



北京交通大学
Beijing Jiaotong University

软件构思综合训练

影视舆情分析系统 详细设计

组 长：汤新宇 17301137

小组成员：王心蕊 17301048

陈嘉琪 17301060

唐 麒 17301138

张钰铎 17301145

贡献分配：同 等 贡 献

指导老师：李 宇

北京交通大学软件学院

2020 年 6 月 15 日

目录

- 1 项目概述 1
 - 1.1 项目背景 1
 - 1.2 业务概述 1
 - 1.3 技术概述 1
 - 1.3.1 开发工具 1
 - 1.3.2 技术路线 1
- 2 总体设计 2
 - 2.1 概述 2
 - 2.2 系统环境描述 2
 - 2.2.1 开发环境 2
 - 2.2.2 运行环境 2
 - 2.3 系统总体结构设计 3
 - 2.3.1 系统业务层次图 3
 - 2.3.2 模块功能介绍 3
 - 2.3.3 模块间接口设计 3
- 3 系统设计 3
 - 3.1 MVC架构 3
 - 3.1.1 表示层 3
 - 3.1.2 业务逻辑层 3
 - 3.1.3 数据层 3
 - 3.2 系统架构图 4
- 4 包设计 4
 - 4.1 整体架构设计 4
 - 4.2 包的设计说明 4
- 5 类设计 5
- 6 接口设计 9
 - 6.1 外部接口设计 9
 - 6.2 内部接口设计 9
 - 6.2.1 注册登录模块 9
 - 6.2.2 搜索模块 10
 - 6.2.3 展示模块 10
 - 6.2.4 记录查询模块 10

6.2.5	支付模块	11
6.2.6	监控与预警模块	11
6.2.7	排行榜模块	11
6.2.8	分析模块	11
7	数据库设计	12
7.1	系统数据流设计	12
7.1.1	数据获取到数据存储数据流设计	12
7.1.2	数据存储到前端展示数据流设计	12
7.2	系统数据库设计	13

1 项目概述

1.1 项目背景

根据中国互联网络信息中心（CNNIC）第40次《中国互联网络发展状况统计报告》显示，截至2017年6月，中国网民规模达到7.51亿，互联网普及率为54.3

经过十多年的发展，以政府部门为主导的舆情管理市场进入了高速成长期，年均增幅50政府在舆情产业中起着主导作用。政府通过制定积极政策对舆情加以引导，同时以自身需要，作为大客户推动舆情领域发展，媒体尤其是官媒、新媒体，在舆情产业的发展中起着某种先锋作用。舆情管理在政策和市场双重因素的刺激下，基本形成了“官商媒教”的多元化产业格局，形成了政府、媒体、教育科研、商业软件四大背景的行业格局。

对于影视作品来说，舆论与口碑在很大程度上决定了影视作品的成功与否。影视舆情信息在互联网上客观存在,形成了与影视剧内容产品相伴而生的衍生信息产品,形成影视剧不同传播时期的先期舆情(影视剧首轮发行期前)、同期舆情(首轮发行期同步)、后期舆情(首轮播映后)、长尾舆情。先期舆情和同期舆情在一定程度上作用于影视剧作品的观影期待,影响观看意向、初步评价、后期评价、口碑评分等,进而影响收视率和票房。因此，舆论导向的监测、分析与预警对于影视行业来说十分重要。

1.2 业务概述

1.3 技术概述

1.3.1 开发工具

硬件：PC端、移动端

软件：浏览器 IE9 及以上等主流浏览器；Windows 操作系统；hadoop 框架，MongoDB 数据库

1.3.2 技术路线

浏览器端：前端采用 Vue.js 实现。

前后端交互：客户端通过Http请求服务器接口，服务器根据对应参数，返回JSON数据，客户端解析JSON数据，进行数据显示。

后台服务器：后端采用 Spring + Spring MVC +Hadoop 框架实现。

页面	序号	功能模块	功能概述
首页	1	登录/注册	用户进行登录/注册
	2	热点舆情实时展示	用户可浏览当前热点舆情事件
	3	功能入口	用户可查看系统功能
个人中心	4	账号信息	用户查看自己的账号信息
	5	历史数据	用户查看自己的历史记录
	6	近期关注	用户查看近期关注的热点
	7	系统消息	用户查看收到的系统消息
	8	历史分析报告	企业用户获得权限后可查看历史分析报告
	9	预警详情	企业用户获得权限可查看预警详情
个人用户	10	筛选影视舆情事件	查询想要了解的舆情事件
	11	添加关注	对感兴趣的热点事件添加关注
企业用户	12	查看分析数据	查看系统对某一事件的分析数据
舆情事件	13	事件详情	当前事件的关系网络，关注度等信息
后台	14	系统管理	管理员对系统进行维护

2 总体设计

2.1 概述

2.2 系统环境描述

2.2.1 开发环境

表 1: 舆情数据获取

	运行环境	版本
Python	Windows10	3.7.4
MongoDB	Windows10	4.2

表 2:

	运行环境	版本
NameNode	Centos7, Java8	2.10.0
SecondNameNode	Centos7, java8	2.10.0
DataNode	Centos7, java8	2.10.0

2.2.2 运行环境

表 3: 舆情数据获取

	运行环境	版本
Python	Centos7	3.7.4
MongoDB	Centos7	4.2

表 4:

	运行环境	版本
NameNode	Centos7, Java8	2.10.0
SecondNameNode	Centos7, java8	2.10.0
DataNode	Centos7, java8	2.10.0

2.3 系统总体结构设计

2.3.1 系统业务层次图

2.3.2 模块功能介绍

2.3.3 模块间接口设计

3 系统设计

3.1 MVC架构

3.1.1 表示层

表示层是平台直接面向用户的界面，直接与用户交互。表示层的主要功能是提供类型的用户操作界面和操作方案,捕捉和收集用户的输入信息通过表示层和给服务器处理的信息,从而给前台反馈。用户应该在表示层的功能组中进行所有功能模块的操作。前端表示层将收集接收到的请求和各种数据，然后将它们传输到应用程序层进行相应的处理。

3.1.2 业务逻辑层

业务逻辑层的主要功能是根据实际的业务规则实现相关的业务逻辑功能。在这个层中，实现了在表示层中相关的目标服务和功能模块。

3.1.3 数据层

数据库管理层的主要职责是管理数据库中不同类型的连接和断开连接，并记录和提示数据库操作中发生的一些异常，以方便开发人员对相关数据库功能进行测试或调试。系统的所有功能都与各种信息配置、业务处理数据、系统运行关系和其他信息有关;数据库文件这些数据和信息，并提供一些基本的搜索接口，以保证数据的可靠性和完整性。

3.2 系统架构图

4 包设计

4.1 整体架构设计

本系统为了能更加有效地进行整合、生产宏观模型，就需要对系统中的类进行分组，以下是本系统中的包设计。

4.2 包的设计说明

下表为服务器的包设计表

表 5: 包设计表

包名	设计说明
com.yuqing.yingshi.service	该包主要存放了高层调用Dao的接口
com.yuqing.yingshi.controller	改包主要存放了业务逻辑相关类，依赖于com.yuqing.yingshi.service，调用Service以完成业务逻辑
com.yuqing.yingshi.result	该包存放了服务器与web端交互时返回的结果对应编码

表 6: com.yuqing.yingshi.dao相关类

com.yuqing.yingshi.dao相关类	
UserDao	封装与User实体类相关的数据访问方法
NewsDao	封装与News实体类相关的数据访问方法
CompanyDao	封装与Company实体类相关的数据访问方法
MovieDao	封装与Movie实体类相关的数据访问方法
TVseriesDao	封装与TVseries实体类相关的数据访问方法
VarietyDao	封装与Variety实体类相关的数据访问方法
StudioDao	封装与Studio实体类相关的数据访问方法
RoleDao	封装与Role实体类相关的数据访问方法
HistoryDao	封装与History实体类相关的数据访问方法
ForkDao	封装与Fork实体类相关的数据访问方法
MonitorDao	封装与Monitor实体类相关的数据访问方法
ReportDao	封装与Report实体类相关的数据访问方法
MarketingDao	封装与Marketing实体类相关的数据访问方法
InfoDao	封装与Info实体类相关的数据访问方法
OrderDao	封装与Order实体类相关的数据访问方法

表13：包设计表

com.yuqing.yingshi.dao	主要负责数据访问，该包中的类封装了对数据库的访问（只包含最原子的数据操作），供高层调用，依赖于com.yuqing.yingshi.entity
com.yuqing.yingshi.config	存放了与前端web网页交互的相关配置
com.yuqing.yingshi.entity	该包主要存放数据库表中所对应的实体类
com.yuqing.yingshi.service.impl	该包存放了service包中定义的接口的具体实现
com.yuqing.yingshi.util	存放了项目相关工具类

5 类设计

本小结以包为单位，对每个包所含有的类及类与类间的调用关系进行说明。

表 7: com.yuqing.yingshi.entity相关类

com.yuqing.yingshi.entity相关类	
User	用户个人信息的实体类
News	新闻信息的实体类
Company	公司信息的实体类
Movie	电影信息的实体类
TVseries	影视剧信息的实体类
Variety	综艺信息的实体类
Studio	工作室信息的实体类
Role	演员/歌手等明星信息的实体类
History	历史记录的实体类
Fork	收藏记录的实体类
Monitor	监控预警项的实体类
Report	报告的实体类
Marketing	营销效果的实体类
Info	单条博文/帖子信息的实体类
Order	用户订单的实体类

表 8: com.yuqing.yingshi.service相关类

com.yuqing.yingshi.service相关类	
UserService	定义了与User实体类相关的业务逻辑接口
NewsService	定义了与News实体类相关的业务逻辑接口
CompanyService	定义了与Company实体类相关的业务逻辑接口
MovieService	定义了与Movie实体类相关的业务逻辑接口
TVseriesService	定义了与TVseries实体类相关的业务逻辑接口
VarietyService	定义了与Variety实体类相关的业务逻辑接口
StudioService	定义了与Studio实体类相关的业务逻辑接口
RoleService	定义了与Role实体类相关的业务逻辑接口
HistoryService	定义了与History实体类相关的业务逻辑接口
ForkService	定义了与Fork实体类相关的业务逻辑接口
MonitorService	定义了与Monitor实体类相关的业务逻辑接口
ReportService	定义了与Report实体类相关的业务逻辑接口
MarketingService	定义了与Marketing实体类相关的业务逻辑接口
InfoService	定义了与Info实体类相关的业务逻辑接口
OrderService	定义了与Order实体类相关的业务逻辑接口

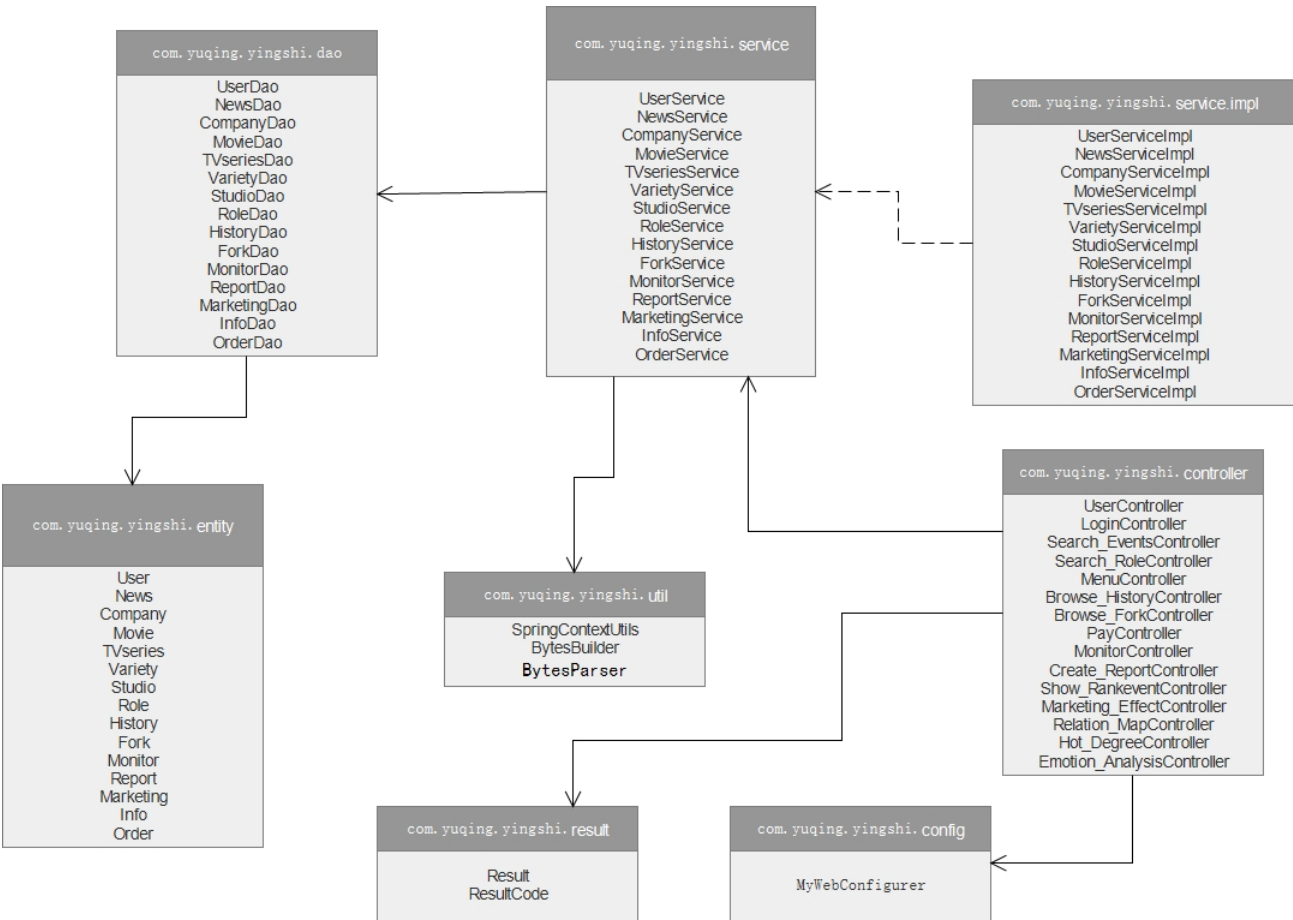


图 1: 舆情系统包的设计图

表 9: com.yuqing.yingshi.util相关类

com.yuqing.yingshi.util相关类	
SpringContextUtils	获取spring容器bean对象工具类
BytesBuilder	字节构造器
BytesParser	字节处理器

表 10: com.yuqing.yingshi.result相关类

com.yuqing.yingshi.result相关类	
Result	结果返回形式定义
ResultCode	定义返回结果编码

表 11: com.yuqing.yingshi. controller相关类

com.yuqing.yingshi. controller相关类	
UserControlle	定义了响应用户相关界面不同点击事件的方法，根据事件的不同调用 UserService的相应方法实现功能，以json格式返回。
LoginController	定义了登陆注册界面不同点击事件的方法，根据事件的不同调用 UserService的相应方法实现功能，以json格式返回。
Search_EventsController	定义了搜索热门影视事件/影视作品相关界面不同点击事件的方法，根据事件的不同调用 NewsService、MovieService、TVseriesService、VarietyService的相应方法实现功能，以json格式返回。
Search_RoleController	定义了搜索明星/公司/工作室相关界面不同点击事件的方法，根据事件的不同调用 Role-Service、CompanyService、StudioService的相应方法实现功能，以json格式返回。
MenuController	定义了展示影视事件/明星/影视作品菜单相关界面不同点击事件的方法，根据事件的不同调用 RoleService、CompanyService、StudioService、NewsService、MovieService、TVseriesService、VarietyService的相应方法实现功能，以json格式返回。
Browse_HistoryController	定义了浏览历史记录相关界面不同点击事件的方法，根据事件的不同调用 HistoryService的相应方法实现功能，以json格式返回。
Browse_ForkController	定义了浏览收藏记录相关界面不同点击事件的方法，根据事件的不同调用 ForkService的相应方法实现功能，以json格式返回。

6 接口设计

6.1 外部接口设计

外部接口是各模块间进行交互访问以及web访问系统时的外部接口，某一模块的具体功能则由内部接口设计实现。

序号	模块功能	外部接口
1	用户注册	User_SignUp
2	用户登录	User_SignIn
3	搜索	User_Search
4	展示	Menu_Show
5	记录查询	Browse_Record_Query
6	支付	Pay
7	监控与预警	Monitor
8	排行榜	Rank_Query
9	舆情分析报告	Generate_Analysis_Report
10	营销效果分析报告	Marketing_Effect_Analysis
11	热度分析	HotDegree_Analysis
12	情感分析	Emotion_Analysis
13	任务关系图谱	Generate_RelationMap

图 2: 外部接口设计

6.2 内部接口设计

6.2.1 注册登录模块

外部接口	内部调用函数	函数说明
User_SignUp	addUser	用户注册，将用户的输入信息插入数据库用户表中。
User_SignIn	getUserById	用户登录，在用户表中比对用户输入的用户名和密码是否存在并匹配，返回值决定弹窗提示信息。

图 3: 注册登录模块

6.2.2 搜索模块

外部接口	内部调用函数	函数说明
User_Search	searchEvents	用户搜索自己感兴趣的事件，用户通过输入关键字来搜索特定事件。
	searchRole	用户搜索自己感兴趣的明星，用户通过输入姓名来搜索特定明星。

图 4: 搜索模块

6.2.3 展示模块

外部接口	内部调用函数	函数说明
Menu_Show	showEvents	首页搜索栏下面为当今热点事件的展示。
	showRoles	首页搜索栏下面为当今热点明星相关信息的展示。

图 5: 展示模块

6.2.4 记录查询模块

外部接口	内部调用函数	函数说明
Browse_Record_Query	browseHistoryQuery	用户查询自己的浏览记录并且展示在界面上。
	browseForkQuery	用户查询自己的收藏记录并且展示在界面上。

图 6: 记录查询模块

外部接口	内部调用函数	函数说明
Pay	sumMoney	用户提交订单后进行支付。

图 7: 支付模块

外部接口	内部调用函数	函数说明
Monitor	monitorWarning	对某特定时间进行监督预警。

图 8: 监控与预警模块

外部接口	内部调用函数	函数说明
Rank_Query	rankByDay	用户查询热点事件日排名并且显示在界面上。
	rankByWeek	用户查询热点事件周排名并且显示在界面上。
	rankByMonth	用户查询热点事件月排名并且显示在界面上

图 9: 排行榜模块

6.2.5 支付模块

6.2.6 监控与预警模块

6.2.7 排行榜模块

6.2.8 分析模块

外部接口	内部调用函数	函数说明
Generate_Analysys_Report	generateAnalysisReport	根据特定的事件系统自动生成舆情分析报告。
Marketing_Effect_Analysis	marketingEffectAnalysis	根据特定的事件进行营销效果分析。
HotDegree_Analysiss	hotDegreeAnalysis	对特定的事件进行热度分析。
Emotion_Analysis	emotionAnalysis	对特定的事件进行情感分析。
Generate_Relation_Map	generateRelationMap	对特定的人物生成人物关系图谱。

图 10: 分析模块

7 数据库设计

7.1 系统数据流设计

热点发现和舆情分析中数据是重中之重。网络数据具有信息量大、更新速度快块的特点，且网络热点话题具有一定的时效性，因此需要从数据获取到结果展示，要有及时、稳定和高效的流程。系统的核心数据流主要为数据获取到数据存储数据流和数据存储到前端展示数据流。本系统使用多种形式的数据库。与市面常见的舆情监控系统相比较，一大优势在于数据源多。传统的舆情监控系统多数只采用媒体新闻数据或单一的新浪微博数据。本系统不仅包含媒体新闻数据、主流门户（sohu、网易、新浪）评论数据、新浪微博数据还使用公司内部的说说、兴趣部落数据和电脑管家、浏览器搜索日志，将社交数据和搜索日志也加到热点舆情发现与分析中。综上，恰当的数据流设计，至关重要。

7.1.1 数据获取到数据存储数据流设计

网络数据的获取来源主要包含外网下载和内部推送。外网下载的数据包括媒体新闻数据、主流门户（sohu、网易、新浪）评论数据和新浪微博数据。外网下载依托spider下载平台，对网页数据进行抓取，并通过通用抽取和特定抽取流程获得结构化信息。Spider下载平台的能够通过下载机器的扩容，提升下载量，每日千万级网页的下载，能够满足本系统的需求。对于普通站点新闻的下载结果使用通用抽取即可满足需求，但对于评论数据和新浪微博数据需要定制特定的抽取程序。内部推送即使用公司内部数据，直接按照指定格式进行推送。内部推送数据包括快报新闻数据、tx新闻评论数据、说说兴趣部落数据和电脑管家、浏览器搜索日志。内部推送的快报新闻数据和tx新闻评论数据，避免了下载、抽取步骤对数据处理的实时性和准确性有了大大的提升。电脑管家、浏览器搜索日志是生成搜索指数的重要来源。说说、兴趣部落数据是生成社交指数的重要来源。数据获取后对新闻数据进行新闻预处理，将处理后的数据一份推送到Hadoop进行存储备份和热点发现使用，另一份推送按指定字段建立索引，并存储到检索系统中，方便之后流程的使用。同样，对新闻评论数据进行评论预处理，按指定字段建立索引存储到检索系统。除此之外，直接对新浪微博数据，兴趣部落数据和搜索日志数据，建立索引，存储到检索系统中。

7.1.2 数据存储到前端展示数据流设计

从Hadoop集群中取得最新的一个小时内的媒体数据，对标题进行热点聚合，得到热点话题和相应关键字列表。在检索系统中按关键字对媒体新闻数据进行检索，并对

检索结果进行计算得到词云、媒体渠道分布、媒体热度趋势数据。对社交数据进行检索计算得到社交渠道分布、社交热度趋势和社交情感趋势数据。对评论数据进行检索计算得到媒体情感分布和用户代表观点。得到上述数据，存入Mysql数据库中，方便前端展示。最后，前端通过对Mysql数据库中数据的使用，展示数据并实现业务逻辑。

7.2 系统数据库设计

热点发现与线索管理系统和舆情系统，数据和业务之间存在一定的耦合。热点发现与线索管理系统主要为个人提供内容生产服务，也会对热点的相关舆情进行分析。舆情系统是对企业提供舆情服务，对企业制定舆情服务，进行舆情监控与报警。针对这些耦合，两个系统后台使用同一数据库，按不同的业务逻辑进行定制和展示。数据库中数据表较多，选取具有代表性的表进行展示。版本控制表是整个数据库中的核心表，通过versionid在不删除和覆盖其他相关表中数据的情况下进行相关数据的更新。版本控制表如下表所示，以versionid为主键，按 versionid或topicid根据不同的需要与其他表做关联。

表 12: 版本控制表

字段	数据类型	主键	是否可控	备注
versionid	int(11)	是	否	id自增
date	varchar(32)	否	否	数据日期
topicid	varchar(100)	否	否	热点id
keyscontent	varchar(128)	否	否	关键词
trade	varchar(40)	否	否	一级行业
strade	varchar(40)	否	否	二级行业
addtime	timestamp	否	否	添加日期

热门话题日数据表如下表所示。与热门话题日数据表结构相类似的还有热门话题周数据表、热门话题月数据表、突发事件数据表、头条新闻数据表和搜索排行数据表等。

表 13: 热门话题数据表

字段	数据类型	主键	是否可控	备注
id	int(11)	是	否	id自增
category	varchar(50)	否	否	一级类别
category2	varchar(200)	否	否	二级类别

Table 13 continued from previous page

classes	varchar(50)	否	否	渠道
topicaid	varchar(64)	否	否	标题
name	varchar(50)	否	否	热度
info	varchar(120)	否	否	URL链接
impact	int(11)	否	否	相关文章个数
timestamp	int(11)	否	否	热点开始时间
reputation	varchar(100)	否	否	热点评论数
heat	int(11)	否	否	热度
locationlist	varchar(64)	否	否	地点列表
versionid	int(8)	否	否	版本数据
detailflag	tinyint(1)	否	否	标识符(有无详情页)
mediahot	int(1)	否	否	媒体热度
socailhot	int(11)	否	否	社交热度
searchhot	int(11)	否	否	搜索热度
imageurl	varchar(256)	否	否	图片地址
personlist	varchar(256)	否	否	人物列表
organizationlist	varchar(256)	否	否	机构单位
cloudwordlist	text	否	否	词云
keywordlist	varchar(256)	否	否	关键词列表
addtime	datetime	否	否	入库时间

用户企业与告警对象一对多联系。一个用户企业可以设置多个告警对象，监测不同方面舆情， 但一个告警对象只能属于特定的用户企业。告警对象与告警配置一对多联系。一个告警对象可以设置多个告警配置， 但一个告警配置只能归属于一个告警对象。告警对象与告警内容一对多联系。一个告警对象可以有多条告警内容， 但一个告警内容只能归属于一个告警对象。告警文章与告警内容多对多联系。一篇告警文章可以对应多个告警内容， 一个告警内容也可以对应多篇告警文章。舆情告警相关数据库设计图如下图所示。

舆情分析数据库设计图如下

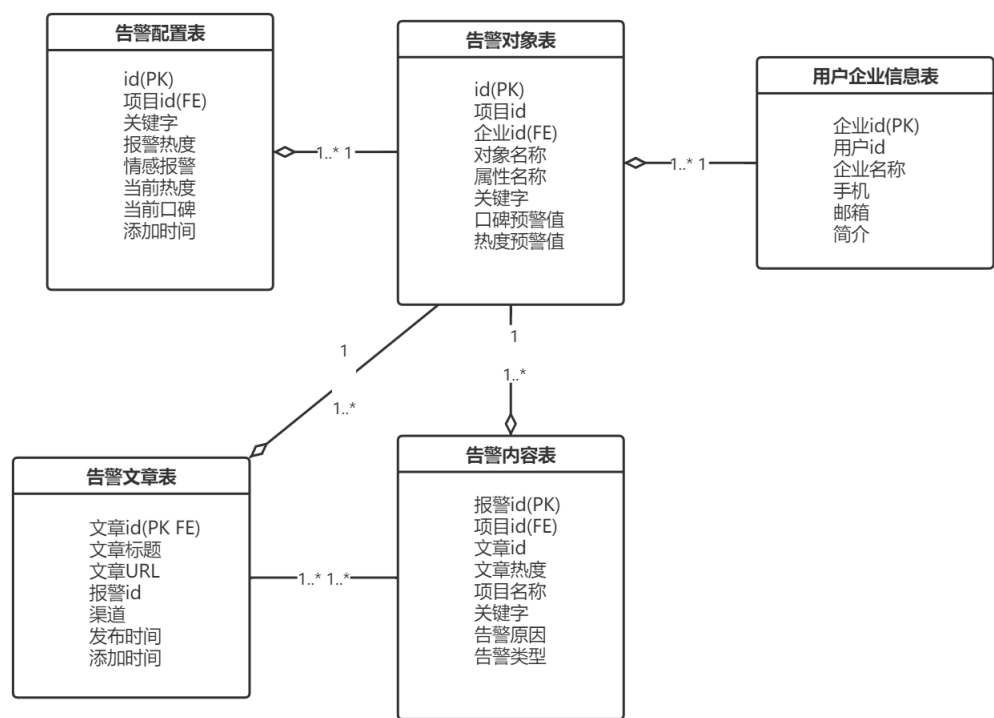


图 11: 舆情告警数据库设计图

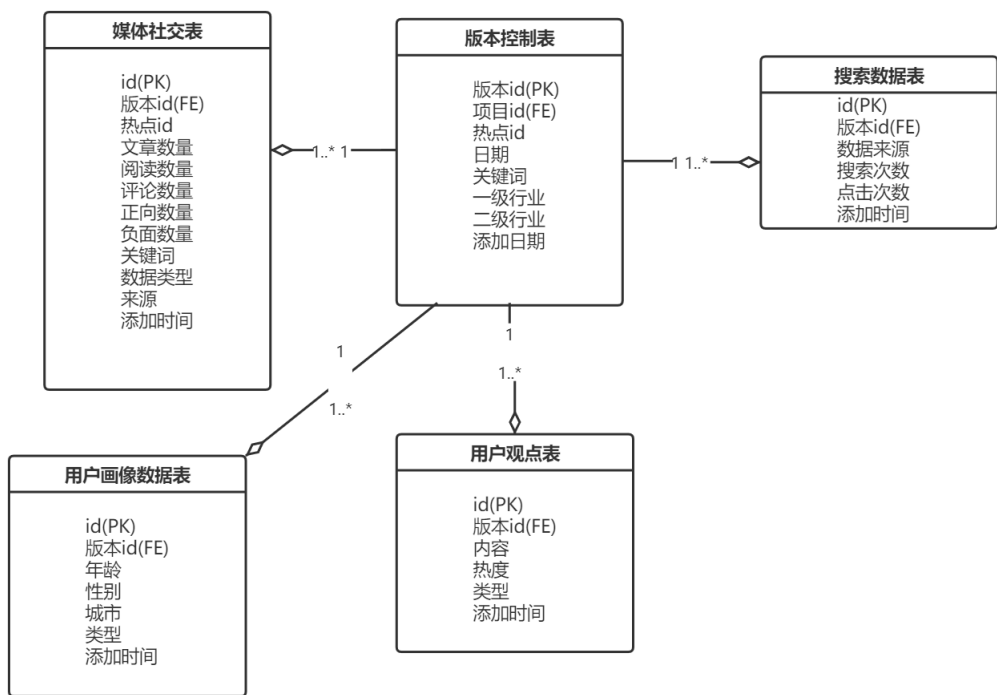


图 12: 舆情分析数据库设计图