Was bedeutet Kodierung?

n verschiedenen Sprachen existieren unterschiedliche Zeichen. Die klassische Zeichenkodierung war auf 255 Zeichen pro *Zeichensatz* begrenzt um Festplattenplatz einzusparen. Der Originalzeichnsatz (ASCII) war sogar auf 128 Zeichen begrenzt und war für die englische Sprache geeignet.

Der internationale Unicode-Standard und die UTF-8 Kodierung wurden entwickelt, um alle Zeichen für alle Sprachen zu unterstützen.

Tiki benutzt seit langem intern die UTF-8 Kodierung.

Kodierungsszenarien

Die MySQL Datenbank erlaubt es, verschieden kodierte Inhalte zu speichern. In der Vergangenheit hat Tiki MySQL häufig nicht darüber informiert in welcher Kodierung Daten in die Datenbank geschrieben wurden. Daher kann es zu verschiedenen Resultaten kommen.

Der UTF-8-Fall

In diesem idealen Szenario wurde die MySQL Datenbank so konfiguriert, dass sie UTF-8 kodierte Daten angenommen und diese als UTF-8 kodierte Daten abgespeichert hat. Alle Funktionen in Tiki haben einwandfrei funktioniert und auf die Daten konnte von anderen Anwendungen problemlos zugegriffen werden. Einige Distributionen waren so voreingestellt, dass sie zu diesem Ergebnis führten. Allerdings war das in den meisten Fällen nicht so, so dass der Tiki Administrator entsprechend eingreifen musste.

Der Latin-Fall

In vielen Fällen war aber MySQL so konfiguriert, dass die Daten in *latin1* Kodierung angenommen und gespeichert wurden. Die Latin1 Kodierung ist die für die meisten westeuropäischen Sprachen geeignet. Bei dieser Konfiguration gab es beim Tiki-Frontend ebenfalls keine Probleme. Aber dies hatte Folgen:

- Wenn man mit anderen Programmen (z.B. phpMyAdmin) Daten, die von Tiki abgelegt wurden, aus der Datenbank abfragte, hat dies zu merkwürdigen Zeichen geführt.
- Der Umzug der Datenbank auf einen anderen Server kann zu unerwarteten Fehlern führen.
- Die Sortierung von Listen kann zufällig erscheinen.

Doppelte Kodierung

Bei den meisten Hostern, haten die Site Administratoren keine Kontrolle über die Standardkodierung auf dem Server. Die Kodierung der Verbindung ist in der Regel latin1. Der Site Administrator erstellt seine Datenbank aber in UTF-8 Kodierung um alle Sprachen zu unterstützen.

Diese Konfiguration führt aber zu einem verdrehten Ergebnis. Immer wenn Tiki Daten in UTF-8 abpeichert "denkt" MySQL die Daten wären in latin1 kodiert. Weil aber die Datenbank die UTF-8 Kodierung verlangt, wurden die Daten nochmals kodiert. Dies führte dann zu einem doppeltkodiertem UTF-8.

In der Realität waren die Folgen ähnlich zu denen im Fall latin1.

Informationsverlust

Wenn der Server so konfiguriert ist, UTF-8 Verbindungen anzunehmen und die Datenbank wurde mit einer latin1 Kodierung angelegt, hat Tiki die Daten richtig als UTF-8 gesendet und so wurden die Daten dann auch von MySQL erkannt. Beim Speichern der Daten in der Datenbank wurden die Daten aber zu latin1 konvertiert. Da UTF-8 über meh verschiedene Zeichen verfügt als andere Kodierungen, wurden unbekannte Zeichen zu einem '?' oder anderen unbekannten Zeichen konvertiert.

Other than not supporting languages not supported by the database encoding, all features would work correctly in Tiki.

Veränderungen

Nach einem teilweise erfolglosem Versuch mit Tiki5 diese Probleme zu lösen, wurden mehr Optionen in ((Tiki51) eingeführt.

Dieser Abschnitt erklärt wie sich das Upgrade auf 5.1 auf die bestehende Installation auswirkt.

Upgrade von 2.x, 3.x oder 4.x

Bei älteren Versionen behält Tiki sein bisheriges Verhalten bei um zu vermeiden, das Daten beschädigt werden. Das Kodierungsproblem lässt sich nur manuell lösen. Die Tiki Installation bringt hierzu Werkzuge mit, um bei der Konvertierung zu unterstützen. Sie **sollten aber auf jeden Fall verlässliche Backups haben, bevor Sie diese Tools benutzen**. Andere Techniken wie z.B. den Export der Daten von MySQL, Abänderung der exportierten Daten und re-import dieser Daten können auch verwendet werden.

Um einen Datenverlust zu verhindert, sollten Sie sicherstellen, dass Ihre Datenbanktabellen die UTF-8 Kodierung verwenden. Dies kann bei der Neuinstallation/dem Upgrade festgelegt werden.

Um die doppelte Kodierung zu entfernen, muss man sich *in seine Tiki einloggen*. Man muss dann den Zeichnsatz des Clients auf UTF-8 umstellen.

Folgen Sie dieser Beschreibungen hier:

http://doc.tiki.org/Upgrade#Fix_the_encoding_issue

Upgrade von Tiki 5.0

Standardmäßig wurde bei der 5.0 Installation der Zeichensatz für die Datenvrbindung automatisch festgelegt und in der Konfigurationsdatei gespeichert.
Neuinstallationen von Tiki 5.0 haben die Kodierung richtig festgelegt und führen zum UTF-8 Fall zu Informationsverlus (s.o.)
Upgrades auf 5.0 führen zu verschiedenen Ergebnissen. Der Upgradeprozess von 5.0 zu 5.1 ändert aber nichts an der Veränderungen durch das vorherige Upgrade. Ggf. bedarf es einer Administratoreingriffs, wenn es zu Kodierungsproblemen kommt. Der Zeichnesatz des Clients kann in der Konfigurationsdatei geändert werden.
Kontaktieren Sie die Entwickler mailing list, wenn Sie Hilfe benötigen.

New 5.1 installations

New 5.1 installations will specify the connection encoding to UTF-8 by default. If the database is not in UTF-8, this will lead to information loss. The installer will propose that you convert your database to UTF-8.

f you have no control over the database encoding and still need to use non-latin characters (ex.: Arabic) , it is best no to force the connection encoding to UTF-8 or to set it as the same encoding as the database itself. It will lead to the latin case, which comes with certain downfalls, but still allow for multiple languages to be used on a single site.

If you can control your database encoding

Pre-Tiki6, make sure to set your MySQL (via PhpMyAdmin for example) so that new tables are also created in UTF-8. In general, for your databases, look for "MySQL connection collation:" and if it's at latin_1, change it to utf8_general_ci. (Generally, this is already the case in typical shared hosting)

Then, in PhpMyAdmin, go to the database you'll be using your Tiki in Server: localhost -> Database: tiki -> Operations -> Collation, and also change from latin_1 to utf8_general_ci before running the installer. (Generally, this needs to be changed in typical shared hosting)

This is the equivalent of the following MySQL statement:

```
ALTER DATABASE DEFAULT CHARACTER SET utf8 COLLATE utf8_general_ci;
```

Tiki 5.2 or 5.3

If you have file upload issues (just upload an image and you will see if it works or not), you can solve by

- applying revision 30522, which fixes it (and will be 5.4)

If you can't do that, you can try to

- add $api_tiki='adodb'; to your db/local.php,

and/or not forcing utf8

In db/local.php

```
$client_charset='utf8';
```

Now commented out

```
// $client_charset='utf8';
```

Tiki 5.4

On a fresh install, Tiki5.4 will by default set the tables to be UTF-8 and commit 30522 will be present.

Tiki6

On a fresh install, Tiki6 will by default set the tables to be UTF-8.

Step by step procedure to fix the encoding issues

See http://doc.tiki.org/Upgrade#Fix_the_encoding_issue