

Re: MySQL Insert Unicode Problem

Source: <http://coding.derkeiler.com/Archive/Python/comp.lang.python/2007-04/msg01338.html>

- *From:* John Nagle <nagle@xxxxxxxxxxxx>
 - *Date:* Mon, 09 Apr 2007 14:59:47 -0700
-

erikcw wrote:

Hi,

I'm trying to insert some data from an XML file into MySQL. However, while importing one of the files, I got this error:

```
Traceback (most recent call last):
File "wa.py", line 304, in ?
main()
File "wa.py", line 257, in main
curHandler.walkData()
File "wa.py", line 112, in walkData
self.cursor.execute(*sql) #try multi-execute to speed up query.
File "build/bdist.linux-i686/egg/MySQLdb/cursors.py", line 151, in
execute
File "build/bdist.linux-i686/egg/MySQLdb/connections.py", line 247,
in literal
File "build/bdist.linux-i686/egg/MySQLdb/connections.py", line 185,
in unicode_literal
UnicodeEncodeError: 'latin-1' codec can't encode character u'\u20ac'
in position 28: ordinal not in range(256)
```

What do I need to change to make this work?

Thanks!
Erik

MySQL does Unicode quite well, but Unicode is not the default. Getting all the defaults right is a pain. It looks like you have MySQL, or MySQLdb, or the connection between them in "latin1" mode.

The database server has to support Unicode, which is a build option and a version issue. Do a

```
SHOW VARIABLES LIKE "character_set%"
```

Ideally, you should see the character sets for client, connection,

Re: MySQL Insert Unicode Problem

database, results, server, and system all as "utf8". You might not, given the problem you're having, but that can be dealt with. Those are just defaults.

The field you're storing into has to be in Unicode mode, which can either be set explicitly with `CREATE TABLE` or can come from the default for the server, database, or table. It's usually easier to use UTF8 for everything in the database, unless you have legacy database issues.

Try this in SQL:

```
ALTER DATABASE yourdatabasename DEFAULT CHARACTER SET UTF8;
```

If you can't execute that statement, assuming you have permissions to do so, then your MySQL isn't configured to support Unicode. (If you're on a shared server, you might have to use `ALTER TABLE` instead; you may not have `ALTER DATABASE` permission.)

Once you've done that, all tables created after that point will have Unicode text fields. You can also use `ALTER TABLE` on existing tables, if you need to convert their format, or create tables with `DEFAULT CHARACTER SET UTF8`.

The connection to the MySQL server has to be in Unicode mode. This is separate from the server's mode and the table mode. And MySQLdb has to be in Unicode mode, which is separate from the connection mode. MySQLdb doesn't know what the server is doing; you have to tell it what you want.

Try something like this to connect to the server:

```
db = MySQLdb.connect(host="localhost",  
use_unicode = True, charset = "utf8",  
user=username, passwd=password, db=database)
```

You need to specify both "use_unicode" (which sets the mode in MySQLdb), and "charset" (which sets the connection mode).

Once you've put all the components into these modes, you should have end-to-end Unicode database capability.

John Nagle

.