

Core - Bug # 14848

Status:	Closed	Priority:	Should have
Author:	Jorgo S.	Category:	
Created:	2005-07-05	Assigned To:	Michael Stucki
Updated:	2006-08-12	Due date:	
TYPO3 Version:	4.0		
PHP Version:	4		
Complexity:			
Is Regression:			
Sprint Focus:			
Subject:	utf-8 data gets converted to utf-8 again under MySQL 4.1.11		
Description	<p>a) MySQL 4.1.11 - typo3 database charset UTF-8; Typo3 backend forced to UTF-8</p> <p>Upon entering text with special characters (accents, etc., in my case, I am using Japanese symbols), the database refuses to accept the entered text with error 102 ("These fields are not properly updated in database: (bodytext) Probably value mismatch with fieldtype.")</p> <p>The reason being most probably MySQL bug #9091 - http://bugs.mysql.com/bug.php?id=9091 - :</p> <p>Upon entering Utf-8 characters, MySQL is not properly notified of the format and tries to convert the characters to UTF-8 again, resulting in a change from 2bit characters to 4bit characters.</p> <p>Likely solution (as suggested in the MySQL bug article): "right after you connect and just before you do any inserts or selects on your tables, do a SET NAMES 'utf8', which will ensure that mysql doesn't reencode already encoded utf8".</p> <p>b) The same setup as above, but with typo3 mysql database set to latin1</p> <p>Doing a database dump without charset parameters and inserting the table again results in conversion/loss of utf-8 characters. This might be because the default charset for MySQL 4.1.x is utf-8, I am not sure if this issue is the same as above.</p> <p>(issue imported from #M1262)</p>		
Related issues:			
related to Core - Bug # 15295: Unable to login into BE when using MySQL 5+	Closed	2005-12-19	
related to Core - Bug # 15821: \$TYPO3_CONF_VARS['SYS']['setDBinit'] is not ha...	Closed	2006-03-14	
related to Core - Bug # 15510: UTF-8 sites display garbled chars in select-fi...	Closed	2006-01-26	

History

#1 - 2005-07-14 22:56 - Michael Stucki

Well, that's definitely a MySQL issue and should not be fixed on TYPO3s side.

However, until they have fixed their bug, here is a workaround patch for you. Completely untested but does exactly what you described above.

Please keep me up to date about any feedbacks from MySQLs side.

#2 - 2005-08-08 03:37 - Jorgo S.

Unfortunately, this patch only solves one half of the problem, namely correct saving. On reading the database, something weird happens again, and utf-8 characters get converted again, resulting in ?? characters (for Japanese language, cf. <http://chie.jorgo.org> . There are also some Spanish special characters that don't work correctly).

Clicking around in the menu and pages sometimes brings up the correct text after a while (I'm guessing whenever the db connects with "setnames")

Not sure if the patch is doing this already, but

```
SET NAMES 'utf8
```

needs to be executed before EVERY read or write operation IMHO.

Starting with MySQL 4.1.13, the patch should probably use `mysql_set_character_set()`

instead (see <http://dev.mysql.com/doc/mysql/en/mysql-set-character-set.html>).

Additional comment:

Funnily enough, the Japanese text we have introduced for LAST UPDATE is always shown correctly. Does it get stored in a different way?

#3 - 2005-08-12 19:07 - Norbert Bernhardt

I think you are on the right way. I've made the following changes to the function 'sql_pconnect' in /t3lib/class.t3lib_db.php (line 827ff):

```
function sql_pconnect($TYPO3_db_host, $TYPO3_db_username, $TYPO3_db_password) {
    global $TYPO3_CONF_VARS;
    if ($GLOBALS['TYPO3_CONF_VARS']['SYS']['no_pconnect']) {
        $this->link = mysql_connect($TYPO3_db_host, $TYPO3_db_username, $TYPO3_db_password);
    } else {
        $this->link = mysql_pconnect($TYPO3_db_host, $TYPO3_db_username, $TYPO3_db_password);
    }
    if ($TYPO3_CONF_VARS['BE']['forceCharset']=='utf-8') {
        mysql_query("SET NAMES utf8",$this->link);
        mysql_query("SET CHARACTER_SET utf8",$this->link);
    }
    return $this->link;
}
```

quote from

<http://dev.mysql.com/doc/mysql/en/charset-connection.html> :

SET NAMES indicates what is in the SQL statements that the client sends. Thus, SET NAMES 'cp1251' tells the server "future incoming messages from this client are in character set cp1251." It also specifies the character set for results that the server sends back to the client. (For example, it indicates what character set column values are if you use a SELECT statement.)

A SET NAMES 'x' statement is equivalent to these three statements:

```
mysql> SET character_set_client = x;
mysql> SET character_set_results = x;
mysql> SET character_set_connection = x;
```

Setting character_set_connection to x also sets collation_connection to the default collation for x.

(end quote)

As far as I have studied the MySQL logfiles, this is basically exactly what phpMyAdmin does when it connects to the MySQL server - and it also works perfectly for Typo3.

#4 - 2005-08-13 01:23 - Sean Ellis

The changes that nbernhardt suggested are working for the strings that I've tried. The characters that were producing the error are being managed correctly as far as I can see.

[edit; some details]

typo3 3.8, mysql 4.1.11, php4-4.3.11, phpmyadmin-2.6.2

mysql variables from mysql client:

character_set_client	latin1
character_set_connection	latin1
character_set_database	utf8
character_set_results	latin1
character_set_server	latin1
character_set_system	utf8
collation_connection	latin1_swedish_ci
collation_database	utf8_general_ci
collation_server	latin1_swedish_ci

#5 - 2005-08-16 07:55 - Jorgo S.

My pages are now saved and displayed correctly, thanks to everybody who contributed here!

Remain 2 small issues:

1. The install tool finds differences with the compare database tool that don't seem to be saved, maybe there's a another screw that needs to be tightened somewhere else.
2. The Visitor tracking extension (VTS) still displays visited pages with Japanese titles as ???, would be nice if a fix could be issued.

Thanks again!

#6 - 2005-08-16 09:00 - Michael Stucki

1. The install tool finds differences with the compare database tool that don't seem to be saved, maybe there's a another screw that needs to be tightened somewhere else.

What are these changes?

2. The Visitor tracking extension (VTS) still displays visited pages with Japanese titles as ???, would be nice if a fix could be issued.

Could be an extension issue if all else works fine.

#7 - 2005-08-16 20:10 - Sean Ellis

The templavoila extension is having trouble after making the edits to /t3lib/class.t3lib_db.php in note 0002729.

Pages that had been working are now producing this error:

Template Object could not be unserialized successfully.

Are you sure you saved mapping information into Template Object with UID "5"?

When you view the templavoila template object in the back end the mapping is gone.

I've tried a number of things with no luck so far.

- SQL ALTER commands on the tx_templavoila_* tables
- ran an extension, convert2utf8
- remapped existing templavoila template objects
- created fresh templavoila template objects, and data structures
- commenting out japanese characters from included templavoila typoscript script. The site is in development, so for now there is only content in english.

no go, so far; swapping in the edited class.t3lib_db.php file toasts my templavoila functionality

#8 - 2005-08-16 22:06 - Norbert Bernhardt

Regarding the Visitor tracking extension (VTS), it uses its own mysql_connect class (class tx_dephpot_Connection_Mysql) and function (PConnect and _pconnect). Adding the lines

```
mysql_query("SET NAMES utf8",$this->link);
mysql_query("SET CHARACTER_SET utf8",$this->link);
```

after each connection to MySQL tells the server that the specified connection is UTF8 (what we need) and not Latin-something. I haven't found a way yet to set up the MySQL server to establish UTF connections by default. This is a very good example why extensions should use the Typo3 functions and classes and not their own.

TemplaVoila and hashes - maybe it helps to clear the cache?

If you have data that was written to the database BEFORE applying the patch, then you've got have a problem: Non-ASCII characters are already messed up in the database, ie. an Ä looks like 㞲¼* or similar in phpMyAdmin and in Typo3 after applying the patch. Or you get the famous ??? because there was no match during conversation.

*㞲 ought to be a A-Tilde...

#9 - 2005-08-16 23:10 - Sean Ellis

after each connection to MySQL tells the server that the specified >connection is UTF8 (what we need) and not Latin-something. I haven't >found a way yet to set up the MySQL server to establish UTF >connections by default.

I think that you might be able to specify this by placing the SET NAMES statement in a file read by the server at startup. Filename defined by --init-file='path'; could also go in my.cnf . I haven't tried this yet.

TemplaVoila and hashes - maybe it helps to clear the cache?

'the' cache? Not sure; I've certainly been clearing all the caches that I'm aware of, including hunting for any other cache entries with phpmyadmin and deleting them.

If you have data that was written to the database BEFORE applying the >patch, then you've got have a problem: Non-ASCII characters are >already messed up in the database, ie. an Ä looks like &000258;¼ or >similar in phpMyAdmin and in Typo3 after applying the patch. Or you*

get >the famous ??? because there was no match during conversation.

**&0000258; ought to be a A-Tilde...*

I tried to eliminate this by using ALTER TABLE ... CHARACTER SET on the templavoila tables and the other steps that I took. And I have tried making entirely new templavoila template objects and data structures with the SET NAMES fix applied, but the same results.

I'll have to keep looking. There are a few things to consider and perhaps I'm overlooking something between dropping and repopulating the database and trying these different remedies.

#10 - 2005-08-17 03:21 - Norbert Bernhardt

1) As far as I've studied the MySQL manual (again), you can set the server default character set in my.cnf with

```
[mysqld]
default-character-set=utf8
default-collation=utf8_general_ci
```

However, this won't help our problem because this does not effect the (default) character set of the connection and result set (what MySQL sends back). It seems that these settings are compiled into MySQL. At least there are no switches that you could set at runtime regarding a "default connection charset".

Keep in mind that "solutions" in the MySQL manual for the [mysql] section of my.cnf only apply to the mysql(.exe) client. If you set

```
[mysql]
default-character-set=utf8
```

in my.cnf then the mysql(.exe) client uses UTF-8 for everything, even the connection. But this doesn't have any influence on the connections made by/from PHP.

2) Another chance trying to set a character set is php.ini since PHP in general connects to the MySQL server. But despite default host, username and password, there is no possibility to select a character set.

3) So the only way to tell MySQL what character set you want to use for each connection is to tell MySQL right after the connection is established.

4) As far as I can oversee, the TemplaVoila extension uses the database class of Typo3. Guess I can't help you as I don't use this extension.

#11 - 2005-08-17 05:32 - Sean Ellis

I found the relevant mysqld option. It's init-connect not init-file.

Setting init-connect='SET NAMES utf8'; in my.cnf gives the same results as editing /t3lib/class.t3lib_db.php; Japanese characters are now accepted in the backend without error, but templavoila no longer works.

So, again, no joy, but closer I think.

#12 - 2005-08-17 18:06 - Norbert Bernhardt

I think this is closest we can get.

I suggest that the

```
if ($TYPO3_CONF_VARS["BE"]["forceCharset"]=='utf-8') {
```

```
SET NAMES utf8;
SET CHARACTER_SET utf8;
}
```

is added to the CVS as solves at least the case for Typo3 and extensions who use the Typo3 classes.

It seems that not only Typo3 has this problem. Googling for solution you find SugarCRM, forums etc.

#13 - 2005-08-24 02:25 - Michael Stucki

Attached is a new patch which also executes the "SET CHARACTER_SET" command.

Did anybody yet find out why there are problems with Templavoilà?

#14 - 2005-08-24 03:19 - Sean Ellis

| *Did anybody yet find out why there are problems with Templavoilà?*

I looked, but my novice abilities prevent me from reaching any useful conclusion. I can only say that templavoila was failing at about line 216 of typo3conf/ext/templavoila/pi1/class.tx_templavoila_pi1.php after an unsuccessful attempt to unserialize(\$TOrec['templatemapping']). This variable is populated; it contains a long string like of html with some kind of extra templavoila meta information interspersed throughout.

Sorry I can't be of more help. Thanks for the feedback.

Sean

#15 - 2005-09-01 09:36 - Eduardas Paulavicius

Hello, nbernhardt

I'm using Lithuanian characters and UTF-8 under Windows XP Pro, Apache 1.3.33, MySQL 4.1.13a, PHP 4.4.

Your patch was not working for me.

When I changed line:

```
if ($TYPO3_CONF_VARS["BE"]["forceCharset"]=='utf-8') {
to:
if ($GLOBALS['TYPO3_CONF_VARS']['BE']['forceCharset']=='utf-8') {
```

it started working correctly.

Thanks,
Eduardas

#16 - 2005-12-03 12:22 - Jochen Rieger

Ran into the described problem, too. Our solution which is working quite fine so far:

Not only to set the default character set for [mysqld] in the my.cnf but also for the [client] section. But I this of cause is just a solution for "fresh" systems with no other applications using that database so far.

I guess this prevents MySQL from reencoding the sent UTF-8 encoded strings.

#17 - 2006-01-26 13:15 - Jorgo S.

This was not fixed in Typo3 4.0beta2 and I just lost all my japanese characters when updating the database.

Manually adding the additional lines suggested by nbernhardt also did not fix the problem (maybe syntax needs to be changed now).

Reverted to 3.8.1 and awaiting fix, which I hope will be included in 4.0 for everybody, since the few extra lines don't hurt anybody and can save utf-8 users a lot of grief.

#18 - 2006-01-26 14:25 - Michael Stucki

Hi Joerg,

indeed it was fixed, but I forgot to explain that here. In the Install Tool, there is a new field for SQL queries which are executed right after the connection was established. There you can now simply add the above commands:

```
SET NAMES utf8
SET CHARACTER_SET utf8
```

For details, have a look at bug #15295.

I will close this bug since I consider it solved. Please reopen if you have any problems left.

#19 - 2006-01-26 17:53 - Jorgo S.

Added line to localconf.php:

```
$TYPO3_CONF_VARS["SYS"]["setDBinit"] = 'SET NAMES utf8'; // Modified or inserted by TYPO3 Install Tool.
```

Result in webpage:

Warning: Invalid argument supplied for foreach() in /home/jorgo/public_html/typo3_src-4.0beta2/t3lib/class.t3lib_db.php on line 890

Warning: Cannot modify header information - headers already sent by (output started at /home/jorgo/public_html/typo3_src-4.0beta2/t3lib/class.t3lib_db.php:890) in /home/jorgo/public_html/typo3_src-4.0beta2/t3lib/class.t3lib_userauth.php on line 260

Warning: Cannot modify header information - headers already sent by (output started at /home/jorgo/public_html/typo3_src-4.0beta2/t3lib/class.t3lib_db.php:890) in /home/jorgo/public_html/typo3_src-4.0beta2/typo3/sysext/cms/tslib/class.tslib_fe.php on line 2698

Also tried to wrap UTF8 in various kinds of colons with the same result.

#20 - 2006-01-26 18:21 - Karsten Dambekalns

The variable is an array, so it must be:

```
$TYPO3_CONF_VARS["SYS"]["setDBinit"][] = 'SET NAMES utf8';
```

```
$TYPO3_CONF_VARS['SYS']['setDBinit'][] = 'SET CHARACTER_SET utf8';
```

#21 - 2006-01-26 18:40 - Jorgo S.

Thanks Carsten, that did the trick and now I am actually seeing the correct characters. However, I get new errors:

```
Warning: fopen(): open_basedir restriction in effect. File(/proc/loadavg) is not within the allowed path(s):  
(/home/jorgo/./usr/lib/php:/usr/local/lib/php:/tmp) in /home/jorgo/public_html/typo3_src-4.0beta2/t3lib/class.gzip_encode.php on line 302
```

```
Warning: fopen(/proc/loadavg): failed to open stream: Operation not permitted in  
/home/jorgo/public_html/typo3_src-4.0beta2/t3lib/class.gzip_encode.php on line 302
```

```
Warning: feof(): supplied argument is not a valid stream resource in /home/jorgo/public_html/typo3_src-4.0beta2/t3lib/class.gzip_encode.php on line  
303
```

I've never seen /proc/loadavg before, is that some part of Typo3?

#22 - 2006-01-27 00:09 - Karsten Dambekalns

You are welcome, Jürg (SCNR).

The open_basedir has nothing to do with the database problem, but is rather a security restriction configured in php.ini. This leads to /proc/loadavg not being accessible, which is used to determine the gzip compression level when it is set to auto in \$TYPO3_CONF_VARS.

#23 - 2006-03-14 14:41 - Jorgo S.

Would you be so kind as to update this thread with the syntax for the string value for "setdbinit" that is required starting with 4.0RC1?

```
$TYPO3_CONF_VARS['SYS']['setDBinit'] = 'SET NAMES utf8 \n SET CHARACTER_SET utf8';
```

somehow doesn't seem to work.

#24 - 2006-03-14 14:53 - Michael Stucki

That was a mistake, sorry.

Currently you need to set the value manually in localconf.php, but notice that it is not an array anymore (changed to a string).

Please check bug #15821 for further details since this is not related with the initial MySQL problem anymore...

#25 - 2006-03-14 14:59 - Jorgo S.

I actually DID add the line manually when I noticed that writing \n produced \n in localconfig.

I am however saying, that additionally, the entry does not seem to have any effect on mysql.

#26 - 2006-03-14 18:04 - Karsten Dambekalns

\n is not recognized as special sequence when inside single quotes. So either use

```
"SET NAMES utf8\nSET CHARACTER SET utf8"
```

or

```
'SET NAMES utf8'.chr(10)'.SET CHARACTER SET utf8'
```

#27 - 2006-03-14 18:37 - Jorgo S.

Works, who'd have thunk it :) Thanks K-arsten!

#28 - 2006-03-26 17:04 - Michael Stucki

Forgot to report here, but the newline-in-textarea issue has been finally solved in 4.0rc2.

Files

bug_1262.diff	632 Bytes	2005-08-24	Administrator Admin
---------------	-----------	------------	---------------------