

全球新聞項目

请为我设计并实现一个基于 Crawl4AI + GitHub Actions + Dify 的自动化全球新闻聚合系统。系统最终效果应类似 NewsMinimalist (<https://www.newsminimalist.com/>)，提供简洁、易读的全球重要新闻摘要。详细需求如下：

1. 系统架构概述

我需要能够自动执行以下流程的系统：

- 定时爬取全球多个新闻源的最新新闻（使用Crawl4AI）
- 处理并存储这些新闻数据
- 通过Dify工作流进行内容聚合、分析和摘要
- 生成一个极简风格的展示页面，类似NewsMinimalist的体验

2. 核心组件需求

2.1 Crawl4AI 爬虫实现

- 实现一个使用Crawl4AI库的Python程序，能够抓取至少5个主要国际新闻源
- 每个新闻条目需要提取：标题、内容摘要、发布时间、来源、URL
- 能够处理不同新闻网站的结构差异
- 将结果保存为JSON和Markdown两种格式
- 注重抓取重大国际新闻，而非娱乐或次要新闻

2.2 GitHub Actions 自动化流程

- 创建一个完整的GitHub Actions工作流配置
- 设置为每天固定时间（UTC+8 早上8:00）自动运行
- 执行爬虫程序并将结果上传到预定义的存储位置
- 触发后续的数据处理和Dify工作流

2.3 数据处理与存储服务

- 实现一个基于FastAPI的服务，用于接收和处理爬取的新闻数据
- 包含数据库设计（使用SQLite）和必要的API端点
- 设计安全的API认证机制
- 提供数据查询和过滤功能
- 实现新闻重要性排序算法，筛选出最重要的全球新闻

2.4 Dify集成

- 详细说明如何将新闻数据集成到Dify平台
- 提供完整的Dify工具配置指南
- 设计用于新闻聚合的Prompt模板
- 确保AI能够：
 - 提取每则新闻的核心事实
 - 用简洁明了的语言总结
 - 去除偏见和情感色彩
 - 按重要性排序

2.5 前端展示（类似NewsMinimalist风格）

- 实现一个极简风格的静态网页，特点包括：

- 黑白/极简配色方案
- 清晰的层次结构，新闻按重要性排序
- 每则新闻只显示简洁的标题和1-2句摘要
- 支持展开查看更多详情
- 顶部显示日期和简短介绍
- 响应式设计，适配移动端和桌面端
- 提供归档功能，可查看往日新闻
- 自动部署到GitHub Pages的完整配置

3. 技术规范

- 使用Python 3.9+实现爬虫和API服务
- 采用异步编程提高爬虫效率
- 前端使用HTML5/CSS3/JavaScript，可选择简单的框架如Alpine.js
- 代码需要有良好的错误处理和日志记录
- 所有配置应使用环境变量或配置文件
- 提供完整的安装和部署文档

4. NewsMinimalist风格特点要求

- **内容极简化**：每则新闻只保留最核心的事实和影响
- **去除干扰**：没有广告、图片少或没有、没有复杂交互
- **排版清晰**：合理的留白、易读的字体、清晰的层次结构
- **新闻筛选**：只展示真正重要的全球新闻，每日不超过10-15条
- **中立客观**：避免情感化和煽动性语言
- **更新时间明确**：清楚标示内容的生成/更新时间

5. 特殊要求

- 考虑不同新闻网站的反爬机制，实现必要的爬虫策略
- 需要对新闻内容进行简单的清洗和规范化
- 系统应具有可扩展性，便于后续添加更多新闻源
- 代码应清晰、模块化，并包含必要的注释
- 必须明确处理跨域问题，确保前端能正常访问API

由於項目比較大，考慮到一次性輸出整個項目代碼比較困難，我希望能夠分步驟進行。 请先提供完整的实现指南，然後我們再基於實現指南，我和你一起分步驟討論實現以下內容， 包括但不限于：

1. 主要Python文件（爬虫程序、API服务）
2. GitHub Actions workflow配置文件
3. 数据库schema设计
4. Dify集成指南与Prompt模板
5. 前端HTML/CSS/JS代码，体现NewsMinimalist风格
6. README.md和部署文档

非常感谢您的帮助！