# 智能问答系统-系统设计说明书

目录

[智能问答系统-系统设计说明书 1](#_Toc518378818)

[一、引言 3](#_Toc518378819)

[（一）编写目的 3](#_Toc518378820)

[（二）命名规范 3](#_Toc518378821)

[（三）背景 3](#_Toc518378822)

[（三）术语定义 3](#_Toc518378823)

[二、总体设计 4](#_Toc518378824)

[（一）需求规定 4](#_Toc518378825)

[（二）运行环境 4](#_Toc518378826)

[（三）系统流程图 4](#_Toc518378827)

[（四）技术框架 5](#_Toc518378828)

[（五）匹配算法 5](#_Toc518378829)

[（六）功能模块清单 6](#_Toc518378830)

[（七）功能需求与系统模块的关系 6](#_Toc518378831)

[（八）人工处理过程 6](#_Toc518378832)

[三、接口设计 7](#_Toc518378833)

[（一）接口说明 7](#_Toc518378834)

[1、extract方法 7](#_Toc518378835)

[2、find\_match方法 7](#_Toc518378836)

[3、find\_answer方法 8](#_Toc518378837)

[（二）已使用框架 8](#_Toc518378838)

[1.word2vec 8](#_Toc518378839)

[2.anytree 9](#_Toc518378840)

[3.TF-INF 9](#_Toc518378841)

[（三）异常处理 9](#_Toc518378842)

[六、系统错误处理 10](#_Toc518378843)

[（一）错误类型 10](#_Toc518378844)

[1、事务内部的故障 10](#_Toc518378845)

[2、系统故障 10](#_Toc518378846)

[3、介质故障 10](#_Toc518378847)

[4、计算机病毒故障 10](#_Toc518378848)

[（二）故障预防及补救 11](#_Toc518378849)

[1、预期的事务内部故障 11](#_Toc518378850)

[2、非预期的事务内部故障 11](#_Toc518378851)

[3、系统故障 11](#_Toc518378852)

[4、介质故障的软件容错 11](#_Toc518378853)

[5、介质故障的硬件容错 11](#_Toc518378854)

[6、计算机病毒故障 11](#_Toc518378855)

[（三）系统维护 11](#_Toc518378856)

[七、测试计划 12](#_Toc518378857)

[（一）单元测试 12](#_Toc518378858)

[（二）集成测试 12](#_Toc518378859)

[（三）系统测试 12](#_Toc518378860)

[1、功能测试 12](#_Toc518378861)

[2、安全性测试 12](#_Toc518378862)

## 一、引言

## （一）编写目的

在完成需求分析的基础上，我们依据智能问答系统的需求以及数据特点对智能问答系统进行设计。为了文档化对系统的设计，编写本文档。本文档主要用于阐明用户的要求的，描述出系统的需求模型、功能和性能要求以及其他约定，为后期的软件设计等工作提供依据。

## （二）命名规范

a. 类名采用大驼峰命名法

b. 变量数据库列名方法名等均采用小驼峰命名法。

## （三）背景

a. 待开发软件系统的名称为智能问答系统

b. 本项目的任务提出者及开发者是三个弹簧手开发小组：

项目负责人：徐家兴

系统设计员：徐家兴

编码员：李昕航、仇善召

软件测试员：仇善召

用户为使用华为云平台的用户，该软件在python3.6、Android5、jdk1.8环境下完成开发。

## （三）术语定义

知识库：是指专家系统设计所应用的规则集合，包含规则所联系的事实及数据，它们的全体构成知识库。

智能客服系统：智能客服系统是在大规模知识处理基础上发展起来的一项面向行业应用的，适用大规模知识处理、自然语言理解、知识管理、自动问答系统、推理等等技术行业，智能客服不仅为企业提供了细粒度知识管理技术，还为企业与海量用户之间的沟通建立了一种基于自然语言的快捷有效的技术手段；同时还能够为企业提供精细化管理所需的统计分析信息。

## 二、总体设计

## （一）需求规定

使用华为云平台的用户是唯一的参与者，参与者通过使用事件与系统进行交互，所有的使用事件综合起来即构成了用户的功能需求。本系统通过机器学习获得词向量模型，进而得到问答对，实现与用户的自动问答。

## （二）运行环境

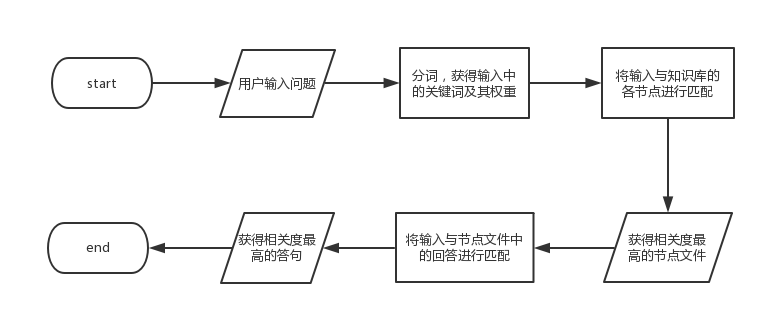
客户端：

* Android5及以上
* jdk 1.8

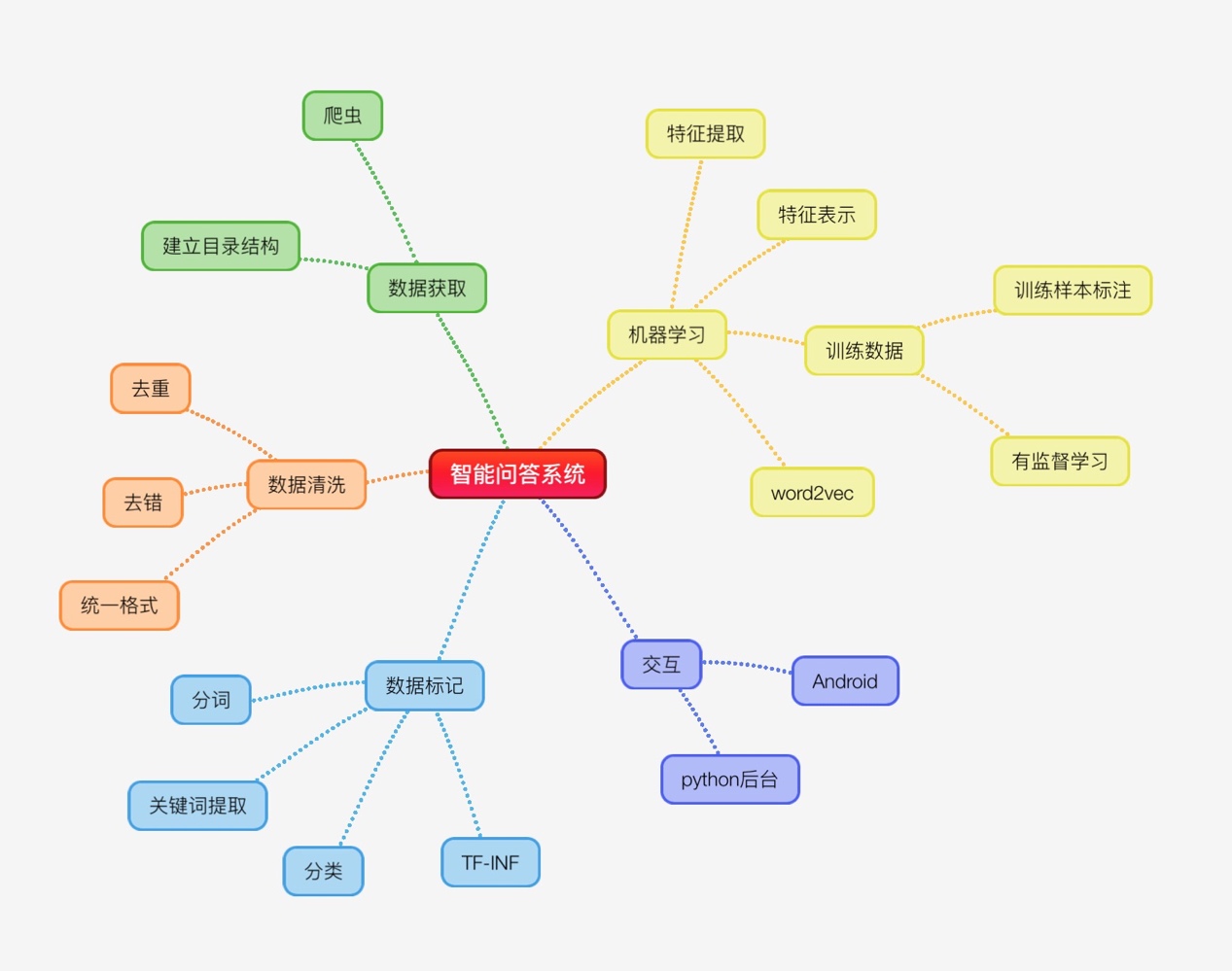
服务端：

* Python3.6

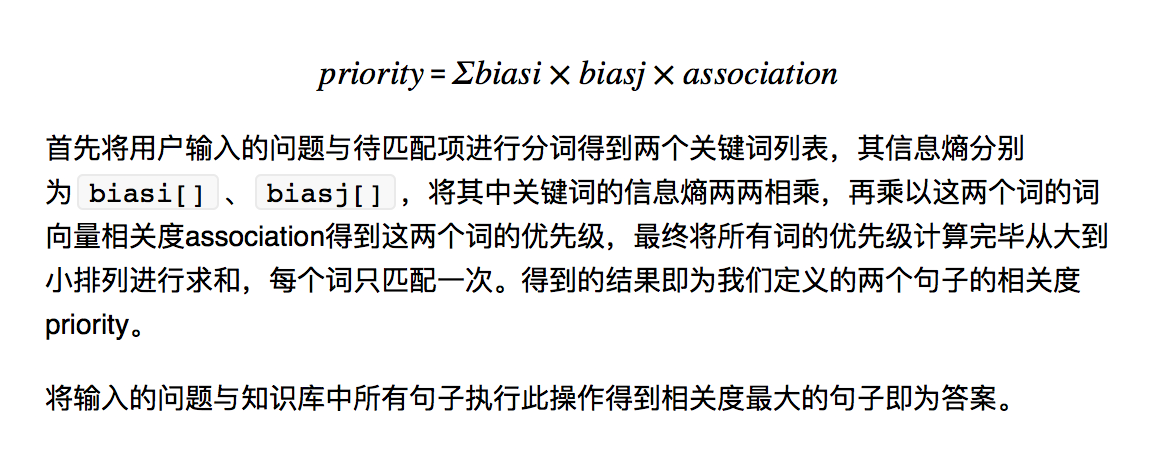
## （三）系统流程图

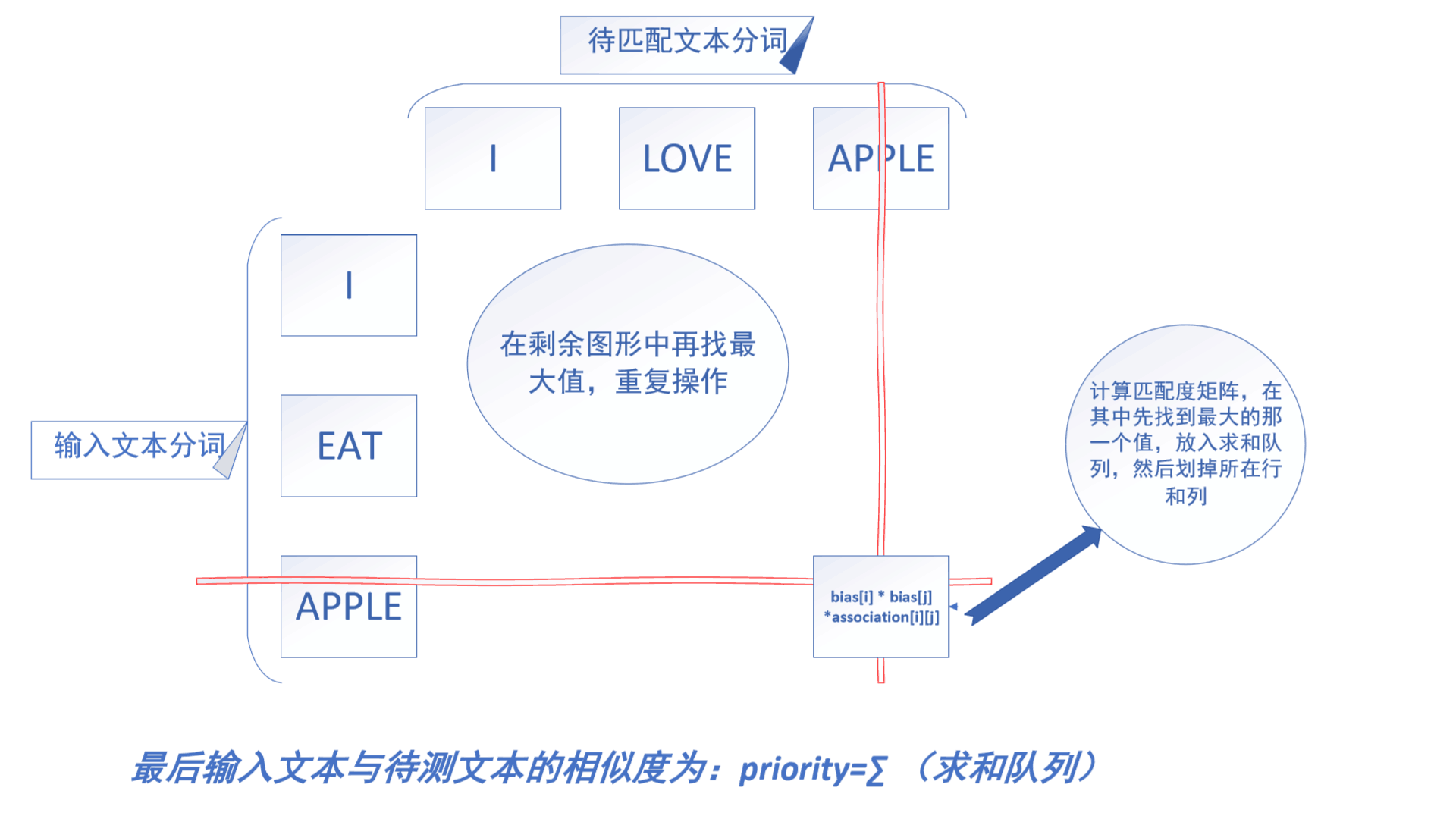


## （四）技术框架



## （五）匹配算法





## （六）功能模块清单

|  |  |  |
| --- | --- | --- |
| **编号** | **名称** | **功能描述** |
| 1 | 模型训练 | 将需要进行学习的语料导入本系统，本系统将自动对其进行建模，得到100维的skip-gram词向量。 |
| 2 | 问答对构建 | 通过自定义算法将用户输入的问句与数据库中的语料集进行对比，取出相关度最高的文件。 |
| 3 | 回答 | 将问句与相关度最高的文件进行对比，取出相关度最高的一句话作为回答。 |

## （七）功能需求与系统模块的关系

|  |  |  |
| --- | --- | --- |
|  | **Train** | **Provider** |
| 模型训练 | √ |  |
| 问答对构建 |  | √ |
| 回答 |  | √ |

## （八）人工处理过程

模型训练时需要人工将语料集放入指定目录下。

## 三、接口设计

## （一）接口说明

智能问答系统服务接口信息如下表所示：

|  |  |
| --- | --- |
| 接口名称 | main.main |
| 接口地址 | 47.75.157.87:8080/home/uftp/QA\_System/QA\_System/NER/main.main |
| 接口协议 | http |

## 1、extract方法

①方法描述：

创建充值请求实例，并发送到数据库。

②请求参数：

| **参数类型** | **参数名** | **参数描述** | **备注** |
| --- | --- | --- | --- |
| String | txt | 用户输入 | 无 |

③响应参数：

| **响应类型** | **响应描述** |
| --- | --- |
| List | 返回输入句子中的关键词及其权重 |

④异常描述：

| **异常类型** | **异常描述** |
| --- | --- |
| RequestException:RuntimeException | 请求异常：输入参数有误或数据库异常 |

## 2、find\_match方法

①方法描述：

将输入与后台知识库的各节点标签进行对比，选出匹配度最高的节点。

②请求参数：

| **参数类型** | **参数名** | **参数描述** | **备注** |
| --- | --- | --- | --- |
| List | keywords | 用户输入句子中的关键词及其权重 | 无 |

③响应参数：

| **响应类型** | **响应描述** |
| --- | --- |
| Object | 返回匹配度最高的节点 |

④异常描述：

| **异常类型** | **异常描述** |
| --- | --- |
| RequestException | 请求异常：输入参数有误或数据库异常 |

## 3、find\_answer方法

①方法描述：

将用户输入的关键词与匹配度最高的语料文件进行匹配，选出相关度最高的一句话。

②请求参数：

| **参数类型** | **参数名** | **参数描述** | **备注** |
| --- | --- | --- | --- |
| String | name | 匹配度最高的节点标签 | 无 |
| List | keywords | 用户输入句子中的关键词及其权重 | 无 |
| String | path | 匹配度最高的文件路径 | 无 |

③响应参数：

| **响应类型** | **响应描述** |
| --- | --- |
| String | 相关度最高的句子 |

④异常描述：

| **异常类型** | **异常描述** |
| --- | --- |
| RequestException | 请求异常：输入参数有误或数据库异常 |

## （二）已使用框架

## 1.word2vec

word2vec 的训练模型是在于其具有一个隐含层的神经元网络。它输入词汇表向量，当看到一个训练样本时，对于样本中的任意一个词，我们把相应的在词汇表中出现的位置的值设置成1，否则设置成0。它输出是词汇表向量，对于训练样本的标签中的每一个词，就把对应的在词汇表中出现所在的值设置成1，反之设置成0。对所有的样本，训练这个神经元网络。收敛后，我们可以把从输入一层到隐含层的那些权值，把它用作每一个词汇表中的词的向量。第一个向量就是（v1,1 v1,2 v1,3 ... v1,p），p是表示向量的维度数。所有虚框中的权值就是所有词的向量的值。有了每个词有限维度数的向量，就可以用到其它的应用中，因为它们就像图，有了有限维度数的统一意义的输入。训练 word2vec的思想，是用一个词在文本中的上下文来表示这个词，这样就可以省去了人工去标注。

word2vec 模型在本系统中用于计算词语相似度，进而得到问答对的相关度。

## 2.anytree

Anytree是一种简单，轻量且可扩展的Tree数据结构，我们可以通过anytree随意分享有关任何树项目的信息。

Anytree在本系统中用于构建文件目录结构。

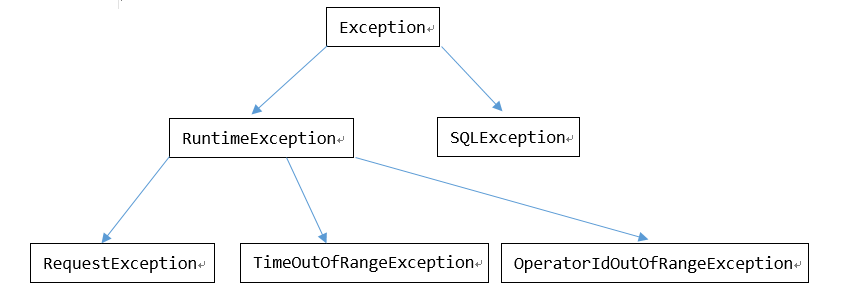
## 3.TF-INF

TF-IDF（term frequency–inverse document frequency）是一种用于检索与探勘的常用的一种加权技术。它是一种统计学方法，以评估字词对于一些文件或语料库中的其中一份文件的紧要程度。字词的重要程度随着它在文件中出现的次数正比增加，但会随着它在语料库中出现的频率次数反比下降。TF-IDF加权形式常被搜寻引擎，作为文件与用户查询之间相关程度的评级。除TF-IDF以外，网上的搜寻引擎还会使用基于连结分析的评级方法，以确定文件在结果中出的顺序。

TF-IDF算法在本系统中用于关键词提取。

## （三）异常处理

各接口对应异常在“接口说明”部分已作详细描述。异常间的继承关系如图2所示：



**图2 异常继承图**

## 六、系统错误处理

## （一）错误类型

## 1、事务内部的故障

事务内部故障可分为预期的和非预期的，其中大部分的故障都是非预期的。预期的事务内部故障是指可以通过事务程序本身发现的事务内部故障；非预期的事务内部故障是不能由事务程序处理的，如运算溢出故障、并发事务死锁故障、违反了某些完整性限制而导致的故障等。

## 2、系统故障

系统故障也称为软故障，是指数据库在运行过程中，由于硬件故障、数据库软件及操作系统的漏洞、突然停电灯情况，导致系统停止运转，所有正在运行的事务以非正常方式终止，需要系统重新启动的一类故障。这类事务不破坏数据库，但是影响正在运行的所有事务。

## 3、介质故障

介质故障也称为硬故障，主要指数据库在运行过程中，由于磁头碰撞、磁盘损坏、强磁干扰、天灾人祸等情况，使得数据库中的数据部分或全部丢失的一类故障。

## 4、计算机病毒故障

计算机病毒故障是一种恶意的计算机程序，它可以像病毒一样繁殖和传播，在对计算机系统造成破坏的同时也可能对数据库系统造成破坏（破坏方式以数据库文件为主）。

## （二）故障预防及补救

## 1、预期的事务内部故障

将事务回滚，撤销对数据库的修改。

## 2、非预期的事务内部故障

强制回滚事务，在保证该事务对其他事务没有影响的条件下，利用日志文件撤销其对数据库的修改。

## 3、系统故障

待计算机重新启动之后，对于未完成的事务可能写入数据库的内容，回滚所有未完成的事务写的结果；对于已完成的事务可能部分或全部留在缓冲区的结果，需要重做所有已提交的事务（即撤销所有未提交的事务，重做所有已提交的事务）。

## 4、介质故障的软件容错

使用数据库备份及事务日志文件，通过恢复技术，恢复数据库到备份结束时的状态。

## 5、介质故障的硬件容错

采用双物理存储设备，使两个硬盘存储内容相同，当其中一个硬盘出现故障时，及时使用另一个备份硬盘。

## 6、计算机病毒故障

使用防火墙软件防止病毒侵入，对于已感染病毒的数据库文件，使用杀毒软件进行查杀，如果杀毒软件杀毒失败，此时只能用数据库备份文件，以软件容错的方式恢复数据库文件。

## （三）系统维护

我们使用数据库的维护功能机制，比如，定期备份数据库，定期检测数据库的一致性，定期查看操作日志等；文件方面：对于下架商品等，定期删除相关文件，减少数据量。编码实现时应采用模块化和分层的思想，提高模块内部的内聚，减少模块间的耦合。使系统逻辑结构清晰，从而增强可读性和可维护性。在编码过程中注意标识符命名的意义，添加适量注释。

## 七、测试计划

测试计划作为本系统的测试指导，测试成员按计划和规定进行测试。本系统的测试将从单元测试、集成测试、系统测试三个方面来设计。

## （一）单元测试

在代码编写之后即可进行，与复审工作结合，根据设计规约选取数据，增大发现各类错误的可能。

|  |  |  |
| --- | --- | --- |
| **编号** | **名称** | **测试内容** |
| 1 | 模型训练 | 传入不同数据，是否能正确工作并生成词向量模型；传入非法的数据，能否处理异常 |
| 2 | 问答对构建 | 发送问答请求，能否执行操作，匹配得到相关度最高的文件。 |
| 3 | 回答 | 发送问答请求，能否执行操作，返回文件中相关度最高的一句话。 |

## （二）集成测试

主要测试各接口调用情况，能否接收来自APP客户端的请求并正确处理。由于本系统多个功能均涉及到接口，则集成测试需在与其他系统对接之后开始测试。在集成测试之前需要先经过单元测试。

主要测试在组合对接之后，此系统能否正常工作以及与其他组模块能否集成起来工作，最后还要一同参与测试整个软件系统的所有模块组合是否能正常工作。

## （三）系统测试

大部分测试都需要与其他系统对接组合之后才能进行，主要测试功能之间的交互是否正常，并测试其在异常状况下处理异常的功能。

## 1、功能测试

与单元测试和集成测试一同进行，检验功能是否能正常运行。

## 2、安全性测试

主要测试数据库的安全性，尽可能消除由于操作不恰当或者超负荷后出现的运行异常。