WorldFootball爬虫文档

# 项目结构

* main.py：主入口，执行联赛和杯赛的抓取与清洗
* wfmaster/
  + scraper/：联赛和杯赛的抓取模块
  + cleaner/：联赛和杯赛的数据清洗模块
  + config.py：配置工具
* scripts/：手动抓取/清洗的辅助脚本
* config/：比赛和联赛的映射 CSV 文件
* output/：生成的 Excel 文件和日志
* .env：文件及路径配置

# 使用方法

1. 安装依赖（如有）：

* pip install -r requirements.txt

1. 更新配置文件

* 修改 .env 文件：
  + Team mapping file  
    球队名称映射文件
  + League & Cup  
    OUT\_FILE：最后输出赛程文件名称（位于 output/） MAP\_FILE：赛事信息文件（位于 config/）
* 修改赛事信息文件（位于 config/）

1. 运行主程序：

* python main.py

# 赛程文件

文件中包含以下四个表格：

1. Sequence：比赛顺序
2. Schedule：赛程
3. Update\_Info：最近一次运行更新信息
4. Summary：赛程汇总信息

# 扩展说明

* 在 wfmaster/scraper/ 或 wfmaster/cleaner/ 中添加新的抓取或清洗模块
* 在 main.py 中增加新的处理流程