

██████ █████ █████ █████ █████ — █████ █████, █████, 5████████
██████ (v1.0)

■■■: 2025-10-04 (KST) · ■■ ■■: 2025-10-04 10:16 KST

§§ — §§§ §§§§/§§ §§§§ §§ §§ §§§ §§§§, §§§§ §§§ §§§ §§§§ §§ §§ §§§ §§§
 § 5§(5§§) §§§ §§§§§§ §§§§§. § §§§§ '§§ § §§§' §§ §§§§§§ §§ §§§§ §§§§ §§→§§→
 §§→§§§ § §§§ §§§§.§§§§§§ § §§§§ §§§§.

[illegible]

1) $\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare (\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare)$

- **爬取策略/规则**：爬取深度 ≥15，`features.require_editor_approval=True` 爬取深度限制。爬取深度限制为 10。
- **数据清洗与存储**：`selected_keyword_articles.json` (关键词) → `selected_articles.json` (文章) → `archive/` 目录。PowerShell 脚本处理文件命名格式。
- **社区来源过滤**：`community_sources.json` 记录来源，过滤包含 `require_keyword`, `min_title_len`, `score_threshold` 的文章。排除特定网站如 Reddit。
- **数据格式转换与UI展示**：将 JSON 数据转换为 QG(图), FC(力导向), GMM(高维映射), UI 展示支持多种视图 (QG/MM/GM/FC), 导出 HTML/MD 文档。
- **数据处理流程**：Collector → Preprocess(QG/FC) → Evaluator → Summary/Translate → Storage → Admin UI → Publisher。集成 Elasticsearch, CI/CD, Docker 部署。
- **接口与输出**：`requests`, `beautifulsoup4`, `lxml` 库用于数据采集和解析。

2) ■■ ■■■■ ■■

- ****[REDACTED] [REDACTED]/[REDACTED] [REDACTED]**:** [REDACTED].[REDACTED] [REDACTED] JSON [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]. [REDACTED] [REDACTED] [REDACTED] [REDACTED].
- ****[REDACTED] [REDACTED] [REDACTED] [REDACTED]**:** Reddit/[REDACTED] API/[REDACTED] [REDACTED], [REDACTED] [REDACTED] '0' [REDACTED]. [REDACTED]/[REDACTED]/[REDACTED] [REDACTED] [REDACTED] [REDACTED].
- ****[REDACTED] [REDACTED] [REDACTED] [REDACTED]**:** [REDACTED] <15 [REDACTED] [REDACTED] [REDACTED] [REDACTED]. [REDACTED] [REDACTED]/[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED].
- ****[REDACTED] [REDACTED] [REDACTED]/[REDACTED] [REDACTED]**:** [REDACTED] [REDACTED] [REDACTED] [REDACTED], [REDACTED]/[REDACTED] [REDACTED] [REDACTED] [REDACTED].
- ****[REDACTED]/[REDACTED] [REDACTED].[REDACTED]**:** [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]/[REDACTED] [REDACTED].[REDACTED] [REDACTED] [REDACTED] [REDACTED].

3) $\square\square\square (\square\square \square\square \square\square\square \square\square)$

- *****rail**:** `## ##/## #####`{'date':..., 'articles':[...]}` / {'keyword':..., 'date':..., 'articles':[...]}`
##. ## `repair_selection_files.py` ## ##### ##.`

- ****[REDACTED] [REDACTED]****: `require_keyword=True` [REDACTED] `min_title_len`, `score_threshold` [REDACTED] '[REDACTED] [REDACTED] [REDACTED]' [REDACTED] [REDACTED]. [REDACTED] [REDACTED](`keyword_synonyms.json`)[REDACTED] [REDACTED] [REDACTED].
- ****[REDACTED] [REDACTED] [REDACTED]****: RSS/[REDACTED] URL [REDACTED] [REDACTED], 404/[REDACTED] [REDACTED] [REDACTED] [REDACTED]/[REDACTED]. [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED].
- ****QG/FC [REDACTED] [REDACTED]****: `editor_rules.json` [REDACTED] QG [REDACTED] [REDACTED]/[REDACTED] [REDACTED] [REDACTED] [REDACTED]. [REDACTED]/[REDACTED].[REDACTED] [REDACTED] [REDACTED] [REDACTED], [REDACTED] [REDACTED] [REDACTED] [REDACTED].
- ****[REDACTED] [REDACTED] [REDACTED]****: [REDACTED] [REDACTED] N([REDACTED]: 20) [REDACTED] [REDACTED] [REDACTED] 'HOLD [REDACTED] + FC PASS [REDACTED]' [REDACTED] [REDACTED].
- ****[REDACTED] [REDACTED] [REDACTED]****: [REDACTED] [REDACTED] [REDACTED].[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]([REDACTED] [REDACTED].[REDACTED] [REDACTED]).
- ****[REDACTED]/[REDACTED] [REDACTED]****: [REDACTED] [REDACTED]→([REDACTED])[REDACTED] [REDACTED], Glossary [REDACTED] [REDACTED]. [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED].
- ****[REDACTED]/[REDACTED] [REDACTED]****: [REDACTED] [REDACTED]([REDACTED], QG/FC [REDACTED], [REDACTED] [REDACTED], [REDACTED] [REDACTED])[REDACTED] [REDACTED] `archive/reports/` [REDACTED] [REDACTED] [REDACTED].

4) [REDACTED] [REDACTED] ([REDACTED].[REDACTED])

- ****[REDACTED] [REDACTED]****: Collector/Preprocessor/Evaluator/Summarizer/Publisher [REDACTED], [REDACTED] DTO(JSON) [REDACTED] [REDACTED].
- ****[REDACTED] [REDACTED] DB****: [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]([REDACTED]/[REDACTED]/[REDACTED] [REDACTED]). Elastic/OpenSearch [REDACTED] [REDACTED].
- ****[REDACTED] UI [REDACTED]****: [REDACTED] [REDACTED] [REDACTED] [REDACTED], [REDACTED]/[REDACTED]/[REDACTED]/[REDACTED], [REDACTED] [REDACTED] [REDACTED], [REDACTED].[REDACTED] [REDACTED] [REDACTED].
- ****[REDACTED]/[REDACTED]****: [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]/[REDACTED]/[REDACTED] [REDACTED]([REDACTED]/[REDACTED] [REDACTED]) [REDACTED] [REDACTED].
- ****CI/CD & [REDACTED]****: [REDACTED].[REDACTED] [REDACTED] [REDACTED].[REDACTED] [REDACTED]. [REDACTED] [REDACTED] [REDACTED]([REDACTED] [REDACTED] [REDACTED]) [REDACTED].
- ****[REDACTED].[REDACTED] [REDACTED]****: [REDACTED] [REDACTED]/[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED], [REDACTED] [REDACTED].[REDACTED], [REDACTED] [REDACTED] [REDACTED] [REDACTED].

5) 5■(5■■) ■■ ■■ — 1■ 1■■

D1 — ■■■■■ ■■■■/■■■

- **■■■**: ■■ ■■■■■ ■■■-■■■ ■■-■■ ■■.
- **■■ ■■**: (a) ■■→■■■→■■→■■■→■■ ■■■ ■■/■■■ ■■, (b) ■■/■■■ JSON ■■■ ■■■ ■■ ■■, (c) `repair_selection_files.py` ■■ ■■ ■■, (d) ■■ ■■ ■■ config ■■ ■■ ■■ ■■.
- **■■■**: ■■ ■■■ ■■, ■■■ ■■ ■■■, ■■■■■ ■■ ■■■■ ■■■.
- **■■ ■■**: ■■ ■■/■■ ■■ ■■■■ ■■■ ■■ ■■ ■■ ■■.

D2 — ■■ ■■■■/■■■■ ■■

- **■■■**: ■■■ ■■ ■■ ■■/■■■■ ■■.
- **■■ ■■**: (a) ■■ ■■ ■■■■/■■ ■■, (b) ■■■■ ■■ ■■(■■ ■■, ■■■ ■■), (c) ■■ ■■ ■■ ■■ ■■■ FC/HOLD ■■■ ■■, (d) ■■ ■■ ■■.
- **■■■**: ■■■■■ collectors ■■ community ■■, `keyword_synonyms.json` ■■ ■■.
- **■■ ■■**: ■■ ■■■■■ ■■ ≥ 20 , ■■ ■■ ≥ 15 ■■ ■■.

D3 — QG/FC/■■■■ ■■■

- **■■■**: QG/FC/■■■■ ■■■■-■■ ■■.
- **■■ ■■**: (a) `editor_rules.json` ■■ ■■■ QG/FC ■■-■■■/■■■ ■■■ ■■, (b) Evaluator ■■■, (c) ■■ ■■ ■■ ■■■.
- **■■■**: evaluator ■■, ■■■ ■■■, ■■■ ■■■(■■■).
- **■■ ■■**: ■■ ■■■ 3■■ ■■ ■■■ ■■-■■ ■■ ■■.

D4 — ■■ ■■■-■■/■■ ■■■

- **■■■**: ■■ ■■■/■■-■■ ■■■ ■■.
- **■■ ■■**: (a) MD/HTML ■■■ ■■(■■■ ■■/■■-■■■ ■■), (b) ■■→■■ ■■ ■■■■ Glossary ■■ ■■, (c) ■■ ■■ ■■.
- **■■■**: ■■■, ■■ ■■■■■, Glossary ■■ ■■■, ■■ ■■■■/■ ■■.
- **■■ ■■**: ■■ 15~20■ ■■ ■■ ■■-■■ ■■ ■■ ■■.

D5 — ■■■ UI/■■/■■■

- **■■■**: ■■■ UI(■■■ ■) ■ ■■ ■■ ■■.
- **■■ ■■**: (a) ■■/■■■ ■■ UI, ■■/■■/■■-■■, ■■■■/■■ ■■, (b) ■■ ■■■ API, (c) ■■■■/■■■ CI, (d) ■■ ■■■.
- **■■■**: /admin/keyword ■, ■■■ API, ■■/CI ■■■■, ■■ ■■■ ■■.
- **■■ ■■**: UI■■ ■■→■■→■■■■ 1■ ■■, ■■■■ ■■ ■■.

6) 数据源

- 数据源: 数据源, 数据源.
- 数据源: 数据源, QG/FC 数据源, 数据源 URL 数据源.
- 数据源: 数据源 ≥ 15 数据源, 数据源, Glossary 数据源.
- 数据源: 数据源 3 (MD/HTML/JSON), 数据源, 数据源 数据源, 数据源 数据源.
- 数据源: 数据源/数据源 数据源 数据源, 数据源 数据源 数据源, 数据源 URL 数据源.

7) 数据源 (数据源)

- `data/selected_keyword_articles.json` (数据源 数据源)
- `selected_articles.json`, `selected_community.json` (数据源 数据源)
- `archive/` 数据源 数据源 数据源 (.md/.html/.json)
- `run_quali_today.ps1` 数据源 数据源 数据源
- 数据源 UI 数据源 数据源/数据源 (数据源)
- 数据源 数据源 (数据源/数据源) 数据源 数据源 数据源

数据源) 数据源/数据源 数据源

- 数据源 数据源: 数据源 数据源 (数据源 ≥ 15), 数据源 数据源, PowerShell 数据源, 数据源 数据源. 数据源 数据源
- 数据源 数据源 数据源: 数据源 数据源 数据源 数据源, QG/FC/数据源, JSON 数据源, UI/数据源 数据源
- 数据源 数据源 数据源: 数据源 数据源 数据源 数据源, 数据源/数据源
- 数据源 数据源 数据源: 数据源 数据源 DB, 数据源 数据源/数据源
- UI 数据源: 数据源. 数据源/数据源/数据源, API
- 数据源 数据源: requests/bs4/lxml 数据源

数据源 数据源 数据源 数据源 数据源 数据源, 数据源 数据源/数据源 数据源 数据源 数据源 数据源 数据源.