

# 文本编码转换 python 脚本

作者: [JoeysCat](#)

原文链接: <https://ld246.com/article/1612370837077>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

《30天自制操作系统》附带的源码，因为使用了CP932字符集，普通的编辑器打开，注释都是乱码的  
以下脚本可以将特定的CP932文本文件转为UTF8编码。

```
import os

def conv(file_path):
    lines = []
    with open(file_path, "r", encoding="CP932") as f:
        try:
            lines = f.readlines()
        except Exception as e:
            print(file_path)
            print(e)
            return
    if not lines:
        return
    with open(file_path, 'w', encoding='utf8') as nf:
        data = "".join(lines)
        nf.write(data)
        nf.flush()

# exts = ['.bat', '.nas', '.txt', '.h', '.c']
# full_names = ['Makefile']
exts = ['.txt', '.h', '.c']
full_names = []

def need_convert(file_name):
    if file_name.startswith('.'):
        return False
    for ext in exts:
        if file_name.endswith(ext):
            return True
    for fn in full_names:
        if file_name == fn:
            return True

    return False

if __name__ == '__main__':
    work_dir = '/xx/30天自制操作系统光盘'
    for parent, dir_names, filenames in os.walk(work_dir, followlinks=True):
        for filename in filenames:
            if not need_convert(filename):
                continue
            file_path = os.path.join(parent, filename)
            conv(file_path)
```