

# MovieLook 电影数据可视化分析系统

## 软件需求规格说明书

名字	学号	贡献描述
姜昊	SZ160110224	服务器系统的调配、服务器环境的配置、数据库管理，实现数据库中数据与网页的连接使用
陈润泽	SZ160110214	制作数据可视化，生成报表界面
但远琦	SZ160110231	爬取功能实现，电影数据数据库功能实现，用户数据的数据库功能实现
赵昕玥	SZ160110216	制作网站首页和登录界面
许浚伟	SZ160110205	实现数据库-浏览器的对接（具体为数据的预处理和传递）

## 文档变更历史记录

[illegible]

# 目录

1. 引言.....	5
1.1 编写目的.....	5
1.2 读者对象.....	5
1.3 软件项目概述.....	5
1.4 文档概述.....	5
1.5 术语定义与说明.....	6
1.6 参考资料.....	6
2. 软件的一般性描述.....	7
2.1 软件产品与其环境之间的关系.....	7
2.2 限制与约束.....	8
2.3 假设与前提条件.....	9
2.3.1 安全性.....	9
2.3.2 可靠性.....	10
3. 功能需求描述.....	11
3.1 系统的划分.....	11
3.2 用户子系统的功能.....	11
3.2.1 子系统用例图.....	11
3.2.2 用户注册用例描述.....	11
3.2.3 用户登录用例描述.....	13
3.2.4 可视化图表生成用例描述.....	14
3.2.5 报表生成用例描述.....	16
3.3 信息管理子系统的功能.....	17
3.3.1 子系统的用例图.....	17
3.3.2 用户信息修改的用例描述.....	17
3.3.3 删除电影信息的用例描述.....	18
4. 其它需求描述.....	19
4.1 性能要求.....	19
4.1.1 内存要求.....	19
4.1.1 响应速度要求.....	19

4.1.2 查询结果要求.....	19
4.2 设计约束.....	20
4.2.1 部署方式.....	20
4.2.2 开发工具.....	20
4.2.3 测试工具.....	20
4.2.4 运行环境.....	21
4.3 界面要求.....	22
4.3.1 清楚一致的设计 .....	22
4.3.2 简洁和美观.....	22
4.3.3 用户操作的安全性.....	22
4.4 进度要求.....	22
4.5 交付要求.....	23
4.6 验收要求.....	23
5. 软件原型.....	24
6. 备注.....	24

# 1. 引言

## 1.1 编写目的

本文档旨在列出一些本次软件系统开发的基本要求与说明，主要说明了电影可视化系统的基本功能、模块划分，以便于各模块开发人员能够更好地了解各自模块的实现需求与详细功能，能够让开发人员可以快速了解该系统的基本情况。

## 1.2 读者对象

本文档主要面向各模块的开发人员编写。

## 1.3 软件项目概述

项目名称：电影数据可视化分析系统

开发人员：姜昊、陈润泽、赵昕玥、但远琦、许浚伟

面向用户：广大电影爱好者，以及电影数据分析师等可能使用到本网站作数据分析的电影行业的从业者。

功能概述：本软件可对部分电影信息网站的数据进行聚合与存储，并且可以做到存储数据的定期更新。注册用户通过关键字查询的方式可获取如电影题材、票房、演员等相关的电影信息，查询所获得的信息可以根据用户的需求以报表的形式或者以动态图表的方式向用户呈现。

## 1.4 文档概述

本文档主要分为五个板块：引言部分、软件一般性描述、软件功能性需求描述、软件的其他需求描述和软件原型组成。引言部分将给出一些最基本的文档介绍以及名词解释，软件一般性描述主要描述；软件功能性需求部分则是从每个功能角度阐释每个功能所满足的需求，并通过 UML 图给出每个功能的详细描述；其他需求描述部分则是描述一些非功能性需求，如可用性、安全性、软件的各种

设计约束，并给出项目的进度、交付和验收等要求；最后的软件原型部分展示了软件的系统早期版本的重要特性，如界面、功能、性能等。

## 1.5 术语定义与说明

下面给出一些本文档所涉及的专门术语、概念、关键词缩写及其他需要解释的内容。

注册：在网站中填写用户个人信息，并将用户信息存储到用户数据库中。

游客：未在本网站中注册过的用户。

用户数据库：存储用户信息的数据库。

电影信息网页：记录有电影的上映时间、票房、参映演员、观众评分等电影信息的 html 网页，网页来源有豆瓣、猫眼、IMDb 等网站。

电影信息数据库：存储电影信息的数据库。

关键字：电影信息中有一些描述电影性质或具体内容的属性，如电影类型、上映时间、参映演员等，用户在查询的时候可以依据这些属性在电影信息数据库中查找对应的电影数据记录。

电影数据图表：依据电影信息数据库返回的数据记录，对页面进行一些处理所生成的动态数据图表。

报表：指用户能够保存导出所需电影可视化图像及数据的文档。

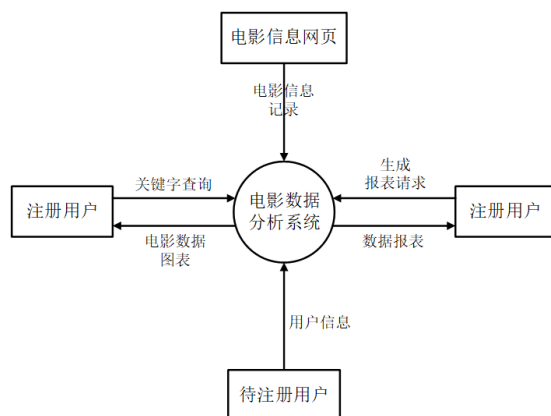
## 1.6 参考资料

[1]. Roger S. Pressman. Bruce R. Maxim 软件工程 实践者的研究方法[M]. 北京市西城区：机械工业出版社，2017.1

## 2. 软件的一般性描述

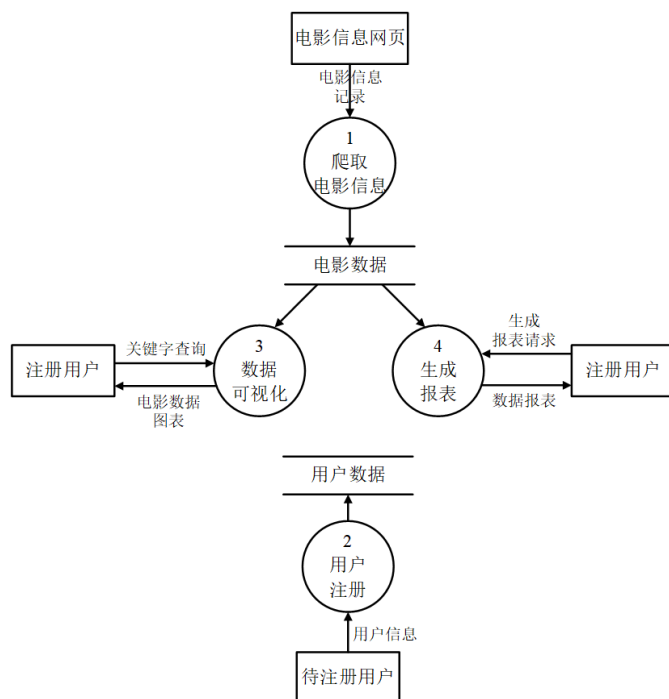
### 2.1 软件产品与其环境之间的关系

对于本系统而言的外部实体可大致分为 3 类：电影信息网页、待注册用户和注册用户，本系统和三者的交互可用如下数据流图表示：



其中，电影信息网页负责向电影数据分析系统提供最基本的电影信息数据，待注册用户可通过电影数据分析系统注册成为注册用户，注册用户可通过电影分析系统查看电影数据的可视化图标以及生成数据报表。这里需要注意的是，非本网站的注册用户不能享受生成可视化图表和生成数据报表的功能。

进一步地，系统可以划分为 4 个具体的子系统，以数据流图的形式表示如下：



其中每一个加工可视为一个子系统，因此本系统可分为电影信息爬取子系统、注册登陆子系统、数据可视化子系统和数据报表生成子系统。电影信息爬取子系统负责从网页上爬取电影信息，并存储到电影信息数据库中；注册登录子系统负责对用户进行权限管理；数据可视化子系统负责接受电影信息数据库返回的数据，并生成电影数据图表；数据报表生成子系统负责接受电影信息数据库返回的数据，并生成电影数据报表。

## 2.2 限制与约束

### (1) 服务器操作系统

为满足项目大型的应用，必须采用 Unix + SQL。

### (2) 开发工具的限制

本系统采 B/S 体系结构，客户端使用常用的浏览器访问网页，而 Webstorm 是开发网页的首选工具。对于 B/S，采用 Nginx + PHP 进行开发。

### (3) 合法性

该软件产品的业务处理必须符合国家的相关法律和法规。例如，用户名密码不可以在数据库中明文储存。

## 2.2 限制与约束修改为：

### (1) 开发环境限制

- 兼容性：选取 Linux x64 作为服务端部署的硬件要求；
- 部署框架：考虑到应用提供服务的规模大小，以及服务端搭建的便捷性，采用“浏览器 + WEB 服务器 + DB”作为本应用服务端的部署框架；
- 网络通信要求：支持 TCP/IP 协议族，支持 HTTP 协议；
- 与网络的接口：采用 Nginx 作为 Web 服务的代理服务器，并提供负载均衡等简单的基本网络通信服务；
- 开发工具：详见 [4.2.2 开发工具](#)。

### (2) 运行环境限制

详见 [4.2.3 运行环境](#)。

### (3) 测试环境限制



本应用采用的测试环境如下：

Server 端	
CPU 型号	Intel(R) Xeon(R) CPU E5-2682 v4 @ 2.50GHz
CPU 核数	1
系统位数	64
操作系统	CentOS Linux release 7.4.1708
内核版本	kernel-3.10.0-693.2.2.el7.x86_64
内存大小	1883496 kB
硬盘空间	42.9 GB
服务器代理	Nginx/1.15.6

Client 端	
操作系统	Windows 7/10 x64
内存大小	≥2GB
硬盘空间	≥50 GB
浏览器类型	Edge/Chrome

#### (4) 合法性限制

应用的业务处理符合国家的相关法律和法规。如：用户名密码不可以在数据库中明文储存。

## 2.3 假设与前提条件

由于在构建服务端的时候只考虑到了基本业务流情况和少量的异常情况，因此需要对服务端的安全性和可靠性做出如下假设：

### 2.3.1 安全性

- (1) 假设服务器上的数据库不会遭受外部攻击和修改。
- (2) 假设不会发生从浏览器页面插入恶意代码的情况。

### 2.3.2 可靠性

- (1) 假设服务器端与客户端的网络通信能够一直保持顺畅，不会出现由于网络无法访问引起的服务超时情况，或者网络拥塞所导致的服务延迟现象。
- (2) 假设服务器不会出现宕机或进程崩溃的情况。

### 3. 功能需求描述

#### 3.1 系统的划分

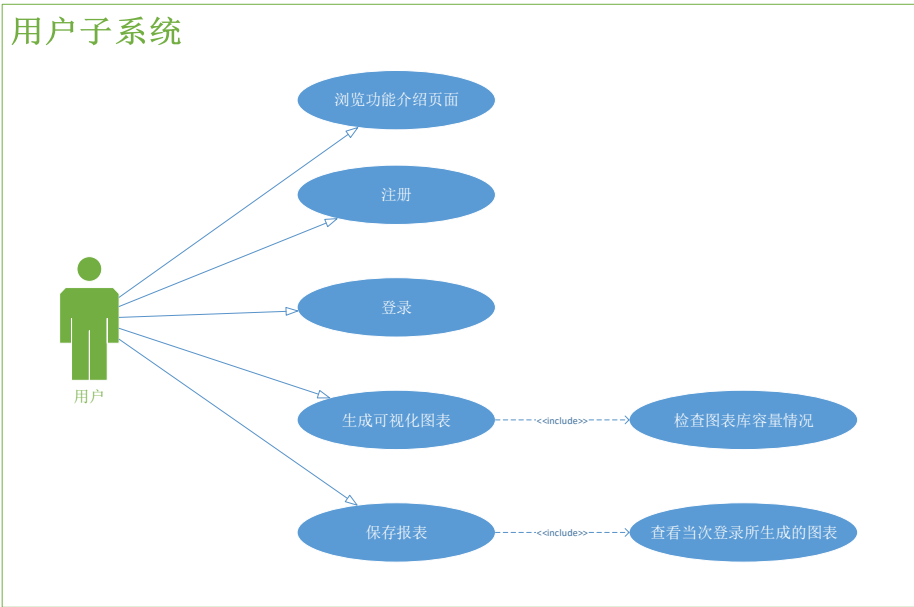
系统分为用户子系统和信息管理子系统两部分。

用户子系统面向用户，包括的功能有浏览、注册、登录、生成可视化图表、保存报表。

信息管理子系统面向管理员，包括的功能有用户和电影信息的维护。

#### 3.2 用户子系统的功能

##### 3.2.1 子系统用例图



##### 3.2.2 用户注册用例描述

用例名称	用户注册
用例描述	本用例允许用户进行注册操作
参与者与关注点	用户, 用户希望注册账号以便得到更完整的网站功能体验
前置条件	用例开始前，用户产生注册需求

主成功场景	<ul style="list-style-type: none"> <li>• 用户点击注册，产生注册需求</li> <li>• <del>用户输入要注册的用户名、密码、注册邮箱</del></li> <li>• 用户输入要注册的用户名、密码、注册邮箱</li> <li>• 系统对用户信息进行识别</li> <li>• 系统检测用户填写的信息是否符合要求</li> <li>• <del>系统将用户信息保存到数据库，增加一个用户记录</del></li> <li>• 系统将用户信息保存到数据库，增加一个未激活用户记录</li> <li>• 用户到注册邮箱内点击网站发送链接，激活注册账号</li> <li>• 系统在数据库将账号激活</li> <li>• 用户注册成功</li> </ul>
后置条件	<p>如果用例执行成功，网站更新相关数据，增加一个用户记录；否则，网站状态不变</p>
扩展	<ul style="list-style-type: none"> <li>• 用户注册用户名已存在 在主流程中，若注册用户名已存在，则网站提示用户名已被使用，用例结束；</li> <li>• 用户两次输入密码不一致 在主流程中，若用户输入密码时两次输入不相同，则网站提示用户密码输入不一致，用例结束</li> <li>• 用户输入的密码长度不合格 在主流程中，若用户输入的密码长度过长或过短，则系统提示用户密码不符合规定，提示用户进行修改，用例结束</li> <li>• 用户输入的邮箱格式不正确</li> </ul>

	<p>在主流程中,若用户输入的邮箱格式不符合要求,则系统提示用户邮箱格式不正确,提示用户进行修改,用例结束</p> <ul style="list-style-type: none"> <li>用户未进行邮箱验证激活</li> </ul> <p>在主流程中,若用户未到填写邮箱内点击激活链接,则系统将用户账号判定为未激活账号,用户无法正常进行登录,用例结束</p>
特殊需求	已注册用户也具备注册新账号功能
发生频率	可能会多次发生
未解决问题	用户注册时输入的密码未进行格式组合限制,用户密码强度没有保障

### 3.2.3 用户登录用例描述

用例名称	用户登录
用例描述	本用例允许已注册用户进行账号登录
参与者与关注点	用户, 用户希望快速准确登录, 流程简易
前置条件	用例开始前, 用户已注册账户
主成功场景	<ul style="list-style-type: none"> <li>用户输入账号密码信息并填写对应的验证码</li> <li>系统对用户信息进行检测</li> <li>系统匹配到该用户所填信息, 提示登录成功</li> <li>用户成功登录</li> <li>用户名显示在网站右上角, 提示处于登录状态</li> <li>用户进行网站完整功能使用</li> </ul>
后置条件	无

扩展	<ul style="list-style-type: none"> <li>用户输入密码错误 在主流程中，若用户登录时输入密码错误，网站弹出密码输入错误提示，登录失败，提示用户重新进行密码输入，用例结束。</li> <li>用户登录时验证码输入错误 在主流程中，若用户登录时输入验证码错误，系统提示用户可重新输入验证码或点击图片更换验证码，用例结束。</li> <li>系统未匹配到对应用户信息 在主流程中，若用户输入的用户名并不存在，系统提示该用户名不存在，提示用户重新输入，用例结束。</li> <li>系统匹配到未激活用户 在主流程中，若用户注册时未到注册邮箱进行账号验证激活，则账号处于未激活状态，系统提示用户先进行激活操作，用例结束。</li> </ul>
特殊需求	未登录用户(游客)仅可访问网站首页进行概览
发生频率	可能会多次发生
未解决问题	无

### 3.2.4 可视化图表生成用例描述

用例名称	用户生成可视化图表
用例描述	系统根据用户输入的图表类型、时间段、查看范围等信息，为用户生成可视化图表
参与者与关注点	用户，用户希望快速生成可视化图表，流程简易，图表美观

前置条件	用例开始前，用户已登录账户
主成功场景	<ol style="list-style-type: none"> <li>1.用户点击页面的可视化功能超链接，进入使用页面，开始一张新图表的生成</li> <li>2.用户填写图表蕴含信息</li> <li>3.判断用户是否正在生成一张其他的图表</li> <li>4.检查账户图表库的使用容量是否达到上限</li> <li>5.系统检测用户填写的信息是否符合要求</li> <li>6.系统为用户生成可视化图表，并展示在页面上</li> <li>7.系统更新用户图表库相关信息</li> <li>8.生成可视化图表成功</li> </ol>
后置条件	无
扩展	<ol style="list-style-type: none"> <li>1. 正在生成其他可视化图表，不能同时生成多张图表 若检测到租户是否已经在生成其他图表过程中，检测到该用户正在生成其他可视化图表，反馈租户，结束可视化流程</li> <li>2. 用户的图表库容量已满 若检测用户图表库容量过程中，检测到用户的图表库容量已达上限，反馈用户，结束可视化流程</li> <li>3. 生成图表信息填写错误 若系统检测用户输入的可视化图表信息时，发现信息不合理，反馈用户，结束可视化流程</li> <li>4. 可视化图表生成/展示失败 若在生成/展示可视化图表的过程中，生成失败，反馈用户，结束可视化流程</li> </ol>
特殊需求	在页面上展示的图表要求具有交互功能
发生频率	可能会多次发生

未解决问题	<ol style="list-style-type: none"> <li>1. 不能提示正确的输入信息范围, 用户出现多次输入不合理信息</li> <li>2. 图表库容量更新不及时, 不能合理控制用户生成的图表数量</li> </ol>
-------	---

### 3.2.5 报表生成用例描述

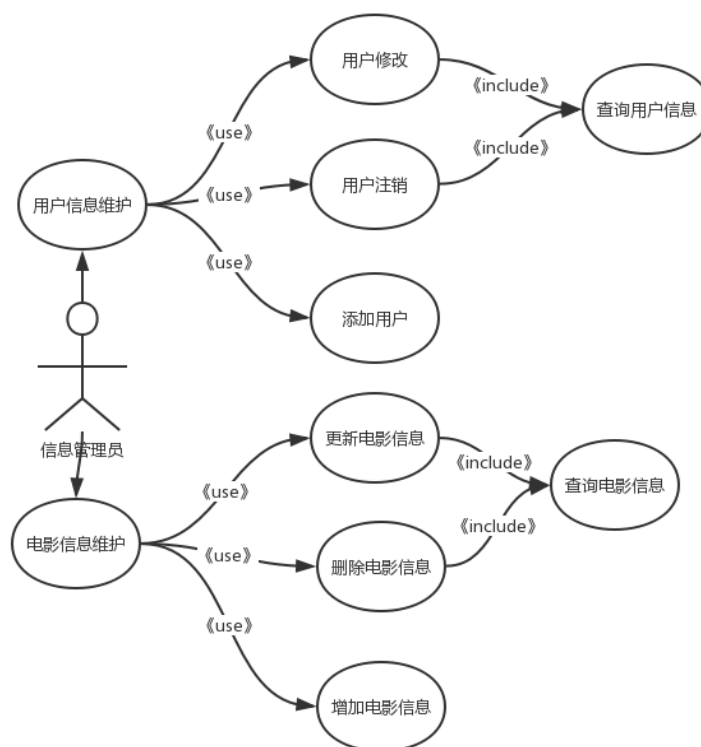
用例名称	用户生成报表
用例描述	用户在已生成的可视化图表中选择需要导出的图表, 导出到报表中
参与者与关注点	用户, 用户希望快速准确导出报表, 流程简易, 格式美观
前置条件	用例开始前, 用户已登录账户, 并在该次登录期间生成一些图表
主成功场景	<ol style="list-style-type: none"> <li>1. 用户点击页面的生成报表超链接, 进入报表生成页面, 开始生成报表流程</li> <li>2. 用户勾选要生成的报表, 点击导出按钮</li> <li>3. 判断用户勾选的图表个数是否大于等于 1 个</li> <li>4. 系统为用户生成报表, 并下载到本地默认目录</li> <li>5. 生成报表成功</li> </ol>
后置条件	无
扩展	<ol style="list-style-type: none"> <li>1. 导出的图表少于 1 个 若系统检测用户勾选的图表个数时, 用户勾选的可视化图表少于 1 个, 反馈用户未选择任何内容, 结束报表生成流程</li> <li>2. 报表生成失败</li> </ol>



	若在生成报表的过程中,生成失败,反馈用户,结束生成报表流程
特殊需求	生成的报表文件格式要美观
发生频率	可能会多次发生
未解决问题	不能提示图表选择情况,用户出现多次选择不合理的情况

### 3.3 信息管理子系统的功能

#### 3.3.1 子系统的用例图



#### 3.3.2 用户信息修改的用例描述

用例名称	用户信息修改
用例描述	信息管理员登录系统对用户信息进行修改
参与者与关注点	信息管理员, 管理员希望准确流畅的修改用户

	信息，过程简便且操作精确
前置条件	管理员信息登陆验证通过
主成功场景	<ol style="list-style-type: none"> <li>1. 信息管理员通过客户端登陆</li> <li>2. 通过验证后输入要修改的用户信息</li> <li>3. 系统检查输入是否正确</li> <li>4. 系统查询是否存在该用户信息</li> <li>5. 对当前用户信息进行修改</li> <li>6. 用户信息更新并保存在用户数据库</li> <li>7. 系统更新数据</li> </ol>
后置条件	用户信息发生变化，用户数据库更新数据
扩展	<ol style="list-style-type: none"> <li>1.管理员登陆验证未通过 系统提示重新验证</li> <li>2.系统检测到输入错误 系统提示输入错误并让管理员重新输入</li> <li>3.经系统查询无该用户信息 系统提示无该用户信息，选择重新输入更改的用户信息或者退出系统</li> </ol>
特殊需求	无
发生频率	可能会多次发生
未解决问题	暂无

### 3.3.3 删除电影信息的用例描述

用例名称	删除电影信息
用例描述	信息管理员对已存在的电影信息进行删除
参与者与关注点	信息管理员，管理员希望准确流畅的删除电影信息，过程简便且操作精确
前置条件	管理员信息登陆验证通过
主成功场景	<ol style="list-style-type: none"> <li>1. 信息管理员通过客户端登录</li> <li>2. 通过验证后</li> <li>3. 通过验证后输入要修改的用户信息</li> <li>4. 系统检查输入是否正确</li> <li>5. 将该条信息从数据库中删除</li> </ol>

	6. 电影数据库更新数据
后置条件	删除该条电影信息，电影数据库更新
扩展	1.管理员登陆验证未通过 系统提示重新验证 2.系统检测到输入错误 系统提示输入错误并让管理员重新输入 3.无该条电影信息 系统提示无要删除电影信息，选择删除其他信息或者退出系统
特殊需求	无
发生频率	可能会多次发生
未解决问题	电影信息查询功能不完善，在电影数据量较大的情况下查询速度较慢

## 4. 其它需求描述

### 4.1 性能要求

#### ~~4.1.1 内存要求~~

~~服务器端需要 1GB 以上的运行内存。~~

#### 4.1.1 响应速度要求

单关键词查询      响应时间 < 1 秒

复合关键词查询    响应时间 < 5 秒。

#### 4.1.2 查询结果要求

关键字查询的无二义性：针对每个关键字或每种关键字组合，只有唯一的查询返回结果。

结果的完整性：返回的是完整的结果，不存在缺失部分数据的情况。

~~（待补充？）~~

其他可讨论的点包括：

- 吞吐量：在一定时间内，系统可完成多少任务
- 并发控制：在一定时间内，系统允许处理多少个并发访问请求
- 安全性
- 可靠性

## 4.2 设计约束

### 4.2.1 部署方式

浏览器 + WEB 服务器 + DB

### 4.2.2 开发工具

网页数据爬虫：Python3.6

后端数据传递：PHP 7+

数据库组织：MySQL5.7

WEB 页面组织：JetBrains WebStorm2018.2.3 edge/chrome/ie 浏览器

#### 4.2.2 开发工具 修改为：

网页数据爬虫：JetBrains PyCharm Professional 2018.2 + Python3.6

后端业务开发：Visual Studio Code 1.30 + PHP 7

数据库组织：MySQL 5.5

WEB 页面开发：JetBrains WebStorm2018.2.3

### 4.2.3 测试工具

爬虫测试：JetBrains PyCharm Professional 2018.2+ Python3.6

数据库测试：MySQL Workbench / Navicat + MySQL 5.5

后端业务测试：Chrome 浏览器 + Visual Studio Code 1.30 + PHP 7

WEB 页面测试：Edge/Chrome/IE 浏览器

本地网站测试：Wampserver / phpMyAdmin

## 4.2.4 运行环境

Server 端	
CPU 型号	Intel(R) Xeon(R) CPU E5-2682 v4 @ 2.50GHz
CPU 核数	1
系统位数	64
操作系统	CentOS Linux release 7.4.1708
内核版本	kernel-3.10.0-693.2.2.el7.x86_64
内存大小	1883496 kB
硬盘空间	42.9 GB

**Server 端运行环境修改为：**

Server 端	
CPU 核数	建议选择可以独占一核的 CPU 使用
系统位数	建议 64 位以上系统
操作系统要求	建议选用 64 位 Linux 系统
内存大小	建议内存大小在 1GB 以上
硬盘空间	建议 40 GB 以上

Client 端	
CPU 要求	建议 CPU 核数为双核或以上
内存大小要求	建议内存大小在 1GB 以上
网络协议	支持 TCP/IP 协议，以及 HTTP 协议
操作系统要求	<ul style="list-style-type: none"> <li>Windows 版本 Windows 2000/XP/2003/Vista/7 及以上</li> <li>Linux 版本 内核要求 Linux kernel - 2.2.14 及以上，并且至少要安装 glibc 2.2.4、gtk+ 2.0 函数库</li> </ul>

	<ul style="list-style-type: none"> <li>• Mac 版本 Mac OS X 10.2.x 及以上</li> </ul>
浏览器要求	<ul style="list-style-type: none"> <li>• Internet Explorer 6.0 及以上版本</li> <li>• Chrome 或 Firefox 浏览器</li> </ul>

## 4.3 界面要求

### 4.3.1 清楚一致的设计

页面的一致性允许用户将已有的知识传递到新的任务中，更快地学习新事物，不必花时间来尝试记住交互中的不同。通过提供一种稳定的感觉，一致使得界面熟悉又可预测。各功能选项和操作的名称、信息的可视表示以及元素在屏幕和窗口内部的放置都应满足一致性。

### 4.3.2 简洁和美观

界面应该简单、易于学习、并且易于使用，他还必须提供对系统所有功能的访问；美观是网页界面的重要部分，可视属性提供了好的印象，并传达特定用户的交互行为的重要线索。

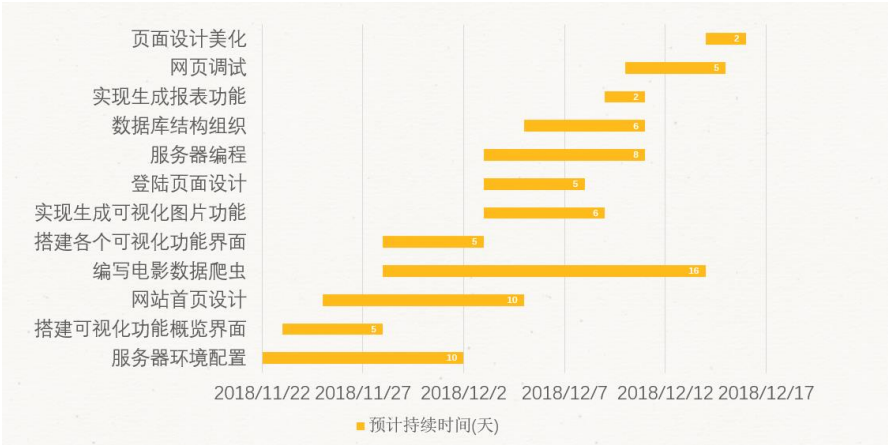
屏幕的布局因功能不同考虑的侧重点不同。各功能区要重点突出、功能明显。各元素应位置平衡、以合理顺序摆放。

### 4.3.3 用户操作的安全性

用户能自由的对界面上的每一项做出选择，且所有选择都是可逆的。在用户做出危险的选择时有信息提示是减少用户错误的有效方法。

## 4.4 进度要求

要求该系统能够在 3 周之内完成所有功能并上线。



## 4.5 交付要求

- 软件需求说明书（本文档）：各迭代版本的电子文件，以及最终版本的打印材料
- 最终版本的测试文档：电子文件
- 网站各功能的 demo：以视频以及实操的方式呈现
- 各功能模块演示：现场操作

### 4.5 交付要求修改为：

交付项	交付形式	描述
程序	源代码	包含服务器端的 html、js、php 等格式的文件
软件需求说明书	电子文件、最终版本的打印材料	描述软件需求、约束以及设计的文档
测试文档	电子文件	最终版本
网站 demo	视频	清晰、完整地展示应用实现的需求和功能

## 4.6 验收要求

- 测试用例全部通过；
- 每个功能都能够正确正常运行，不会发生错误；
- 网站界面美观，性能较优。

## 5. 软件原型

## 6. 备注

请修改人员注意：

1. 对于只需要删除的内容，请用~~红色+删除线~~标出待删除的内容；
2. 对于小部分需要修改的内容，请用~~加粗删除线~~标出需要修改的内容，并在删除内容旁边用红色标出修改后的内容；
3. 对于需要大段增加或修改的内容，将修改内容置于所修改部分的下方，并用红色字体标出，注意不要删除原来的内容。