**所选题目：**

爬取技术类博客，初步计划以CSDN为例。

**题目内容：**

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。

将技术博客进行网页内容抓取，并将其存入数据库，最后进行数据分析。

**项目所涉及的技术：**

初步计划以webmagic进行网页内容的抓取。

计划以MySQL数据库进行网页内容的存储。

计划以网页进行数据分析结果的展示。