

thanks!

Attachments

filename	created	hits	comment	version	filetype

No attachments for this item

The original document is available at https://dev.tiki.org/item929-database-charset-issue