需求规格说明书

**酒店数据收集分析系统**

**小组：** 第五组

**姓名： 黄锦波 学号： 2017324236**

**姓名： 谢兴悦 学号： 2017324204**

**姓名： 郭奇东 学号： 2017324217**

**姓名： 沈小玉 学号： 2017324218**

**姓名： 谢金英 学号： 2017324225**

**版本：V1.0**

**发布时间：2020-09-10**

**岭南师范学院 信息工程学院**

**深圳信盈达科技有限公司 人才实训学院**

**文件变更记录**

\*A - 增加 M - 修订 D - 删除

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **版本** | **作者** | **日期** | **修订描述** | **变更记录** |
| V1.0 | 谢金英 | 2020/09/14 |  | 撰写初始版本 |
| v1.1 | 谢金英 | 2020/09/15 | M | 修订需求说明书格式 |
| V2.0 | 沈小玉 | 2020/09/16 | M | 修订项目背景，功能需求概述，功能列表说明 |
| V2.1 | 谢金英 | 2020/09/17 | M、A | 修订项目背景，增加项目总结 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**目录**

[1 引言 1](#_Toc52207139)

[1.1 编写目的 1](#_Toc52207140)

[1.2 读者对象 1](#_Toc52207141)

[1.3 编写原则 1](#_Toc52207142)

[1.4 术语与缩写解释 1](#_Toc52207143)

[2 项目概述 1](#_Toc52207144)

[2.1 项目背景 1](#_Toc52207145)

[2.2 项目目标 2](#_Toc52207146)

[2.3 项目范围 2](#_Toc52207147)

[2.4 客户与最终用户介绍 2](#_Toc52207148)

[3 功能需求概述 2](#_Toc52207149)

[4 功能需求详细说明 3](#_Toc52207150)

[4.1 功能列表说明 3](#_Toc52207151)

[4.2 业务分析—数据收集 3](#_Toc52207152)

[4.2.1 功能概述 3](#_Toc52207153)

[4.2.2 功能点描述 3](#_Toc52207154)

[4.3 业务分析—全量筛选 4](#_Toc52207155)

[4.3.1 功能概述 4](#_Toc52207156)

[4.3.2 操作流程 4](#_Toc52207157)

[4.3.3 功能点描述 4](#_Toc52207158)

[4.4 业务分析—系统筛选 5](#_Toc52207159)

[4.4.1 功能概述 5](#_Toc52207160)

[4.4.2 功能点描述 5](#_Toc52207161)

[4.5 业务分析—评分分析 6](#_Toc52207162)

[4.5.1 功能概述 6](#_Toc52207163)

[4.5.2 功能点描述 6](#_Toc52207164)

[5 性能需求说明 6](#_Toc52207165)

[系统运行指标 6](#_Toc52207166)

[高可用性要求 7](#_Toc52207167)

[6 其它需求说明 7](#_Toc52207168)

[其它需求 7](#_Toc52207169)

[新系统开发、运行条件与限制 7](#_Toc52207170)

[7总结 7](#_Toc52207171)

# 引言

## 编写目的

编写本文档是为了帮助项目小组明确本酒店数据收集分析项目的详细需求，以细化项目计划，评估项成本，实现系统设计和开发，编写测试用例，减少需求变更，保证项目开发成功。

## 读者对象

本文档适合以下人员阅读：

为项目的开发人员，用户

## 编写原则

技术在革新，时代在变迁，在智能信息时代，人们的生活在“数据化”，而酒店的运营管理活动也在被“数据化”。分析和挖掘数据是酒店经营者必须要掌握的基本技能，更充分利用数据掌握酒店内部和外部信息，提高个人的影响力和控制力，从而为酒店创造更大的价值。“信息即权力”，在酒店掌握信息越多的人，就会拥有更多的话语权。而信息来自对数据的加工，如果不能从数据中萃取信息用于经营决策，那么采取的数据只会成为酒店的成本；如果能够充分发挥数据的力量，在数据中挖掘价值，那么数据就会成为酒店的资产。

整个系统围绕着获取所需的酒店数据(酒店名称、地理位置、价格、评分)进行开发设计，从而实现了整个酒店数据收集分析系统的设计。

## 术语与缩写解释

|  |  |
| --- | --- |
| **缩写、术语** | **解 释** |
| 评分 | 客户评分在某个区间占的比例 |
|  |  |

# 项目概述

## 项目背景

艺龙酒店通过特有的系统联接上国际互联网向全球多姿多彩、声情并茂地展示自己的风貌、特色，推销自己客房和各种服务，并可依此组成酒店连锁业，结成战略联盟，以劲灵活的营销手段向广大市场进军。它可以向众多的客户提供面对面的营销方式。它开拓了市场的广度和深度，这些都是平常方式下的人力物力所无法与之比拟的。它的数据可以为酒店开发客源市场带来了限的商机。

## 项目目标

该项目主要是爬取数据，然后进行分析。

1. 在广东省的区域内，根据城市的名字，获取该城市下的评分4.5以上的酒店的名称。
2. 根据酒店名称获取该酒店的价格，地理位置，评价（score）。
3. 通过获取的数据来分析酒店的价格。
4. 酒店评分：评价从0~5.0，我们划分为五个区域（0~1，1~2，2~3，3~4，4~5），客户评分在某个区间占的比例(1~5(取整))

## 项目范围

艺龙酒店数据分析，关注于价格、评分的内容分析。

艺龙酒店数据分析：针对现在酒店需求日渐增多，提高客服的服务质量，给用户更好的酒店信息，为酒店的价格、评价进行的深入分析。

为了满足以上需求实现以下功能：数据收集、全量筛选、系统筛选、评分分析。

## 客户与最终用户介绍

本项目所建设的艺龙酒店数据收集分析系统，主要用户为住客。

# 功能需求概述

在本项目中，首先需要得到艺龙酒店的网址，然后依据其爬取酒店数据，要获取的内容有：酒店名称、地理位置、价格、评价。再通过筛选，分析在所有的酒店中，某一类酒店价格是多少，以及客户评分在某个区间占得比例。

# 功能需求详细说明

## 功能列表说明

|  |  |
| --- | --- |
| 酒店数据收集分析系统功能清单 | |
| 一级菜单 | 功能说明 |
| 数据收集 | 快捷查询目标酒店 |
| 全量筛选 | 根据配置条件以及数据收集范围筛选数据 |
| 系统筛选 | 根据配置条件以及采集周期和运行周期来配置任务 |
| 评分分析 | 分析多条件维度下酒店评分 |

## 业务分析—数据收集

### 功能概述

快捷查询目标酒店，同一页面输入筛选条件，点击查询即可查看结果，显示结果可以选择按详情展示和按数据显示。一般可用于在别的地方已得知某一查询条件，如酒店地理位置，价格或者评分的情况下。

运用场景：当你从其他相关系统接收到了一份酒店信息，只填写了酒店名称和地理位置，这时候就可以运用综合查询直接按酒店名称和地理位置查找对应价格及评分即可。

### 功能点描述

* **查询**

按照特定的条件和抽样方式进行酒店抽样，比如酒店名称属性、地理位置属性等。抽样结果可以选择直接分配给加入样本池。

* **保存策略**

当筛选条件选择完成后，创建csv文件或者json格式保存数据，得到对应的酒店清单。

* **导出**

按条件筛选出酒店清单，可以选择导出酒店清单的信息，导出excel表，包含酒店名称、地理位置、价格、评分等详细信息。

* **详细信息**

输入查询条件显示酒店清单后，住户可以查看酒店的详细信息，详细信息页左侧显示了酒店的基本信息和价格、评分特征。

## 业务分析—全量筛选

### 功能概述

全量筛选可理解为“即时任务”，可根据配置条件以及酒店评分范围筛选酒店。全量筛选按照预设的全量筛选策略对选定范围内的全部酒店进行全量筛选，找出有问题的疑似酒店清单，对有问题的酒店进行标识。

### 操作流程



### 功能点描述

* **新建倒是筛选**

进入全量筛选，可以选择新建全量筛选项目，配置全量筛选的条件及筛选模板，保存后新建项目。

* **新建全量筛选模板**

新建全量筛选项目之前，可以先配置全量筛选模板，通过设定关键词、价格、评分标准值完成模板。

设置全量筛选模板的路径：业务配置——全量筛选模板。

* **执行任务**

在全量筛选条件设置完成后运行程序，系统开始按条件进行筛选酒店

* **执行结果**

全量筛选“执行状态”为“已完成”后，“执行结果”在下方“项目执行情况”区域，显示已检出酒店详情。

* **加入样本池**

筛选出来的酒店清单，可以将全选或部分放入样本池暂存，后续进行统一分配质检等。

* **模板操作**

对已经执行的任务，可以进行“复制”“修改”“删除”操作，“复制”是新建一条与当前相似筛选条件的任务时的便捷操作，“修改”是对当前任务的筛选条件进行修改，“删除”是删除当前筛选任务。

## 业务分析—系统筛选

### 功能概述

系统筛选可理解为“循环任务”，可根据配置条件以及采集周期和运行周期来配置任务，任务执行开始之后会根据任务配置的运行周期循环执行，找出有问题的疑似酒店清单，对有问题的酒店进行标识。

### 功能点描述

* **新建筛选任务**

进入系统筛选，可以选择新建系统筛选任务，配置系统筛选的条件及筛选模板，保存后完成新建筛选策略。下次可以直接调用，在“查看已保存策略”选择，然后点击“启用”后执行。

运行周期：定义该筛选策略多少天运行一次。

采集周期：该策略的酒店采集的时间，是在今天之前的多少天。

* **新建筛选模板**

新建系统筛选任务之前，可以先配置系统筛选模板，通过设定关键词、价格、评分标准值完成模板。

设置全量筛选模板的路径：业务配置——系统筛选模板。

* **查询**

成员可根据时间范围对系统筛选项目进行查询。

* **筛选任务详情**

查询结果展示筛选任务详情，包括项目执行基本情况（检出数量、状态、检出率）和具体的酒店清单，酒店清单可以选择按数据展示，点击酒店查看酒店的详细信息。

## 业务分析—评分分析

### 功能概述

主要用于分析多条件维度下酒店中的评分。

运用场景：希望查询针对住户对某酒店的高评价，在此场景下主维度可选择某酒店，副维度选择投诉，即可展示出此场景下出现的分数以及他们的环比和在此时间范围内的趋势变化。

### 功能点描述

* **查询**

成员可以根据主维度和副维度对酒店进行周期性查询，查询结果用词频统计、趋势图和词语信息直观显示。

* **评分统计**

成员根据主维度和副维度对酒店进行周期性查询，查询后筛选出评分区间所占比例，统计出各区间范围的评分比例。

# 性能需求说明

## 系统运行指标

每物理核CPU每小时转译3小时的酒店数据量。

提供房型比例、评分统计功能，可以自动发现酒店中存在客户和客服的热词。

提供自定义模型建立的工具，可根据不同的业务要求，通过关键词或组合的配置，实现分析模型的建立。

## 高可用性要求

指在规定条件下和给定时间内[计算机系统](http://www.hudong.com/wiki/è®¡ç®æºç³)正确运行（①程序不为故障所破坏或停止;②结果不包括由故障所引起的错误；③执行时间不超过一定的限度；④程序运行在允许的领域内。）的概率。

系统能够满足：平均无故障间隔时间MTBF大于90天。

# 其它需求说明

## 其它需求

涉及算法：排序算法，各文本按日期排序，精确时间查询定点分析

## 新系统开发、运行条件与限制

暂无。

# 7总结

该文档确定对酒店数据收集分析的用户需求和功能需求。具体的功能需求可参考文档的功能描述部分。文档确定了系统所需要完成的工作以及目标系统提出完整、准确、清晰、具体的要求，为系统功能开发提供指导方向。