

Поддержка UTF-8	2
Введение	2
Основные настройки	2
php.ini	2
Файл конфигурации сайта на базе сервера Apache vhost.conf.....	2
Файл конфигурации my.cnf	3
Установки TYPO3	3
Установки TypoScript	3
Настройка localconf.php	3
Настройки TYPO3 Install Tool	3
Настройка расширений	3
RealURL	4
Расширения, использующие strlen вместо t3lib_cs	4
Дополнительная информация.....	4
Проблемы с базами данных.....	4
Использование t3lib_cs	5
Использование HTML Tidy	5
Используемые материалы:	6

ПОДДЕРЖКА UTF-8

ВВЕДЕНИЕ

Здесь собрана информация по теме поддержки кодировки UTF-8. Существует много параметров, влияющих на это. Для начала, хорошо бы удостовериться, что все параметры в цепочке, начинающейся с Apache, php.ini, my.cnf и до Ваших параметров настройки TYPO3, установлены в UTF-8. В некоторых случаях необходимы не все параметры настройки, и все будет работать прекрасно даже без определенных изменений. Но здесь, по крайней мере, Вы найдете контрольный список того, что может быть ответственным за ужасные символы или вопросительные знаки при загрузке Вашего сайта.

Примечание: самую новую информацию Вы можете взять на англоязычной странице http://wiki.typo3.org/index.php/UTF-8_support#Apache_vhost.conf.

ОСНОВНЫЕ НАСТРОЙКИ

PHP.INI

Здесь приводятся установки, которые должны присутствовать в файле php.ini.

РАЗРЕШЕННЫЕ РАСШИРЕНИЯ PHP

```
extension=php_iconv.so
```

Если Вы используете php5 iconv, то эта установка сделана по умолчанию, и Вы не нуждаетесь в расширении.

Также требуется установка в localconf.php (см. ниже).

Не устанавливайте mbstring.func_overload с параметром "php_mbstring.so". В то время как это вообще полезно в настройках UTF-8, но находится в противоречии с внутренним набором символов Typo3, обрабатывающим t3lib_cs.

ПРОБЛЕМЫ С КОДИРОВКОЙ ПОЧТОВЫХ СООБЩЕНИЙ

Следующая установка внутри php.ini поможет Вам разобраться с проблемами, с которыми Вы могли бы столкнуться при пересылке почты через direct_mail:

```
mbstring.internal_encoding = UTF-8
```

Если же невозможно отредактировать php.ini, можно установить этот параметр в своем PHP-скрипте (то есть в двух Index.php-файлах TYPO3-Dummy и TYPO3-Source):

```
mb_internal_encoding("UTF-8");
```

ФАЙЛ КОНФИГУРАЦИИ САЙТА НА БАЗЕ СЕРВЕРА APACHE VHOST.CONF

Нужно убедиться в присутствии следующей установки:

```
AddDefaultCharset UTF-8
```

А также обратить внимание на следующие строчки в файле конфигурации сайта:

```
<IfModule mod_charset.c>
```

```
CharsetPriority utf-8 windows-1251 koi8-r ISO-8859-5 ibm866
```

```
CharsetDefault utf-8
```

```
CharsetSourceEnc utf-8
```

```
</IfModule>
```

ФАЙЛ КОНФИГУРАЦИИ MY.CNF

Будьте внимательны с этой установкой! Так как она воздействует и на существующие базы данных в кодировке LATIN1. Удостоверьтесь, что на сервере находятся только лишь базы данных UTF-8. Также эта установка не нужна, если Вы устанавливаете эти параметры в инструменте настройки TYPO3.

```
[mysqld]
default_character_set = utf8
```

УСТАНОВКИ TYPO3

УСТАНОВКИ TYPOSRIPT

Для правильного представления Вашего внешнего интерфейса следует установить следующие два параметра в поле настройки Вашего корневого шаблона сайта TypoScript:

```
config.renderCharset = utf-8
```

Примечание: когда Вы устанавливаете config.renderCharset, config.metaCharset, будет по умолчанию установлен в то же самое значение. Если Вы установите оба значения, TYPO3, будет использовать renderCharset внутри и преобразовывать сгенерированные страницы прежде, чем отправить их браузеру.

Примечание: если Вы устанавливаете в файле конфигурации localconf.php forceCharset в utf-8 (см. ниже), тогда это и будет значением по умолчанию для renderCharset. Таким образом эта настройка вообще не строго необходима.

НАСТРОЙКА LOCALCONF.PHP

```
// Настройка для кодировки внутреннего интерфейса
$TYPO3_CONF_VARS['BE']['forceCharset'] = 'utf-8';

// Для поддержки GIBUILDER
// Установить на 'iconv' или 'mbstring'
$TYPO3_CONF_VARS['SYS']['t3lib_cs_convMethod'] = 'iconv';
// Для поддержки 'iconv' нужен PHP 5!
$TYPO3_CONF_VARS['SYS']['t3lib_cs_utils'] = 'iconv';
```

Примечание: установив свою базу данных в UTF-8, не используйте установку \$TYPO3_CONF_VARS['SYS']['multiplyDBfieldSize'] = 3 для азиатских языков - это не нужно. Просто удалите индекс из этого поля и перезагрузитесь.

НАСТРОЙКИ TYPO3 INSTALL TOOL

[setDBinit] String (textarea): Команды, посылаемые в базы данных после соединения, разделяются переходом на новую строку. Игнорируются расширением DBAL за исключением 'родного' типа!

```
SET CHARACTER SET utf8;
SET NAMES utf8;
SET SESSION character_set_server=utf8;
```

В большинстве случаев достаточно добавить в localconf.php лишь эти строчки:

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

Помните, что каждая команда в setDBinit - это оператор SQL, который должен иметь точку с запятой после каждой команды.

НАСТРОЙКА РАСШИРЕНИЙ

REALURL

Одна из проблем состоит в том, что RealURL не в состоянии понять название страницы, если в нем присутствуют необычные (то есть не латинские) символы. Например, с название страницы на японском языке. Используя псевдонимов (латинскими символами) решает эту проблему (например, установив "home" как псевдоним начальной страницы, эта страница будет предоставлена как `yoursite.com/home.html`).

РАСШИРЕНИЯ, ИСПОЛЬЗУЮЩИЕ STRLEN ВМЕСТО T3LIB_CS

Информация: `strlen` несовместим с UTF-8, так как в UTF-8 используется от 1 до 3 байт для одного символа.

ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

ПРОБЛЕМЫ С БАЗАМИ ДАННЫХ

Хотя использовать UTF-8 в базе данных и не строго необходимо (например в MySQL), но рекомендуется. Иначе функции сортировки базы данных, не будут работать правильно.

Примечание: если Вы устанавливаете свою базу данных в UTF-8, не обязательно использовать `multiplyDBFieldSize`.

Вы можете столкнуться со следующей ошибкой:

```
SQL=Specified key was too long; max key length is 1000 bytes:
```

Эта специфическая проблема происходит, когда Вы используете кодирование UTF-8. UTF-8 использует до 3 байтов для одного символа, а максимальная длина индекса составляет 1000 байт, таким образом максимальная возможная длина индекса - $1000/3 = 333$ символа. Некоторые таблицы используют более длинные индексы, вследствие чего возникает эта ошибка (из за этого могут возникнуть и другие проблемы).

ПРЕОБРАЗОВАНИЕ СУЩЕСТВУЮЩЕЙ БАЗЫ ДАННЫХ В UTF-8

Некоторые ссылки на тему преобразования:

- <http://dev.mysql.com/doc/refman/4.1/en/charset-convert.html> (преобразование баз данных MySQL)
- <http://www.typo3-media.com/blog/article/utf8-and-typo3-updated.html>
- <http://m.tacker.org/blog/64.script-to-convert-wordpress-content-encoding.html> (полезный PHP для конвертации кодировок)
- Php скрипт для преобразования базы данных можно найти здесь http://wiki.typo3.org/index.php/Talk:UTF-8_support
- <http://bugs.typo3.org/view.php?id=6098>

Пример:

Преобразование базы данных typo3 в набор символов utf8.

URL: <http://tlug.dnho.net/?q=node/276>

Требования:

- Shell доступ на основной unix сервер
- "Sed" пакет, установленный на сервере

Внимание (для этого примера мы принимаем):

имя хоста: myhost.freedomson.com

база данных: typo3

```
#Соединение с сервером через ssh (этот пример для пользователей linux, пользователи windows должны использовать putty.exe)
ssh -l (user) myhost.freedomson.com
#архивирование базы данных
mysqldump -u (user) -p(pass) --max_allowed_packet=10000000 typo3 > typo3_backup.sql
#архив базы данных (без таблицы typo3.sys_refindex *)
mysqldump -u (user) -p(pass) --max_allowed_packet=10000000
--ignore-table=typo3.sys_refindex typo3 > typo3_utf8.sql
#Преобразование всех вхождений символов в кодировке latin1 (или другой кодировки)
в typo3_utf8.sql в utf8
sed -e 's/latin1/utf8/g' -i typo3_utf8.sql
#импорт базы данных
mysql -u (user) -p(pass) --default-character-set=utf8 typo3 < typo3_utf8.sql
# набор символов базы данных и их сопоставление изменены
mysql -u (user) -p(pass) -e "ALTER DATABASE typo3 DEFAULT CHARACTER SET utf8 DEFAULT
COLLATE utf8_bin"
```

ИСПОЛЬЗОВАНИЕ T3LIB_CS

Для разработчиков: Используйте эти функции например, чтобы получить длину строки. Функция `strlen` не получает правильную длину строки, потому что символы UTF-8 могут иметь от 1 до 3 байтов.

В PHP 5.3 PECL/intl будут доступны, таким образом, возможно что основные разработчики TYPO3 переключаются на него.

ИСПОЛЬЗОВАНИЕ HTML TIDY

Если у Вас есть проблемы с объектами `html`, такими как `&nbsp` показываемый как ? в браузере, добавьте параметр `-utf8` к HTML переменной `tidy_path` например в инструменте установки.

```
$TYPO3_CONF_VARS['FE']['tidy_path'] = 'tidy -i --quiet true --tidy-mark true -wrap 0 -raw --output-xhtml true -utf8'
```

ИСПОЛЬЗУЕМЫЕ МАТЕРИАЛЫ:

1. Официальный сайт сообщества TYPO3: <http://typo3.com/>
2. Официальный сайт WEC: <http://webempoweredchurch.com>
3. «Typo3: Enterprise Content Management», Copyright © 2005 Packt Publishing.
4. «Building Websites with TYPO3: A practical guide to getting your TYPO3 website up and running fast», Copyright © 2007 Packt Publishing.
5. «Typo3 Installation and Upgrading», Copyright 2000-2004, Kasper Skårhøj.
6. Страницка: http://wiki.typo3.org/index.php/UTF-8_support#Apache_vhost.conf