

**分类8新闻分类系统**

**软件需求规格说明书**

项目名称： 分类8新闻分类系统系统

申 请 者： 刘兴瑞 张潇潇 王亚涛 王方

指导教师： 代祖华

学 院： 计算机科学与工程学院

目录

[1、引言 2](#_Toc15854)

[1.1编写目的 3](#_Toc30129)

[1.2背景 3](#_Toc26016)

[2、任务概述 4](#_Toc25049)

[2.1目标 4](#_Toc5132)

[2.2用户的特点 4](#_Toc10250)

[2.3假定和约束 4](#_Toc20158)

[3、需求分析 5](#_Toc19271)

[3.1处理需求 5](#_Toc31151)

[3.2安全性和完整性需求 6](#_Toc3921)

[3.3 数据流图 7](#_Toc31195)

[1.基本系统模型： 7](#_Toc3705)

[2.分类系统的功能级数据流图： 7](#_Toc11357)

[3.4 数据字典 7](#_Toc30248)

[3.5 软件系统状态图 10](#_Toc2746)

[4、团队WBS 11](#_Toc28781)

[5、对性能的规定 12](#_Toc29008)

[5.1精度 12](#_Toc28000)

[5.2时间特性要求 13](#_Toc32504)

# 1、引言

## 1.1编写目的

作为用户和软件开发人员达成的技术协议书，作为着手进行设计工作的基础和依据，故而编写新闻分类系统需求规格说明书其目的是为了进一步定制软件开发的细节问题,便于用户需求与需求分析人员，软件开发人员以及相关利益设计人员的需求达成共识，以此保证软件开发的质量和需求的完整性。本文面向的读者主要是项目委托单位的管理人员，希望能使本软件开的工作更具体。

## 1.2背景

新闻发展越来越快，每天各种各样的新闻令人目不暇接，对新闻进行科学的分类既能够方便不同的阅读群体根据需求快速选取自身感兴趣的新闻，也能够有效满足对海量的新闻素材提供科学的检索需求。新闻发展越来越快，每天各种各样的新闻令人目不暇接，对新闻进行科学的分类既能够方便不同的阅读群体根据需求快速选取自身感兴趣的新闻，也能够有效满足对海量的新闻素材提供科学的检索需求。文本作为信息的重要载体，对文本数据的提取与表示是解决文本信息管理问题的关键手段。其中，新闻数据作为日常最为常见的文本数据之一,对新闻的精准分类,有助于缩短查阅时间,提升阅读体验。文本表示是文本分类的基础，传统的文本表示主要基于统计计数。这种表示方法默认为单词与单词之间是相互独立的，这会导致文本语义信息的丢失;另外，提取的特征具有高纬度和高稀疏的特点，不利于文本信息的表示。伴随着深度学习提出与推广，通过层次神经网络模型，可以有效地完成从低层特征到高层特征的逐层提取融合，对于重要特征给予高比重权值。深度学习提出为文本的特征提取与文本分类提供了有力的理论与技术支持。

项目名称: 新闻分类算法系统

项目产品名称：分类8新闻分类系统

项目的任务开发者：刘兴瑞 张潇潇 王亚涛 王方；

项目的用户：每位需要对新闻进行分类的人员；

# 2、任务概述

## 2.1目标

1.提出了基于单词级和字符级混合特征的文本分类模型.针对中文短文本具有长度短,稀疏性和上下文依赖强等特点,本文分别使用以单词级向量和字符级向量作为输入的两个长短时记忆网络或双向长短时记忆网络对句子语义进行编码,然后将两个输出向量相加结合作为整个句子向量表示.利用NLPCC 2017中文新闻标题分类数据集进行实验,结果表明,对于中文短文本分类,单词嵌入和字符嵌入的结合使用可以在中文短文本的句子语义表示中相互补充,有助于提高中文短文本的分类性能。

2.完成了新闻文本分类系统具体的功能实现和测试.首先部署系统实现所需要的系统环境,然后详细阐述新闻分类模块一个系统核心模块的功能实现过程,给出关键函数,对实现结果进行展示,包括数据爬取,对数据库的操作和文本分类模型的构建等,并展示系统的整体运行结果.通过对系统进行功能和性能测试,说明各模块的实现都符合系统设计的要求。

## 2.2用户的特点

1. 本软件系统的最终用户为希望对新闻文本进行分类的人员，软件设计等符合该类群体的使用习惯。
2. 操作人员的教育水平和技术专长：软件开发

维护人员的教育水平和技术专长：软件开发

1. 本软件的预期使用频度：该软件的最终使用者中对需要对新闻文本进行分类的人员使用值是最高的。

## 2.3假定和约束

进行本软件开发工作的假定和约束，例如经费限制、开发期限等如下表1.2所示：

表1.2 项目的约束与具体描述

|  |  |
| --- | --- |
| 约束 | 描述 |
| 可靠性需求 | 由于安卓手机版本不同，可能会导致下载的APP不能使用 |
| 经费限制 | 主要花费为资料费，实验费，打印费，交通费等，项目经费为1000元，按照预期计划与进度进行，才能保证经费足够的情况下完成项目的开发。 |
| 开发期限 | 开发期限为三个月，截止期限为2021.6.30。 |

# 3、需求分析

分类8新闻文本分类系统，其主要实现用户对新闻文本需要分类的需求。让用户可方便快捷的得到自己需要的分类。

应解决的问题：

(1)实现用户管理（注册，登录，查询，信息管理）；

(2)实现用户提交表格的保存；

(3)实现对文本的分类。

## 3.1处理需求

本系统主要由用户模块和管理员模块构成，主要功能如下：

### （1）用户功能模块：

登录：通过登录系统获得相关的功能；

分类：对于单条文本和多条文本及文件进行分类；

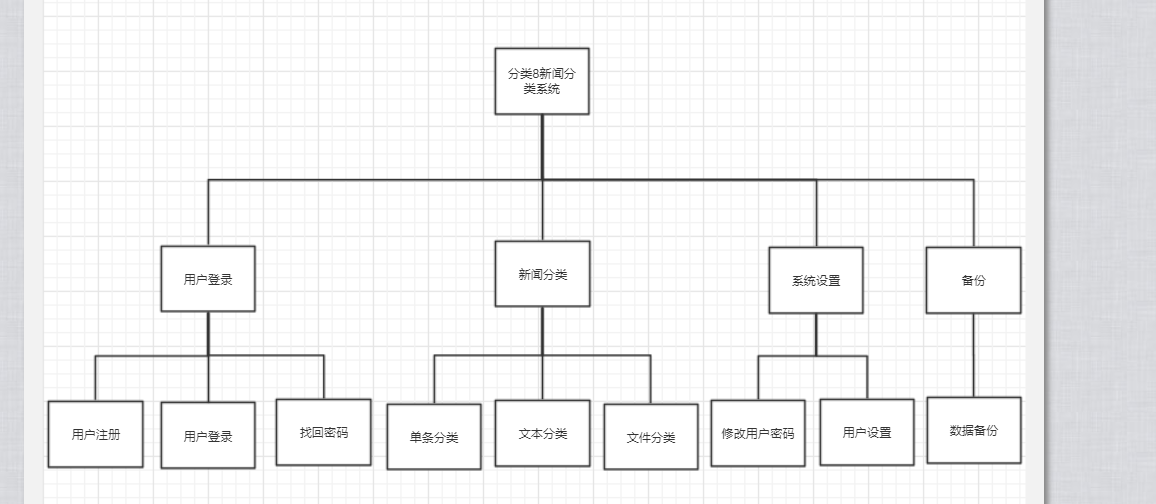
### （2）商家功能模块：

用户信息管理：用户实体是对用户信息进行管理，保存用户的所有信息；

分类文件管理模块：管理用户提交的需要分类的文件；

系统的主要功能如图1.1所示。

图1.1 系统主要功能用例图

图1.1 系统主要功能用例图

## 3.2安全性和完整性需求

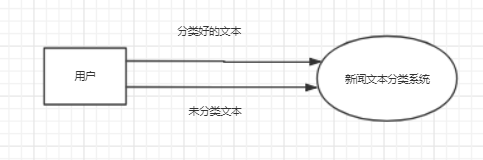
数据库的安全性是指在信息系统的不同层次保护数据库，防止未授权的数据访问，避免数据的泄漏、不合法的修改或对数据的破坏。安全性问题不是数据库系统所独有的，它来自各个方面，其中既有数据库本身的安全机制如用户认证、存取权限、视图隔离、跟踪与审查、数据加密、数据完整性控制、数据访问的并发控制、数据库的备份和恢复等方面，也涉及到计算机硬件系统、计算机网络系统、操作系统、组件、Web服务、客户端应用程序、网络浏览器等。只是在数据库系统中大量数据集中存放，而且为许多最终用户直接共享，从而使安全性问题更为突出，每一个方面产生的安全问题都可能导致数据库数据的泄露、意外修改、丢失等后果。

数据库的安全性和完整性是数据库安全保护的两个不同的方面。数据库的安全性保护数据库以防止不合法用户故意造成的破坏，数据库的完整性保护数据库以防止合法用户无意中造成的破坏。从数据库的安全保护角度来讲，完整性和安全性是密切相关的。

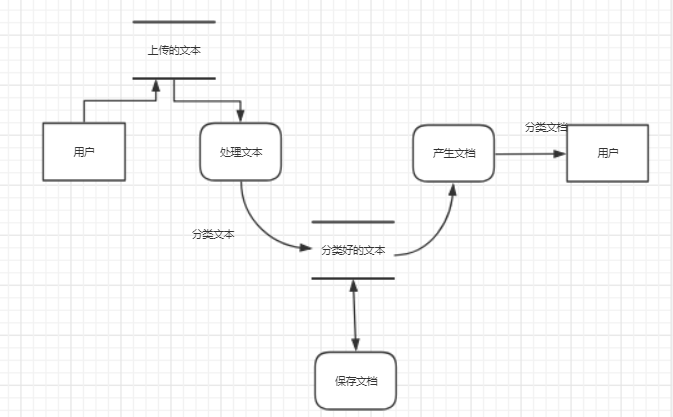
数据库的完整性的基本含义是指数据库中数据的正确性、有效性和相容性，其主要目的是防止错误的数据进入数据库。正确性是指数据的合法性,例如数值型数据只能含有数字而不能含有字母。有效性是指数据是否属于所定义域的有效范围。相容性是指表示同--事实的两个数据应当一致，不一致即是不相容。

## 3.3 数据流图

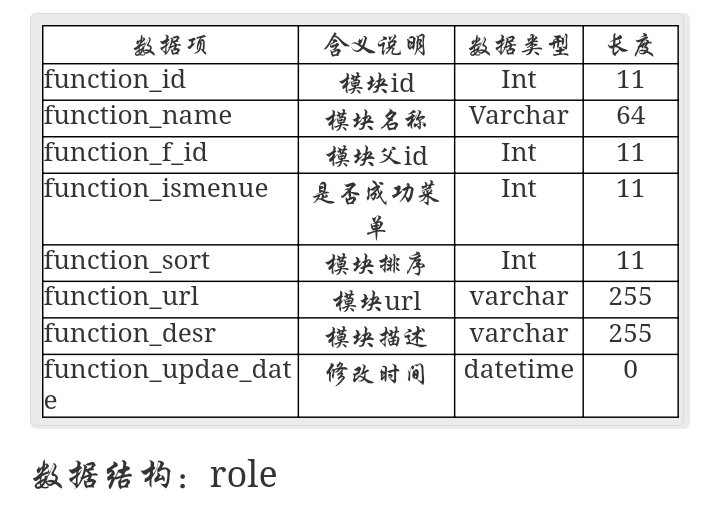
### 1.基本系统模型：

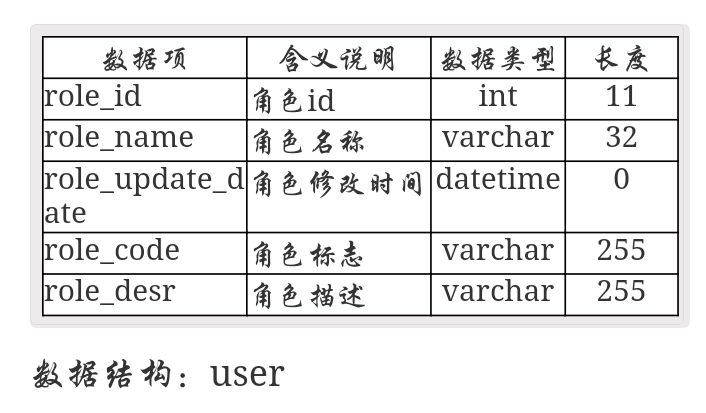


### 分类系统的功能级数据流图：

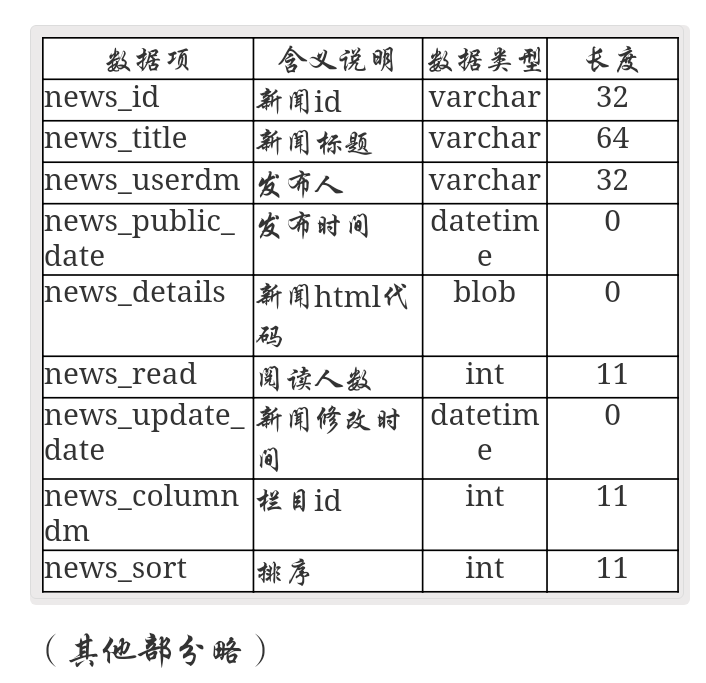


## 3.4 数据字典

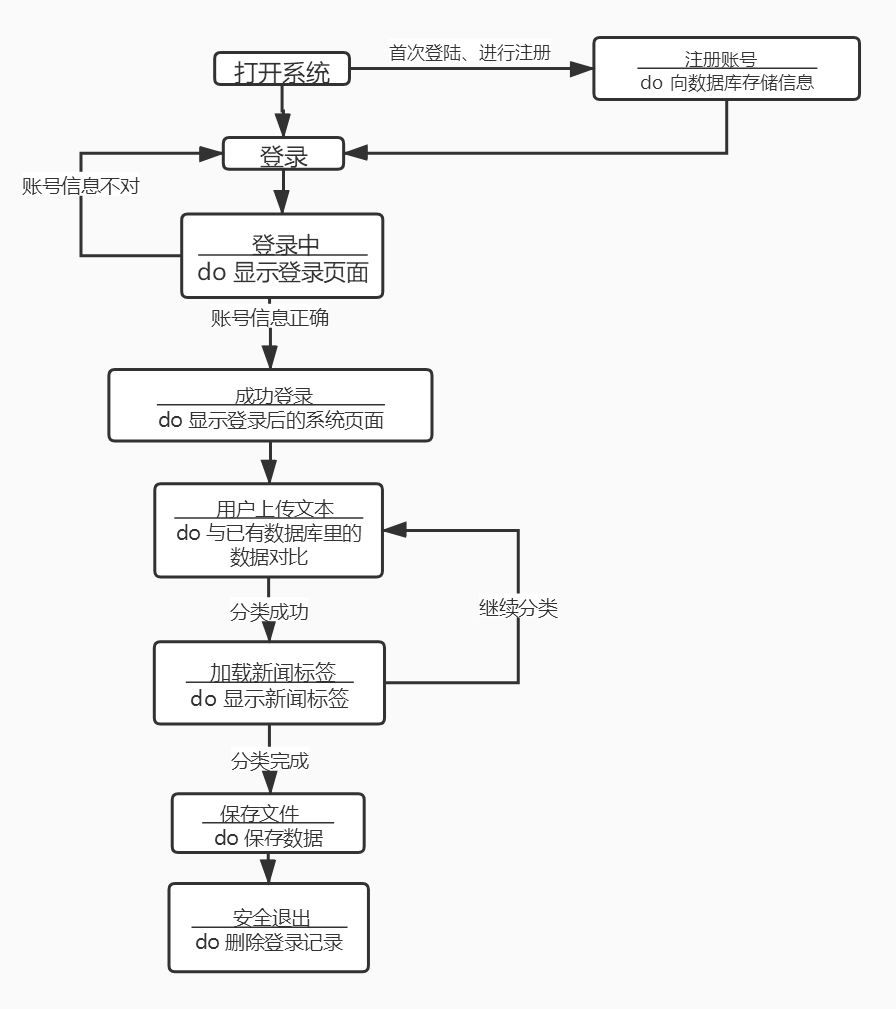








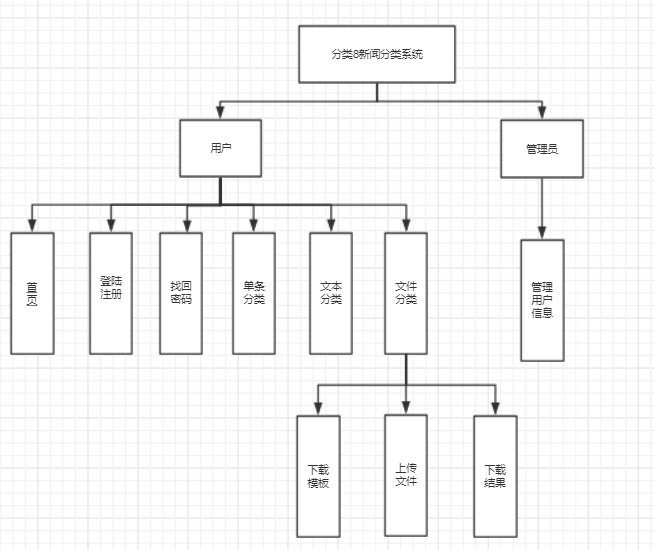
## 3.5 软件系统状态图



# 4、团队WBS

完成需求分析后，若要开发能够满足用户需求的软件，持续有序实现团队目标，一个团队软件项目要在一段时间内完成诸多任务，编制项目WBS（Work Breakdown Structure），即工作分解结构，是根据项目目标把工作分解成层次分明的、可交付成果的工作任务，用逻辑图形或树形结构表示出来），是团队项目有序管理的工作依据。

按照最后想提交给用户的软件功能来做WBS，我们将系统分割成：



# 5、对性能的规定

## 5.1精度

该软件数据的输入、处理、输出所要数据达到的精度说明如下：本软件涉及到需要手工输入数据的字段及说明如下：

表1.3 软件数据精度说明表

|  |  |  |
| --- | --- | --- |
| 字段 | 精度 | 备注 |
| 账号 | 授权微信账号登录或者手机号注册登录 | 需要同意《生态梨园圈子用户登录协议》 |
| 密码 | 6个字符以上 | 至少包含英文和数字，保证安全性 |
| 用户是否存在 | Int类型 | 0或1 |

## 5.2时间特性要求

响应时间：用户查看分类结果及对分类文本进行下载。

要求时间迅速，信息及时更新。