

北京理工大学

毕业设计（论文）任务书

题目类别： 毕业设计

题目性质： 工程设计

毕业设计（论文）题目： 短租房应用平台服务质量数据采集分
析系统

学 院： 管理与经济学院

专 业： 信息管理与信息系统

班 级： 21111301

姓 名： 文 良

指导教师： 金 丹

ABSTRACT

[Click here and input abstract in English]

Key words: ; ;

摘要

[单击此处输入中文摘要]

关键词：

目录

第一章. 绪论.....	12
1.1. 课题背景.....	12
1.2. 课题研究的目的和意义.....	13
1.3. 国内外相关研究现状.....	15
1.4. 个人工作内容.....	17
1.5. 论文的结构.....	17
第二章. 爬虫相关技术介绍	18
2.1. 爬虫技术概述.....	18
2.2. Python 语言介绍	19
2.3. 正则表达式.....	20
2.4. Scrapy 框架	21
2.5. PyQt 介绍	22
2.6. 本章小结.....	23
第三章. 短租房应用平台服务质量数据采集分析系统的需求分析	24
3.1. 系统业务需求.....	24
3.1.1. 系统特性分析	24
3.1.2. 用户需求分析	25
3.2. 系统功能性需求.....	26
3.2.1. 数据抓取功能需求分析	26
3.2.2. 数据处理功能需求分析	27
3.2.3. 数据分析功能需求分析	28
3.2.4. 用户管理功能需求分析	31
3.3. 系统非功能性需求.....	34
3.4. 本章小结.....	34
第四章. 短租房应用平台服务质量数据采集分析系统的系统概要设计	35
4.1. 系统架构设计.....	35
4.2. 系统功能设计.....	36
4.2.1. 数据抓取功能	37
4.2.2. 数据处理功能	38
4.2.3. 数据分析功能	38
4.2.4. 用户管理功能	39
4.3. 系统数据库设计.....	40
4.4. 本章小结.....	40
第五章. 短租房应用平台服务质量数据采集分析系统的系统详细设计与实施 ...	41
5.1. 主要功能设计与实施.....	41
5.1.1. 数据抓取功能	41
5.1.2. 数据处理功能	41
5.1.3. 数据分析功能	41
5.1.4. 用户管理功能	41

5.2. 本章小结.....41

第六章. 系统验证 41

6.1. 系统业务功能验证.....41

6.2. 系统性能验证.....41

6.3. 本章小结.....41

第七章. 结论 41

7.1. 全文总结.....41

7.2. 不足与展望.....41

参考文献 42

致谢 43

第一章. 绪论

本章作为论文的绪论，分为五个小节，阐述了此论文研究的课题背景、课题研究的目的及其意义，短租房应用平台服务质量国内外的相关研究现状，作者在该课题中主要进行的工作，并描述了论文的组织结构和各章的职责划分。

1.1. 课题背景

房价高、买房难，楼市持续低迷，租房需求持续攀升，租房市场异常活跃。互联网、新媒体、第三方交易平台的推广使用实现了传统行业线上、线下的融合，租房市场的 O2O 模式变革了传统的门店式租房中介运营成本高、信息不透明的低效率状态，“去门店化”赫然成为中介行业发展的新趋势。

短租房，又名“移动公寓”、“日租公寓”、“自助公寓”、“日租房”，是将房屋短期出租给客人，即传统意义上的“家庭旅馆”。与传统租赁时间为半年至几年不等不同，“短租房”的租期一般是一日至几个月。它可以满足假日旅游、度假或者短期出差等短期异地住宿需求，短租的住宿费用甚至在同档次品质的前提下比三星级酒店的费用要低到 30%到 50%左右，同时能为租客营造“家”的感觉，满足个性化需求，即经济实惠，又灵活多变。

由于短租市场日益成熟，国内互联网企业开始涉足短租房市场，开办短租房信息网站。为房东的短租房提供一个展示和出租平台，同时，也为商务旅游、家庭出游等人群短租提供方便。

网络短租平台是租房市场 O2O 模式近几年催生的一种新生事物。短租平台概念来源于国外的 Airbnb，为房东和房客提供一个交流和交易的平台，平台本

身充当房东和租客之间的担保方，房东在平台销售短租房源，租客在平台购买，然后去线下消费。

在日常生活的应用中，大型比赛、交易会、旅游、开学季春节探亲、临时出差、考研考公务员等短期租住需求使得经济实惠、灵活方便的短租市场异常火爆，短租行业正风声水起，不少投资家很看好这个行业，近年来，像小猪短租、途家、蚂蚁短租、游天下、美租网等大批网络短租平台竞相上线，加入到短租市场的争夺之中。

1.2. 课题研究的目的是和意义

在过去，传统的酒店住宿有以下的特点：标准化、资源紧缺、缺少人情味儿。一方面，城市中产阶级规模不断扩大，购买力增强，传统的酒店住宿已经难以满足较高收入人群对于住宿的需要。另一方面，国内住房资源紧缺，尤其是在一线城市更为严重，共享经济形式的住宿可以通过让渡房屋的使用权，解决房源紧缺的问题。

近年来，以短租房为主要信息源建立的短租网站受到风投的热切关注，大咖们纷纷注资，参与到这场竞逐中来，市场格局变化风起云涌。不久前，58同城战略伙伴“小猪短租”完成1500万元B轮融资，途家也已完成3亿美元D及D+轮融资，估值超10亿美元，拥有的房源数量为45万套，房租支付平台“租房宝”获得华映资本1000万元注资，“蚂蚁短租”获3大VC千万美元注资，“YOU+国际青年公寓”获雷军1亿投资，优客逸家完成B轮2200万美元融资，“住百家”获联想百万美元融资，租房搜索服务Zumper获650万美元融资，专注于都市白领合租的O2O服务平台“蘑菇公寓”完成B轮2500万美元的融资。

短租市场虽备受追捧，但是背后的风险却不容人们忽视，这几年短租市场暴露的问题也越来越多：国内短租鼻祖“爱日租”关闭，爱屋吉屋租房佣金减半首次在业内掀起价格竞争。此外，关于民宿房屋的安全性问题，自始至终都受到关注。不久前 Airbnb 信任危机爆发，其短租服务违反纽约私人房产出租法规，引起热烈讨论。尤其是去年一篇题为《曝光一个上戏学生，他用 Airbnb 毁了我整个家》的文章，更是将民宿短租推向了舆论的制高点。究竟“该不该租？租给谁？”的话题引发了热议。一直游走在监管灰色地带的短租应用平台网站能否顺利的走下去，如何在众多的租房网站中建立自己的核心竞争力，从众多的短租企业中脱颖而出，持久发展？短租行业将如何洗牌？行业的未来走向值得关注。

这时候，信用体系建设成为一个至关重要的议题。包括途家在内的诸多短租应用平台，都选择与芝麻信用进行合作，希望依托这些公开的信用数据，提供更标准化的服务，能够对用户进行画像和判断，从而避免租客破坏房屋的事情发生。

在这样的情形下，短租房应用平台希望通过信用体系考察消费者；而对于消费者而言，也需要有对应的服务质量评价系统，去对市场上林林总总的短租房平台以及其房源信息做出观测和评估，以期能在选择和使用短租房平台时有所参考，更好地享受短租房应用平台为他们提供的服务。除此之外，该服务质量评价系统的搭建还能对信用评价体系的建立起到很大的促进作用，一方面，企业可对接服务质量评价系统，及时得到其下产品和房源信息的基于服务质量评价体系的打分，以及用户对于其服务的反馈，以方便短租房平台企业通知房东或自身进行整合调整；另一方面，信用评价体系能参考该服务质量评价系统的数据及结果，帮助构建信用指标和用户画像判断。

因此，作者研究的课题希望通过建立一个关于短租房应用平台的数据采集系统，以期对采集到的数据运用服务质量相关体系打分，同时满足用户对于短租房应用平台某一类特定数据的获取和参考，以及满足企业管理者对于其提供的服务的自我评估，推动房东和消费者双方更好地做出选择。

1.3. 国内外相关研究现状

由于短租房应用平台是一个较新的概念，国内外对于在此基础上建立服务质量信息采集的数据分析系统这方面的能参考的直接研究较少；但在短租房市场方面，由于其已经发展了相当长的一段时间，在行业分析和服务质量方向上，许多学者已经做出了相当的研究成果。

在短租房的行业方面，薛奕妹研究了我国短租房旅游市场，指出目前国内在线短租主要有三种模式。客户到客户（C2C）模式、管理公司到客户（B2C）模式、托管公司到客户（N2C）模式；沈珂通过对我国的在线短租市场的研究，指出了短租行业研究的意义：短租行业是伴随着中国经济的快速增长、休假制度的完善、大众旅游模式的转变以及在国家宏观调控下资金重寻增值方式等众多因素的影响下产生。它的出现不仅增加了社会的就业，促使传统酒店模式的转型，也使社会闲置资源得到了更加有效的利用；Steffi则对小猪短租的模式进行了深入的研究，发现了其核心优势是牵手建筑师，带动乡村旅游经济；而它成功的关键因素有很多：一是价格优——省钱划算，居家体验。二是个性化——特色房源，个性选择。三是便捷——操作简单，轻松入住。四是安全——房客保障计划，安心入住。以上学者的研究表明，短租房是当下联系房地产、互联网以及共享经济领域的热门模块，建立一个针对短租应用平台的服务质量评价打分体系十分必要。

此外，许多学者也致力于运用不同的方法去推动模型的建立，以求评价标准客观性和可衡量性都较高。Limam N, 和Boutaba R认为，信任通常通过信誉系

统来评估；然而，许多现有系统依赖于消费者提供的评级，这引起了许多涉及服务评级的主观性和不公平性的问题。他们两人的研究介绍了声誉感知软件服务选择和评级的框架，为服务推荐设计了一种选择算法，为SaaS消费者提供了基于质量，成本和信任的最佳选择。李东娟,熊胜绪在已有顾客感知服务质量研究基础上，结合现实观察和消费体验，提出酒店其他顾客行为对顾客感知服务质量影响的研究构想，并构建了酒店其他顾客行为对顾客感知服务质量影响的模型。他们通过因子分析得出7个测量指标，利用结构方程模型验证了酒店其他顾客行为对顾客感知服务质量影响的因果关系。他们的研究在理论上突破已有顾客感知服务质量研究从顾客到服务提供者的两维框架，构建了从评价主体的顾客到正在消费的其他顾客再到服务提供者的三维框架，从而丰富了顾客感知服务质量的研究体系。从现实意义来讲，他们研究的结论表明，酒店进行服务质量管理不能仅仅局限于酒店自身，更要将顾客纳入质量管理的范围之内，借助顾客的力量实现全面质量管理的目标。

更多的学者运用模糊数学等方法做定量研究。隋婷婷通过调查用户对短租房体验的评价，并结合对旅行社、房屋中介、短租网站的访谈信息，运用模糊数学评价模型，综合探析了短租房市场目前的发展状况行业内部存在的问题：房源质量缺乏保证，客户服务有待完善，三方（房主、租客、短租网）的有效沟通不足。栗娟,麻学峰,李佩耕,向国娟以美国顾客满意指标(ACSI)模型为基础，建立一个以顾客价值为核心要素的酒店顾客满意度七要素测评模型，对树立提高酒店服务质量和酒店品牌形象具有重要的应用价值。该研究采用模糊数学理论，以张家界市酒店顾客满意度的实调查数据对模型应用研究，以了解张家界市酒店顾客满意度水平。张德然,何鹏光针对酒店服务质量的构成要素，体现在有形设施及无形服务上主要是其有形设施要方便、舒适、安全，无形服务要素则有:友谊、好客、相助；通过量化建立综合评价模型，较好地解决了现代酒店服务质量的综合评价问题。因此，本文在爬取到的短租房应用平台数据的基础上，结合当下短租房行业参与者关心的问题，进行数据分析和可视化处

理，以求从更直观的角度去对短租房领域进行更深入的分析 and 研究。

1.4. 个人工作内容

短租房应用平台服务质量数据采集分析系统主要由数据抓取系统、数据处理系统、数据分析系统和用户管理系统四个子系统构成。在课题的研究过程中，作者分析了整个数据采集分析系统的系统需求，针对需求完成了系统的架构设计；完成了系统中上述四个子系统的功能设计和对应的系统数据库的搭建；最终完成了该系统主要工作流程的实现，并完成系统的业务功能的验证和系统效率的验证。主要工作分列如下：

- 1) 结合市场需求和用户的需要，整理分析系统初级阶段的业务需求。
- 2) 参与项目的总体设计，通过查阅资料 and 对比分析，选用合适的技术框架做系统支撑。设计系统架构，使系统各个部分职责明确；参照数据库设计要求，搭建数据库并做好数据库字典设计。
- 3) 按照项目里程碑，按计划按时完成系统功能的设计和实现。
- 4) 对于在课题研究实施中遇到的问题，通过查阅相关资料，选用相应的技术解决，保证系统的稳定性和执行的效率。
- 5) 在系统的实施过程中，日志记录实时遇到的问题，不断优化和修改现有问题，使系统整体趋于优化。
- 6) 及时记录和总结在系统分析、设计和实施阶段的不足，为日后的学习提供经验。

1.5. 论文的结构

本论文根据系统设计开发方法中面向对象的思想，结合实际工作中的环节，分为七个部分。

第一章为论文的绪论部分，主要介绍课题研究的背景，阐述了课题研究的目的和意义。同时回顾 and 介绍了国内外短租房应用平台服务质量信息系统相关研究

的现状，对个人的工作内容作了介绍，并对论文的组织结构进行梳理。

第二章对系统实施过程中可能会用到的技术做了介绍。

第三章为系统的需求分析部分。主要对系统的业务性需求、功能性需求和非功能性需求分别做了分析和整理说明。

第四章介绍了系统的概要设计，从系统的架构设计、系统功能设计和系统的数据库设计及搭建三个方面进行说明。

第五章为系统的详细设计及实施。分别实现了该系统中的数据抓取、数据处理、数据分析和用户管理四个子系统的功能，并对功能辅以设计文档图表说明。

第六章为系统验证部分，主要进行了系统的业务功能验证和系统性能验证，为系统的稳定流畅运行提供保障。

第七章为结论部分，对论文做了总体的回顾和总结，回顾了系统从分析到实施、验证的流程。并对这一过程中的不足做了总结和反思，进一步展望系统下一步可行的发展方向，完善系统的结构。

第二章. 爬虫相关技术介绍

本章主要介绍了在项目的实施过程中主要运用的技术、框架，以及为解决相关问题需要用到的第三方的库和插件等等。由于短租房应用平台服务质量数据采集分析系统涉及到数据的爬取和分析两大块，在爬虫和数据处理、分析方面 Python 具有优势，因此整个开发过程中选用 Python 作为开发语言。为了更好地契合系统实现目标中对稳定性和效率的要求，开发过程中整合了很多成熟的第三方库及技术，减少很多重复的工作。

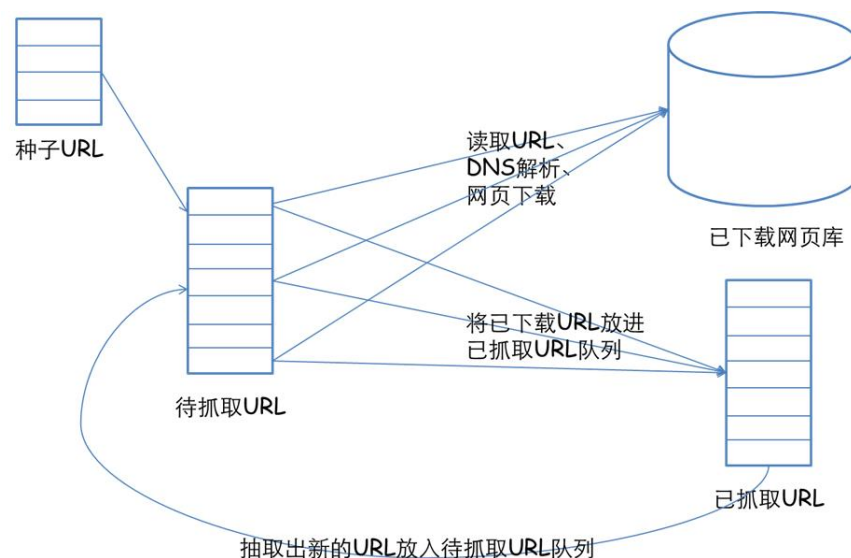
2.1. 爬虫技术概述

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。另外一些不常使用的名字还有蚂蚁、自动索引、模拟程序或者蠕虫。

网络爬虫是搜索引擎抓取系统的重要组成部分。爬虫的主要目的是将互联网上的网页下载到本地形成一个或联网内容的镜像备份

一个通用的网络爬虫的框架如图所示:



网络爬虫的基本工作流程如下:

- 1) 首先选取一部分精心挑选的种子 URL;
- 2) 将这些 URL 放入待抓取 URL 队列;
- 3) 从待抓取 URL 队列中取出待抓取在 URL, 解析 DNS, 并且得到主机的 ip, 并将 URL 对应的网页下载下来, 存储进已下载网页库中。此外, 将这些 URL 放进已抓取 URL 队列;
- 4) 分析已抓取 URL 队列中的 URL, 分析其中的其他 URL, 并且将 URL 放入待抓取 URL 队列, 从而进入下一个循环。

2.2. Python 语言介绍

Python 是一种面向对象的解释型计算机程序设计语言, 由荷兰人 Guido van Rossum 于 1989 年发明, 第一个公开发行人版发行于 1991 年。Python 是纯粹的自由软件, 源代码和解释器 CPython 遵循 GPL(GNU General Public License)协议。

Python 本身被设计为可扩充的。并非所有的特性和功能都集成到语言核心。Python 提供了丰富的 API 和工具，以便程序员能够轻松地使用 C 语言、C++、Cython 来编写扩充模块。Python 编译器本身也可以被集成到其它需要脚本语言的程序内。因此，很多人还把 Python 作为一种“胶水语言”（glue language）使用。使用 Python 将其他语言编写的程序进行集成和封装。

Python 在执行时，首先会将.py 文件中的源代码编译成 Python 的 byte code（字节码），然后再由 Python Virtual Machine（Python 虚拟机）来执行这些编译好的 byte code。这种机制的基本思想跟 Java，.NET 是一致的。然而，Python Virtual Machine 与 Java 或.NET 的 Virtual Machine 不同的是，Python 的 Virtual Machine 是一种更高级的 Virtual Machine。这里的高级并不是通常意义上的高级，不是说 Python 的 Virtual Machine 比 Java 或.NET 的功能更强大，而是说和 Java 或.NET 相比，Python 的 Virtual Machine 距离真实机器的距离更远。或者可以这么说，Python 的 Virtual Machine 是一种抽象层次更高的 Virtual Machine。

基于 C 的 Python 编译出的字节码文件，通常是.pyc 格式。

除此之外，Python 还可以以交互模式运行，比如主流操作系统 Unix/Linux、Mac、Windows 都可以直接在命令模式下直接运行 Python 交互环境。直接下达操作指令即可实现交互操作。

2.3. 正则表达式

正则表达式是对字符串操作的一种逻辑公式，就是用事先定义好的一些特定字符、及这些特定字符的组合，组成一个“规则字符串”，这个“规则字符串”用来表达对字符串的一种过滤逻辑。

给定一个正则表达式和另一个字符串，我们可以达到如下的目的：

- 1) 给定的字符串是否符合正则表达式的过滤逻辑（称作“匹配”）；
- 2) 可以通过正则表达式，从字符串中获取我们想要的特定部分。

正则表达式的特点是：

- 1) 灵活性、逻辑性和功能性非常的强；

- 2) 可以迅速地用极简单的方式达到字符串的复杂控制。
- 3) 对于刚接触的人来说, 比较晦涩难懂。

由于正则表达式主要应用对象是文本, 因此它在各种文本编辑器场合都有应用, 小到著名编辑器 EditPlus, 大到 Microsoft Word、Visual Studio 等大型编辑器, 都可以使用正则表达式来处理文本内容。

本次我们使用 Python 开发系统, Python 提供 re 模块, 包含所有正则表达式的功能。由于 Python 的字符串本身也用\转义, 所以要特别注意。

2.4. Scrapy 框架

Scrapy 是一个为了爬取网站数据, 提取结构性数据而编写的应用框架。可以应用在包括数据挖掘, 信息处理或存储历史数据等一系列的程序中。其最初是为了网络抓取所设计的, 也可以应用在获取 API 所返回的数据(例如 Amazon Associates Web Services) 或者通用的网络爬虫中。

Scrapy 的一个最主要的优势: 请求(request)是被异步调度和处理的。这意味着, Scrapy 并不需要等待一个请求(request)完成及处理, 在此同时, 也可以发送其他请求或者做些其他事情。同时这也意味着, 当有些请求失败或者处理过程中出现错误时, 其他的请求也能继续处理。

在允许使用者可以以非常快的速度进行爬取时(以容忍错误的方式同时发送多个 request), Scrapy 也通过一些设置来允许使用者控制其爬取的方式。例如, 可以为两个 request 之间设置下载延迟, 限制单域名(domain)或单个 IP 的并发请求量, 甚至可以使用自动限制插件来自动处理这些问题。

Scrapy 的其他特性:

- 1) 对 HTML, XML 源数据选择及提取的内置支持, 提供了 CSS 选择器(selector)以及 XPath 表达式进行处理, 以及一些帮助函数(helper method)来使用正则表达式来提取数据。
- 2) 提供交互式 shell 终端, 为测试 CSS 及 XPath 表达式, 编写和调试爬虫提供了极大的方便
- 3) 通过 feed 导出提供了多格式(JSON、CSV、XML), 多存储后端(FTP、S3、本地

文件系统)的内置支持。

- 4) 提供了一系列在 spider 之间共享的可复用的过滤器(即 Item Loaders), 对智能处理爬取数据提供了内置支持。
- 5) 针对非英语语系中不标准或者错误的编码声明, 提供了自动检测以及健壮的编码支持。
- 6) 高扩展性。可以通过使用 signals , 设计好的 API(中间件, extensions, pipelines) 来定制实现功能。
- 7) 内置的中间件及扩展为下列功能提供了支持: * cookies and session 处理 * HTTP 压缩 * HTTP 认证 * HTTP 缓存 * user-agent 模拟 * robots.txt * 爬取深度限制。
- 8) 内置 Telnet 终端, 通过在 Scrapy 进程中钩入 Python 终端, 使用者可以查看并且调试爬虫
- 9) 以及其他一些特性, 例如可重用的, 从 Sitemaps 及 XML/CSV feeds 中爬取网站的爬虫、可以自动下载爬取到的数据中的图片(或者其他资源)的 media pipeline、带缓存的 DNS 解析器, 等等。

2.5. PyQt 介绍

本课题将要实现的短租房应用平台服务质量数据采集分析系统将使用 PyQt 工具包创建 GUI 用户界面。

PyQt 是一个创建 GUI 应用程序的工具包。它是 Python 编程语言和 Qt 库的成功融合。Qt 库是目前最强大的库之一。PyQt 是由 Phil Thompson 开发。

PyQt 实现了一个 Python 模块集。它有超过 300 类, 将近 6000 个函数和方法。它是一个多平台的工具包, 可以运行在所有主要操作系统上, 包括 UNIX, Windows 和 Mac。PyQt 采用双许可证, 开发人员可以选择 GPL 和商业许可。在此之前, GPL 的版本只能用在 Unix 上, 从 PyQt 的版本 4 开始, GPL 许可证可用于所有支持的平台。

因为可用的类有很多, 他们被分成几个模块。QtCore 模块包含核心的非 GUI 功能。该模块用于时间、文件和目录、各种数据类型、流、网址、MIME 类型、

线程或进程。**QtGui** 模块包含图形组件和相关的类，例如按钮、窗体、状态栏、工具栏、滚动条、位图、颜色、字体等。**QtNetwork** 模块包含了网络编程的类，这些类允许编写 TCP/IP 和 UDP 的客户端和服务端，他们使网络编程更简单，更轻便。**QtXml** 包含使用 XML 文件的类，这个模块提供了 SAX 和 DOM API 的实现。**QtSvg** 模块提供显示的 SVG 文件的类。可缩放矢量图形（SVG）是一种用于描述二维图形和图形应用程序的 XML 语言。**QtOpenGL** 模块使用 OpenGL 库渲染 3D 和 2D 图形，该模块能够无缝集成 Qt 的 GUI 库和 OpenGL 库。**QtSql** 模块提供用于数据库的类。

2.6. 本章小结

本章主要介绍了在系统实现过程中运用到的各种技术框架。首先介绍了爬虫技术的发展和原理。其次在数据爬取方面，作者选用了当下比较热门的 Python 编写爬虫。运用爬虫框架 Scrapy 和正则表达式确保爬虫代码的高效和稳定性。最后整个系统的用户交互界面 GUI 用 PyQt 工具包编写。

因为篇幅有限，很多技术理论没有做更具体的介绍和详细的体现，会在后面的章节里运用时做说明和补充。

第三章. 短租房应用平台服务质量数据采集分析系统的需求分析

该项目主要通过爬虫获取短租房应用数据,进而存储和处理数据保存到本地;再通过制定的分析规则实现服务质量的数据分析和分析的可视化呈现。本章主要是对该短租房应用平台服务质量数据采集分析系统的需求分析。先从整体角度梳理系统的业务需求,进而进行系统功能性需求和非功能性需求的分析。在分析过程中,具体到每一个子系统的用例模块、指出系统在设计和实现时的目标和行为,根据具体的短租房应用平台的网页结构及数据特点,阐述用例的流程,为系统的设计和 implement 提供方向和思路。

3.1. 系统业务需求

系统的业务需求是根据系统本身要达到的特性或要求,以及系统的使用者(操作者、用户)来决定的。从这两个维度去梳理系统的业务需求,才能更周全地综合各方面的要求,形成比较全面的系统需求分析计划。

3.1.1. 系统特性分析

1) 可复用性

一般来说,网络爬虫的设计是根据具体要爬取数据来源的网站的页面结构来设计编写的,当想要爬取其他的网站时爬虫往往需要重新设计,可复用性较低。本系统在设计时应该尽量提高可复用性,即使用比较成熟的第三方爬虫库如 `Scrapy` 等,一方面减少自己代码编写的工作量,另一方面提高核心部分的可复用性,使系统后续的迭代更新更方便。

2) 分布式爬虫的设计

目前简单的爬虫一般采用单节点的方式来部署,效率并不高,并不能支持多线程的并发的爬宠任务,在数据量较大的爬虫工作中速度很慢。因此本系统可以考虑采用分布式的爬虫设计。

3.1.2. 用户需求分析

短租房应用平台服务质量评价系统数据采集分析系统的使用人员组成比较复杂，有系统的开发者，可能的使用人员还有短租房的消费者、短租房应用平台的企业管理人员、提供短租房的房东和行业内的分析人士。因此在系统分析时应当充分考虑到满足这些角色需求上的差异。

1) 配置简单

除了系统的开发者之外，其他使用人员在使用系统的过程中并不关心具体配置的操作，因此系统在安装和更新时应当尽量简单。若系统需要使用到一些第三方的插件或打包的库，最好能集成到系统中提供一站式的安装和配置环境，使得使用人员能够花最少的时间去完成系统的初始化和配置环节，集中精力去解决他们关注的核心环节。

2) 交互简单

清晰的交互逻辑、简洁美观有序的操作界面能给使用者良好的印象和使用体验。让使用者明确每一个按钮和操作具体能达到什么目的，下一步指向哪一块的操作，这样就要求程序在设计时逻辑清晰，界面的设计遵循设计原则。交互简单能在很大程度上减少可能出现的问题。

3) 用户权限

因为使用人员的组成较为复杂，在使用过程中可能会出现想要限制部分人员使用部分功能的情景，这对于系统的封装和一体化也是有利的。因此考虑加入用户管理，要求用户登录验证权限后使用相关功能。

3.2. 系统功能性需求

本节从系统功能的分析方向出发，分析系统的功能性需求。本课题将要完成的短租房应用平台服务质量数据采集分析系统是一个集数据爬取、数据处理和数据分析的系统，并针对使用群体提供用户管理的功能。因此该短租房应用平台服务质量数据采集分析系统主要分为几大功能模块：数据抓取功能、数据处理功能、数据分析功能和用户管理功能。本节为这几个功能模块做了细致的划分，对于其中的用例和功能流程做了阐述，对于功能要点的具体含义和职责进行了设计和分析说明。

3.2.1. 数据抓取功能需求分析

数据抓取功能模块（子系统）主要实现用户在系统中输入特定的 URL，通过系统发起 URL 爬虫请求。系统判定 URL 通过后开始获取数据。

3.2.1.1. 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 3.2.1 提交 URL 爬虫请求用例描述

系统管理员进入系统，进入数据抓取子系统界面，在文本框中输入将要抓取数据的 URL。系统需要对 URL 脚本进行校验检查，如果系统中有该 URL 对应的爬虫脚本则校验通过，若没有则提示校验失败，需要用户重新输入 URL。校验通过后开始爬取，显示爬取等待界面，系统开始在后台进行 URL 和对应页面数据的爬取。（如表 3.2.1）

3.2.2. 数据处理功能需求分析

3.2.2.1. 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 3.2.2 存储转化数据用例描述

在开始爬取数据后，系统根据爬虫代码下载 URL 对应页面的内容，对于页面里的内容进行分类解析。解析出新的 URL 和图片、字符串和文本等信息。对于爬取到的 URL，系统将记录下来并暂时存储，等待循环结束后对于这些 URL 进行判重和多次解析，若判重通过则继续对该 URL 进行爬取，直到爬虫脚本执行完毕，获取到了相应的信息。

对于图片、字符串和文本等爬取到的信息按照正则表达式进行匹配，系统

将匹配后的信息进行下载保存到数据库中。数据库接收到相应的信息后对该数据进行整理，并拆分成若干对应的数据表。（表 3.2.2）需要说明的是以上步骤都是系统在后台自动完成的，除了中途的报错等信息需要用户确认后中断爬取和数据处理，环节之间的衔接对于用户来说是不可见的。数据处理完成后，系统返回相应信息。

3.2.3. 数据分析功能需求分析

数据分析功能是关系到短租房应用平台服务质量数据采集分析系统结果可视化输出的子系统模块。主要实现用户对于现有的分析规则的维护（包括分析规则的创建、删除和修改）以及指定数据，提交现有的分析规则进行数据分析并输出可视化结果。

3.2.3.1. 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 3.2.3 创建分析规则用例描述

拥有特定权限的用户进入系统的数据分析子系统界面，选择创建分析规则。在系统提供的分析规则选项中依次选择，以组合成新的分析规则。点击验证以确保分析规则的合理性，验证通过后，系统返回相应消息提示分析规则创

建成功。(表 3.2.3)

3.2.3.2. 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 3.2.4 修改分析规则用例描述

拥有特定权限的用户进入数据分析子系统界面, 在要修改的分析规则后选择修改, 对于其中具体的分析规则项进行调整改动, 然后进行系统验证保证分析规则的合理性。验证通过后保存, 系统返回相应提示消息以说明分析规则修改成功。(表 3.2.4)

3.2.3.3. 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 3.2.5 删除分析规则用例描述

拥有特定用户权限的用户进入数据分析子系统界面，在要删除的分析规则后点击删除按钮，确认删除后，系统返回相应的消息，以说明分析规则删除成功。

3.2.3.4. 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 3.2.6 提交分析请求用例描述

拥有特定权限的用户进入数据分析子系统界面，在要提交分析的分析规则后点击提交，系统接下来判断数据是否有对应的数据并且已经处理完成。判断通过后，进入等待界面，系统在后台根据选择的分析规则进行分析处理。分析结束后输出数据分析的可视化结果（图片或表格数据等），以供用户参考使用。

3.2.4. 用户管理功能需求分析

用户管理功能主要是对用户信息和用户权限的管理,以实现系统多元化的使用群体能够使用系统的不同功能,并对于一些涉及系统安全 and 信息安全的数据和信息起限制浏览的保护作用。主要功能有用户信息的维护(包括新用户的注册、用户信息的删除和更新)以及用户权限的管理。

3.2.4.1. 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 3.2.7 注册新用户用例描述

系统管理员或用户进入用户管理子系统界面, 选择注册新用户。在注册新用户的界面中填写对应的用户信息(用户名、密码等), 点击注册, 系统返回对应的消息提示, 以表明新用户注册成功。

3.2.4.2. 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 3.2.8 删除用户信息用例描述

系统管理员进入用户管理子系统界面，进入查看系统中已注册用户界面中，在要删除的用户的后面点击删除按钮，系统返回确认删除的对话框。确定删除后，系统返回删除成功消息。

3.2.4.3. 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 3.2.9 更新用户信息用例描述

系统管理员或用户进入用户管理子系统界面。系统管理员在查看系统中已注册用户中对要修改的用户进行信息的修改、保存更新。用户则在个人信息界面进行信息的修改、保存和更新。

3.2.4.4. 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 3.2.10 用户权限管理用例描述

超级管理员进入用户管理子系统界面，选择用户权限管理，在界面中对应的权限选项后添加或者删除用户的 ID 以分配或者撤销相应用户的权限。一个权限可以分配给多个用户，同时，一个用户可以拥有多个权限。在权限以外的界面，用户将无法访问和修改。点击保存后，系统返回保存成功的对话框，以说明用户权限修改完成。

3.3. 系统非功能性需求

该短租房应用平台服务质量数据采集分析系统作为一个完成的爬虫和数据分析的信息系统，在设计时还需要考虑满足以下非功能性需求。

- 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。

3.4. 本章小结

本章主要讨论了短租房应用平台服务质量数据采集分析系统的需求分析，为系统的设计和后续实施提供了思路和指导方向。内容上，本章从业务需求、功能性需求和非功能性需求三个方面进行需求的分析。其中系统业务需求分析过程中结合系统的特性和用户的需求两个方面的要求。在功能性需求分析过程中，对系统的数据抓取、数据处理、分析规则管理和用户管理这四个系统最主要的功能模块，结合用例描述做了功能的说明。在非功能性需求方面，提出了效率、安全和健壮性的三个优化建议，以便系统后续流程更好地进行。

第四章. 短租房应用平台服务质量数据采集分析系统的系统概要设计

本章是短租房应用平台服务质量数据采集分析系统的概要设计，在写作思路本章采用从总体到部分的方式去展开，先给出该短租房应用平台服务质量数据采集分析系统的整体系统架构设计，然后根据上一章中的需求分析要求，对具体的系统功能做设计阐述。最后根据系统数据存储方面的要求对系统的数据库设计做了说明。

4.1. 系统架构设计

应用架构是用于实现信息系统的技术规范的描述。在系统设计过程的初期，需要开发一个架构蓝图，以作为后面的外部和内部设计的提纲。

该课题研究设计的短租房应用平台服务质量数据采集分析系统是为更好地推动短租房应用平台的服务质量评价体系建立，以及更好地满足短租房应用的消费者和使用者的服务需求而设计的。当前短租房市场上的应用琳琅满目，房源信息也良莠不齐。对于这部分的数据的信息采集，初期限定为从途家这款短租房的应用的 PC 端来获取。

为了应对和满足多样化的网络数据存储和更新要求，以及使用者对于该系统应用的便捷性要求，系统的网络架构采用基于因特网的计算架构。这是一种多层的架构设计方案，其中表现层和表现逻辑层是在客户端的 Web 浏览器中使用从某一个 Web 服务器下载的内容来实现。在这之后，表现逻辑层连接到运行在应用服务器上的应用逻辑层。最终，它连接到后台的数据库服务器。

该短租房应用平台服务质量数据采集分析系统的网络架构图如下所示（图 4.1.1）。它说明了用户通过客户端或浏览器与服务器之间的连接行为，以及用户与处理器交互的方式。

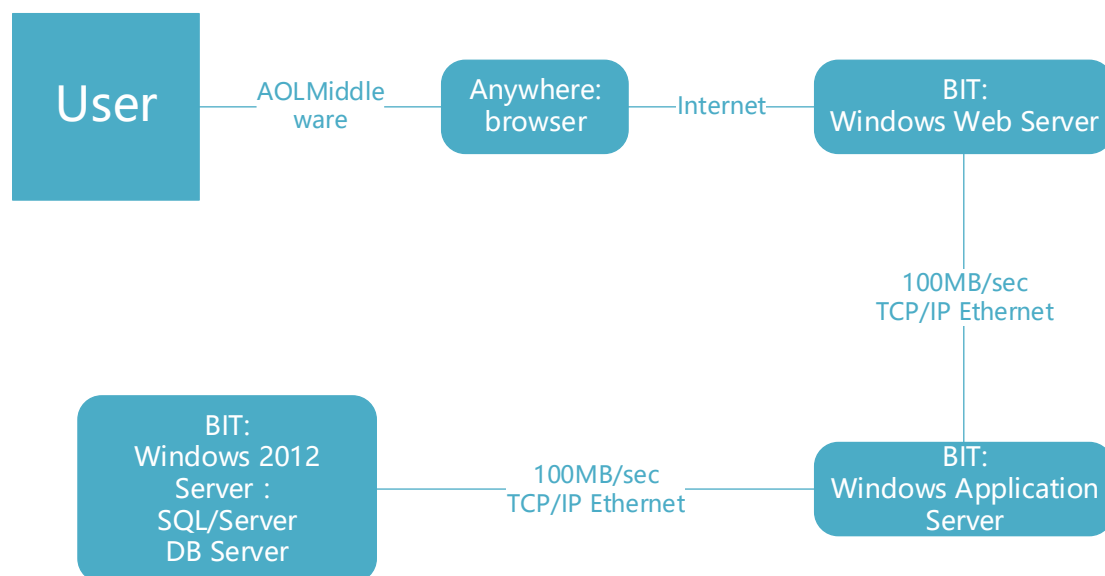


图 4.1.1 系统网络架构图

如上图所示，系统的用户通过终端在任意地点登录浏览器，浏览器通过访问因特网访问部署在 BIT 的网页服务器。网页服务器通过百兆光纤和 TCP/IP 协议与同样部署在 BIT 的应用服务器交互，从而传输和处理指令。应用服务器同样通过百兆光纤和 TCP/IP 协议访问数据库服务器，获取与指令相对应的数据并在应用服务器中处理，返回给网页服务器，最终呈现给用户。该系统还具有后续的延展和迭代的空间，部署在 BIT 的数据库服务器可以通过 WAN 联入数据库广域网，即若未来系统的现有数据存储已经无法满足大规格数据的处理要求，建立起分布式的数据库服务器后，该架构的末端可以接入成为这个大的分布式系统的一部分而工作。

4.2. 系统功能设计

本小节从系统功能的角度出发，结合上一章系统功能性需求的分析，将该短租房应用平台服务质量数据采集分析系统划分为若干个子系统，对主要的功能流程进行说明描述，对功能点的具体含义及其职责进行了设计。本节主要根据系统设计类图，对系统的各个功能点进行展开说明。

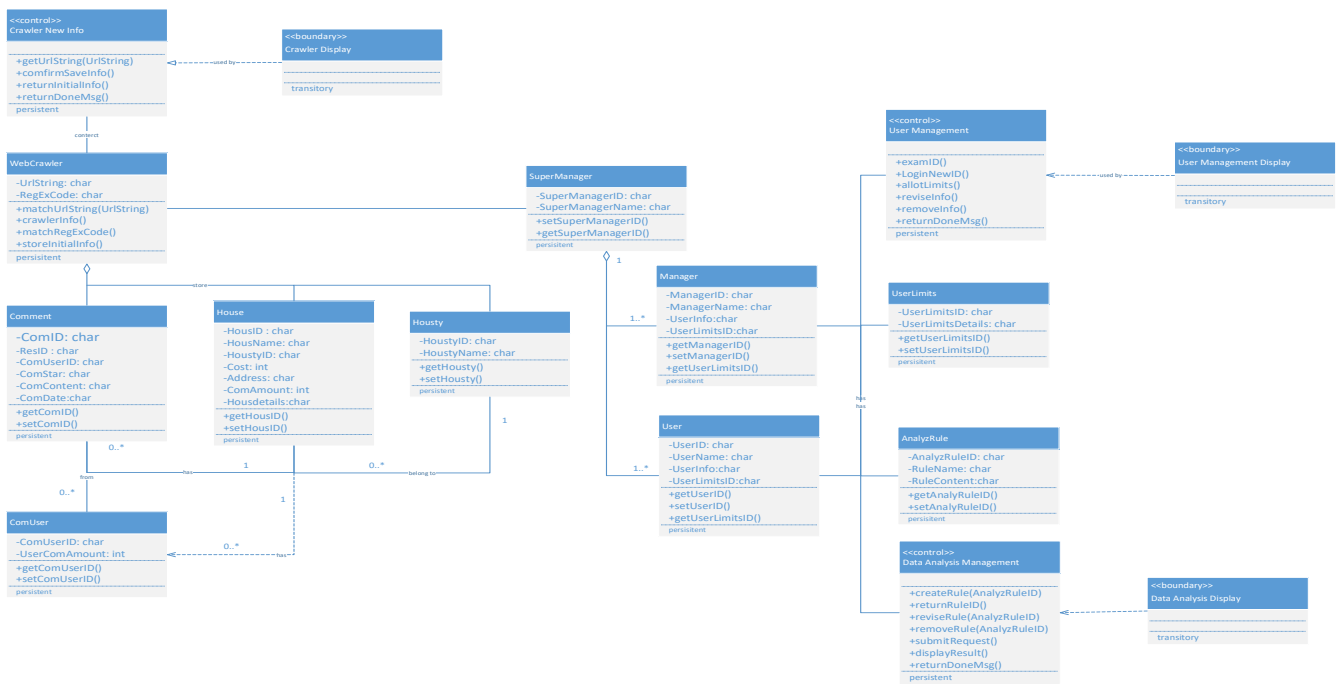


图 4.2.1 系统设计类图

4.2.1. 数据抓取功能

数据抓取功能主要实现该短租房应用平台服务质量数据采集分析系统应用爬虫，从输入系统的 URL 中获取对应的信息文本。系统中与实现数据抓取功能有关的类主要是 WebCrawler。

WebCrawler 为数据爬取子系统的数据爬取类，主要属性有：

- 1) UriString 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

主要方法有：

- 1) matchUriString 该方法用于匹配系统中是否有对应的 URL 的爬虫脚本
- 2) crawlerInfo 该方法主要用于匹配通过后，系统执行爬虫命令从对应的网页获

取数据

- 3) **matchRegExCode** 该方法主要用于在爬虫过程中匹配正则表达式得到对应格式的数据
- 4) **storeInitialInfo** 该方法用于储存爬取到的信息文本等数据

4.2.2. 数据处理功能

数据处理功能主要实现系统在执行了数据抓取之后，自动存储到数据库中的系统行为。具体在数据库中的存储方式见本章的 4.3 小节系统数据库的设计。在系统设计类图中相关的方法为上一小节中的 **storeInitialInfo**，该方法用于储存爬取到的信息文本等数据。

4.2.3. 数据分析功能

数据分析功能主要实现系统数据分析规则的维护（包括创建规则、删除规则、更新规则）和分析请求的提交与分析的实现。分析规则是基于爬取到的用户评论的词频的统计学量化指标，或者是基于信息的基本字段的筛选与串接。系统中与数据分析功能相关的类主要是 **AnalyzRule**、控制类 **Data Analysis Management** 与界面交互类 **Data Analysis Display**。

- 1) **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) **Data Analysis Management** 为控制类，主要提供分析规则的增删改和提交分析请求的具体方法，以及返回对应的提示消息的方法。
- 3) **Data Analysis Display** 为界面交互类，主要实现在界面呈现提交分析请求后，系统数据分析的可视化结果的呈现，以供用户参考。

4.2.4. 用户管理功能

用户管理功能主要为该短租房应用平台服务质量数据采集分析系统的使用者提供用户管理(包括新用户的注册、用户信息的修改、删除和用户权限的修改)功能,给不同的使用群体呈现不同的系统功能,提高系统功能的安全性和完整性。系统中,与用户管理功能对应的类主要有 **SuperManager**、**Manager**、**User**、**UserLimits** 和控制类 **User Management**。

- 1) **SuperManager** 为系统的超级管理员,一般为系统的开发者和维护者,负责配置系统的各项内置属性(例如正则表达式、爬虫脚本的维护、分析规则选项的设置等),主要属性有超管的 ID 和其姓名。
- 2) **Manager** 为系统的管理员,负责用户信息的管理,可以参与修改用户信息和分配用户的权限。管理员也是用户的一员,主要属性有管理员 ID、管理员的姓名、管理员的信息和其权限。
- 3) **User** 为系统的用户,作为系统中最普遍人数最多的角色,用户在默认情况下可以参与系统最基本、最主要的功能,即参考数据分析的结果。用户在第一次登陆时需要注册。主要的属性有用户 ID、用户姓名,用户信息和用户对应的权限。
- 4) **UserLimits** 为用户权限类,用户权限 ID 由一个数值表示,数值越高权限的等级越高,能够访问的系统功能,参与的系统流程也越多。一般默认一般用户的权限最低,超级管理员的系统权限最高。该类主要的属性有用户权限的 ID 和权限对应的说明。
- 5) 控制类 **User Management** 主要实现用户信息的维护和权限的分配,提供方法给对应权限的操作者以实现注册新用户、用户登录、权限分配、信息修改、信息删除和返回相应的系统提示消息。

4.3. 系统数据库设计

短租房应用平台服务质量数据采集分析系统采用关系型设计库，该系统对于数据库的应用主要是数据抓取模块抓取到的数据，在数据处理模块的存储，以及系统用户管理功能模块涉及到的用户信息的存储。

本小节主要说明数据处理模块存储过程中，建立的数据库表。

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表，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 为房屋表，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 类为房型表，主要属性有：房型 ID 和房型名称
- 4) ComUser 类为评论者表，主要属性有：评论者的 ID 和该评论者发布过的评论数。

4.4. 本章小结

本章作为短租房应用平台服务质量数据采集分析系统的概要设计，从系统架构设计，系统功能设计，系统的数据库设计三个方面进行了阐述。系统架构设计主要从系统的网络架构图，对整个系统应用的技术和架构做了描述，系统功能设计从设计类图为视角出发，对数据抓取、数据处理、数据分析和用户管理这四个主要功能的子功能做了梳理。系统的数据库设计部分主要介绍了系统的数据处理模块存储的数据表。本章作为承上启下的一章，为系统的实施和后续设计指明了方向。

第五章. 短租房应用平台服务质量数据采集分析系统的系统详细设计与实施

5.1. 主要功能设计与实施

5.1.1. 数据抓取功能

5.1.2. 数据处理功能

5.1.3. 数据分析功能

5.1.4. 用户管理功能

5.2. 本章小结

第六章. 系统验证

6.1. 系统业务功能验证

6.2. 系统性能验证

6.3. 本章小结

第七章. 结论

7.1. 全文总结

7.2. 不足与展望

参考文献

- [1]Limam N, Boutaba R. Assessing Software Service Quality and Trustworthiness at Selection Time[J]. IEEE Transactions on Software Engineering, 2010, 36(4):559-574.
- [2]Steffi. “小猪短租”:开启共享经济新时代[J]. 商业文化, 2016, (19):42-47.
- [3]崔博. Airbnb 模式在中国能否成功?[J]. 经济, 2015, (10):21-25.
- [4]马君海. 布瑞恩与 Airbnb.com[J]. IT 经理世界, 2014, (19):102-104.
- [5]薛奕妹. 我国短租房旅游市场发展初探[J]. 旅游纵览(下半月), 2014, (09):44-45.
- [6]李东娟, 熊胜绪. 酒店顾客感知服务质量的影响维度[J]. 统计与决策, 2014, (08):185-188.
- [7]李东娟. 酒店其他顾客行为对顾客感知服务质量的影响[J]. 旅游学刊, 2014, (04):48-54.
- [8]隋婷婷. 从用户体验角度探析在线短租的发展情况及对策[J]. 时代金融, 2014, (02):71-72.
- [9]沈珂. 探析我国的在线短租市场[J]. 今日中国论坛, 2012, (11):35-36.
- [10]熊伟, 高阳, 吴必虎. 中外国际高星级连锁酒店服务质量对比研究——基于网络评价的内容分析[J]. 经济地理, 2012, (02):160-165.
- [11]文吉, 曾婷婷. 主题酒店顾客感知服务质量与购后行为的关系研究——基于深圳市主题酒店的实证研究[J]. 人文地理, 2011, (04):127-131.
- [12]粟娟, 麻学峰, 李佩耕, 向国娟. 基于顾客价值的酒店顾客满意度测评[J]. 统计与决策, 2008, (05):73-76.
- [13]张德然, 何鹏光. 现代酒店服务质量标准的量化及综合评价[J]. 统计与决策, 2005, (18):137-139.
- [14]Su C S, Sun L H. Taiwan's hotel rating system: a service quality perspective.[J]. Cornell Hospitality Quarterly, 2007, 48(4):392-401.
- [15]沈涵. 基于 ACSI 的经济型酒店顾客满意度测评模型[J]. 旅游学刊, 2011, (01):58-62.

- [16] 沈涵, 郭旻. 经济型酒店顾客满意度指标体系研究[J]. 消费经济, 2010, (04):30-33.
- [17] 丁于思, 肖轶楠. 基于网络点评的五星级酒店顾客满意度测评研究[J]. 经济地理, 2014, (05):182-186+192.
- [18] 焦明宇. 基于顾客价值的经济型酒店顾客满意度测评研究[J]. 旅游学刊, 2014, (11):80-86.
- [19] 熊伟, 高阳, 吴必虎. 中外国际高星级连锁酒店服务质量对比研究——基于网络评价的内容分析[J]. 经济地理, 2012, (02):160-165.
- [20] 沈涵, 吴文庆. 服务企业顾客满意度与再购买倾向的模型研究——以四星级酒店为例[J]. 旅游学刊, 2011, (09):85-89.
- [21] 吴雪飞. 经济型酒店顾客满意度测量模型及实证研究[J]. 浙江学刊, 2010, (03):192-195.

致谢