**项目题目及简短介绍爬取技术类博客：**

将技术类博客进行汇总分类，如 java，python，hadoop，spark等。java目录又分为算法，多线程，分布式等.

**项目所涉及的技术：**

WebMagic （国人作品，推荐）垂直、全栈式、模块化爬虫。更加适合抓取特定领域的信息。它包含了下载、调度、持久化、处理页面等模块。每一个模块你都可以自己去实现，也可以选择它已经帮你实现好的方案。这就有了很强的定制性。