

Изменение кодировки DBF файла

[Обсудить в форуме](#) Комментариев — 22

Эта страница опубликована в основном списке статей сайта по адресу <http://gis-lab.info/qa/dbf-encode.html>

Описание использования библиотеки dbfpy для изменения кодировки данных и скрипт.

Неотъемлемой частью данных в формате share является атрибутивная таблица в формате DBF. Атрибутивные данные могут храниться в разных кодировках, часто встречаются такие кодировки как Windows-1251 (CP1251) и UTF8, KOI8-R. Некоторое, особенно относительно давно появившиеся ГИС, такие как Arcview GIS 3.x, некоторые версии Mapinfo, не умеют работать с более новой кодировкой UTF8. Соответственно, часто встречается задача перевода DBF из одной кодировки в другую.

Для преобразования dbf из одной кодировки в другую можно использовать скрипт на языке Python ([скачать скрипт](#)). Эта статья описывает подготовку к работе, идею скрипта и как с ним работать.

Содержание

- [1 Подготовка к работе](#)
- [2 Идея](#)
- [3 Использование](#)
- [4 Пример использования](#)
- [5 Ошибки конвертации](#)
- [6 Ссылки по теме](#)

Подготовка к работе

Для работы скрипта необходимо иметь Python 2.5/2.6 и один раз скачать и установить библиотеку [dbfpy](#). После загрузки библиотеки нужно выполнить:

```
python setup.py install
```

Правильность установки библиотеки можно проверить запустив Python и выполнив:

```
from dbfpy import dbf
```

Если все установлено нормально, ошибок эта команда выдать не должна.

Идея

В процессе работы, наш скрипт должен построчно проверять тип поля в исходном файле DBF и, если он строковый, т.е. есть вероятность того, что данные не в той кодировке, что нужно, осуществлять преобразование. Упрощенная версия скрипта на языке Python для перевода данных из кодировки в кодировку выглядит следующим образом:

```
#!/usr/bin/python
#-*- encoding: utf-8 -*-

import sys
from dbfpy import dbf
from types import *

db = dbf.Dbf(sys.argv[1])
```

```

newDB=dbf.Dbf(sys.argv[2], new=True)

for f in db.header.fields:
    newDB.addField(f)

for rec in db:
    r=newDB.newRecord()
    newData=[]
    for f in rec.fieldData:
        if type(f)==StringType:
            f=unicode(f,'utf-8')
            f=f.encode('windows-1251')
        newData.append(f)
    r.fieldData=newData
    r.store()

db.close()
newDB.close()

```

Использование

Утилита вызывается из командной строки и принимает несколько параметров:

```
python dbf-encode.py -f encin -t encout -n input output
```

- input — имя исходного файла dbf, который требуется преобразовать, обязателен
- output — имя конечного файла dbf, куда будет записан результат, обязателен
- -f (--from-code) — кодировка исходного DBF файла, не обязателен, по умолчанию UTF-8
- -t (--to-code) — кодировка выходного DBF файла, не обязателен, по умолчанию CP-1251
- -n (--no-subs) — переключатель, если установлен, то неизвестные символы пропускаются, если не установлен — заменяются на знаки вопроса («?»)
- -h (--help) — справка
- -v (--version) — версия программы

Пример использования

Конвертировать adygeya-admin-a.dbf в adygeya-admin-a2.dbf из кодировки UTF-8 в кодировку CP-1251.

```
python dbf-encode.py -f UTF8 -t CP1251 adygeya-admin-a.dbf adygeya-admin-a2.dbf
```

или упрощенная запись параметров командной строки (эти кодировки и так являются кодировками на входе и выходе по умолчанию):

```
python dbf-encode.py adygeya-admin-a.dbf adygeya-admin-a2.dbf
```

Ошибки конвертации

Если конвертер сталкивается с символом, который не может перекодировать, он по-умолчанию заменяет его на вопросительный знак («?»), если к строке запуска добавлен --no-subs (-n), то символ пропускается.

Если возникает другая ошибка при конвертации, конвертер выводит содержание записи на которой возникла ошибка в консоль и прекращает работу. Конечный файл создается неполным и для подмены оригинального в share-файле непригоден.

Ошибка **invalid syntax** может быть связана с неправильным запуском скрипта, под Windows запуск нужно осуществлять из командного процессора («Пуск → Выполнить → cmd») после чего вводить команду указанную выше.

[Скачать](#) полную версию dbf-encode.py

Ссылки по теме

- [Конвертация DBF в CSV](#)

[Обсудить в форуме](#) Комментариев — 22

Последнее обновление: 2014-05-15 00:45

Дата создания: 16.05.2010

Автор(ы): [Максим Дубинин](#)