3 系统分析

本系统框架主要面向网站管理员，使采用单页面应用技术的网站能够轻松地使用本框架达到搜索引擎优化的效果。本系统主要的功能是对搜索引擎采取的优化，除此之外，还允许用户进行一些自定义的优化，例如执行页面特定的JavaScript函数等。另外，系统还要向管理员提供监控页面和接口，让本系统方便管理和维护。

3.1 可行性分析

当前，使用vue，react，angular等技术开发单页面web应用已成为一种趋势，但是单页应用使用了大量的ajax技术，如何解决SEO就成了一个难题。

phantomjs本是用来进行自动化测试的一套工具，它可以在没有浏览器的情况下执行页面中的JavaScript代码（包括发送ajax请求）。使用它我们就像一个用户一样，执行页面的js代码，填充页面的内容，然后将解析后的内容发送给搜索引擎。这样我们就可以解决单页引用的搜索引擎优化难题了。

3.2 问题定义

本系统的核心模块是对页面中JavaScript的执行以及爬虫，这是基本的功能。何时进行对网站更新，保证与原网页的实时性又不对网站的访问太过频繁是系统优化的重点。本系统将利用mongo数据库配合爬虫和phantomjs优化抓取策略以及存储策略。

3.3 需求分析

3.3.1 功能需求

1. 系统可以对源网站自动进行抓取。
2. 如果提供sitemap，系统可以根据sitemap进行抓取。
3. 系统可以将保存下来的网页与新抓取的网页进行快速对比，判断网页是否已经更新。
4. 系统可以将网页提交给搜索引擎，被搜索引擎正常收录。
5. 系统提供方便，直观的管理界面，方便监控和维护各个服务。
6. 守护进程，确保程序的正常运行。

3.3.2性能需求

能够对新产生的网页进行快速收录。

对搜索引擎爬虫发送过来的请求，要保证及时响应（100ms以内）。

健壮性：对自身网站存在的错误，要能够处理，不能导致系统崩溃。

要尽量减少对源网站的请求次数，避免使原网站的负载压力过大。

易监控，提供监控页面，可以快速、直观地查看系统运行情况。

易维护，系统出现问题后，可以快速定位到问题并排查。

3.4 开发环境

硬件要求： 苹果笔记本电脑。

软件： zsh shell（配合oh-my-zsh），tmux终端工具，vim编辑器，git版本控制工具

开发语言：JavaScript，Python

3.5 本章小结

本章对本系统和框架进行了详细介绍。详细地分析了系统的功能，并且定义了性能要求。证明了该方案是可能的，为后面的开发增加了信心，奠定了基础。此外，还介绍了开发环境。

4.单页应用SEO优化系统的设计与实现

4.1 系统的结构设计

本系统为单页应用的SEO优化而设计，整个系统主要分为两部分。第一部分是爬虫引擎，主要负责从目标网站抓取页面内容，将爬取的任务信息存到数据库。第二部分是网页服务器部分，主要负责给搜索引擎返还真实的页面内容。

为了达到“部署快捷”的原则，整个系统运行在docker之上，以镜像的形式发布，做到了一键安装。总体的系统架构如图4-1所示。

