Scripts Python

Corriger/convertir l'encodage d'un texte utf-8

Alternative sous MySQL/MariaDB

Lien : https://stackoverflow.com/questions/20151835/how-to-convert-wrongly-encoded-data-to-utf-8)

```sql select convert(binary convert(field_name using latin1) using utf8) from table_name ```

ftfy (fixes text for you) est une librairie python spécialisée dans la correction des erreurs d'encodage utf-8

Installer pip et ftfy sous ubuntu

```
apt install python3-pip
pip3 install ftfy
```

Corriger l'encodage d'un fichier (par exemple la sauvegarde d'une base mysql)

```
#!/usr/bin/python3
# coding: utf-8

import ftfy

# Set input_file
input_file = open('c1alfahnet.dump', 'r', encoding="utf-8")
# Set output file
output_file = open ('c1alfahnet.utf8.dump', 'w')

# Create fixed output stream
stream = ftfy.fix_file(
    input_file,
    encoding=None,
    fix_entities='auto',
    remove_terminal_escapes=False,
    fix_encoding=True,
    fix_latin_ligatures=False,
    fix_character_width=False,
    uncurl_quotes=False,
    fix_line_breaks=False,
    fix_surrogates=False,
    remove_control_chars=False,
    remove_bom=False,
    normalization='NFC'
)

# Save stream to output file
```
```

```
stream_iterator = iter(stream)
while stream_iterator:
    try:
        line = next(stream_iterator)
        output_file.write(line)
    except StopIteration:
        break
```

From:
https://wiki.pielo.net/ - **Pielo.net - Wiki**

Permanent link:
https://wiki.pielo.net/scripts-python?rev=1686482349

Last update: **2023/06/11 13:19**