

## 依赖项

### 前端

- jQuery-min
- Echarts-min
- node.js 18.13.0

### 后端

- Java8, Eclipse (2020-03)
- Jackson, 提供了便捷的 Java 容器转换为 json 的通道

### 服务器

- Tomcat9.0, classpath 需要添加 jxl-2.6.jar (读取 xls 文件) 与 mysql-connector-java-5.1.47.jar (连接 MySQL 数据库)

### 数据库

- MySQL 5.7.32, 使用前需要手动在 MySQL 中创建一个名为 JavaFinal 的 Database

## 数据库设计

### ER 图

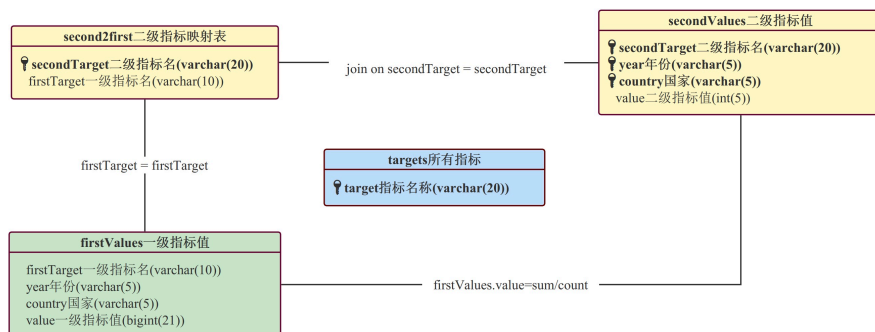


Figure 1: 数据库

## 数据库结构

数据库中共有三张表：

- second2first，存储二级指标与一级指标的对应关系，二级指标作为主键
- secondValues，以二级指标名、年份、国家作为索引存储对应的二级指标值
- targets，存储所有指标的名称（包括一级和二级指标），指标名作为主键要求指标名不能重复
  - 增加此表是因为在数据库管理操作中，“添加一级指标”会导致一级指标暂时没有对应的二级指标，无法写入 second2first；“删除一级指标”也会导致二级指标独立于一级指标存在。

数据库中还有一个视图 firstValues，存储一级指标名、年份、国家和对应的一级指标值。由于二级指标值可以做增删改的操作，一级指标值又是根据其包含的二级指标值计算得到，需要动态更新，故设计其为视图。

## 输入文件结构

### 论文数据

本系统只接收“.xls”格式的文件，支持将数据分“sheet”存储。文件内部结构应保持与“信息化高被引论文/数据 1.txt”字段顺序一致。

为了指标值的计算结果准确，“PY”字段应以“yyyy”格式记录论文发表年份，“RP”字段相同通讯地址作者姓名间以“；”间隔，作者姓名与通讯地址间应有“(通讯作者)”，通讯地址应包含国家全称 {"China","USA","England","France","Russia"}。

### 专利数据

本系统只接收“.xls”格式的文件，支持将数据分“sheet”存储。文件内部结构应保持与专利数据“Demo.xls”字段顺序一致。

为了指标值的计算结果准确，“申请日”和“授权日”两个字段缺失值应用“-”填充，不缺失则应以“yyyy-mm-dd”格式记录日期；“发明人地址”字段应包含国家简称 {"CN","US","GB","FR","RU"}。

## 功能实现说明



Figure 2: 系统

### 文件上传

由于数据的格式是 Excel, 而在本项目处理过程中使用 MySQL, 需要提供让用户把 Excel 表格转化为 SQL 语句的功能。

前端用 html 表单获取上传文件, 后端用 `javax.servlet.http.Part` 处理前端上传的文件, 放到指定路径, 处理后生成 SQL 语句在 localhost 上建立数据库。

注意使用 `javax.servlet.http.Part` 时需要在 web.xml 里指定 multipart-config 属性, 提供“最大传输文件大小”、“保存相对路径”等信息, 此处因为只是演示所以并未选定具体的信息, 但仍需要在 `<servlet>` 属性内添加 `<multipart-config/>`。

### 计算指标

构建 4 个数组 `int[3][5]` 存储 4 个二级指标的值。其中, 3 代表各个年份 (2019-2021 年份具有天然的顺序); 5 代表各个国家 (构建了 `HashMap nation2id` 保证顺序固定为为中、美、英、法、俄)。

对于研究情况, 遍历论文数据文件中的每一个 sheet。由于要求中定义顶级学者当年发文至少 2 篇, 故以年 year 为单位进行统计, 对每一年构建两个 `HashMap name2nation`、`name2num` 分别存储学者名与国家的映射一级学者名与发文的映射。按行遍历获取论文发表年份, 属于当前统计年份则继续获取通讯作者信息, 提取出第一位通讯作者的姓名, 若 `name2nation` 尚无该作者国家信息, 则继续提取通讯地址中包含的国家 (可能提取国家失败) 写入 `name2nation`; 若在 `name2num` 中尚无该作者信息则添加 (`key=name, num=0`), 然后找到该作者当前值增加 1。最后, 遍历 `name2num` 中的作者姓名 name, 若可以在 `name2nation` 中找到该作者姓名对应的国家 nation, 则将 `D112[year][nation]` 当前值增加该作

者当年发文数 num; 对于发文量大于等于 2 的作者, 将 D111[year][nation] 当前值增加 1。

对于应用情况, 遍历专利数据文件中的每一个 sheet, 内部按行遍历处理。首先提取地址信息, 接着将非空的申请时间提取前 4 位处理成年份, 循环判断地址中是否包括特定国家。若包括, 则将 D121[year][nation] 当前值增加 1。授权时间同理。

最后通过静态方法将 4 个数组按要求归一化。注意, 由于要求保留两位小数, 此处为了方便处理, 实际归一化到 0~10000 保留精度, 前端展示时再除以 100。

## 构建数据库

连接到 database JavaFinal, 按照数据库结构创建 targets、second2first、secondValues 三张表, 并向 targets 和 second2first 写入已有信息。遍历数组 D111、D112、D121、D122 得到 (secondTarget, year, country, value) 写入 secondValues 表。按照一级指标得分计算规则创建视图 firstValues。

## json 生成

前端的可视化工具 Echarts 生成统计图的数据源是 json 数组, 为了方便在 java 环境中操作各种不同格式的 json (一共可以生成 7 种不同的统计图, 最多需要 7 种形态不同的 json), 在后端 ResponseGenerator.java 中定义了 ResponseGeneratorBase 抽象类, 以及 SingleValueGenerator, MultiValueGenerator 两个子类, 可以用于生成形态各异的 json 数组。

两个子类的区别是分别使用了一维和二维的 ArrayList<Integer> 作为容器。

返回的 json 数组内容如下:

- 如果正常返回, 则形如 {"result": "success", "scores": []} 其中 scores 是一维或二维的 json 数组, 表示前端需要的值
- 如果前端需要图标中 X 轴上的标签, 则形如 {"result": "success", "scores": [...], "names": [...]}, names 是一维的 json 数组, 表示标签
- 如果前端传入了错误的请求, 则不会返回 json 数组, 并设置 ResponseGeneratorBase 基类的 errMsg 成员为报错信息

使用外部包 jackson 提供的 ObjectMapper 类可以将 java 容器方便地转换为 json。

## FeedSelect

注意到在前端可视化界面存在一个下拉框, 内容是从数据库中查询到的所有指标的指标名, 此功能的实现需要在后端执行一次 SQL 查询, FeedSelect.java 即在网页加载之后执行这次查询。

## 数据库管理

Web 系统可以实现指标的增加、删除、修改，包括指标本身和指标数值的增删改操作。

- 添加一级指标
  - 输入：待添加的一级指标名
  - 操作：判断不为空且与 targets 表中现有指标名不重复后，向 targets 表中添加一级指标名
- 添加二级指标
  - 输入：待添加的二级指标名、该二级指标所属一级指标名
  - 操作：判断一级指标已存在且二级指标名与 targets 表中现有指标名不重复后，向 targets 表中添加二级指标名，向 second2first 表中添加从属关系
- 删除指标
  - 输入：待删除的指标名
  - 操作：判断待删除的指标名存在后，将三张表中的指标名等于待删除的指标名的行删除
- 修改指标名称
  - 输入：修改前指标名，修改后指标名
  - 操作：判断不修改前指标名不等于修改后指标名、修改前指标名存在且修改后指标名不存在后，将三张表中的指标名等于修改前指标名的都修改为修改后指标名
- 添加二级指标值
  - 输入：待添加值的二级指标名、年份、国家，待添加的值
  - 操作：判断值非负、二级指标存在且（指标名，年份，国家）与 secondValues 表中现有索引不重复后，向 secondValues 表中添加二级指标值
- 删除二级指标值
  - 输入：待删除值的二级指标名、年份、国家
  - 操作：判断 secondValues 中存在要删除的索引（指标名，年份，国家）后，删除这条索引对应数据
- 修改二级指标值
  - 输入：待修改值的二级指标名、年份、国家，修改后的值
  - 操作：判断值非负且 secondValues 中存在要修改的值的索引（指标名，年份，国家）后，修改这条索引对应数据的值

## Apache Echarts

ECharts，一个使用 JavaScript 实现的开源可视化库，可以流畅的运行在 PC 和移动设备上，兼容当前绝大部分浏览器（IE9/10/11，Chrome，Firefox，Safari 等），底层依赖矢量图形库 ZRender，提供直观，交互丰富，可高度个性化定制的数据可视化图表。支持传入多种数据格式，在前端展示各种可视化图表。

在引入 echarts-min.js 后可以在 jsp 中生成 Echarts 图表，用户选定可视化内容后点击提交就可以看到所选项目的可视化结果。实现上使用 ajax 从后台取出每次渲染所需的 json，把前一次生成的图表清除后生成新的图表，实现了不同样式（条形图/折线图），不同下标格式（下标可以为年份/国家/指标名），不同数据量（单系列统

计图/多系列统计图) 的可视化。

### 前端美化: **ajax** 和 **css**

ajax 可以实现单个网页内容的局部刷新, 在 jsp 与后台的交互中本项目大量使用了 ajax, 让页面结构更简洁, 项目内容更清晰。

使用简单的 css 美化了前端。

## 项目总结

本项目完成了作业的全部要求 (1-5)。

代码层面, 使用了本学期课程强调过的重要概念, 抽象类、静态内部类、方法的重写、Java Collection、异常处理、文件输入输出流、数据库连接和 Web Servlet 开发, 起到了很好的巩固效果。

合理安排复杂的项目结构, 第一次面对跨语言和高代码量项目, 为我增加了宝贵经验。