**《数据挖掘实践任务教学平台》**

**需求规格说明书**

|  |  |  |  |
| --- | --- | --- | --- |
| **作者：** | **马腾** | **日期：** | 2018.04 |
| **审核：** | **刘祥德** | **日期：** | 2018.05 |
| **批准：** |  | **日期：** |  |

**变更记录**

|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | **版本** | **变更说明** | **作者** |
| 2018.04.06 | V1.1 | 文档创建 | 马腾 |
| 2018.04.24 | V1.2 | 需求的详细说明 | 马腾 |
| 2018.05.04 | V1.3 | 修订了功能性需求 | 刘祥德 |
|  |  |  |  |
|  |  |  |  |

目录

[1. 引言 5](#_Toc523819919)

[1.1 目的 5](#_Toc523819920)

[1.2 背景 5](#_Toc523819921)

[1.3 参考资料 6](#_Toc523819922)

[1.4 术语和缩略语 7](#_Toc523819923)

[2. 系统概述 8](#_Toc523819924)

[2.1 产品描述 8](#_Toc523819925)

[2.2 产品目标 9](#_Toc523819926)

[2.3 产品功能 10](#_Toc523819927)

[2.3.1 系统用例模型 10](#_Toc523819928)

[2.4 一般约束 10](#_Toc523819929)

[2.5 平台开发约束 11](#_Toc523819930)

[2.5.1 Web三层架构 12](#_Toc523819931)

[2.5.2 表现层 12](#_Toc523819932)

[2.5.3 业务层 14](#_Toc523819933)

[2.5.4 数据访问层 15](#_Toc523819934)

[3. 功能性需求 15](#_Toc523819935)

[3.1 学生端 15](#_Toc523819936)

[3.1.1 登陆学生端 16](#_Toc523819937)

[3.1.2 学生浏览数据集 17](#_Toc523819938)

[3.1.3 学生浏览算法库 17](#_Toc523819939)

[3.1.4 学生查看分配任务 17](#_Toc523819940)

[3.1.5 学生提交任务 17](#_Toc523819941)

[3.1.6 查看提交结果 17](#_Toc523819942)

[3.2 教师端 18](#_Toc523819943)

[3.2.1 数据集管理 18](#_Toc523819944)

[3.2.2 学生信息管理 20](#_Toc523819945)

[3.2.3 分组信息管理 22](#_Toc523819946)

[3.2.4 算法库管理 22](#_Toc523819947)

[3.2.5 实践任务管理 24](#_Toc523819948)

[4. 非功能性需求 25](#_Toc523819949)

[4.1 接口需求 25](#_Toc523819950)

[4.1.1 用户接口： 25](#_Toc523819951)

[4.1.2 软件接口： 26](#_Toc523819952)

[4.1.3 通信接口 26](#_Toc523819953)

[4.2 硬件需求 26](#_Toc523819954)

[4.3 数值需求 26](#_Toc523819955)

[4.3.1 静态数值需求： 26](#_Toc523819956)

[4.3.2 动态数值需求： 26](#_Toc523819957)

[4.4 性能需求 26](#_Toc523819958)

[4.5 服务需求 27](#_Toc523819959)

[4.6 可用性需求 27](#_Toc523819960)

# 引言

## 目的

编写此文档的目的是进一步定制《数据挖掘实践任务教学平台》的需求细节问题,希望能使本软件开发工作更具体。为使用户、软件开发者及分析人员对该软件的初始规定有一个共同的理解，它说明了本产品的各项功能需求、性能需求和数据要求，明确标识各功能的实现过程，阐述实用背景及范围，提供客户解决问题或达到目标所需的条件或权能，提供一个度量和遵循的基准。

## 背景

云计算时代的强势来袭，许多高校已经陆续对数据挖掘专题学科的探索、设立。但是，数据挖掘是从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不知道的、但又是潜在有用的信息和知识的过程，这个过程是业务理解、数据理解、数据准备、模型建立等多方面的综合。这给高校开展数据挖掘方面的教学任务带来很大的挑战。在数据挖掘教学的实践操作中，会涉及到通知发布、任务建立、数据集准备、挖掘算法发布等一系列复杂流程。如何正确组织这些数据，执行这些教学流程，需要花费师生一定的精力。《数据挖掘实践任务教学平台》的设计与开发就是围绕着数据挖掘教学目的进行的。

这个系统意在给老师、学生提供一个实践任务教学平台，管理员可以在平台轻松导入学生信息，给学生分配指定类型的数据发掘任务，并同步追踪学生的数据发掘进度，进行有效的教学评估。学生在平台领取一系列发掘任务，根据任务的要求执行数据挖掘算法，产生挖掘结果，完成数据挖掘流程的分类、估计、预测、分析、总结工作，并上传期望进度，汇报发掘结果。当然，系统集成了数据集管理模块，是数据挖掘教学实践任务展开的核心。

系统会提供诸如图形、表格的数据呈现手段，从而完成数据理解、任务追踪的可视化帮助教学任务的较好完成。

《数据挖掘实践任务教学平台》的功能性需求是由具备《数据挖掘》等课程讲授资格的高校教授提出，面向的用户主要是高校师生及相关培训机构使用。开发者为具备科研训练任务的本科生三名。

## 参考资料

|  |  |  |  |
| --- | --- | --- | --- |
| **资料名称[标识符]** | **出 版 单 位** | **作 者** | **日 期** |
| Hibernate实战（第2版） | 清华大学出版社 | Christian Bauer  Gavin King  Gary Gregory | 2016 |
| Spring MVC实战 | 电子工业出版社 | 国华 | 2016 |
| Spring Boot 2精髓：从构建小系统到架构分布式大系统 | 电子工业出版社 |  | 2017 |
| 数据挖掘 概念与技术（原书第3版） [Data Mining Concepts and Techniques Third Edition] | 机械工业出版社 |  | 2012-08-01 |
| JavaEE开发的颠覆者：Spring Boot实战 | 电子工业出版社 |  |  |
| Vue.js 前端开发 快速入门与专业应用 |  |  |  |
| Vue.js权威指南 |  |  |  |

## 术语和缩略语

|  |  |
| --- | --- |
| **术语、缩略语** | **解 释** |
| REST风格 | REST即表述性状态传递（英文：Representational State Transfer，简称REST）是Roy Fielding博士在2000年他的博士论文中提出来的一种软件架构风格。它是一种针对网络应用的设计和开发方式，可以降低开发的复杂性，提高系统的可伸缩性。目前在三种主流的Web服务实现方案中，因为REST模式的Web服务与复杂的SOAP和XML-RPC对比来讲明显的更加简洁，越来越多的web服务开始采用REST风格设计和实现。例如，Amazon.com提供接近REST风格的Web服务进行图书查找；雅虎提供的Web服务也是REST风格的。 |
|  |  |
|  |  |

# 系统概述

## 产品描述

《数据挖掘实践任务教学平台》可以让教师对学生信息进行管理，包括一键导入、查询、修改、删除；提供数据集管理功能，让老师通过平台实时上传数据集的文件，数据集可以被平台的用户浏览和下载。教师可以方便建立实践任务，任务跟数据集有密切关联，且实践该任务的分组会加入到任务追踪去，让教师实时获取到学生当前的数据挖掘状况。最后，学生可以在平台进行数据的简单分析，如提取数据的密度、挖掘的数据的特征，让学生能够明确采用哪一种数据挖掘算法，执行数据挖掘任务。

## 产品目标

打造一个具有高性能、可拓展的数据挖掘实践任务教学平台。



图 2.2‑1 系统功能模块图

## 产品功能

### 系统用例模型

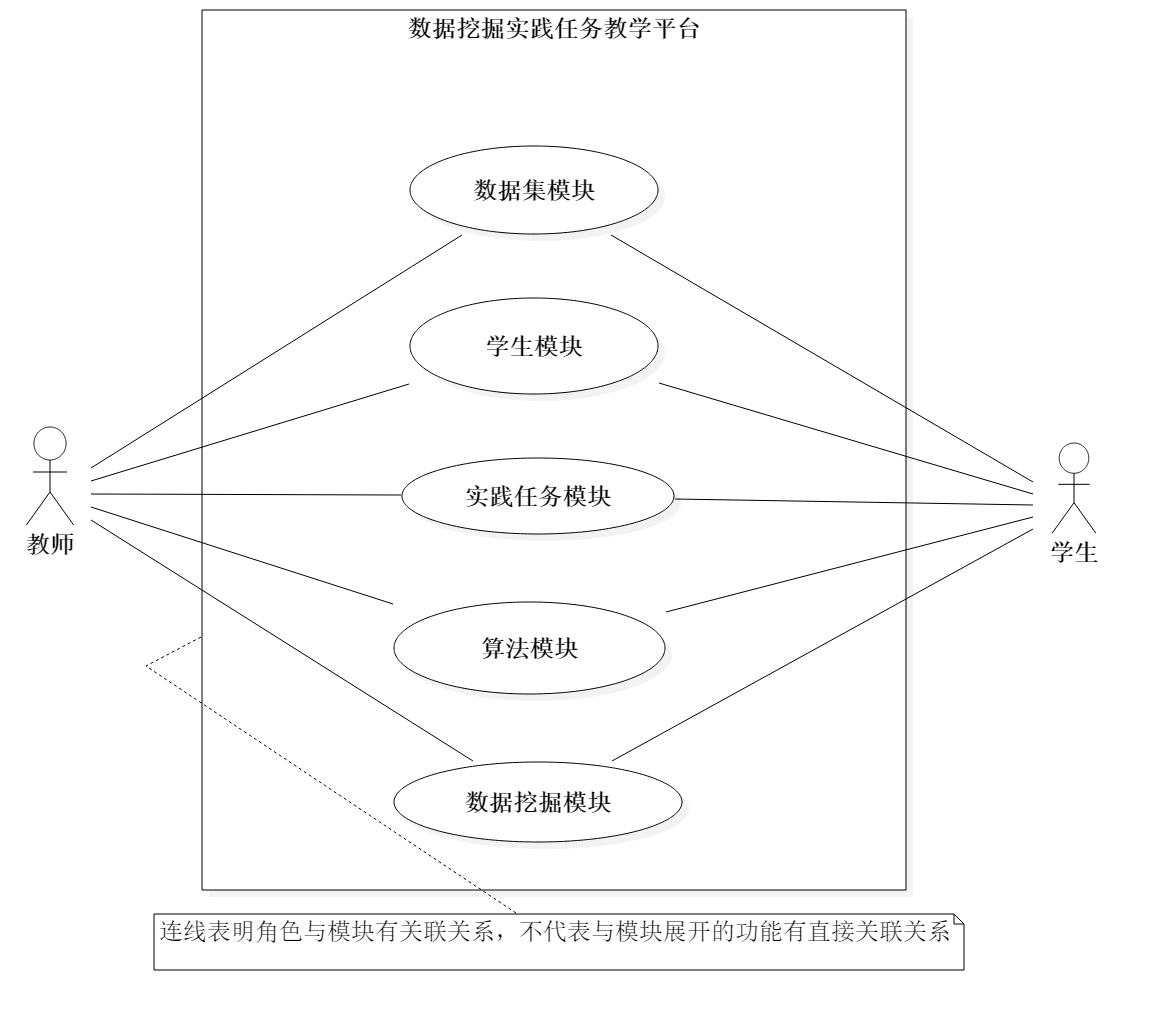


图 2.3‑1 系统用例

## 一般约束

#### 假定：用户能够提供交付测试的环境；用户能够参与到需求的核准工作；。

#### 约束：本系统的最后的交付日期是2018-9-30；

## 平台开发约束

服务器核心功能规定为Java语言，以及其衍生的框架、类库。其他次要或者或者的功能可接入Python、C#等语言开发的接口；

前端UI基于JavaScrpip、HTML开发，及其衍生的框架、类库技术。

### Web三层架构

图 2‑5 系统构架层次图

### 表现层

#### 模型（Model）：在MVC中模型保持着一个应用程序的状态，和相应视

#### 图中来自用户交互的状态变化。在上图中我们可以看到model会接受来之控制器的状态变化响应和视图的显示状态查询view渲染的数据来源。同时model还有通过事件机制（观察者模式）通知view状态的改变要求view渲染响应。view和模型之间存在一定的耦合，view必须了解model，这也是MVP模式出现原因之一。正在这里的模型model可以是来之分布式soap或者resetfult的DTO（数据传输对象），也可以直接是我们的领域对象（DO）或者数据层返回的数据集，并不严格的要求。

#### 控制器（Controller）：控制器由view触发，响应用户界面的交互，并根据表现层逻辑改变model状态，以及中间层的交互，最终修改model，controller不会关心视图的渲染，是通过修改model，model的事件通知机制，触发view的刷新渲染。控制器和模型的交互式一种“发出即忘”，或者分布式OneWay的调用。控制器Controller不会主动了了解view和view交互，除了唯一的视图选择外，控制器需要选择下一次显示的视图view是什么。一般控制器会请求全局应用程序路由下一个需要显示的view，在使其呈现出来。

#### 视图（View）：视图是表现层模式出现的原因，因为他的多样性和变化的

#### 频繁性，不易测试，所以理想的视图尽可能透明，被动，视图只负责渲染呈现给用户交互。视图由一系列GUI组件组成，响应用户行为触发控制器逻辑，修改model状态使其保持view同步。视图并需要相应model的变化被动的接受model状态变化刷新相应给用户。

本系统应使用以MVVM为设计理念的vue.js新型渐进式框架打造轻便、简洁、流畅的前台系统UI。表现层作为前端主要的内容，只负责页面的渲染与数据可视化加工。在如今RESTful微服务大行其道的大潮下，小组内负责表现层前端的人员只通过aixos封装的ajax向后台发送GET、PUT、DELETE、PATCH等异步请求，与后端进行数据交互。其中的API与后端约定好swagger生成文档，spring mvc会统一处理所有合法的http请求，以json或xml的数据格式返回、或响应response。

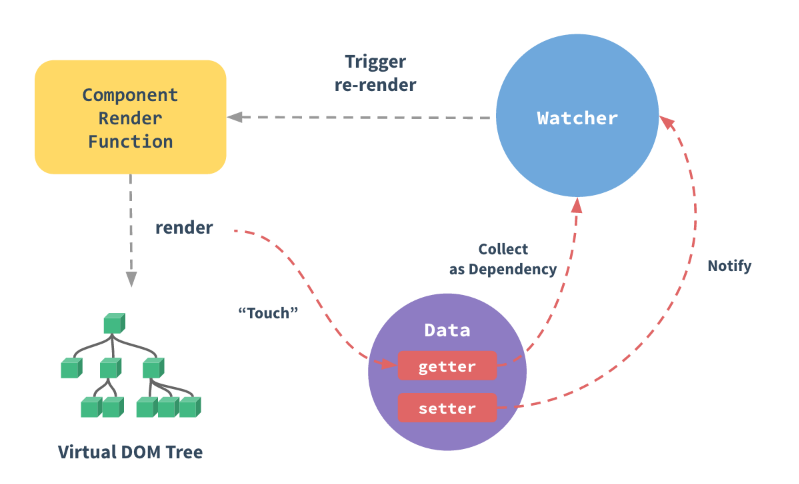


图 2‑6 vue.js框架渲染UI流程图

### 业务层

系统业务层的实现主要体现在各种组合的JavaBean。为体现系统的高内聚与低耦合特性，系统采用spring container管理所有系统的bean,将bean的生命周期托付给spring。在系统启动后，所有组件都被构造生成。利用spring的依赖注入和控制反转的先进技术，诸如DataSetService、StudentService、DataMiningTaskService等承载着复杂业务逻辑的服务性组件会被注入Controllers中去。同时业务层开发亦约定了一系列开发原则。



图 2.5‑1 MVC模式逻辑控制图

#### (1)面向接口编程：对于所有service的开发，都有之对应接口声明规范，每个方法体应该提供对应的joc文档注释；

#### (2)限定数据库事务边界：对于所有与业务逻辑的代码应该定义适当的事务边界，事务的管理应交付给spring 与对应ORM工具匹配的Transation Manager管理。一旦在处理业务逻辑中出现异常，要进行事务回滚。

#### (3)注解优先：所有业务代码压缩xml的使用比率，统一使用新型简洁的java annotation进行编码；如在后台中所有框架（诸如shiro、spring mvc、hibernate）的整合都应摒弃传统的xml的繁杂分离配置，使用java config全注解代码的优化系统的启动与维护；

#### 全局异常捕捉，返回统一状态：在service层以下发生的异常应该在controllers处进行异常捕捉，每个异常提供专一的处理方法，并在response中纳入对应的错误码或状态码，方便前端根据状态码进行页面逻辑的跳转与渲染。

### 数据访问层

数据访问利用全自动的hibernate对象——关系映射框架，构造整个系统后台的数据库支持，舍弃传统的jdbc繁杂的模板访问代码。

#### 数据库选用：利用开源免费MySQL，主要考虑其强大的事务支持特性，为数据挖掘准备良好的支撑；后期或可加入Redis键——值形式的数据库模型，处理相应的图数据；

#### 数据库连接池：选用c3p0连接池

#### 数据访问：对于DAO层的代码，一主要利用hql面向对象的语句执行CRUD等操作，二看齐JAP统一类型安全的查询规范，少用Criteria已被deprecated的非类型安全的Hibernate API。

#### 性能优化：在重复利用较多数据的上考虑使用二级缓存；对于批量插入数据量比较大的数据库访问操作，考虑绕开Hibernate API，直接使用原生的jdbc语句；建立必要的主键约束跟表索引；

#### 级联控制：用代码进行控制对象实体的关联与级联关系，取代数据库底层的数据表的级联；

# 功能性需求

## 学生端

### 登陆与注册模块

#### 注册：学生可以通过页面注册自己的账号及密码，也可以使用教师统一分配的账号和密码。

#### 登录：输入账号密码统一进入到系统首页，首页展示用户近期的活动情况，系统根据账号动态生成页面。



图 3.1‑1 登录客户端流程图

### 数据集模块

#### 浏览数据集库：可以查看所展示的所有数据集的种类、数量、主题等概括性基本信息，支持根据不同的信息进行分类查看;

#### 浏览数据集：浏览数据集基本描述信息及具体内容，基本的描述信息包括：数据集的名称、年份、实例数、所属的类型（分类、聚类）等。数据集的具体内容可以查询，并且能够按照不同的条目（如年份、类型）进行分类排序浏览;

#### 参与者：学生；

### 学生浏览算法库

#### 算法浏览：展示系统内数据挖掘算法。支持分类、聚类算法分类浏览；

#### 算法呈现：对具法算法进行阅读。包括算法参数、类型、接口的综合描述；

### 任务模块

#### 学生查看系统分配给自己的任务；并且通过任务浏览详情窗口看到当前任务的里程碑进度，历史上传过的数据挖掘结果，随时更新当前的结果。更新的结果会在状态上进行反映

#### 对阶段性的任务进行提交、总结；

### 数据挖掘模块

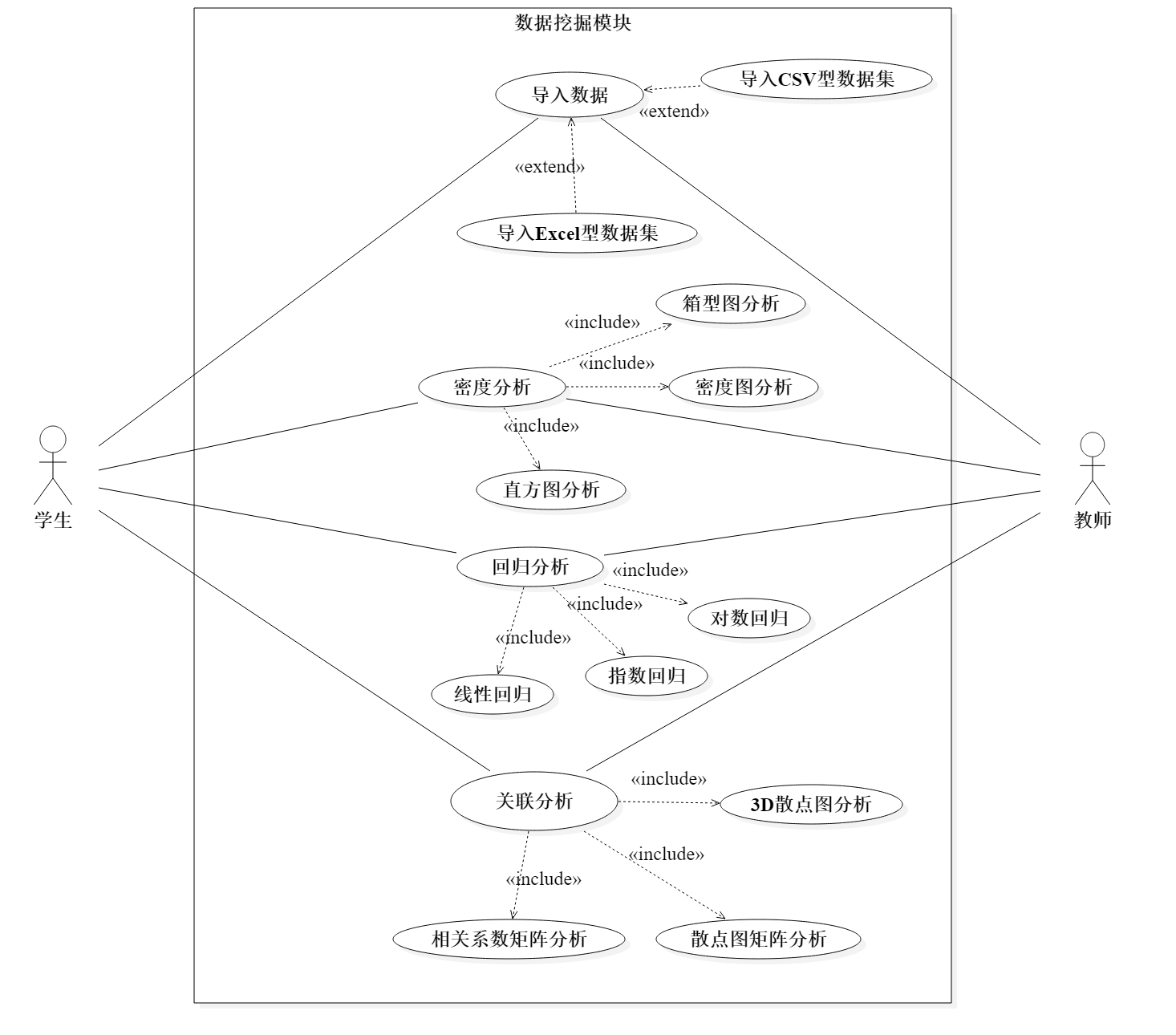


图 3.1‑2 数据挖掘模块用例图

#### 对于比较规整的excel、csv格式数据集，平台提供可视化分析手段；

#### 密度分析：平台提供直方图、箱线图、密度图对数据密度进行分析；

#### 回归分析：提供线性回归、指数回归、对数回归方程对用户选择的一对一属性进行回归分析，分析数据的相关性；

#### 关联分析：用相关系数矩阵直观方向数值型数据的相关性；用散点图矩阵描绘数据集中的一个属性跟其他多个属性的关联性；用3D散点图展示数据在三维空间的特征，方便用户分析数据在三维空间上面的分布状况；

## 教师端

### 数据集管理

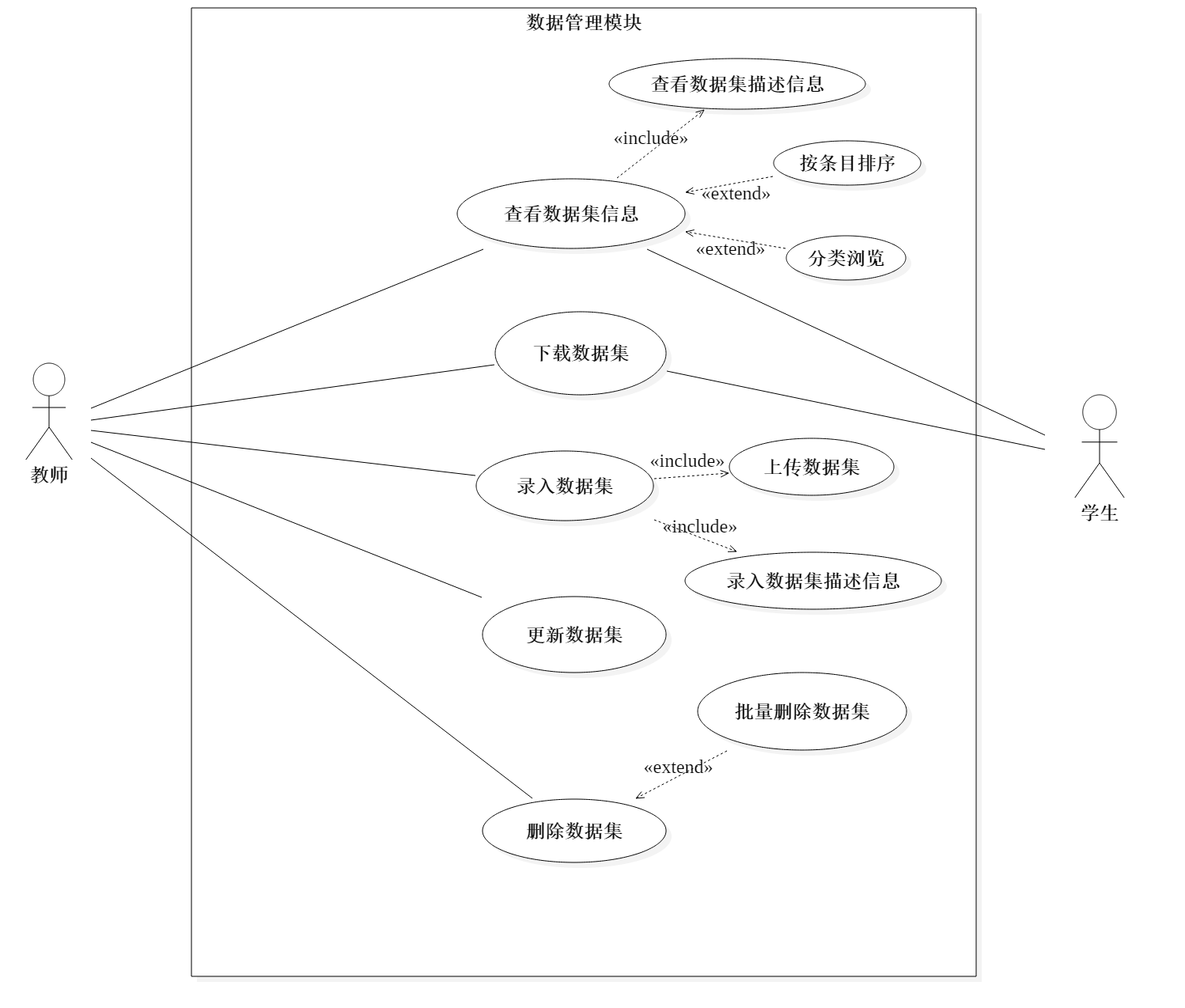


图 3.2‑1 数据集管理用例

#### 上传数据集：上传数据集的文件，以文件或二进制流的方式保存在服务器或mysql数据库中去；

#### 下载数据集：提供数据集的下载功能；

#### 删除数据集：删除已上传的数据集。提示相应的删除提示：如已被关联的数据集删除警告、数据量大的评估警告；

#### 更新数据集：可更新数据集的基本信息、属性、批量删除部分实例；

#### 文件读入数据集信息：根据管理员上传的属性描述文件，读入数据集的基本信息；

#### 浏览数据集：可以查看所展示的所有数据集的概括性基本信息，支持根据不同的信息进行分类查看；



图 3‑2数据集管理流程图

### 学生信息管理

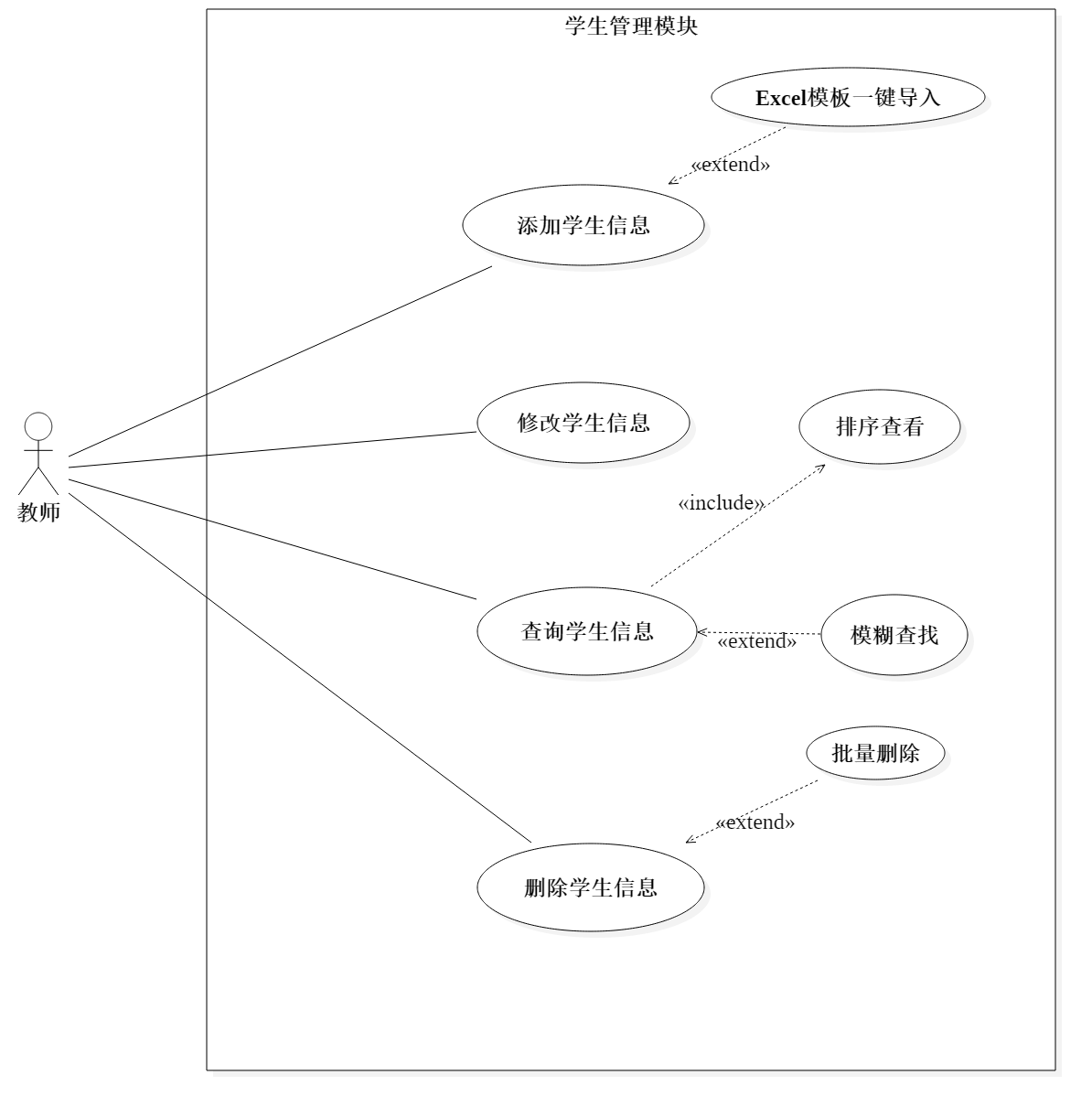


图 3.2‑2 学生管理模块用例图

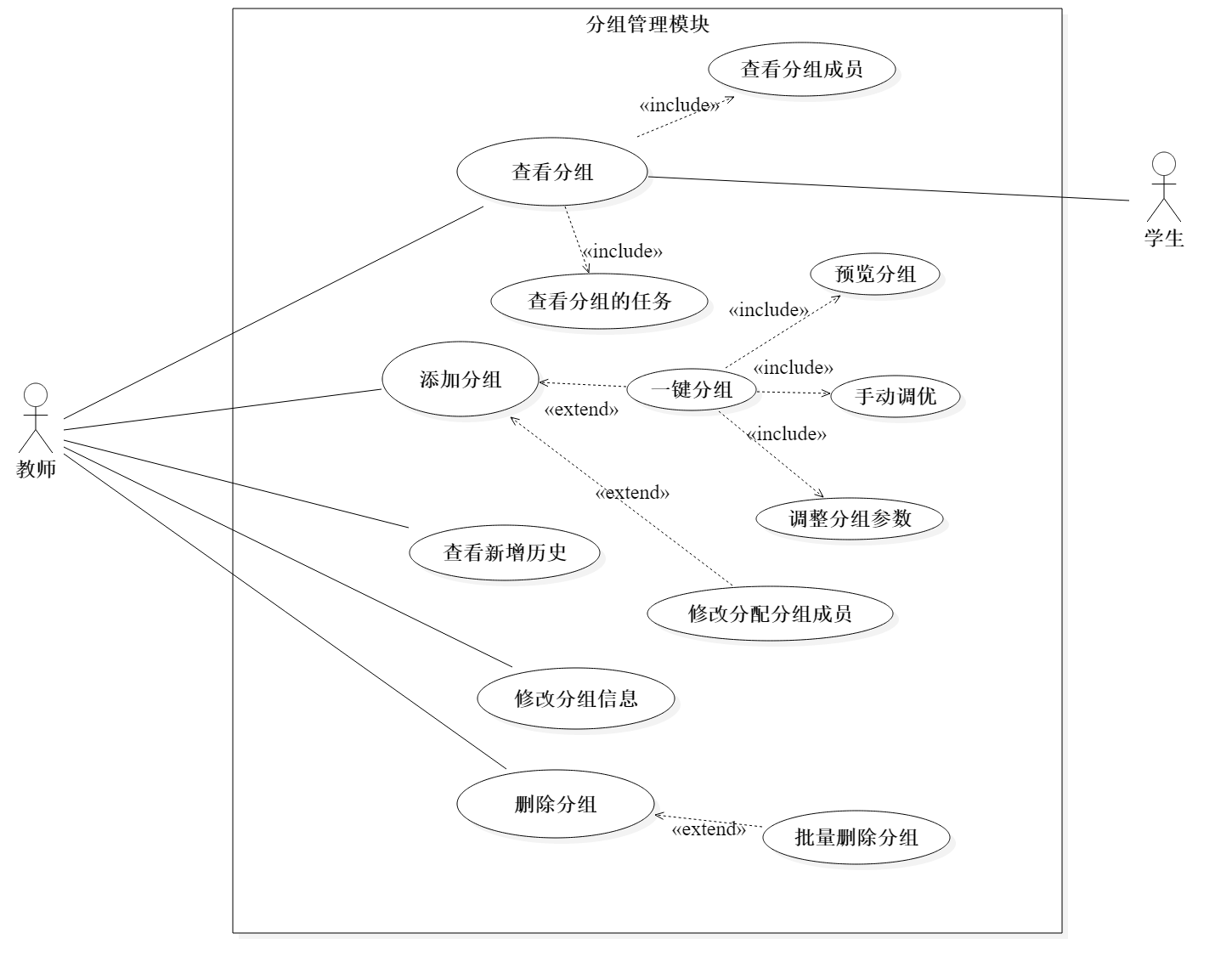
#### 管理学生信息：管理员对学生信息进行添加，具体内容包括学生的姓名、学号、班号、专业等。可以对学生信息进行增、删、改。

#### 一键导入：提供用户导入学生信息的模板；用户上传符合模板格式的学生信息后，这些信息会被自动录入系统，每个学生被分配默认登录的账号和密码，初始化为学生的学号。



图 3‑3学生信息管理流程图

### 分组信息管理



#### 查看分组：主要查看每个任务队伍的信息，主要包括队伍的队名、分组编号、组长、关联的任务名称、当前的任务状态等。

#### 管理分组：当前列表框支持用户进行增、删、改操作。点击操作上的按钮，会弹出相应的对话框。点击编辑按钮，可以更新分组的参数，以及关联的分组。

#### 一键分组：提供几种分组策略，勾选多个任务，批量分配给分组。分组的学生可以由用户筛选提供；每次分组完毕，提供预览视图让用户确认当前分组是否合适；允许用户对分组间的人员进行调整。

#### 新增历史：查看当前新增的分组信息。

### 算法库管理

#### 上传算法：上传数据挖掘算法的文件，以文件或二进制流的方式保存在服务器或mysql数据库中去；

#### 删除算法：删除已上传的算法。提示相应的删除提示：如已被关联的数据集删除警告、数据量大的评估警告。

#### 更新算法：可更新算法的描述、参数、结果示例；



图 3.2‑3算法库管理流程图

### 实践任务管理

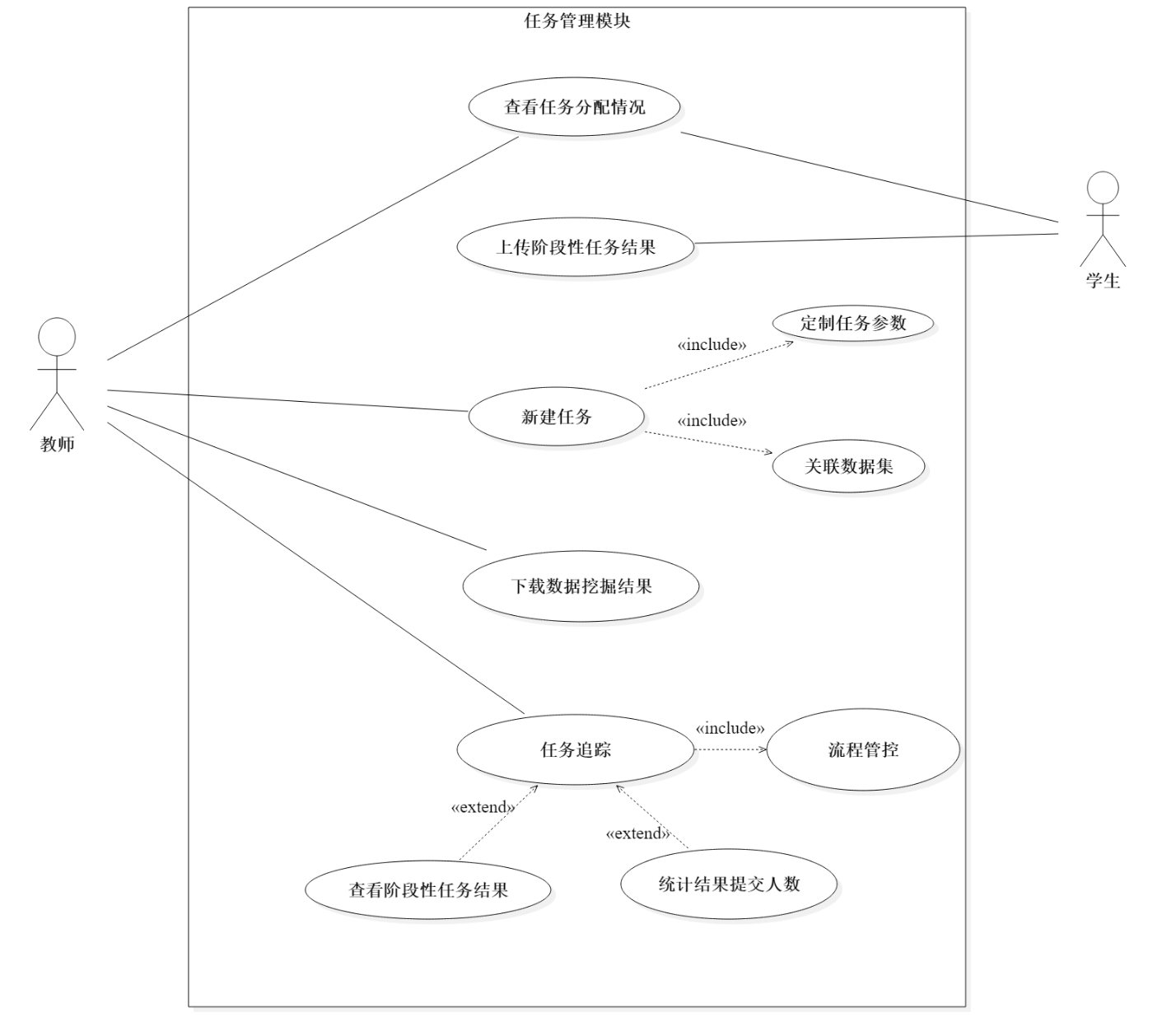


图 3.2‑4 任务管理用例

#### 任务管理：浏览库内基本的任务概要，如任务名称、任务编号、语言要求、任务里程碑、关联的分组、任务的状态等信息；提供浏览、修改操作。

#### 任务数据集指定：合理默认为几个任务指定数据集库里的数据集。

#### 任务学生分组指定：合理为默认任务指定几组不同的学生，每组学生每次只能分配一组任务，一个任务可以由多组学生参与。

#### 任务算法配置：为不同类型任务配置适量的算法。

#### 新增任务：新增数据挖掘任务。

#### 修改任务：对任务所分配的数据集、学生、算法进行添加、删除、修改。

#### 查看详情：查看特定任务下的分组提交数据挖掘结果的情况；统计已交、未交、新提交的结果数目；查看任务关联的数据集信息、算法信息、分组信息。

#### 跟踪进度：记录学生执行数据挖掘任务的进度，呈现学生任务的阶段；查看任务的里程碑情况，以及当前里程碑学生的数据挖掘结果的提交状态；

#### 收集挖掘结果：一键收集学生的上传记录，下载学生上传的挖掘文件。



图 3.2‑5实践任务管理流程图

# 非功能性需求

## 接口需求

系统建设采用先进的成熟技术，建立严密、体系化的系统管理、应用平台，应具有良好的分层设计，整体系统扩充性能良好，能够根据业务的发展或变更，在保持现有业务处理不受影响的前提下，具有持续扩充功能、适度变化的能力。

系统应提供符合Web 2.0规范语义的接口，通过REST可以方便的与客户现用客户端进行交互，交换的信息格式采用规范的JSON式，可以很方便地与其他系统进行信息交换，以满足信息化不断发展和系统集成需要。更严格的接口分析定义如下:

### 用户接口：

提供用户使用软件产品时的接口需求。例如，如果系统的用户通过显示终端进行操作，就必须指定如下要求：

#### 对屏幕分辨率要求：1080P及其以上；

#### 报表或菜单的页面打印格式和内容：A4；

#### 输入输出的相对时间：误差2分钟；

#### 程序功能键要求：返回、关闭、返回桌面。

### 软件接口：

提供可以跟本校教务系统对接的软件接口，引用校方开放的功能服务。

### 通信接口

数据源认识的请求类型映射为HTTP URL。服务器必须是一个web app，用以分析客户端请求的URL字符串并翻译为数据源认识的操作。约定HTTP通信头为PUT、DELETE、PUT、GET、PATCH，通过JSON数据流约束通信交互动作，以严格状态码（如404、421）区分通信异常，前段响应与后端响应分离，后端对通信操作自处理。

## 硬件需求

#### 显示器尺寸：笔记本15.4寸，台式显示屏27寸；

#### 处理器：四核及以上；

#### 内存：4G及以上；

#### 硬盘容量：100G及其以上；

## 数值需求

### 静态数值需求：

#### 支持的终端数：150

#### 支持并行操作的用户数：200

#### 系统响应的时间特性：用户对系统请求的反应时间控制在5秒内，如果超过5秒考虑设计交互友好的方式显示等待信息

### 动态数值需求：

#### 欲处理的事务和任务的数量：100事务并行，超越阈值回滚或阻塞。

#### 正常情况下一定时间周期中处理的数据速率：200MB/s。

#### 峰值工作条件下一定时间周期中处理的数据总量：1GB。

## 性能需求

#### 网络环境下的多用户系统：信息主要存储在服务器端的数据库中，由各使用者在规定的权限下在各自的客户端上录入，修改，删除相关的内容，进行各自的相关操作，不可跨权限经常操作。各用户还可进行查询，调用，达到信息共享。

#### 数据的完整性与准确性保护：录入数据采用表格方式，对录入的数据进行相关的限制，限制录入数据类型及取值范围以保证数据的完整性及准确性。

#### 服务器响应：服务器一般响应时间（除报表统计、数据导入）不超过2秒。

|  |  |  |
| --- | --- | --- |
| 时间段 | 种类 | 响应时间(秒) |
| 平时 | 新增教务数据 | 2 |
| 查询高峰 | 4 |
| 平时 | 简单查询 | 2 |
| 复杂查询 | 10 |
| 查询高峰 | 简单查询 | 8 |
| 复杂查询 | 20 |

图 4.4‑1 数据库不同时段查询时间需求

## 服务需求

#### 服务的安全性：保证教务管理系统中的数据进行相应的安全限制，管理员与用户的操作必须经过的严格安全体系校验，不能随便进行更改，保密性要高，以防发生随意泄露事件发生。

#### 服务的灵活性：调用相关数据操作要简单，从数据中调出来的数据要方便打印，录入，修改，删除等相关的操作。软件允许用户使用屏幕触摸使用，也可以对手机接入的鼠标、键盘做出反应。软件的设计和实现需要考虑到运行环境的变化，并能够在运行环境变化的情况下正常使用。同时，软件需要兼容其他软件接口的变化，以保证在不同运行环境，不同软件接口的情况下的正常使用。

## 可用性需求

#### 方便操作，操作流程合理：尽量从用户角度出发，以方便使用本产品。如：键入公告信息时，敲入回车键光标的自动跳转、输入法的自动转换，信息检索时输入汉语简拼快速检索到结果等。可以通过快速键方便用户录入信息，所有操作可仅通过键盘完成。

#### 系统合理：支持没有计算机使用经验、计算机使用经验较少及有较多计算机使用经验的用户均能方便地使用本系统。

#### 控制必录入项：本系统能够对必须录入的项目进行控制，使用户能够确保信息录入的完整。同时对必录入项进行有效的统一的提示。

#### 容错能力：系统具有一定的容错和抗干扰能力，在非硬件故障或非通讯故障时，系统能够保证正常运行，并有足够的提示信息帮助用户有效正确地完成任务。

#### 统一规范的提示信息：例如删除操作时，系统可提示警示框“您确认删除记录吗？操作不可恢复！”，用户点击确认后，系统才执行删除操作，删除后可直接返回相关页面。