

인코딩 이해하기

다른 언어는 다른 문자를 필요로 합니다. 예전의 캐릭터 인코딩은 텍스트를 저장하기 위하여 필요한 디스크 공간의 양을 줄이기 위하여 "캐릭터 세트" 당 255 캐릭터로 제한되어 있었습니다. 본래의 캐릭터 세트, **ASCII**, **sm 128** 캐릭터로 제한되어 있었고, 영어에는 적당했습니다.

유니코드 활동과 **UTF-8** 인코딩은 모든 언어에 대하여 모든 캐릭터를 최대한 효과적으로 지원하기 위하여 생성되었습니다.

티키는 내부적으로 모든 작동에 대하여 **UTF-8** 을 사용하고 있으며, 매우 오랜 시간 동안 그렇게 해왔습니다.

인코딩 시나리오

하지만, MySQL 은 보dana은 효율성을 위하여 다른 인코딩을 사용하는 것을 허용합니다. 과거에는, 티키가 MySQL 에게 제공하는 콘텐츠에 대하여 잘못된 정보를 제공하였습니다. 다양한 결과가 초래될 수 있었습니다.

UTF-8 사례

이상적인 시나리오에서는, **MySQL** 데이터베이스 서버가 **UTF-8** 인코딩을 받아들이고 데이터를 **UTF-8** 로 저장하도록 설정되는 것입니다. 모든 기능이 정상적으로 작동하였고 데이터가 다른 응용프로그램으로부터 자유롭게 접근될 수 있었습니다. 어떠한 배포판은 이런 설정으로 도달하게 되는 세팅을 제공하였습시다만, 대부분의 경우, 시스템 관리자가 개입을 해야만 했습니다.

latin 사례

많은 경우, MySQL 이 콘텐츠를 대부분의 서유럽 언어에 적절한 인코딩인 *latin1* 으로 받아들이고 저장하도록 설정되었습니다. 이러한 설정 하에서는 티키가 전면에서는 제대로 작동했습니다, 하지만, 다음과 같은 부작용이 있었습니다:

- 티키에 의하여 저장된 데이터를 phpMyAdmin과 같은 다른 응용프로그램에서 추출하는 것이 이상하게 인코드된 캐릭터들이 발생하게 되었다.
- 데이터베이스를 다른 서버로 이동하는 것이 예측하지 못한 오류를 야기하였다.
- 목차에 순서를 부여하는 것이 무작위처럼 보였다.

더블 인코딩

대부분의 공유호스팅에서는 사이트 관리자가 서버 상의 기본 연결 인코딩에 대하여 제어권이 없습니다. 연결 인코딩은 **latin1** 으로 사용되고 있을 수도 있습니다. 하지만, 데이터베이스를 생성할 때, 사이트 관리자는 **ahems djsdjfmf wldnjsgkrl dnlgkdu** 데이터베이스를 **UTF-8** 으로 만들게 됩니다.

이러한 설정은 기괴스러운 효과를 낳게 되었습니다. 티키가 **UTF-8** 데이터를 전송할 때, **MySQL** 은 이것이 **latin1** 이라고 받아들이는 것입니다. 데이터베이스가 **UTF-8** 을 요청하였기 때문에, 변환을 시도하고 이중으로 인코딩된 **UTF-8** 을 야기하게 되었습니다. 현실에서 이러한 결과는 **latin** 사례와 매우 유사하게 되었습니다.

정보 손실

만약 서버가 기본적으로 **UTF-8** 연결을 받도록 설정되었고 데이터베이스가 **latin1**으로 생성되었다면, 티키는 올바르게 **UTF-8** 데이터를 보낼 것이고 **MySQL** 은 그를 인식할 것입니다. 하지만 저장할 시간이 되었을 때, 데이터는 **latin1** 으로 변환될 것입니다. **UTF-8** 이 기타 인코딩 종류보다 더 많은 캐릭터를 보유할 수 있기때문에, 알려지지 않은 낯은 캐릭터는 ? 표시 혹은 다른 알려지지 않은 캐릭터로 변환되었습니다.

데이터베이스 인코딩에 의하여 지원되지 않는 비지원 언어 외에는, 모든 기능은 티키에서 올바르게 작동할 것입니다.

변경내역

Tiki5 에서 이러한 문제점들을 해결하려했고, 부분적으로는 실패한 시도 이후, Tiki51 에서 더 많은 선택사항이 추가되었습니다.
이 섹션에서는 5.1 로의 업그레이드가 귀하의 설치에 어떤 영향을 미칠 지 설명합니다.

2.x, 3.x 혹은 4.x 에서 업그레이드하기

이전 버전들에서 티키는 데이터가 변질되는 것을 피하기 위하여 설치된 그대로 계속 작동할 것입니다.

인코딩 문제를 수정하는 것은 수동 개입을 필요로 할 것입니다. 티키 설치관리자에는 변환을 지원하기 위한 도구가 있습니다. 하지만, **도구를 사용하기 전에 신뢰할 수 있는 백업을 해두셔야 합니다.** 데이터 내보내기, 덤프파일 수정하기, 그를 다시 불러오기 등이 관련되는 다른 방식도 사용될 수 있습니다.

데이터 손실을 피하기 위하여, 테이블이 UTF-8 을 사용하고 있다는 것을 확인하셔야 합니다. 이는 설치프로그램의 설치/업그레이드 페이지에서 진행할 수 있습니다.

이중 인코딩을 삭제하는 것은 *Enter Your Tiki* 에서 진행가능합니다. 이는 클라이언트 캐릭터셋을 UTF-8 로 강제로 설정하는 것이 필요합니다.

다음에 설명된 단계를 따라서 진행하여 주십시오:

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

5.0 에서 업그레이드

In default 5.0 installations and upgrades, the connection character set was specified automatically in the configuration file.

New 5.0 installations correctly specified the connection encoding and will lead to either the UTF-8 case or the information loss case.

Upgraded 5.0 installations will lead to different results. However, the upgrade procedure to 5.1 does not impact the changes that were made at that time. Some administrator judgment is advised if encoding issues appear. The client character set can be altered in the configuration file. Contact the developer mailing list if support is required.

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. If you have no control over the database encoding and still need to use non-latin characters (ex.: Arabic) , it is best not 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.

인코딩 문제를 해결하기 위한 단계별 절차

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

참조

다음에서는 인코딩 문제를 심층적으로 다루고 있으며 데이터를 새 UTF8 으로 인코드된 데이터베이스로 불러오기를 한 후, 문제를 성공적으로 해결하는 방법을 설명합니다.

Getting out of MySQL Character Set Hell