

Project-Spider

软件项目计划书

周林、游宇杰、徐遵杰

2017-3-13

目录

1 引言	1
1.1 编写目的	1
1.2 背景	1
1.3 定义	1
1.4 参考资料	2
1.5 标准、条约和约定	2
2 项目概述	2
2.1 项目目标	2
2.2 产品目标与范围	2
2.3 假设与约束	3
2.4 项目工作范围	3
2.5 应交付成果	3
2.5.1 需完成的软件	3
2.5.2 需提交用户的文档	4
2.5.3 须提交内部的文档	4
2.5.4 应当提供的服务	4
2.6 项目开发环境	4
2.7 项目验收方式与依据	4
3 项目团队组织	4
3.1 组织结构	4
3.2 人员分工	5
3.3 协作与沟通	5
3.3.1 项目团队内部协作	5
3.3.2 项目接口人员	5
3.3.3 项目团队外部沟通与协作模式	5
4 实施计划	5
4.1 风险评估及对策	5
4.2 工作流程	5
4.3 总体进度计划	6
4.4 项目控制计划	7
4.4.1 质量保证计划	7
4.4.2 进度控制计划	7
4.4.3 预算监控计划	7
4.4.4 配置管理计划	7
5 支持条件	7
5.1 内部支持	7
5.2 客户支持	8
5.3 外包（可选）	8
6 预算	8
6.1 人员成本	8
6.2 设备成本	8
6.3 其它经费预算	8
7 关键问题	8
8 专题计划要点	9

Project-Spider 项目计划书

根据《GB8567—88 计算机软件产品开发文件编制指南》中项目开发计划的要求，结合实际情况调整后的《项目计划书》内容如下：

1 引言

1.1 编写目的

为了保证项目团队按时保质地完成项目目标，便于项目团队成员更好地了解项目情况，使项目工作开展的各个过程合理有序，有必要以文件化的形式，把对于在项目生命周期内的工作任务范围、各项工作的任务分解、项目团队组织结构、各团队成员的工作责任、团队内外沟通协作方式、开发进度、经费预算、项目内外环境条件、风险对策等内容以书面的方式描述出来，作为项目团队成员以及项目干系人之间的共识与约定，项目生命周期内所有项目活动的行动基础，项目团队开展和检测项目工作的依据。

本项目开发计划面向项目组全体成员。

1.2 背景

(1) 项目的名称

Project-Spider 项目开发。

(2) 项目的委托单位

四川大学 2015 级软件学院项目开发小组。

(3) 项目的用户（单位）：

插画家、专业的漫画师或者漫画爱好者。

(4) 项目的任务提出者

周林。

(5) 项目的主要承担部门

项目的承担部门主要有：开发部门、测试部门、集成部门等三个部门。

(6) 项目建设背景

从业务环境来看本项目能够使广大喜爱漫画的人员能够更简单的寻找到当前热点漫画和高点击量漫画，而且本爬虫项目还能够完善相关网站的搜索功能，让使用者能够更加方便的获取自己喜爱类型的相关漫画。

(7) 软件系统与其他系统的关系

本系统依赖于 node.js 平台、腾讯云、Mongodb 数据库等服务

(8) 软件系统与机构的关系

由于本项目是基于爬虫的一类项目，因此，对于目标网页的原式数据和相关程序不能篡改和破坏，而且对于我们获取的目标网页的数据我们要严格保护，在没经过对方同意的情况下坚决不用于商业行为。

本项目不需要外包，且不用经过专业测试机构的相关测试。

1.3 定义

网络爬虫：

1.定义：网络爬虫（又被称为网页蜘蛛，网络机器人，在 FOAF 社区中间，更经常的称为网页追逐者），是一种按照一定的规则，自动地抓取万维网信息的程序或者脚本。

2.英文表示：web crawler

1.4 参考资料

【GB8567—88 计算机软件产品开发文件编制指南】

1.5 标准、条约和约定

【GB8567—88 计算机软件产品开发文件编制指南】

2 项目概述

2.1 项目目标

详细模块任务如下表所示：

主要任务描述：

	模块	功能点	估计工作量
前台	Index Page	导航、页头及整体布局等	4(时/人)
	Main Page	主页面设计	4(时/人)
	Picture Page	插画展示页面、详细展示	4(时/人)
	Download Page	下载页面、下载功能	8(时/人)
	Search Page	检索页面、数据分析决策功能	8(时/人)
	Crawler Page	爬虫功能、获取数据	4(时/人)
后台	Crawler Management	爬虫管理	4(时/人)
	Request Management	Cookie、请求头的获取	4(时/人)
	File Management	文件模块，记录数据	5(时/人)
	Data Management	数据分析模块	8(时/人)
	Database Management	数据库模块	8(时/人)
	Route Management	路由管理	4(时/人)
	Test Management	测试模块	10(时/人)
	Middleware Management	中间件的设计	8(时/人)
	About Management	主页面后台编辑、管理	4(时/人)
	News Management	插画分类的后台编辑、管理	5(时/人)
	Search Management	检索功能后台编辑、管理	8(时/人)
	Log Management	日志的收集和管理	4(时/人)
	User Management	收集用户信息	8(时/人)
	Download Management	下载后台编辑、管理	4(时/人)

2.2 产品目标与范围

产品目标：该项目产品是网站，通过对爬到的数据进行的分析为对插画、原画有浓厚兴趣的人以及热爱美术的人提供检索、下载收集的便利。比如，可以为感兴趣的人员提供当前网站最火热的画师，或者提供又需要的问题答案。

产品范围：该项目产品是依据爬虫技术以及数据分析，通过爬虫技术获取数据，进而对收集到的数据进行分类、计算，从而为用户提供检索便利，同时能为用户做决策提供帮助。

2.3 假设与约束

假设：

项目进度安排：

编号	任务阶段	开始时间	结束时间
1	需求分析	2017/03/05	2017/04/04
2	系统开发	2017/04/04	2017/04/20
3	前后台联调、功能测试	2017/04/21	2017/05/02
4	部署测试、Bug Fix	2017/05/03	2017/05/15
5	Release	2017/05/16	2017/05/23

约束：若数据来源网站有反爬虫机制，更换数据来源网站。

2.4 项目工作范围

产品范围界定：整个系统分为站点展示部分和后台管理部分，每个模块的开发任务都包含了界面设计、数据库设计相关环节。

前台：导航、页头及整体布局等、主页面设计、插画展示页面、详细展示、下载页面、下载功能、检索页面、数据分析决策功能、爬虫功能、获取数据。

后台：爬虫算法、数据库管理、主页面后台编辑、管理；插画分类的后台编辑、管理；检索功能后台编辑、管理；检索功能后台编辑、管理；下载后台编辑、管理。

工作范围界定：

工作范围	时长/天
需求分析	7
系统开发	30
前后台联调、功能测试	3
部署测试、Bug Fix	2
Release	2

2.5 应交付成果

2.5.1 需完成的软件

项目名称：project-Spider

编程语言：JavaScript

数据库：Mongodb

开发平台：Node.js

存储程序的媒体形式：web

2.5.2 需提交用户的文档

列出需要移交给用户的每种文档的名称、内容要点及存储形式，如需求规格说明书、帮助手册等。此处需要移交用户的文档可参考合同中的规定。

2.5.3 须提交内部的文档

无

2.5.4 应当提供的服务

无

2.6 项目开发环境

操作系统：Windows10

开发工具：Sublime Text 3

开发平台：Node.js v7.1.0

数据库系统：Mongodb v3.4.1

配置管理工具：Git 2.10.2.windows.1

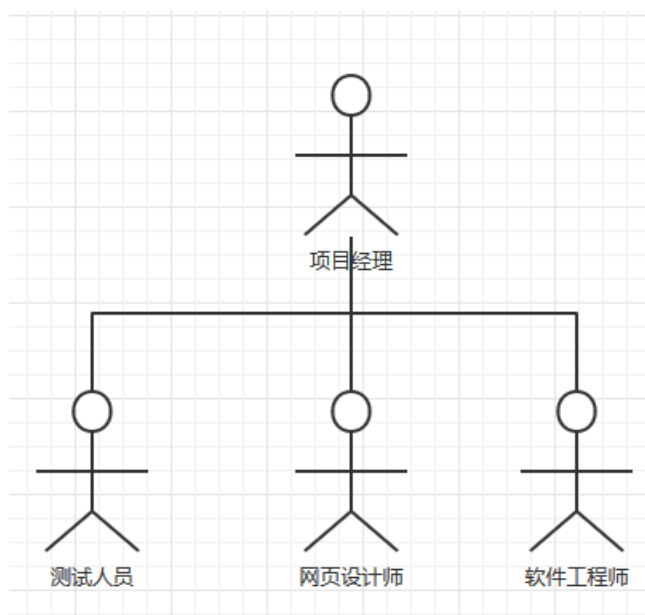
2.7 项目验收方式与依据

项目验收分为内部验收和外部验收两种方式。在项目开发完成后，首先进行内部验收，由测试人员根据用户需求和项目目标进行验收。项目在通过内部验收后，交给客户进行验收，验收的主要依据为需求规格说明书。

3 项目团队组织

3.1 组织结构

为了完成项目的开发，人员分工由项目经理、测试人员、网页设计师、软件工程师构成



3.2 人员分工

姓名	技术水平	角色	工作描述
周林	初级软件工程师	项目经理，软件工程师	负责项目的分析，框架的搭建，项目开发进度的跟踪；软件编码工作
游宇杰	初级软件工程师	软件工程师，网页设计师	负责网页风格的确定，负责网站图片的设计；承担一定的编码工作
徐遵杰	初级软件工程师	软件工程师，测试人员	对软件进行测试，编写软件测试文档；承担一定的编码工作

3.3 协作与沟通

3.3.1 项目团队内部协作

为了保证项目开发的顺利，特要求如下沟通内容：

1. 每天晚上 9—9:30 点成员间口头沟通交流
2. 每周六晚上 10 点前提交周报告，并制作格式模板
3. 每周六上午十点召开周例会，报告项目的进展以及存在的问题。会后做好记录

3.3.2 项目接口人员

姓名	职责	联系方式	沟通方式	协作者
周林	项目管理，系统构架	手机 15208171708	短信及电话联系	游宇杰，徐遵杰
游宇杰	编码，前端设计	手机 15828108644	短信及电话联系	周林
徐遵杰	编码，测试，文档	手机 18692077799	短信及电话联系	周林

3.3.3 项目团队外部沟通与协作模式

无

4 实施计划

4.1 风险评估及对策

1、网站反爬虫的风险。若原网站修改反爬虫策略，可能导致爬虫无法获取数据，或者无法实现原先预定的需求。

处理对策：更换网站进行爬取

2、技术开发风险。本网站的开发主要用到了 javascript 作为前台开发和 node.js 后台相结合的结构体系，并设计到了 Mongodb 的数据库部分。组件和构件技术都是为了提高网站的可靠性和网站的友好性而采

用的技术手段。从技术成熟度上说不存在风险，但为了实现良好的网站构架和用户体验度，与传统开发方法比较，有相当的多的额外工作需要做，这会给项目工期带来一定的风险，但影响较小。

处理对策：使用其他技术手段实现，如：使用 Python 或者 PHP 实现后台

3、用户体验风险。在项目开发的过程中可能会因为了一种功能进而想到继续开发更多的关联功能，最终导致产品验收不能顺利地完完成，也有可能是由于功能的多次变更，导致原先开发的项目需要被重做，从而导致开发的延期提交。

处理对策：减缓

4、原网站崩溃的风险。在项目开发过程中，原网站可能由于天灾、人祸、或者政策原因禁止或者限制访问，导致无法实现需求

处理对策：更换网站进行爬取

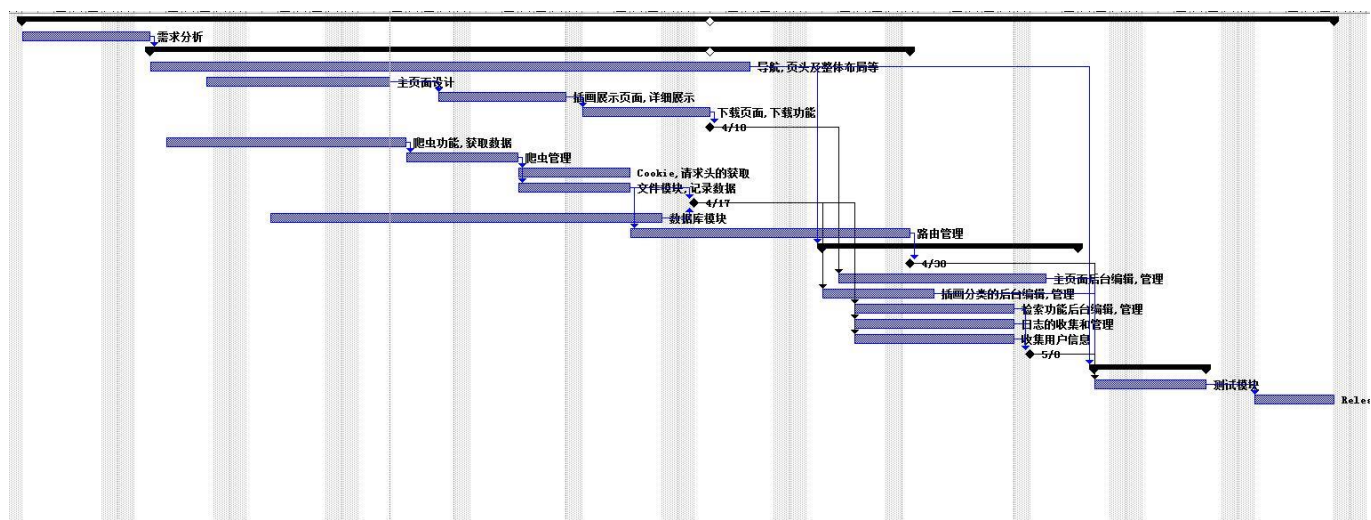
4.2 工作流程

本项目采用增量模型，遵守基本的 需求分析——设计——建模——编码——部署及维护 软件开发流程。



模块	功能点	负责人	协作者
Index Page	导航、页头及整体布局等	游宇杰	周林
Main Page	主页面设计	游宇杰	徐遵杰
Picture Page	插画展示页面、详细展示	游宇杰	徐遵杰
Download Page	下载页面、下载功能	游宇杰	周林
Search Page	检索页面、数据分析决策功能	游宇杰	周林
Crawler Page	爬虫功能、获取数据	徐遵杰	周林
Crawler Management	爬虫管理	周林	游宇杰
Request Management	Cookie、请求头的获取	周林	游宇杰
File Management	文件模块，记录数据	周林	游宇杰
Data Management	数据分析模块	徐遵杰	游宇杰
Database Management	数据库模块	徐遵杰	游宇杰
Route Management	路由管理	周林	徐遵杰
Test Management	测试模块	徐遵杰	周林
Middleware Management	中间件的设计	周林	徐遵杰
About Management	主页面后台编辑、管理	徐遵杰	周林
News Management	插画分类的后台编辑、管理	徐遵杰	周林
Search Management	检索功能后台编辑、管理	周林	徐遵杰
Log Management	日志的收集和管理	徐遵杰	周林
User Management	收集用户信息	游宇杰	徐遵杰
Download Management	下载后台编辑、管理	徐遵杰	周林

4.3 总体进度计划



4.4 项目控制计划

4.4.1 质量保证计划

无

4.4.2 进度控制计划

无

4.4.3 预算监控计划

无

4.4.4 配置管理计划

无

5 支持条件

说明为了支持本项目的完成所需要的各种条件和设施。

5.1 内部支持

◎硬件:

服务器 CPU: 1 核 1Mbps 或更高

内存: 1GB 或更高

硬盘: 50G 或更高

网络: 至少一台服务器

使用 TCP/IP 协议的网络服务

◎软件:

服务器操作系统: centOS6.5 或以上

开发平台操作系统: Windows10

后端平台: Node.js
开发工具: Sublime Text 3
数据库系统: Mongodb v3.4.1
配置管理工具: Git 2.10.2.windows.1

5.2 客户支持

无

5.3 外包（可选）

无

6 预算

6.1 人员成本

无

6.2 设备成本

无

6.3 其它经费预算

无

7 关键问题

1、网站进度风险：即网站是否能够按工期的要求完成。网站构建是一个大型的开发项目，涉及到网站的前端和后台开发，也涉及到整个网站的工作运行流程，而该项目的上线时间是不能改变的，所以项目进度是一个签字的风险。如果，不能按时完成，则大量需要文档进行协调的工作时，致使开发进度越来越慢。网站的开发不同于其他的工程，在不同的工程阶段，需要的人员不同，需要配合的方面也不同，所有这些都需行之有效的软件管理的保证。

2、网站需求风险：开发是以用户的需求开始，在大多数情况下，用户需求要靠网站开发方诱导才能保证需求的完整，再以书面的形式形成《用户需求》这一重要的文档。需求分析更多的是开发方确认需求的可行性和一致性的过程，在此阶段需要和用户进行广泛的交流和确认。需求和需求分析的任何疏漏造成的损失会在软件系统的后续阶段被一级一级地放大，因此本阶段的风险很大。

3、技术开发风险：本网站的开发主要用到了 javascript 作为前台开发和 node.js 后台相结合的结构体系，并设计到了 Mongodb 的数据库部分。组件和构件技术都是为了提高网站的可靠性和网站的友好性而采用的技术手段。从技术成熟度上说不存在风险，但为了实现良好的网站构架和客户体验度，与传统开发方法比较，有相当的多的额外工作需要做，这会给项目工期带来一定的风险，但影响较小。

4、网站质量体系风险：任何网站管理忽略软件质量监督环节都将对网站的生产构成巨大的风险。网站的质量体系集中在开发中的测试阶段和最后的维护阶段。随着网站系统的不断扩大，需要不断的对网站的体系进行维护和改进，如果监督不到位，技术支持体系将会无效运转，造成损失。

5、网站设计风险：本身的风险主要来自于系统分析人员。分析人员在设计网站结构时过于定制，网站的友好型和实用性较弱，会给后期维护带来巨大的负担，和维护成本的激增。对用户来说网站的点击比例会有明显的折扣，甚至造成访问量急剧减少。反之，网站结构的过于灵活和通用，必然引起网站开发的难度增加，网站的复杂度会上升，这又会在实现和测试阶段带来风险，网站的稳定性也会受到影响。所以，

该风险巨大。

8 专题计划要点