需求规格说明书

**酒店数据收集分析系统**

**小组：**

**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 |
| V3.0 | 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 |
| V3.1 | 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 4](#_Toc51456504)

[1.1 编写目的 4](#_Toc51456505)

[1.2 读者对象 4](#_Toc51456506)

[1.3 编写原则 4](#_Toc51456507)

[1.4 术语与缩写解释 4](#_Toc51456508)

[2 项目概述 4](#_Toc51456509)

[2.1 项目背景 4](#_Toc51456510)

[2.2 项目目标 5](#_Toc51456511)

[2.3 项目范围 5](#_Toc51456512)

[2.4 客户与最终用户介绍 5](#_Toc51456513)

[3 功能需求概述 5](#_Toc51456514)

[4 功能需求详细说明 6](#_Toc51456515)

[4.1 功能列表说明 6](#_Toc51456516)

[4.2 业务分析—数据收集 6](#_Toc51456517)

[4.2.1 功能概述 6](#_Toc51456518)

[4.2.2 功能点描述 6](#_Toc51456519)

[4.3 业务分析—全量筛选 7](#_Toc51456520)

[4.3.1 功能概述 7](#_Toc51456521)

[4.3.2 操作流程 7](#_Toc51456522)

[4.3.3 功能点描述 7](#_Toc51456523)

[4.4 业务分析—系统筛选 8](#_Toc51456524)

[4.4.1 功能概述 8](#_Toc51456525)

[4.4.2 操作流程 9](#_Toc51456526)

[4.4.3 功能点描述 9](#_Toc51456527)

[4.5 业务分析—评分分析 10](#_Toc51456528)

[4.5.1 功能概述 10](#_Toc51456529)

[4.5.2 功能点描述 10](#_Toc51456530)

[4.6 业务分析—聚类规则 11](#_Toc51456531)

[4.6.1 功能概述 11](#_Toc51456532)

[4.6.2 操作流程 12](#_Toc51456533)

[4.6.3 功能点描述 12](#_Toc51456534)

[5 性能需求说明 13](#_Toc51456535)

[系统运行指标 13](#_Toc51456536)

[高可用性要求 14](#_Toc51456537)

[6 其它需求说明 14](#_Toc51456538)

[其它需求 14](#_Toc51456539)

[新系统开发、运行条件与限制 14](#_Toc51456540)

[7总结 14](#_Toc51456541)

# 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

本文档将作为测试和需求变更的依据。对于本文档指明应达到的需求系统未实现的，属于缺陷；对于本文档未细化的需求，若开发完成后需要调整，属于需求变更，需发起项目变更管理流程（小于一人天的调整，诸如界面字体大小颜色样式等，无需发起项目变更流程）。整体开发过程中，需求变动不得超过10%。

## 术语与缩写解释

【简要说明本文档中涉及的专门术语、容易引起歧义的概念、关键词缩写及其他需要解释的内容。】

|  |  |
| --- | --- |
| **缩写、术语** | **解 释** |
| 房型 | 在所有的酒店中，某一类房型占的比例是多少 |
| 评分 | 客户评分在某个区间占的比例 |
| 关键词 |  |

# 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房型，地理位置，评价（score）。
3. 通过获取的数据来分析酒店的房型——在所有的酒店中，某一类房型占的比例是多少。
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，关注于房型、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，为酒店的房型、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、聚类规则。

## 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、房型、评价。再通过筛选，分析在所有的酒店中，某一类房型占的比例是多少，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 | 二级菜单 | 功能说明 |
| **分析** |  |  |
|  | 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 |
|  | 聚类规则 | 根据数据样本和关键字产出不同的主题 |

## 业务分析—数据收集

### 功能概述

快捷查询目标录音，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，如录音编号，手机号码或者少数关键字的情况下。

运用场景：当你从其他相关系统接收到了一份工单信息，只填写了客户来电号码和来电时间，这时候就可以运用综合查询直接按录音时间和来电号码查找对应录音即可。

### 功能点描述

* **查询**

质检主管按照特定的条件和抽样方式进行录音抽样，比如录音属性、客户属性、员工属性及语音特征等。抽样结果可以选择直接分配给质检人员或加入样本池。

* **保存策略**

当筛选条件选择完成后，点击“策略保存”，为筛选策略命名，下一次可以直接从“查看已保存策略”中选择，然后点击“查询”，得到对应的录音清单。

* **导出**

按条件筛选出录音清单，可以选择导出录音清单的信息，导出excel表，包含录音编号、录音时间、通话时长等详细信息。

* **详细信息**

输入查询条件显示录音清单后，质检主管可以点击录音查看录音的详细信息，详细信息页左侧显示了语音的基本信息和语音特征，右侧显示客户信息和事后录入单。

## 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录音时间范围筛选录音。全量筛选按照预设的全量筛选策略对选定范围内的全部录音进行全量筛选，找出有问题的疑似录音清单，对有问题的录音进行标识。

### 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、静默、语速、重叠以及情绪值完成模板。

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点击“开始任务”，系统开始按条件进行筛选录音

* **执行结果**

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

* **加入样本池**

筛选出来的录音清单，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，找出有问题的疑似录音清单，对有问题的录音进行标识。

### 操作流程



### 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：该策略的录音采集的时间，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、静默、语速、重叠以及情绪值完成模板。

设置全量筛选模板的路径：业务配置——系统筛选模板。

* **查询**

质检主管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）和具体的录音清单，录音清单可以选择按详情展示或者按数据展示，点击录音查看录音的详细信息。

## 业务分析—评分分析

### 功能概述

主要用于分析多条件维度下录音中的高频词汇。

运用场景：希望查询针对A产品的投诉录音中出现的高频词，在此场景下主维度可选择A产品，副维度选择投诉，即可展示出此场景下出现的高频词汇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**

质检员主管可以根据主维度和副维度对录音进行周期性查询，查询结果用词频统计、趋势图和词语信息直观显示。

* **词频统计**

质检主管根据主维度和副维度对录音进行周期性查询，查询后筛选出词库里面相关的关键字，统计出关键字的词频和环比进行显示。

* **添加关键词**

质检主管可以在“业务配置-词库管理”中添加、修改以及删除业务关键字和停词水词。在词频统计中，可以对关键字添加关键字描述。

* **添加水词**

质检主管可以在“业务配置-词库管理”中添加、修改以及删除停词水词。在词频统计中，可以对关键字添加水词，添加水词之后该词不会再出现在词频统计中。

* **趋势图**

系统根据主维度和副维度对选定时间周期的录音进行分析，筛选出高频率的词语，用折线统计图直观显示，横坐标是时间，纵坐标是词频。质检主管可以选择性对词汇不进行显示。

* **词语信息**

词语信息对词语词频情况进行显示，鼠标置于词语上面能看到该词语的词频。

## 业务分析—聚类规则

### 功能概述

以录音和关键字为基础，结合算法，挖掘产出录音中关联性较高的主题以及典型问法。

运用场景：根据录音样本和关键字产出不同的主题，例如录音样本为营销成功类，执行聚类规则后可产出营销成功场景下相关的主题和典型问法。

### 操作流程



### 功能点描述

* **新建聚类规则**

质检主管选择新建聚类规则，给聚类规则命名，设定显示主题词数（5-9），限定语句数，选择分析对象（全部/客户/客服）。

* **训练样本**

质检主管可以对聚类规则进行样本训练，样本训练分为加入训练样本和移除训练样本两部分。训练样本必须大于50条才可以进行聚类。

* **呼入录音**

进入呼入录音，质检主管可以从录音属性、员工属性、客户属性以及语音特征四个方面进行录音查询，查询结果可以按详情显示或者按数据显示。录音清单可以选择性加入训练样本。

* **呼入已选录音**

进入呼入已选录音，质检主管可以根据客户姓名、坐席姓名以及坐席工号进行录音查询，查询结果可以按详情显示或者按数据显示。录音清单可以选择性进行移除训练样本，同时也可以选择导出录音（excel）。

* **执行样本**

点击“开始”，系统执行训练样本。训练样本的数量必须在50条以上，启用状态的词表个数需在[100,2000]范围内。

* **执行结果**
* **主题聚类**

主题聚类根据聚类规则对训练样本进行聚类，显示结果格式为“主题：词1-词2-词3-词4-词5-词6-词7（词频）”。

主题聚类的结果后面的数字，表示主题词在多少个句子中出现（不是多少条录音）。

* **典型问法**

典型问法根据聚类规则对训练样本进行聚类，显示结果格式为“词1-词2-词3-词4-词5-词6-词7（语句数量）”。

* **词表维护**

再进入“词表维护”启用/停用关键词。词表中的关键词受词库管理中的“业务关键词”、“停词水词”的影响。加入到“停词水词”的词，不会出现在词表中。

* **查看日志**

日志记录用户对聚类的操作记录。

# 性能需求说明

## 系统运行指标

每物理核CPU每小时转译3小时的录音数据量。

提供热词统计功能，可以自动发现录音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。

支持语音转化成文本能力，系统上线后，经过多轮优化迭代，对于业务录音汉语普通话语音转写效果全文与关键字正确率分别达到80%和80%以上。

## 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，各语音文本按日期排序，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

本系统融合了云计算、语音处理、商业智能和互联网技术，以实现语音数据的高效转写、分析统计和数据挖掘等广泛应用为目的，面向语音数据密集型的话务行业，提供数据处理所需资源、能力和应用服务的云平台，能够提升企业的客服质量。

目前机器处理的文本情感分析还只是仅仅从数学统计角度融合一些自然语言的语法规则, 完成极性分类、语法分析的任务, 而情感是一个相对复杂的研究客体, 交给机器与算法处理只是用数学量化的基本模型完成简单任务，对大量数据隐含的情感特征和语言表达的统计和排查只是初步解决了分析中的小问题，如何让机器对人类情感进行精确的把握和分析，甚至是通过分析让机器模仿、创造出与人类的情感交流方式是文本分析更深入的问题。