高性能网站批量智能检查与访问系统

## **项目概述**

工作中需要定期检查50+甚至上百的网站的可访问性，传统手动方式耗时长达八分钟以上，同时，一次性打开所有网站，会导致浏览器崩溃，无法区分有效和无效链接，为提高效率并优化用户体验，开发了此智能网站处理系统。

该工具是一个高效的网站批量检测系统，用于自动化检查大量网站的可用性并自动在浏览器中打开可访问的网站。工具支持多线程并发检测、智能去重、实时状态显示和详细的报告生成功能。

我的职责

1.全栈设计并实现网站批量处理解决方案

2.设计核心算法优化性能瓶颈，设计更高可用的代码

3.实现错误处理机制和用户交互系统

4.设计并实现结果报告生成模块

技术逻辑和核心作用

1.多线程检测：

采用线性池(ThreadPoolExecutor)实现多线程并发检查，支持最多30个线程并发检查网站可用性

2.智能URL处理：

自动去除重复URL，自动补全URL协议（添加https://前缀）

3.实时状态监控：

实时显示检测进度，显示每个网站的检测结果和状态，预估剩余处理时间

4.自动浏览器操作：

成功检测后自动在浏览器中打开可用网站

可配置的打开延迟（防止浏览器崩溃）

5.详细报告生成：

自动生成包含成功/失败网站的详细报告

按时间戳命名报告文件

分类展示不同状态的网站

6.错误处理：

全面捕获各种网络错误（超时、DNS解析失败、证书错误等）

分类显示错误信息便于排查

我在运行代码出现的问题

1 .性能瓶颈

原始方案：顺序检查(50站点×10秒等于500秒)

优化方案：并行处理(50站点÷20线程×3秒)≈15秒

2.浏览器崩溃风险

新增网站打开间隔控制

失败网站自动过滤不打开

实时资源监控

3.错误诊断困难

生成六类错误自动分类

生成详细诊断报告

保留原始URL顺序便于定位