

# ■■■■ PowerShell ■■■■ ■■■■ (■■■ + ■■■■■)

■ ■■■ Windows PowerShell■■■ ■■■■ ■■■■■■ ■■■■ ■ ■■ '■ ■■■' ■■■■ ■■ ■■■■ ■ ■■■■ ■■■■ ■■■■. ■ ■■■■ ■■■■(■■■ ■■ ■■)■ ■■■■■■.

## 1) ■■■

• ■■■■ ■■ ■■■■ ■■(■: qj\_paths → rel/cj\_rel)■ ■■■■ ■■ ■■■■ ■■■■. • ■■■→■■■→■■■→■■■→■■■ ■■■■ ■■ ■■■■■■ ■■■■. • ■■ ■■■■(archive/\* .md|.html|.json)■ ■■ ■(≥15)■ ■■■■ ■■■■■■ ■■.

## 2) ■■ ■■(■■■■■ ■■■■ ■■)

```
# 2-1. run_all.py(■■ ■■ ■■■■■■■■) ■■ ■■  
Get-ChildItem -Path . -Filter run_all.py -Recurse
```

```
# 2-2. run_all.py■ 'qj_paths' ■■■■ ■■■■ ■■  
Select-String -Path ".\run_all.py" -Pattern "qj_paths" -SimpleMatch
```

■ '■■■ ■■ ■ ■■'■ ■■ ■■ ■■■■ ■■■■ ■■■■ ■■■■ ■■ ■■■■. ■■■■ ■■■■ ■ ■■ ■■■■■■.

## 3) ■■■■ ■■■■ ■■ ■■(■■■■ ■■ → ■■ ■■)

```
# 3-1. ■■■■ ■■■■ ■■■■(■■ ■■■■ ■■ ■■)  
Get-Content ".\run_all.py" |  
Where-Object { $_ -notmatch '^s*from\s+qj_paths\s+import\s+rel.*\scj_rel' } |  
Set-Content ".\run_all_new.py" -Encoding UTF8
```

```
# 3-2. ■■■■/■■■■■ ■■■■ ■■ ■■■■ ■■ ■■ ■■  
Remove-Item ".\run_all.py"  
Rename-Item ".\run_all_new.py" "run_all.py"
```

```
# 3-3. ■■(■ ■■ 'qj_paths'■ ■■■■ ■■ ■■)  
Select-String -Path ".\run_all.py" -Pattern "qj_paths" -SimpleMatch  
■■ ■■: ■■■■ Select-String■ ■■ ■■ ■■■■ ■■■■ ■■■■■■.
```

## 4) ■■ ■■■■■■ ■■ ■■■■(■■■■)

```
# 4-1. ■■ ■■(■■■/■■■■■/■■■■ ■■■■ ■■■■ ■■ ■■ ■■1 ■■ ■■)  
py orchestrator.py --collect  
py orchestrator.py --collect-community  
py orchestrator.py --collect-keyword "IPC-A-610"
```

```
# 4-2. ■■■■/■■■  
py .\tools\rebuild_kw_from_today.py  
py .\tools\augment_from_official_pool.py
```

```
# 4-3. ■■ ■■ ■■(≥ 15■ ■■■■)  
(Get-Content -Raw .\selected_keyword_articles.json | ConvertFrom-Json).articles |  
Where-Object { $_.approved -eq $true } | Measure-Object | Select-Object -ExpandProperty Count
```

```
# 4-4. ■■ ■■ + ■■■■ ■■■■  
py .\tools\repair_selection_files.py  
py .\tools\sync_selected_for_publish.py
```

```
# 4-5. ■■■■ ■■(■■■■■ ■■■■ ■■)  
py orchestrator.py --publish-keyword "IPC-A-610"
```

## 1. 数据源

- selected\_keyword\_articles.json: 包含 keyword 和 approved=True 的 15 条数据。
- selected\_articles.json: 包含 10 条数据。
- archive/ 下的 \_ 开头的 .md/.html/.json: 包含 10 条数据。
- 从 1000 条数据中随机抽取 10 条数据。

## 5) 异常处理

- • IOError: 在 run\_all.py 中处理 IOError。使用 try-except 捕获并打印错误信息。
- • f-string/here-string: 使用 f-string 或 here-string 来格式化字符串。
- • UnicodeDecodeError: 在 UTF-8 编码时遇到错误，使用 encoding='utf-8', errors='replace' 来处理。
- • 使用 logging 模块记录错误信息。

## 6) 数据清洗

- ① 在 run\_all.py 中定义清洗函数。
- ② 使用正则表达式清洗文本。
- ③ 在 run\_all.py 中调用清洗函数。
- ④ 在 run\_all.py 中定义清洗函数。
- ⑤ 使用 pandas 库进行数据清洗。

## 7) 数据导出

文件/目录	内容
selected_keyword_articles.json	包含 keyword 和 approved=True 的 15 条数据
selected_articles.json	包含 10 条数据
archive/ 下的 _ 开头的 .md/.html/.json	包含 10 条数据

8. 数据导出到数据库。