# 本文件用于合并软著代码

import re

import os

files\_list = [

 r"server\main.py",

]

# 主要逻辑： 读取文件清除代码注释，清除空白行，最后合并文件内容

files\_suffix = ["py", "js"] # 需要合并的文件后缀

pass\_files = ["config.toml", "requirements.txt", "README.md", "LICENSE", "index.html"] # 不需要合并的文件

pass\_path = [".venv"] # 无需扫描的文件夹

if not any(files\_list):

 print("未自定义文件列表，开始自动搜索")

 for root, dirs, files in os.walk("."):

 if any([pass\_dir in root for pass\_dir in pass\_path]):

 continue

 for file in files:

 if file.endswith(tuple(files\_suffix)) and file not in pass\_files:

 files\_list.append(os.path.join(root, file))

 print(f"共找到 {len(files\_list)} 个文件")

exit()

contents = ""

for file in files\_list:

 with open(file, "r", encoding="utf-8") as f:

 content = f.read()

 # 将非独立一行的注释转移到新的一行

 content = re.sub(r'(?<!\n)(#[^\n]\*)', '\n<1>', content)

 # 清除URL

 content = re.sub(

 r'http[s]?://(?:[a-zA-Z]|[0-9]|[$-\_@.&+]|[!\*\\(\\),]|(?:%[0-9a-fA-F][0-9a-fA-F]))+', 'URL\_PLACEHOLDER', content)

 # 清除注释

 # content = re.sub(r"(#[^\n]\*)|(\'\'\'.\*?\'\'\')|(\"\"\".\*?\"\"\")", '', content, flags=re.DOTALL) # 清除单行和多行注释

 print(f"文件 {file} 合并成功，共 {len(content)} 字符")

 contents += content

# 清除空白行

content = re.sub(r'\n\s\*\n', '\n', contents)

if os.path.exists("code.txt"):

 os.remove("code.txt")

with open("code.txt", "w", encoding="utf-8") as f:

 f.write(content) # 将合并后的内容写入文件

lines = content.split("\n")

print(f"总共有 {len("".join(lines))} 字，有 {len(lines)} 行，可写 {len(lines)/50} 页")

# 没有更多了

# 没有更多了

# 没有更多了

# 没有更多了