

**[短租房应用平台服务质量数据采集
分析系统]
系统设计文档
—面向对象方法**

学生姓名： 文良

学 号： 1120133099

指导教师： 金丹

班 级： 21111301

2017 年 4 月

1 系统需求

1.1 系统功能性需求

1.1.1 用例清单

用例名称	用例描述	参与者
提交 URL 爬虫请求	系统管理员在系统中输入要爬取数据的页面的 URL 并提交，获取爬虫原始数据。	管理员
存储转化数据	数据处理子系统将爬虫原始数据整理并写入数据库中储存。	时间类（Time）
创建分析规则	该用例描述一个用户自定义一个分析规则满足需求	管理员 用户
修改分析规则	该用例描述一个用户修改已有的分析规则	管理员 用户
删除分析规则	该用例描述一个用户删除一个已有的分析规则	管理员 用户
提交分析请求	该用例描述一个用户提交所需要的数据分析请求	管理员 用户
删除用户信息	该用例描述一名管理员删除用户的信息。	管理员
用户权限管理	该用例描述超级管理员对每名用户分配用户权限。	超级管理员
注册新用户	该用例描述一名用户提交注册信息。	管理员 用户
更新用户信息	该用例描述管理员更改用户信息或用户更改自己的信息。	管理员 用户

表 1.1.1 用例清单

1.1.2 用例图

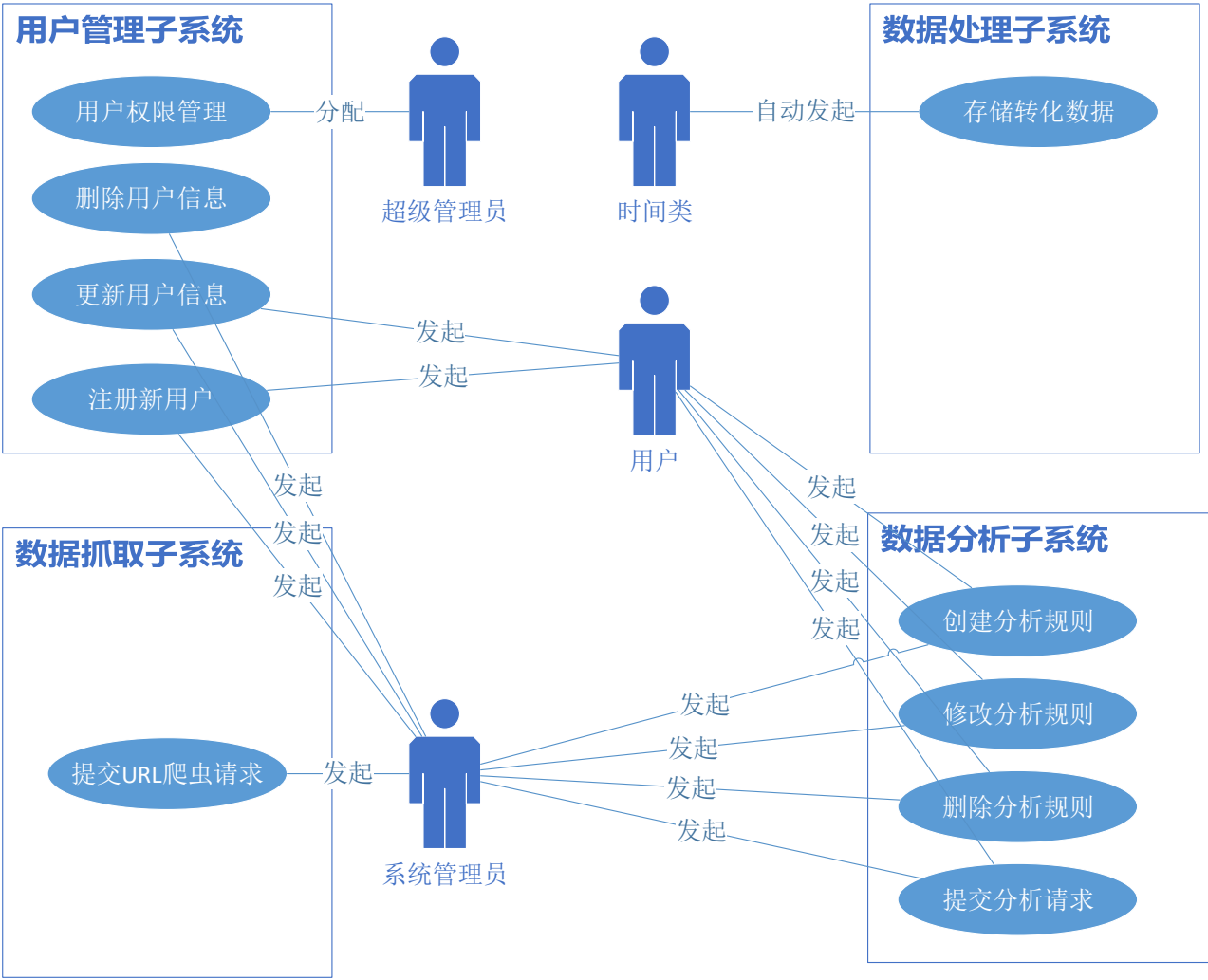


图 1.1.2 系统用例图

1.1.3 用例的详细描述

提交 URL 爬虫请求

Author (s): Wen Liang

Date: 2017/04/05

Version: 1.0

USE CASE NAME:	提交 URL 爬虫请求	USE CASE TYPE Abstract: <input type="checkbox"/> Extension: <input checked="" type="checkbox"/>
USE CASE ID:		
PRIORITY:	高	
INVOKED BY:	无	
PARTICIPATING ACTORS:	系统管理员	
DESCRIPTION:	系统管理员在系统中输入要爬取数据的页面的 URL，获取爬虫原始数据。	
PRE-CONDITION:	对于该网页的爬虫逻辑及代码已初始化完成	
TYPICAL COURSE OF EVENTS:	Step 1: 系统管理员进入【数据抓取】界面，输入将要爬取的网页的 URL	
	Step 2: 点击【检查】按钮检查系统中是否有对应 URL 的爬虫脚本，并返回对应的提示对话框（验证通过或重新输入）。	
	Step 3: 系统中有对应 URL 的爬虫脚本，验证通过，点击【确定】返回上一级菜单后点击【开始爬取】，系统开始爬取网页，返回原始数据	
ALTERNATE COURSES:	Step 3：系统中没有对应的爬虫脚本，点击【确定】重新输入 URL 或退出系统	
POST-CONDITION:	对爬取的数据确认保存、丢弃或重新爬取	

表 1.1.3 提交 URL 爬虫请求用例描述

存储转化数据

Author (s): Wen Liang

Date: 2017/04/05

Version: 1.0

USE CASE NAME:	存储转化数据	USE CASE TYPE Abstract: <input type="checkbox"/> Extension: <input checked="" type="checkbox"/>
USE CASE ID:		
PRIORITY:	高	
INVOKED BY:	无	
PARTICIPATING ACTORS:	时间类	
DESCRIPTION:	数据处理子系统将爬虫原始数据整理并写入数据库中储存。	
PRE-CONDITION:	系统管理员对于数据爬取子系统的原始数据已确认保存	
TYPICAL COURSE OF EVENTS:	Step 1: 系统按照预设的正则表达式对原始数据进行匹配和筛选，生成原始表单	
	Step 2: 系统对原始表单进行整理（生成 ID、拆分成若干对应的表等）	
	Step 3: 系统将整理后的表单保存到数据库中，返回【数据处理完毕】提示对话框	
ALTERNATE COURSES:	无	
POST-CONDITION:	对处理后的数据，选取相关的分析规则进行查询和分析。	

表 1.14 存储转化数据用例描述

创建分析规则

Author (s): Wen Liang

Date: 2017/04/05

Version 1.0

USE CASE NAME:	创建分析规则	USE CASE TYPE Abstract: <input type="checkbox"/> Extension: <input checked="" type="checkbox"/>
USE CASE ID:		
PRIORITY:	高	
INVOKED BY:	无	
PARTICIPATING ACTORS:	系统管理员、用户	
DESCRIPTION:	系统管理员或用户在系统中创建一个数据可视化的分析规则	
PRE-CONDITION:	有分析规则的选项并可以组合	
TYPICAL COURSE OF EVENTS:	Step 1: 系统管理员或用户进入数据分析子系统，选择【创建分析规则】	
	Step 2: 点击【创建】，并依次选择分析规则选项	
	Step 3:点击【验证】检查规则是否通过，若通过，点击【确定】返回上一级菜单	
ALTERNATE COURSES:	Step 3：检查不通过，点击【确定】重新创建分析规则	
POST-CONDITION:	对分析规则的修改、删除或提交分析请求	

表 1.15 创建分析规则用例描述

修改分析规则

Author (s): _____Wen Liang_____

Date: __2017/04/05_____

Version: _____1.0_____

USE CASE NAME:	修改分析规则	USE CASE TYPE Abstract: <input type="checkbox"/> Extension: <input checked="" type="checkbox"/>
USE CASE ID:		
PRIORITY:	高	
INVOKED BY:	无	
PARTICIPATING ACTORS:	系统管理员、用户	
DESCRIPTION:	系统管理员或用户修改已有的分析规则并保存	
PRE-CONDITION:	系统中有能够修改的分析规则	
TYPICAL COURSE OF EVENTS:	Step 1: 系统管理员或用户进入数据分析子系统，在要修改的分析规则后选择【修改】	
	Step 2: 选择并修改分析规则中具体的项	
	Step 3: 点击【保存】，返回【分析规则修改成功】对话框	
ALTERNATE COURSES:	无	
POST-CONDITION:	对分析规则的创建、修改、删除或提交分析请求	

表 1.1.6 修改分析规则用例描述

删除分析规则

Author (s): _____Wen Liang_____

Date: __2017/04/05_____

Version: _____1.0_____

USE CASE NAME:	删除分析规则	USE CASE TYPE Abstract: <input type="checkbox"/> Extension: <input checked="" type="checkbox"/>
USE CASE ID:		
PRIORITY:	高	
INVOKED BY:	无	
PARTICIPATING ACTORS:	系统管理员	
DESCRIPTION:	系统管理员或用户删除已有的分析规则并保存	
PRE-CONDITION:	系统中有能够删除的分析规则	
TYPICAL COURSE OF EVENTS:	Step 1: 系统管理员或用户进入数据分析子系统，在要删除的分析规则后选择【删除】	
	Step 2: 返回【是否删除】对话框，点击【确定】	
	Step 3:返回【删除成功】对话框	
ALTERNATE COURSES:	无	
POST-CONDITION:	对分析规则的创建、修改、删除或提交分析请求	

表 1.1.7 删除分析规则用例描述

提交分析请求

Author (s): _____Wen Liang_____

Date: _____2017/04/05_____

Version: _____1.0_____

USE CASE NAME:	提交分析请求	USE CASE TYPE Abstract: <input type="checkbox"/> Extension: <input checked="" type="checkbox"/>
USE CASE ID:		
PRIORITY:	高	
INVOKED BY:	创建分析规则	
PARTICIPATING ACTORS:	系统管理员、用户	
DESCRIPTION:	系统管理员或用户提交现有的某一条分析规则以处理数据并输出结果	
PRE-CONDITION:	系统中有对应的分析规则与数据	
TYPICAL COURSE OF EVENTS:	Step 1: 系统管理员或用户进入数据分析子系统，在要提交分析规则后选择【提交】	
	Step 2: 系统判断是否有对应数据	
	Step 3: 系统呈现数据分析的输出结果，返回【数据分析完毕】对话框	
ALTERNATE COURSES:	Step 2: 无对应数据，系统返回【无数据！】对话框	
POST-CONDITION:	输出可视化的数据分析结果	

表 1.1.8 提交分析请求用例描述

用户权限管理

Author (s): _____Wen Liang_____

Date: _____2017/04/05_____

Version: _____1.0_____

USE CASE NAME:	用户权限管理	USE CASE TYPE Abstract: <input type="checkbox"/> Extension: <input checked="" type="checkbox"/>
USE CASE ID:		
PRIORITY:	高	
INVOKED BY:	注册新用户	
PARTICIPATING ACTORS:	超级管理员	
DESCRIPTION:	超级管理员分配或撤销用户的权限	
PRE-CONDITION:	系统中有已经注册的用户	
TYPICAL COURSE OF EVENTS:	Step 1: 超级管理员进入用户管理子系统，选择【用户权限管理】	
	Step 2: 在对应的权限后添加或删除用户 ID 以分配或撤销用户权限	
	Step 3:点击【保存】，返回【保存成功】对话框	
ALTERNATE COURSES:	无	
POST-CONDITION:	用户信息的删除或修改	

表 1.1.9 用户权限管理用例描述

删除用户信息

Author (s): _____Wen Liang_____

Date: _____2017/04/05_____

Version: _____1.0_____

USE CASE NAME:	删除用户信息	USE CASE TYPE Abstract: <input type="checkbox"/> Extension: <input checked="" type="checkbox"/>
USE CASE ID:		
PRIORITY:	高	
INVOKED BY:	注册新用户	
PARTICIPATING ACTORS:	系统管理员	
DESCRIPTION:	系统管理员删除一名或多名用户的信息	
PRE-CONDITION:	系统中有已经注册的用户	
TYPICAL COURSE OF EVENTS:	Step 1: 系统管理员进入用户管理子系统	
	Step 2: 在对应的用户后选择【删除】，返回【确认删除该用户信息？】对话框	
	Step 3:点击【确定】，返回【删除成功】对话框	
ALTERNATE COURSES:	无	
POST-CONDITION:	用户信息的删除或更新	

表 1.1.10 删除用户信息用例描述

更新用户信息

Author (s): _____Wen Liang_____

Date: _____2017/04/05_____

Version: _____1.0_____

USE CASE NAME:	更新用户信息	USE CASE TYPE Abstract: <input type="checkbox"/> Extension: <input checked="" type="checkbox"/>
USE CASE ID:		
PRIORITY:	高	
INVOKED BY:	注册新用户	
PARTICIPATING ACTORS:	系统管理员、用户	
DESCRIPTION:	系统管理员或用户修改用户的信息并更新	
PRE-CONDITION:	系统中有已经注册的用户	
TYPICAL COURSE OF EVENTS:	Step 1: 系统管理员或用户进入用户管理子系统，在要修改的用户后选择【修改】	
	Step 2: 修改对应的用户信息	
	Step 3:点击【保存】，返回【用户信息更新成功】对话框	
ALTERNATE COURSES:	无	
POST-CONDITION:	用户信息的删除或更新	

表 1.1.11 更新用户信息用例描述

注册新用户

Author (s): _____Wen Liang_____

Date: _____2017/04/05_____

Version: _____1.0_____

USE CASE NAME:	注册新用户	USE CASE TYPE Abstract: <input type="checkbox"/> Extension: <input checked="" type="checkbox"/>
USE CASE ID:		
PRIORITY:	高	
INVOKED BY:	无	
PARTICIPATING ACTORS:	系统管理员、用户	
DESCRIPTION:	系统管理员或用户在系统中注册新的用户	
PRE-CONDITION:		
TYPICAL COURSE OF EVENTS:	Step 1: 系统管理员或用户进入用户管理子系统，选择注册新用户	
	Step 2: 填写对应的用户信息	
	Step 3:点击【注册】，返回【用户注册成功】对话框	
ALTERNATE COURSES:	无	
POST-CONDITION:	用户信息的删除或更新	

表 1.1.12 注册新用户用例描述

1.2 非功能性需求

- 1) 效率：数据爬取时间与写入数据库时间不能过长，需要设计一个高效的分布式爬取和存储模块以解决数据处理的效率问题。
- 2) 安全：某些场合需要将操作记录到操作日志中，日志的保留期限的设定，以便随时查看和记录问题，方便产品的后续更新迭代。
- 3) 健壮性：因停电、网站的代理池更换等其他外力因素导致的爬虫过程中断时，应该能保证后台数据存储的完整性，并且在系统排错重启之后能够迅速恢复工作现场，因此必须保证系统的健壮性和高容错率。

2 系统分析

2.1 系统概念模型

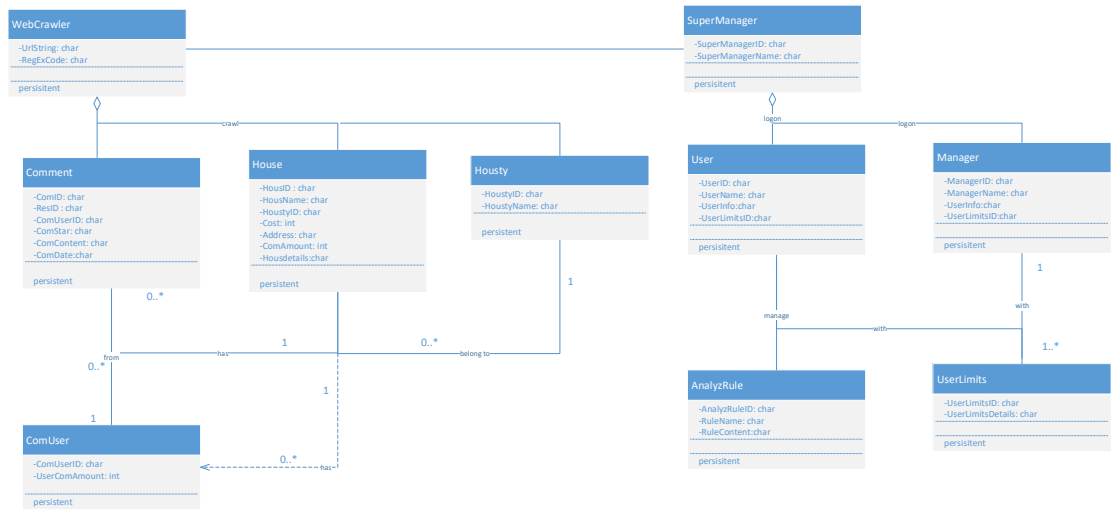


图 2.1.1 系统分析类图

说明：

1. WebCrawler 为数据爬取子系统的数据爬取类，主要属性有：
 - (1) UrlString 用于匹配输入 URL 的字符串的值
 - (2) RegExCode 系统内置的对应 URL 的正则表达式
2. Comment、House、Housty 分别为数据处理子系统将原始数据处理完毕后，存储到数据库中的三个类。ComUser 类为评论者类，由以上几个类计算出每个评论者发布过的评论数量。
 - (1) Comment 类为某一房屋用户评论类，主要属性有：评论者 ID，房屋 ID，评论者给出的星级评价，评论者给出的评分
 - (2) House 为房屋类，主要属性有：房屋 ID，房屋名称，房型 ID，房屋详情，房屋地址，参考费用，评论数
 - (3) Housty 类为房型类，主要属性有：房型 ID 和房型名称
 - (4) ComUser 类为评论者类，主要属性有：评论者的 ID 和该评论者发布过的评论数。
3. SuperManager，User，Manager 为用户类，分别代表使用该系统的超级管理员、普通用户和系统管理员。三个类均有该人员的 ID 及名称；其中，用户类和系统管理员类还有用户信息和用户权限属性。
4. UserLimits 为用户权限类，主要属性有用户权限的 ID 与权限 ID 对应的具体内容
5. AnalyzRule 为分析规则类，主要属性有分析规则的 ID，分析规则名称与规则的内容

2.2 类职责定义

类名	职责
WebCrawler	1. 匹配输入的 URL 连接并返回对应的提示 2. 匹配正确，则爬取对应网页的数据 3. 获取系统内对应 URL 的正则表达式 RegExCode 并处理原始数据，在数据库中储存成若干个表（对应以下几个类）
Comment	该类主要储存某一房屋评论相关数据，主要属性有：评论者 ID，房屋 ID，评论者给出的星级评价，评论者给出的评分
House	该类主要储存某一房屋相关数据，主要属性有：房屋 ID，房屋名称，房型 ID，房屋详情，房屋地址，参考费用，评论数
Housty	该类主要储存房屋类型，主要属性有：房型 ID 和房型名称
ComUser	该类主要提供方法计算某一评论者发布的评论数，以供数据分析使用，主要属性有：评论者的 ID 和该评论者发布过的评论数。
SuperManager	该类主要储存系统中超级管理员的信息，主要属性有：超级管理员的 ID，超级管理员的名称。
User	该类主要储存系统中一般用户的信息，主要属性有：用户 ID，用户名称，用户信息及用户权限 ID
Manager	该类主要储存系统中管理员的信息，主要属性有：管理员 ID，管理员名称，管理员信息及管理员的用户权限 ID
UserLimits	该类主要储存系统中用户权限的等级及权限等级对应的内容，并提供方法分配权限，主要属性有：权限 ID，权限内容
AnalyzRule	该类主要储存系统的分析规则，并提供方法创建、修改、删除和更新分析规则，主要属性有：规则 ID，规则名称及规则内容

表 2.2.1 类职责描述

2.3 系统流程定义

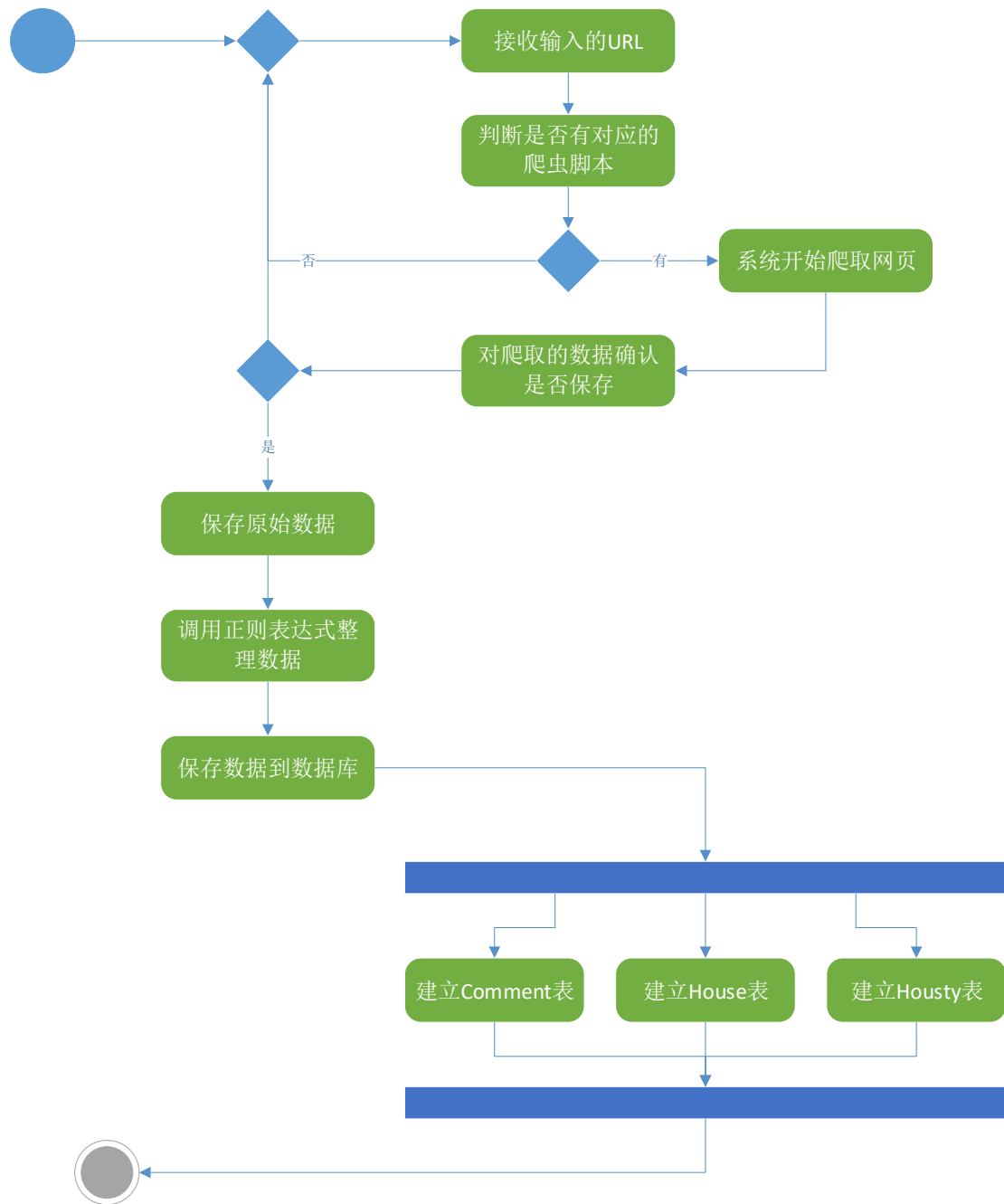


图 2.3.1 数据抓取、数据处理系统活动图

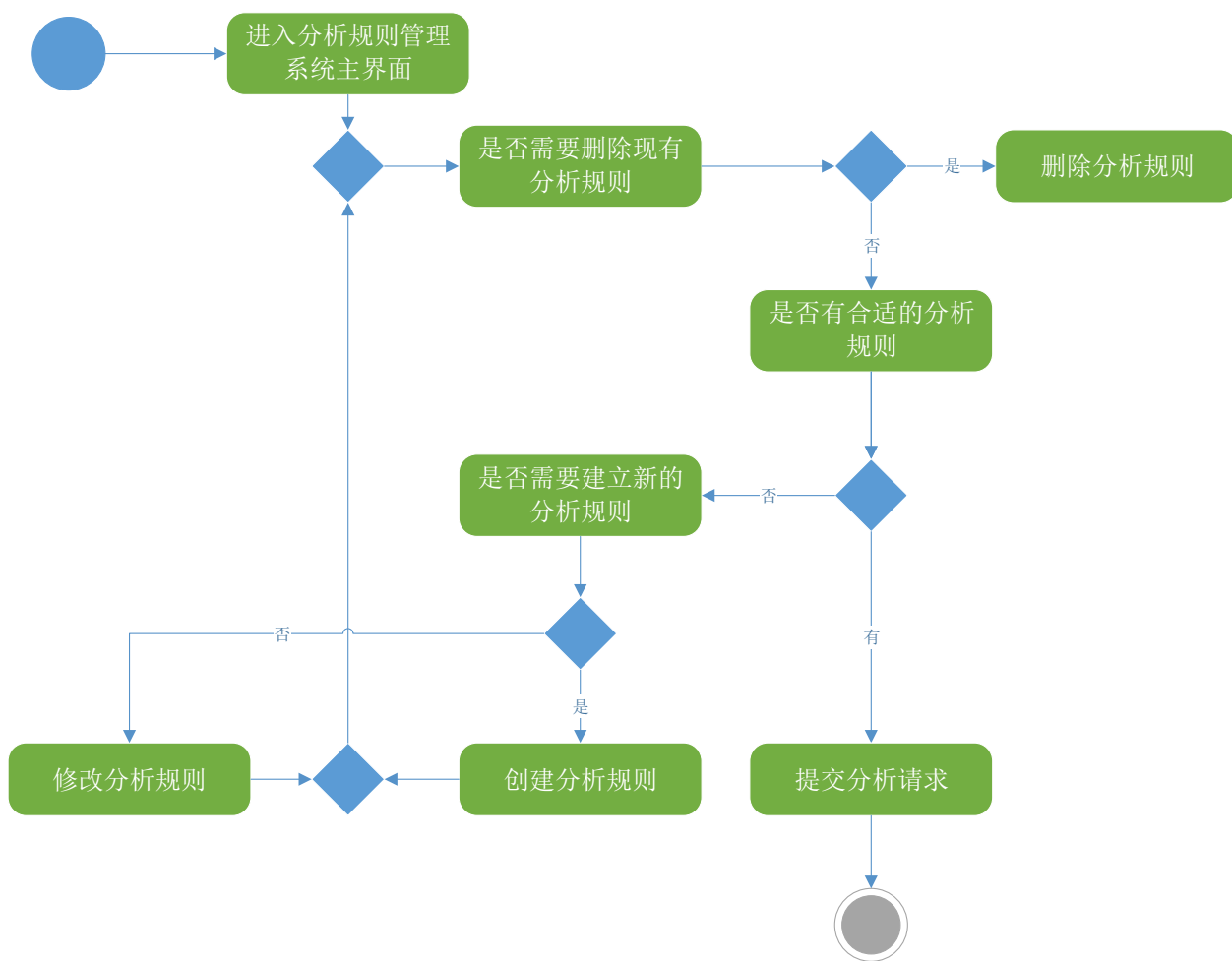


图 2.3.2 数据分析系统活动图

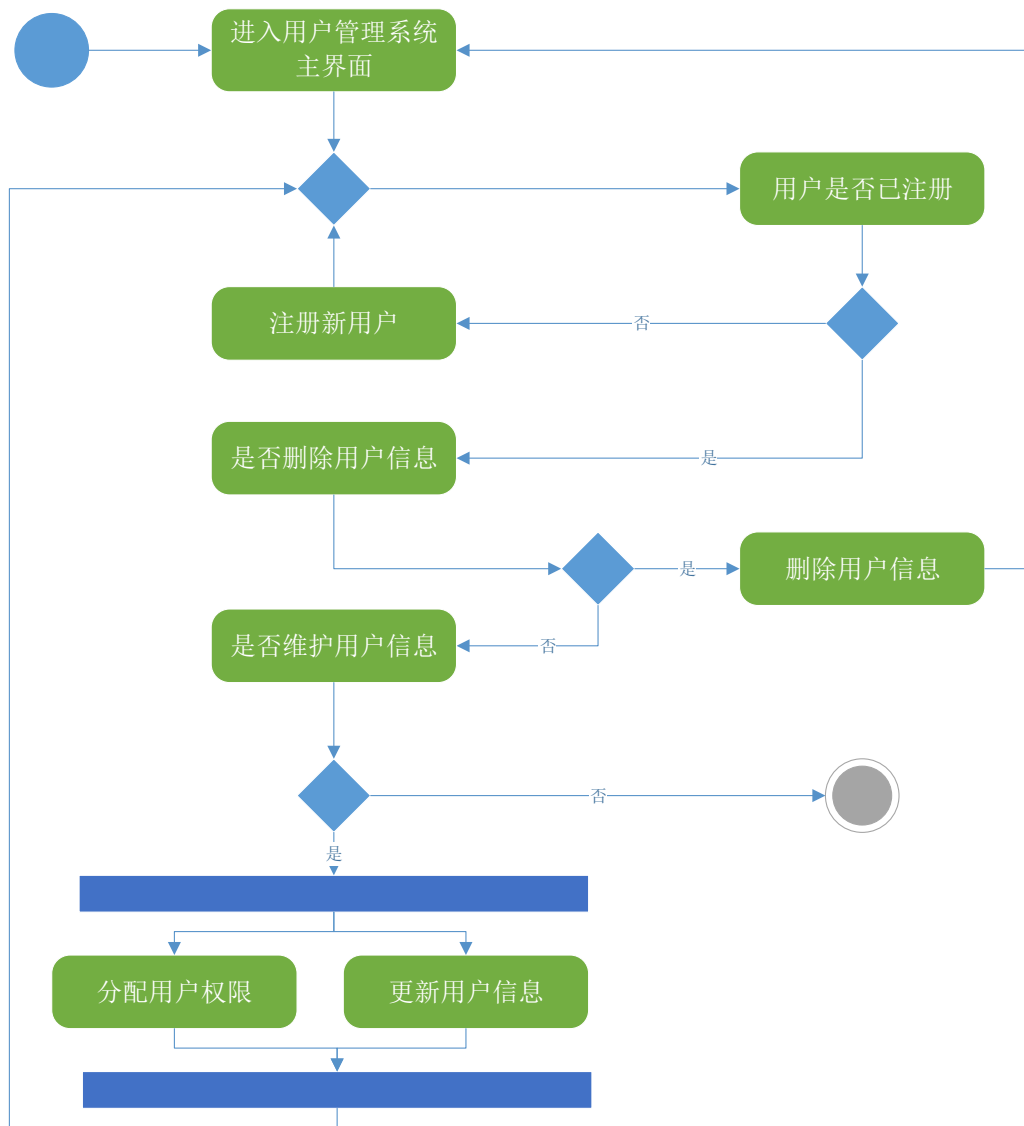


图 2.3.3 用户管理系统活动图

2.4 系统行为定义

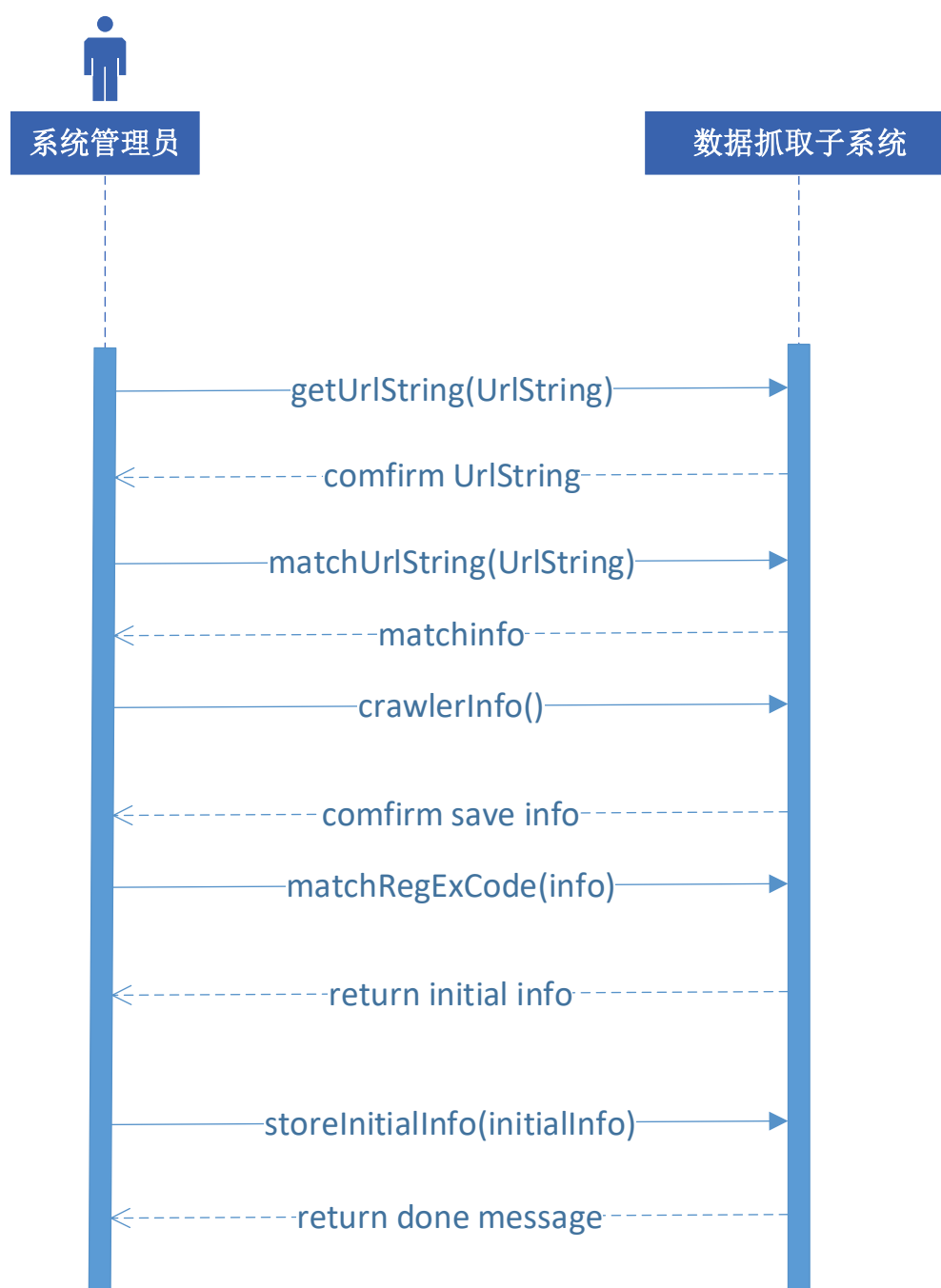


图 2.4.1 数据抓取系统顺序图



图 2.4.2 用户管理系统顺序图



图 2.4.3 数据分析系统顺序图

3 系统设计

3.1 系统总体布局设计

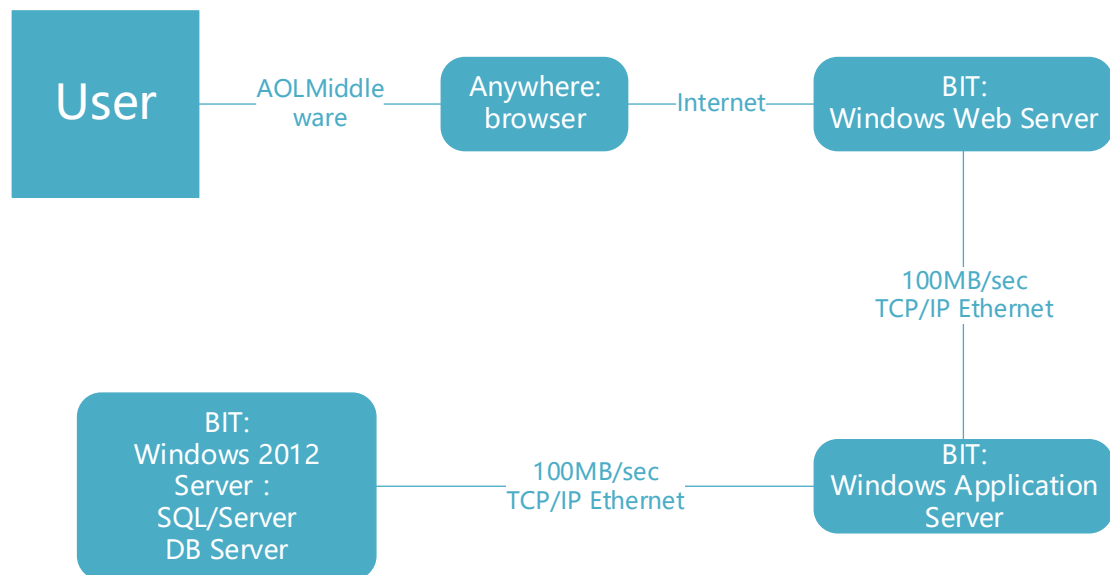
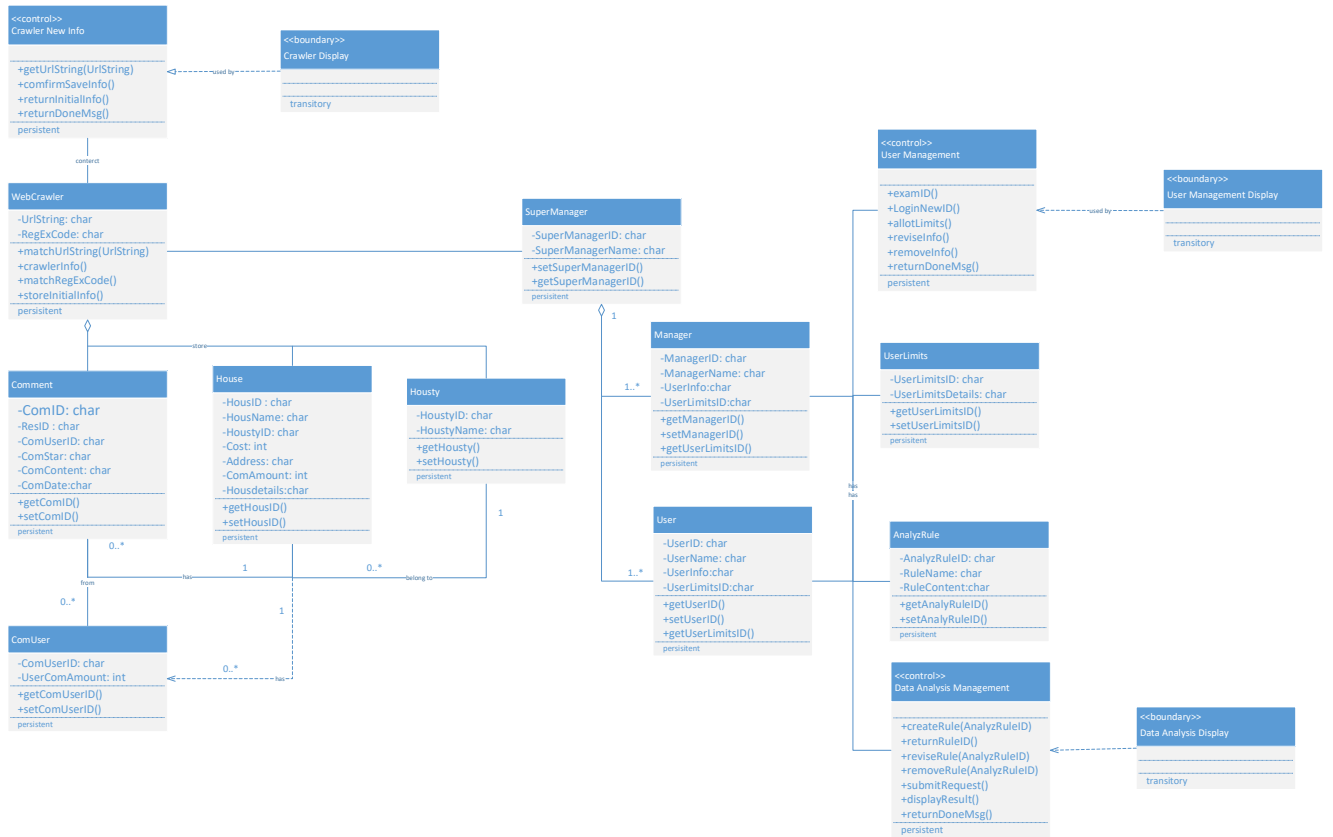


图 3. 1. 1 系统架构图

3.2 结构设计



3.3 交互设计

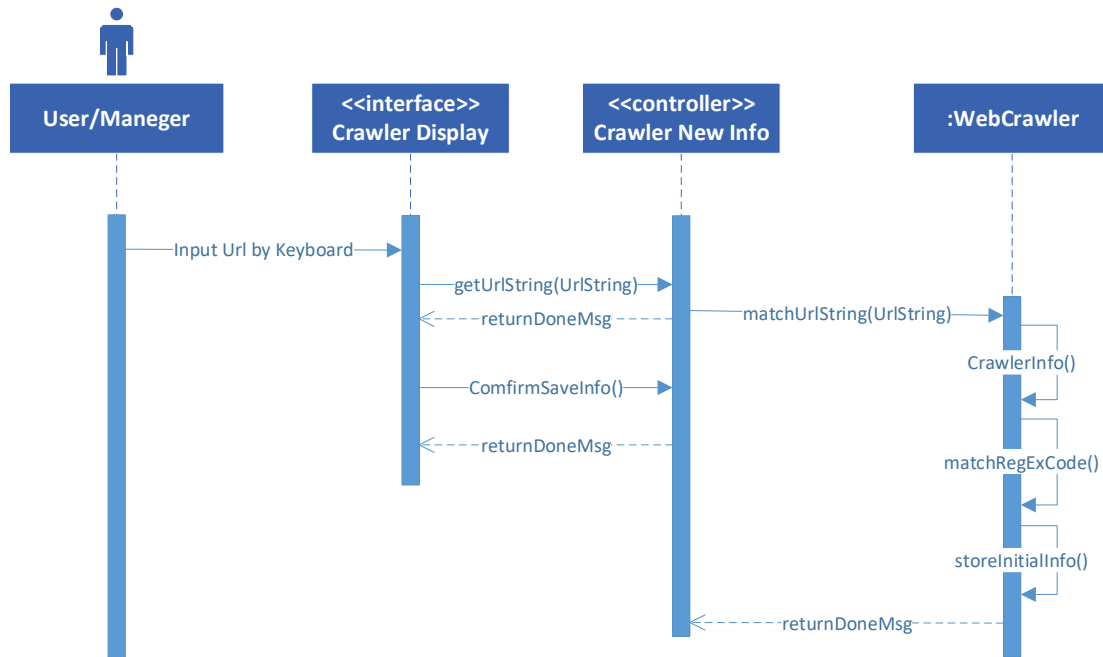


图 3.3.1 数据抓取用例顺序图

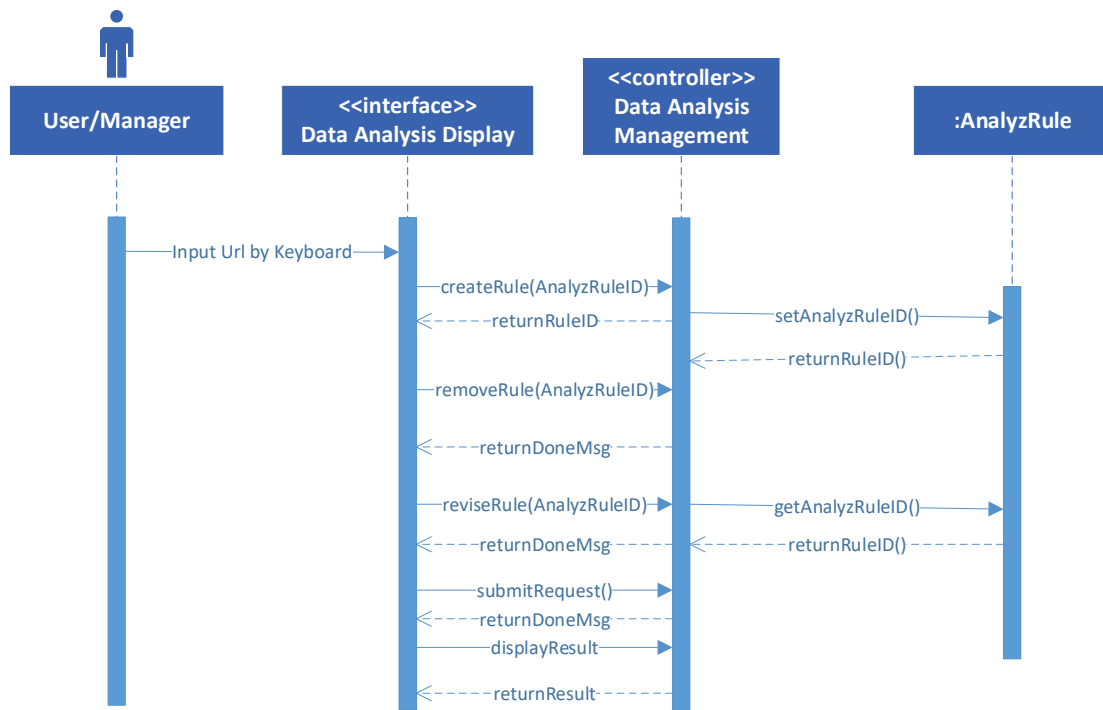


图 3.3.2 数据分析用例顺序图

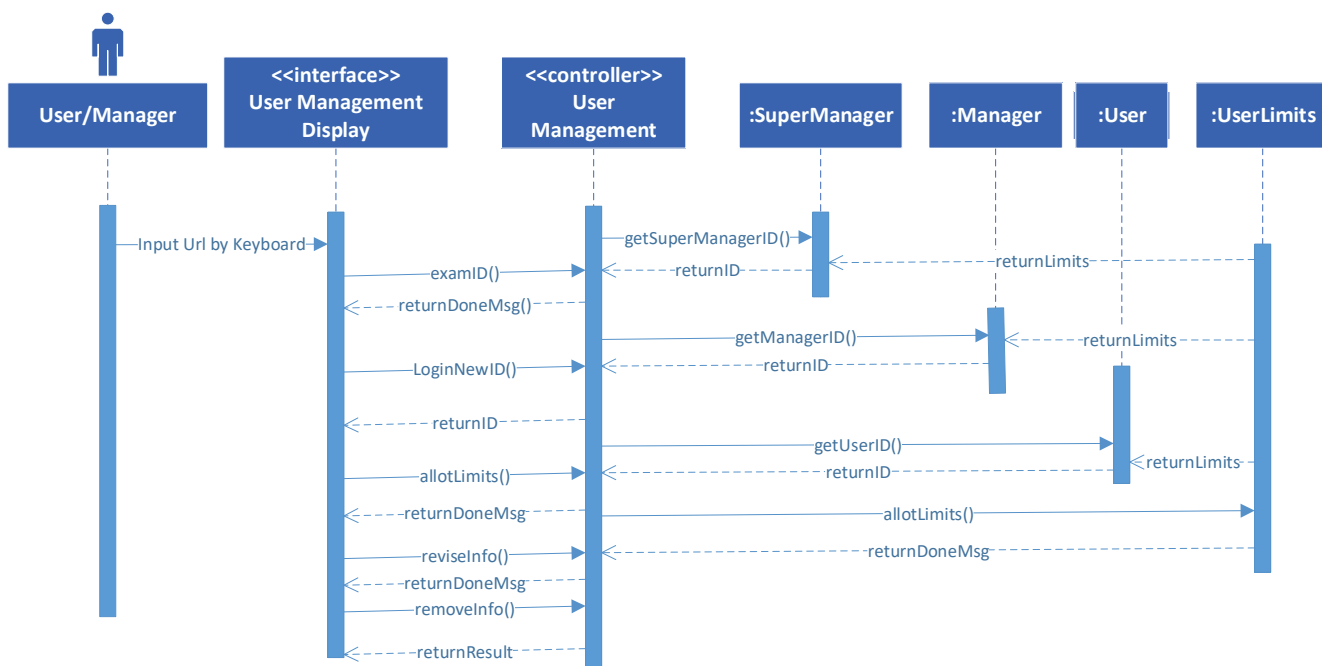


图 3.3.3 用户管理用例顺序图

3.4 状态设计

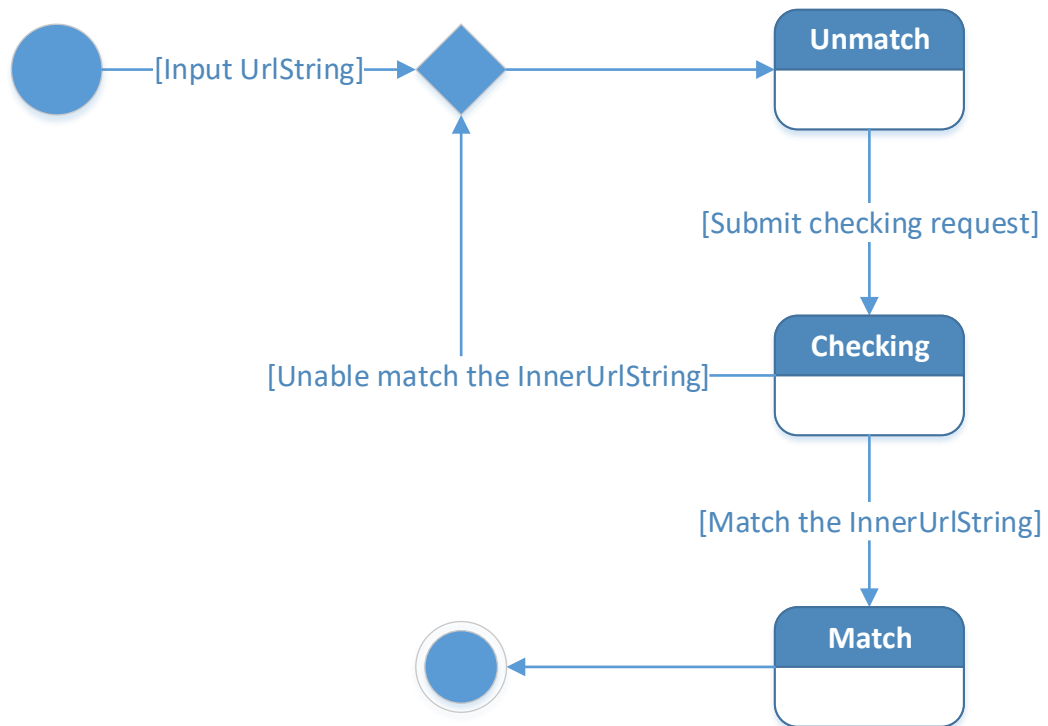


图 3.4.1 Ur1 状态描述

3.5 界面原型设计



图 3.5.1 系统主界面原型设计



图 3.5.2 数据处理界面原型设计



图 3.5.3 数据分析界面原型设计



图 3.5.4 数据分析界面原型设计