**软件需求规格说明书**

**基于弹幕评论的大数据分析平台**

**团队：开拓者开发团队**

目录

[1引言 3](#_Toc517985682)

[1.1编写目的 3](#_Toc517985683)

[1.2背景 3](#_Toc517985684)

[1.3定义 3](#_Toc517985685)

[1.4参考资料 3](#_Toc517985686)

[2任务概述 4](#_Toc517985687)

[2.1目标 4](#_Toc517985688)

[2.2用户的特点 7](#_Toc517985689)

[2.3假定和约束 7](#_Toc517985690)

[3需求规定 8](#_Toc517985691)

[3.1对功能的规定 8](#_Toc517985692)

[3.1.1主要功能 8](#_Toc517985693)

[3.1.2功能描述 9](#_Toc517985694)

[3.1.3用户界面 10](#_Toc517985695)

[3.1.4 用户场景 12](#_Toc517985696)

[3.2对性能的规定 15](#_Toc517985697)

[3.2.1精度 15](#_Toc517985698)

[3.2.2 灵活性 15](#_Toc517985699)

[3.3故障处理要求 15](#_Toc517985700)

[4运行环境规定 16](#_Toc517985701)

[4.1设备 16](#_Toc517985702)

[4.2开发环境 16](#_Toc517985703)

# [1引言](https://blog.csdn.net/royyeah/article/details/4808116#_Toc521463250)

## [1.1编写目的](https://blog.csdn.net/royyeah/article/details/4808116#_Toc521463251)

本文档的编写目的在于详细介绍基于弹幕评论的大数据分析平台所包含的需求，以便客户能够确认产品的确切需求以及开发人员能够根据需求设计编码，以下叙述将结合文字描述、UML用例图、数据库实体关系图、Er图、类图来描述斗鱼弹幕大数据分析平台的功能、性能、用户界面、运行环境、外部接口以及针对用户操作给出的各种响应。

本文档的预期读者有用户、开发人员以及跟该项目相关的其他竞争人员和无关人员。

## [1.2背景](https://blog.csdn.net/royyeah/article/details/4808116#_Toc521463252)

本文档介绍的产品是基于弹幕评论的大数据分析平台，该软件面向直播平台主播和直播平台用户。主要是利用国内知名的直播平台——斗鱼，通过其弹幕服务器第三方接入协议来爬取海量弹幕数据，对于弹幕数据做一个大数据的分析。

## [1.3定义](https://blog.csdn.net/royyeah/article/details/4808116#_Toc521463253)

1) 用户访谈:事先与客户沟通，见面分析产品需求的过程称用户访谈。

2) 开发人员:开发本文档所介绍的产品的程序员。

## [1.4参考资料](https://blog.csdn.net/royyeah/article/details/4808116#_Toc521463254)

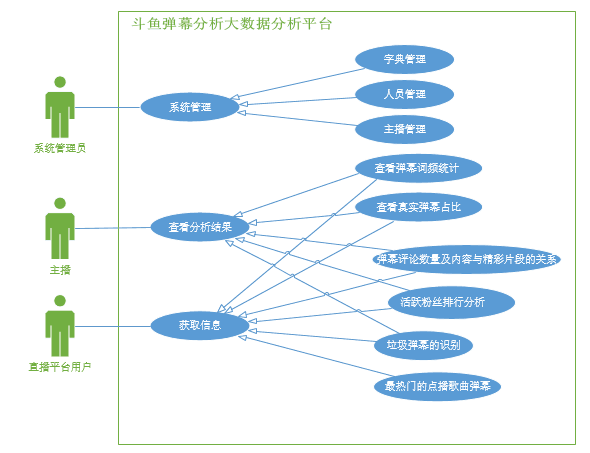
计算机软件需求规格说明书 （GB/T 9385-2008）

# [2任务概述](https://blog.csdn.net/royyeah/article/details/4808116#_Toc521463255)

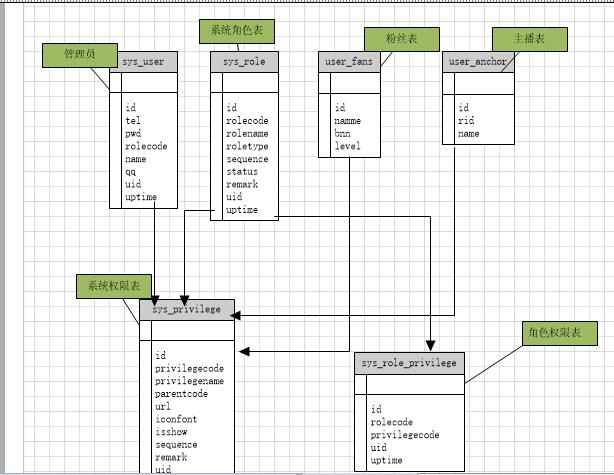
## [2.1目标](https://blog.csdn.net/royyeah/article/details/4808116#_Toc521463256)

基于弹幕评论的大数据分析平台是利用国内知名的直播平台-斗鱼，通过其弹幕服务器第三方接入协议来爬取海量弹幕数据，对于弹幕数据做一个大数据的分析，并将实验数据通过echarts展示在大数据分析平台上，主要适用于直播平台主播和直播平台用户。

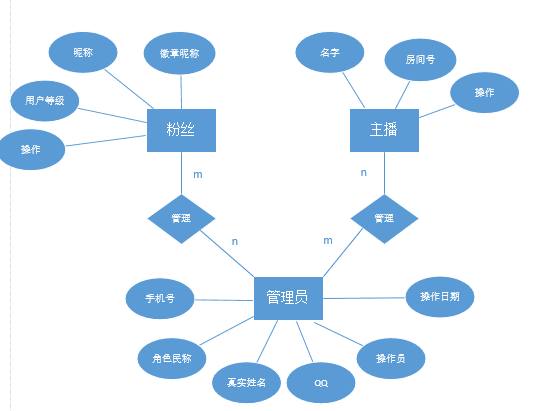
直播平台主播根据该平台提供的功能查看弹幕中高频率出现词语了解到观众对当前直播内容的看法，查看弹幕中真实弹幕的所占的比例可以知道真正关心直播内容的观众所占比例，查看相对活跃粉丝的排行情况可以适当的对活跃粉丝进行反馈回报，可以将垃圾弹幕进行识别并且屏蔽，了解到观看人群的偏好和兴趣所在。直播平台用户根据该平台提供的功能查看弹幕中高频率出现词语然后判断当前直播是否为自己感兴趣的内容，可以将垃圾弹幕进行识别并且屏蔽。



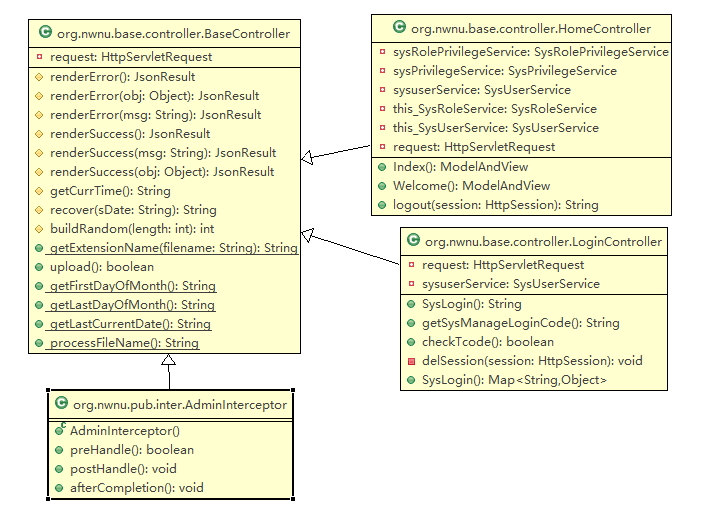
UML用例图

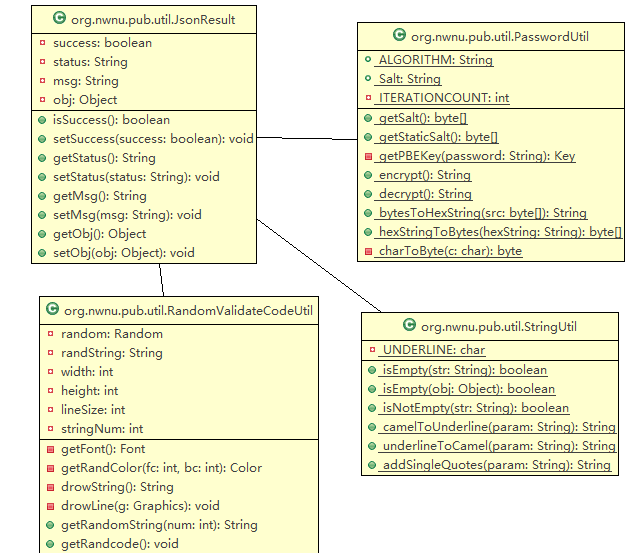


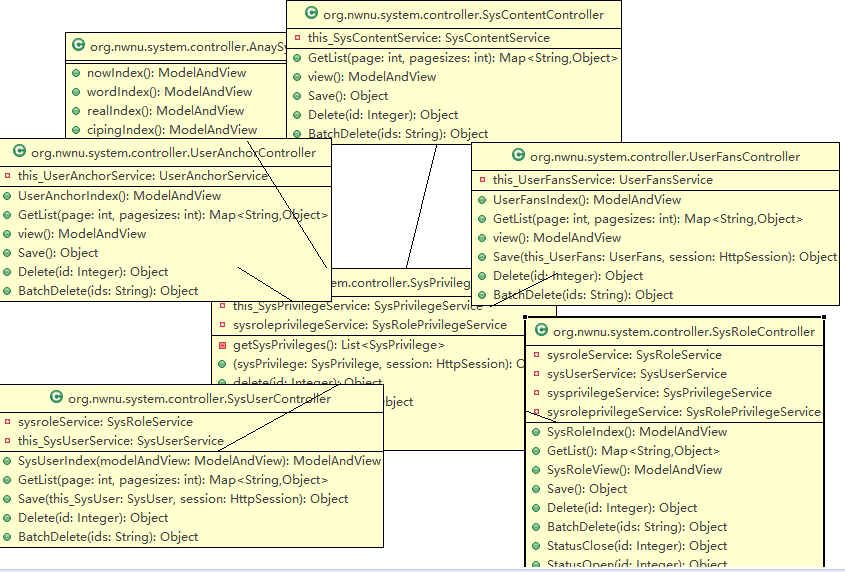
数据库实体关系图



Er图







类图

## [2.2用户的特点](https://blog.csdn.net/royyeah/article/details/4808116#_Toc521463257)

|  |  |
| --- | --- |
| 用户 | 描述 |
| 直播平台主播 | 直播平台主播是本软件使用人数最多的群体之一，直播平台主播可以通过本软件实时查看主播房间内的弹幕分析情况和活跃粉丝排行情况，了解当前观看人群的需求，故需要该软件的弹幕分析功能。 |
| 直播平台用户 | 直播平台用户可以通过本软件查看主播房间内的弹幕分析情况，由此判断该主播直播的内容是否是自己的兴趣所在，由此筛选自己感兴趣的直播内容，所以需要本软件的弹幕分析功能。 |

## [2.3假定和约束](https://blog.csdn.net/royyeah/article/details/4808116#_Toc521463258)

假定方面：

（1）开发团队均掌握开发系统所需的相关知识。 

（2）开发时间充足。 

约束方面：

|  |  |
| --- | --- |
| 约束 | 描述 |
| 高级语言 | 该软件首先利用Java多线程编程来实时爬取海量斗鱼弹幕数据，其次利用Spginmvc，Spring data，H-ui搭建本次项目的主体框架并集成数据可视化工具Echarts体现大数据分析结果，最后利用Python制作数据诊断工具来了解特征的分布及基本的统计指标值，再有目的性的选择一些聚类、分类、文本挖掘算法对于弹幕数据做一个大数据的分析，并将实验数据通过echarts展示在大数据分析平台上。 |
| 可靠性需求 | 该软件涉及利用国内知名的直播平台-斗鱼，通过其弹幕服务器第三方接入协议来爬取海量弹幕数据，因此需要长期维护，服务器的质量也需要有保障。 |
| 并行操作 | 该系统涉及到的用户有直播平台主播和直播平台用户，使用量大，可能存在同时在线人数过多导致系统崩溃等问题 |

# [3需求规定](https://blog.csdn.net/royyeah/article/details/4808116#_Toc521463259)

## [3.1对功能的规定](https://blog.csdn.net/royyeah/article/details/4808116#_Toc521463260)

### 3.1.1主要功能

|  |  |  |
| --- | --- | --- |
| 功能 | 概述 | 用户 |
| 弹幕词频统计分析 | 直播平台主播和用户查看软件对弹幕中高频率出现词语的统计分析 | 直播平台主播、直播平台用户 |
| 真实弹幕占比分析 | 直播平台主播查看软件对弹幕中真实弹幕的所占比例的分析 | 直播平台主播 |
| 活跃粉丝排行分析 | 直播平台主播查看软件对相对活跃粉丝的排行分析 | 直播平台主播 |
| 垃圾弹幕的识别 | 直播平台主播和用户可以通过软件对垃圾弹幕的识别对垃圾弹幕进行屏蔽 | 直播平台主播、直播平台用户 |
| 弹幕评论数量及内容与精彩片段的关系 | 直播平台主播可以利用软件对弹幕评论数量及内容与精彩片段的关系了解到观看人群的偏好和兴趣所在 | 直播平台主播 |

### 3.1.2功能描述

直播平台主播：

查看高频词汇统计分析：直播平台主播通过查看高频词汇统计分析了解到观看人群的偏好和兴趣，可据此修改直播内容，以便吸引更多用户观看。

查看真实弹幕所占比例：直播平台主播可以通过查看真实弹幕所占比例分析得出真实用户比例。

查看活跃粉丝排行：直播平台主播查看活跃粉丝排行得出粉丝的身份、年龄等资料，可以了解自己直播所吸引的观看人群的特点。

对垃圾弹幕进行识别：直播平台主播可以使用该软件将垃圾弹幕进行识别，进而将垃圾弹幕屏蔽。

查看弹幕评论数量及内容与精彩片段的关系：直播平台主播根据弹幕评论数量及内容与精彩片段的关系分析得出吸引观看人群的直播内容的特点。

直播平台用户：

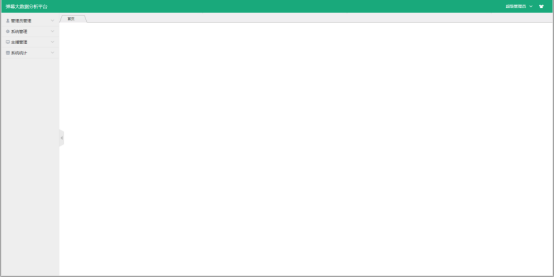
查看高频词汇统计分析：直播平台用户通过某个直播的高频词汇统计分析可以辨别当前直播内容是否是自己感兴趣的，可以据此筛选自己喜欢的直播内容。

对垃圾弹幕进行识别：直播平台用户可以使用该软件将垃圾弹幕进行识别，进而将垃圾弹幕屏蔽。

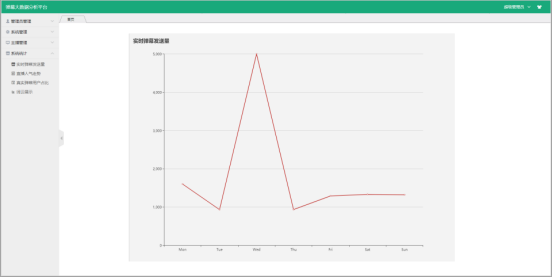
### 3.1.3用户界面



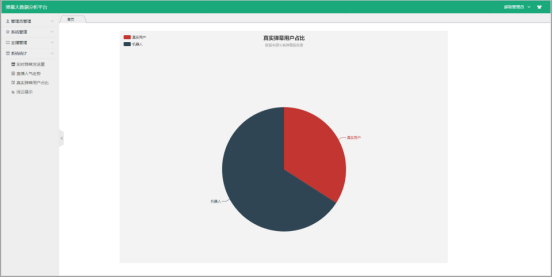
登录界面



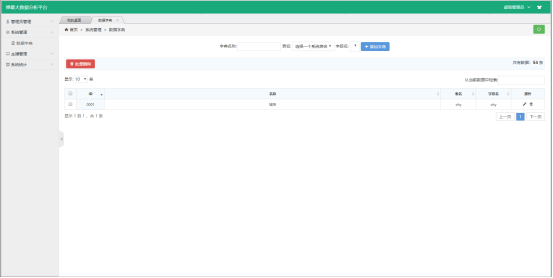
系统主界面



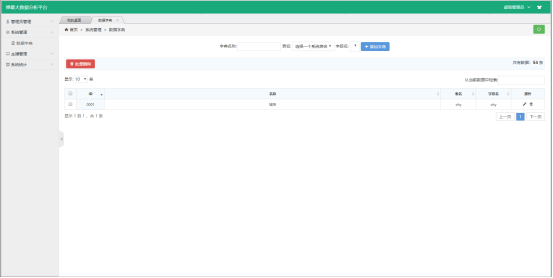
实时弹幕发送量



真实用户占比分析

****

管理员管理（角色管理，权限管理，管理员管理），系统管理（数据字典）等。

****

弹幕词云展示

### 3.1.4 用户场景

#### 3.1.4.1 典型用户

斗鱼主播

|  |  |
| --- | --- |
| 姓名 | 王佳佳 |
| 性别、年龄 | 女、21 |
| 职业 | 西北师范大学音乐学院大三学生 |
| 知识层次与能力 | 艺术系本科生，对直播平台比较了解 |
| 动机、目的、困难 | 在进行直播时，由于垃圾弹幕的影响使观看人数下降；无法摸清观众喜好 |
| 用户偏好 | 观看粉丝评论 |
| 典型场景 | 直播时总会出现各种各样的垃圾弹幕，让关注直播的人不可避免的看到，造成不好的影响；直播观看量时高时低，不知道比较会吸引观众的是哪方面的内容 |
| 典型描述 | 垃圾弹幕很烦人、不知道应该从哪方面入手对直播内容进行改进 |

斗鱼用户

|  |  |
| --- | --- |
| 姓名 | 罗希 |
| 性别、年龄 | 男、20 |
| 职业 | 西北师范大学计算机系大二学生 |
| 知识层次与能力 | 计算机专业学生，能够熟练使用电脑，对斗鱼直播平台有一定的了解和关注。 |
| 动机、目的、困难 | 筛选自己喜欢的直播内容，对垃圾弹幕进行识别 |
| 用户偏好 | 听歌、看游戏直播 |
| 典型场景 | 在斗鱼直播众多的视频里无法在短时间内精确找到自己感兴趣的内容推荐、观看直播时垃圾弹幕太多很影响心情 |
| 典型描述 | 想要平台能够精确推荐感兴趣内容、自动屏蔽垃圾弹幕 |

#### 3.1.4.2 背景

1）用户需要/迫切需要解决的问题

a.斗鱼主播王佳佳：在进行直播时，由于垃圾弹幕的影响使观看人数下降；无法摸清观众喜好

b.斗鱼用户罗希：筛选自己喜欢的直播内容，对垃圾弹幕进行识别

2）假定以下功能已经实现

a.弹幕词频统计分析功能已经实现

b.真实弹幕占比分析功能已经实现

c.活跃粉丝排行分析功能已经实现

d.垃圾弹幕的识别功能已经实现

e.弹幕评论数量及内容与精彩片段的关系分析功能已经实现

#### 3.1.4.3 场景

Part one

晚上七点，王佳佳打开斗鱼直播开始直播唱歌，观看人数因主播歌声独特优美而逐渐上涨，此时不良观众在评论区传播不良言语，恶意辱骂主播，让本来并不多的观看人数一下子下降了好多，王佳佳使用了基于弹幕评论的大数据分析平台，自动识别了垃圾弹幕并且进行了屏蔽，让新的观众可以专心倾听她的歌声并因此称为她的粉丝。

Part two

罗希同学结束了忙碌的一天，回到寝室一边泡脚一边打开了斗鱼直播，打算看自己喜欢的游戏主播——韩金龙的英雄联盟比赛讲解和直播，看的正起劲，双方交战正酣，突然屏幕被很多垃圾广告占领：咨询小编QQ咨询【39739693】也可加我微信号： dh7040（手机加QQ好友请认准第一个人）小编没有QQ群，其他都是骗子，切勿上当受骗！！瞬间占领了整个屏幕，而这种垃圾弹幕霸屏的场景已经不是第一次出现了，罗希一边想继续看比赛进展，在弹幕上跟其他的相关弹幕进行交流，但又恼火与层出不穷的垃圾弹幕，体验极差，心情也变得有点烦。本来好好的睡前放松就被乱七八糟的垃圾弹幕搅扰了。罗希同学使用了基于弹幕评论的大数据分析平台，自动识别了垃圾弹幕并且进行了屏蔽，他再也不用受垃圾弹幕满天飞的困扰了。

## [3.2对性能的规定](https://blog.csdn.net/royyeah/article/details/4808116#_Toc521463261)

### 3.2.1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：

|  |  |  |
| --- | --- | --- |
| 字段 | 精度 | 备注 |
| 账号 | 11 个字符,只包括数字 | 账号跟手机号一样 |
| 密码 | 5-20 个字符，包括大小写字母，数字，下划线 | 密码由用户自己设置 |
| 姓名 | 2-20 个字符，包括中文以及“·” | 用户可能会有外国人 |
| 昵称 | 2-20 个字符，包括中文以及“·” |  |
| 用户等级 | 1-6 个字符，只包括数字 |  |
| 房间号 | 2-20 个字符，只包括数字 |  |
| 手机号码 | 11 个字符,只包括数字 |  |

### 3.2.2 灵活性

1）对用户信息进行保存，如果用户发生变化，可更改注册用户的基本信息。

2）需求发生变化时，可以在现行代码的基础上稍作修改便可满足需求。

## [3.3故障处理要求](https://blog.csdn.net/royyeah/article/details/4808116#_Toc521463267)

1）考虑到多人同时访问服务器，服务器是否能正常运行，是否会发生崩溃

2）多人同时操控数据库是否会产生冲突等都需要在后面的设计文档中详细分析

3）用户和密码错误时反馈错误

# [4运行环境规定](https://blog.csdn.net/royyeah/article/details/4808116#_Toc521463269)

## [4.1设备](https://blog.csdn.net/royyeah/article/details/4808116#_Toc521463270)

移动终端，PC。

## [4.2开发环境](https://blog.csdn.net/royyeah/article/details/4808116#_Toc521463271)

java

python

mysql

maven

Tomcat