

Scripts Python

Corriger l'encodage d'un texte utf-8

`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('clalfahnet.dump', 'r', encoding="utf-8")
# Set output file
output_file = open ('clalfahnet.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=1501002288>

Last update: **2017/07/25 19:04**

