安徽财经大学字

**本科毕业设计**

|  |  |
| --- | --- |
| **题 目** | **基于Python的北上广住房数据分析与可视化** |
| **学 院** | **管理科学与工程学院** |
| **专 业** | **计算机科学与技术** |
| **班 级** | **20计科四班** |
| **学 号** | **20202568** |
| **姓 名** | **韩志鹏** |
| **指导老师** | **张晓春** |

**2021 年 5 月**

安徽财经大学管理科学与工程学院

本科生毕业论文（设计）诚信承诺书

本人承诺：

1.所呈交的毕业论文（设计）《 毕业设计过程管理系统》，是在认真学习理解《安徽财经大学学位论文作假行为处理办法》和《管理科学与工程学院本科毕业论文（设计）工作管理办法》后，保质保量独立完成的，没有弄虚作假，没有抄袭别人的内容；

2.毕业论文（设计）所使用的相关资料、数据、观点等均真实可靠，文中所有引用的他人观点、材料、数据、图表均已注释说明来源；

3.毕业论文（设计）中无抄袭、剽窃或不正当引用他人学术观点、思想和学术成果，伪造、篡改数据的情况；

4.本人已被告知并清楚：学院对毕业论文（设计）中的抄袭、剽窃、弄虚作假等违反学术规范的行为将严肃处理，并可能导致毕业论文（设计）成绩不合格，无法正常毕业、取消学士学位资格或注销并追回已发放的毕业证书、学士学位证书等严重后果；

5.若在省教育厅、学校、学院组织的毕业论文（设计）检查中，被发现有抄袭、剽窃、弄虚作假等违反学术规范的行为，本人愿意接受学院按有关规定给予的处理，并承担相应责任。

学生（签名）：

年 月 日

指导老师（签名）：

年 月 日

**中文标题**

**（宋体 二号 加粗 1.5倍行间距 段前段后0.5行 居中）**

**摘 要**

正文：小四 宋体 1.5倍行距 段落首行缩进2字符

**关键词：xx**；xx

**Title**

**（times new roman 二号 加粗 1.5倍行间距 段前段后0.5行 居中）**

**Abstract**

正文：小四 times new roman 1.5倍行距 首行缩进2字符

**Keywords： xx; xx**

# 目 录

**一级标题三号加粗 行间距1.5倍 段前段后0.5行或6磅**

**中文宋体，英文和字符 times new roman**

# 1 引言

主要写开发该系统的背景、目的和意义（比摘要要详细和全面）、软件的定义、软件的开发工具，阐述课题设计的创新点、实用价值等。

(一级标题 宋体 三号 加粗 1.5倍行间距 段前段后各0.5行 左对齐)

正文：小四 宋体 1.5倍行距 首行缩进2字符

# 2 可行性研究

概述段落

## 2.1 需求概述

北上广城市住房对中国其他城市的住房及以房地产为主导的经济起到定位和引导作用，研究北上广房价关系到中国基层人民的生活质量。

此系统通过收集北上广地区住房信息，结合用户确切的地区需求，通过数据清洗和可视化将当地住房信息具象化展现给用户。（用户需要先进行登录才能操作）。



系统流程图

粗略分析系统操作流程以及功能和特色。

(二级标题 宋体 小三号 加粗 1.5倍行间距 段前段后各0.5行 左对齐)

结合系统流图或业务流程图概述系统需求及物理组成。

## 2.2 系统可行性

### **2.2.1经济可行性**

(三级标题 宋体 四号 加粗 1.5倍行间距 段前段后各0.5行 左对齐)

**（1）工作量估算**

在软件开发阶段需要使用到的人力工作量百分比如下表2-1所示。

表2-1毕业设计过程管理系统各个开发阶段的人力百分比

（表格标题： 中文黑体 英文times new roman 五号 不加粗 表格上方 居中对齐）

|  |  |
| --- | --- |
| 任务 | 人力（%） |
| 可行性研究 | 5 |
| 需求分析 | 10 |
| 概要设计和详细设计 | 25 |
| 编码和测试 | 60 |
| 总计 | 100 |

**（2）成本估算**

在软件开发阶段需要其他一次性支出如下表2-2所示。

表2-2软件开发过程中各个开发阶段的一次性支出

|  |  |
| --- | --- |
| 项目 | 费用（元） |
| 系统前期需求研究 | 500 |
| 开发计划与测试基准研究 | 500 |
| 数据库的建立与数据字典 | 1500 |
| 检查费用和管理性费用 | 1500 |
| 培训费及软件开发人员所需的一次性支出 | 1000 |
| 总计 | 5000 |

表2-3软件开发过程中成本估算

|  |  |  |  |
| --- | --- | --- | --- |
| 项目 | 单价 | 数量 | 费用 |
| PC机 | 5000 | 2 | 10000 |
| 打印机 | 1000 | 1 | 1000 |
| 管理员工资 | 2000元/月 | 12 | 24000 |
| 不可预知费用 | \*\*\* | \*\*\* | 10000 |
| 一次性支出 | \*\*\* | \*\*\* | 5000 |
| 总计 | | 50000 | |

**（3）效益**

表2-4软件开发过程中各个开发阶段的效益

|  |  |
| --- | --- |
| 项目 | 收益（元/年） |
| 一次性收益 | 无 |
| 经常性收益（下载文献） | 80000 |
| 不可定量收益 | 无 |
| 企业定制服务 | 20000 |
| 总计 | 100000 |

**（4）收益/投资比**

一次性支出：26000元

经常性支出：24000元/年

收益：100000元/年

收益/投资比：100000\*5/(26000+24000\*5)=3.424657

**（5）货币的时间价值**

五年预计收益：100000\*5-（26000+24000\*5）=354000（元）

住房数据分析系统的前期软件开发阶段，需要耗费巨大的人力，投资的成本和未来获得的效益会随着软件每年的货币时间价值有所改变。住房数据分析系统五年收益为354000元。实际收益需要通过货币时间价值来进行核算。用利率的形式表示货币的时间价值。假设年利率为i，如果现在存入P元，则n年后可以得到的钱数为：

H:\temp\ksohtml14132\wps1.jpg (1)

这也就是P元钱在n年后的价值。反之，如果n年后能收入F元钱，那么这些钱的现在价值是：

H:\temp\ksohtml14132\wps2.jpg (2)

假定年利率为12%，利用上面计算货币现在价值的公式可以算出系统5年预计收益的现在价值，如下表2-5所示。

表2-5 将来收入折算成现在值

|  |  |  |
| --- | --- | --- |
| 将来值（元） | （1+i)^n | 现在值（元） |
| 354000 | 1.7623 | 200873 |

**（6）投资回收期估算**

住房数据分析系统基础建设成本约为26000元，第一年内收入为100000元，软件投资的回收期约为

 （3)

因此软件的投资回收期约为0.26年，本系统开发成本较低，维修成本适中，软件投资回报期较短可以较快获得利润，用户群体丰富且需求量大，经济利益客观值得投资。

### **2.2.2技术可行性**

爬取技术网络爬虫（Web Crawler）是基于 http 协议中的请求与响应机制，按照一定的规则自动提取 Web 网页的应用程序或者脚本，可以完整爬取整个网页的源代码。开发者通过相关 设备发送请求并传输 url 到目标服务器，目标服务器通过相关处理和优化分析后返回访问者所需的信息，访问者得到相关 HTML 信息后可利用 Beautiful Soup、正则表达式、Xpath 等技术获取想要的网页固定信息。

数据清洗和可视化技术 Python 作为一种新兴语言为开发者们提供了两种常用的数据处理方式，即 Pandas 和 Matplotlib。Pandas 除了可以提取和保存数据以外，还支持不同数据结构间进行相互转换。Matplotlib 可以将数据以图表形式呈现出来，更为直观地展示给用户，进 而将数据的特征显示出来。在爬取数据前，第一步选择进行 数据采集的网站，第二步对选择的目标网站结构进行分析，进 而找到合适的请求方式解析请求返回的响应，同时将爬取的数据保存到本地，从而为进行数据分析做好下一步准备工作。

### **2.2.3 操作可行性**

操作可行性是对开发系统在一个特定的工作环境中能否运行或运行的好坏程度的衡量。该系统采用前端web页面显示可视化结果，界面友好，操作简单，用户无需掌握相关技术，即可进行相关操作，故该系统具有操作可行性。

## 2.3 项目进度计划

示例：

表2-5软件开发过程中各个开发阶段的一次性支出

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **项目阶段** | **时间** | **工作内容** | **成果** | **负责人** | **审核人** |
| 需求调研 | 2022.10.1-2022.10.15 | 对项目需求进行详细调研 | 系统需求规格说明书 | 韩志鹏 | 韩志鹏 |
| 系统设计 | 2022.10.16-2022.10.31 | 在需求调研的基础上对系统构架、安全体系、功能等进行系统设计 | 系统设计说明书 | 韩志鹏 | 韩志鹏 |
| 系统开发 | 2022.11.1-2022.11.15 | 进行各个子系统迭代开发，完成单元测试 | 不同迭代版本的可运行系统 | 韩志鹏 | 韩志鹏 |
| 系统集成 | 2022.11.16-2022.11.30 | 系统集成和对各模块集成测试 | 测试报告，形成可完整运行的系统 | 韩志鹏 | 韩志鹏 |
| 系统初验 | 2022.12.1-2022.12.15 | 项目初验 | 初验报告 | 韩志鹏 | 韩志鹏 |
| 系统试运行 | 2022.12.16-2022.12.31 | 1. 系统试运行 2. 系统持续优化 | 升级版本的可运行系统并安装部署到用户本地 | 韩志鹏 | 韩志鹏 |
| 用户培训 | 2023.1.1-2023.1.15 | 对各级用户进行培训 | 使用报告 | 韩志鹏 | 韩志鹏 |
| 项目终验 | 2023.16-2013.2.1 | 项目终验 | 验收报告 | 韩志鹏 | 韩志鹏 |

# 3 需求分析

概述段落

## 3.1 系统综合需求

### **3.1.1功能需求概述**

分析并完成用户指定地区(北上广)住房数据的可视化

### **3.1.2 系统性能需求概述**

用户能够登录网站并得到分析结果

### **3.1.3 其他需求概述**

无

## 3.2 数据流图分析

### **3.2.1 顶层数据流图**



系统的源点和终点都是用户，用户通过登录进入住房数据分析系统，系统处理用户指令后将分析结果返回给用户

### **3.2.2 数据流图细化与分解**

#### （1）0层数据流图

管理员具有以下功能：

1. 用户信息管理功能：对用户基本信息进行检索、增删改和权限修改；
2. 电影数据系统中所有信息进行查询和增删改管理；
3. 快捷操作和帮助模块：一些系统的辅助操作，便于用户使用系统；

普通用户则可以通过登录该系统，实现电影信息的查询。VIP用户在普通用户的基础上增加住房推荐功能。



#### （2）1层数据流图







#### (3)2层数据流图





## 3.3 数据字典

（由于数据字典定义内容非常多，可针对每一小类选择代表性的示例展示）

### **3.3.1数据流条目定义**

实例：

表3-X 管理员信息数据字典定义

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 数据流 | | | | |
| 系统名：AUFE学生选课系统 | 编号：D-001 | | | |
| 条目名：管理员信息 | 别名： | | | |
| 来源：管理员 | 去处：管理员 | | | |
| 数据流结构：  管理员：{管理员账号+密码+邮箱+手机号码}所有管理员 | | | | |
| 简要说明：管理员数据流记录着管理员的基本信息，可用于管理员的信息注册、查看、修改，登录验证等处理 | | | | |
| 修改记录 | 编写 | 喻杰 | 日期 | 2022.10.7 |
| 审核 | 刘士强 | 日期 | 2022.10.8 |

根据3.2节数据流图分析情况可知，系统主要涉及的数据项包括：xx, xx, xx, xx和xx分别定义如图3-X~3-Y所示。

（每个类型的数据字典定义前应有一段总结性的开场白。）

### **3.3.2数据存储条目定义**

实例：

表3-X 学生记录数据字典定义

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 数据存储 | | | | |
| 系统名：AUFE学生选课系统 | 编号：D-006 | | | |
| 条目名：学生记录 | 别名： | | | |
| 存储组织：每个学生一条记录 | 记录数：10^8 | | 主关键字：学生学号 | |
| 记录组成：  项名：学号 密码 手机号码 邮箱 学院 专业 班级 姓名  长度： 8 20 11 30 15 15 10 10 | | | | |
| 简要说明：存储每个学生的基本信息 | | | | |
| 修改记录 | 编写 | 喻杰 | 日期 | 2022.10.7 |
| 审核 | 刘士强 | 日期 | 2022.10.8 |

### **3.3.3数据处理储条目定义**

实例：

表3-X 用户登录管理数据字典定义

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 数据处理 | | | | |
| 系统名：AUFE学生选课系统 | 编号：D-010 | | | |
| 条目名：用户登录管理 | 别名： | | | |
| 输入数据流：学生信息 | 输出数据流：学生信息及消息反馈 | | | |
| 加工逻辑：   1. 账号信息合法性检查； 2. 账号匹配； 3. 根据用户类别禁用或开放相应功能。 | | | | |
| 简要说明：对学生信息进行管理 | | | | |
| 修改记录 | 编写 | 喻杰 | 日期 | 2022.10.7 |
| 审核 | 刘士强 | 日期 | 2022.10.8 |

### **3.3.4数据项条目定义**

## 3.4实体联系分析

### **3.4.1 实体提取及实体图**

表3-18 实体提取结果

|  |  |  |
| --- | --- | --- |
| 实体 | 主码 | 其他属性 |
| 用户 | 用户ID | 用户密码，姓名，性别，用户权限等级 |
| 文献 | 文献编号 | 文献编号，名称，作者，文献来源，文献类别，文献发表时间，文献关键词 |
| 管理员 | 管理员编号 | 姓名，性别，级别，密码 |
| 检索记录 | 检索序号 | 检索日期 |

图 3-2 住房数据分析系统-文献属性图



图 3-3 住房数据分析系统-用户属性图



图 3-4 住房数据分析系统-管理员属性图

### **3.4.2 实体联系图**

# 4系统设计

概述段落

## 4.1总体设计

概述段落

### **4.1.1 系统功能结构**



图4-1 变换型数据流图示例



图4-2 变换型数据流图对应的软件结构图



图4-3 事务型数据流图示例



图4-4 事务型数据流图对应的软件结构图

按照变化型数据流图和事务型数据流图分析软件结构

### **4.1.2 系统层次图**

## 4.2 详细设计

概述段落

### **4.2.1 代表性模块设计**

#### 4.2.1.1 xx模块设计简介

（四级标题： 中文黑体 英文times new roman 1.5倍行间距 段前段后0.5行）

**（1）xx模块简介**

**（2）xx模块流程图（2，3, 4）可选择一种**

**（3）xx模块NS图**

**（4）xx模块PAD图**

**（5）针对复杂组合逻辑的判定表或判定树分析**

#### 4.2.1.2 YY模块设计简介

**（1）yy模块简介**

**（2）yy模块流程图（2，3, 4）可选择一种**

**（3）yy模块NS图**

**（4）yy模块PAD图**

**（5）针对复杂组合逻辑的判定表或判定树分析**

### **4.2.2 系统数据库设计**

#### 4.2.2.1 数据库的逻辑设计

#### 4.2.2.2 数据库的表设计及表与表之间的关联

# 5 编码与测试

概述段落

## 5.1 编码

### **5.1.1 编码规则简介**

### **5.1.2代表性模块示例**

## 5.2测试

概述段落

### **5.2.1 白盒测试**

#### 5.2.1.1 xx模块独立路径分析

#### 5.2.1.2 xx模块测试用例设计

#### 5.2.1.3 xx模块测试情况分析

### **5.2.2 黑盒测试**

#### 5.2.2.1 xx模块的等价类分析

#### 5.2.2.2 xx模块测试用例设计

#### 5.2.2.3 xx模块测试情况分析

# 6 系统使用说明

概述段落

## 6.1 系统运行环境和配置

## 6.2 系统操作说明（按照结构图或层次图的框架依次介绍）

### **6.2.1 XX1模块说明**

### **6.2.2 XX2模块说明**

### **6.2.3 XX3模块说明**

### **6.2.4 XX4模块说明**

# 7 总结

# 参考文献

按国家标准GB3469执行。

# 致谢