

基于xml等技术实现异构数据的集成

方案二：基于多个电影信息平台的电影信息整合



高毓彬 161250028

陈迪新 161250007

葛宇 161250029

潘羽 161250096

于瑞召 161250185

2019-5-19

目录

[需求 1](#_Toc9171060)

[框架 1](#_Toc9171061)

[实现 2](#_Toc9171062)

[效果 3](#_Toc9171063)

## 需求

1. 对外接口需求
2. 用户界面

展示电影信息

1. 豆瓣电影平台接口

|  |  |
| --- | --- |
| 接口地址 | 描述 |
| https://douban.uieee.com/ | 获取豆瓣电影平台的信息 |
| https://douban.uieee.com/v2/movie/top250 | 获取豆瓣电影top250的信息 |

1. 聚合数据平台接口

|  |  |
| --- | --- |
| 接口地址 | 描述 |
| http://v.juhe.cn/movie/cinemas.search | 按关键字检索影院（省略请求参数描述） |
| http://v.juhe.cn/movie/cinemas.movies | 获取影院上映影片信息 |
| http://v.juhe.cn/movie/citys | 获取支持城市的列表 |
| http://v.juhe.cn/movie/movies.today | 获取当天电影的信息 |

1. 功能需求
2. 基于互联网应用的数据集成：

选择多个电影信息平台整合电影信息、票价等，为用户提供全局视野数据。

1. 查看整合后的电影信息：

选择豆瓣电影和聚合数据两个平台，通过平台提供的API，分别获取这两个平台的电影信息，进行整合和数据集成，并统一展示。

1. 其他非功能需求
2. 约束

已有的互联网应用有数据异构的特征。

1. 数据需求

尽量使用XML作为元数据管理方案。

## 框架

采用SpringBoot框架，简化编码、配置、部署和监控过程，自动配置Spring。

Python数据服务不使用框架

## 实现

1. 业务逻辑层
2. DoubanService
3. JuheService

提供的服务

|  |  |  |
| --- | --- | --- |
| JuheService.getTodayMovie | 语法 | public List<Movie> getTodayMovie() |
| 前置条件 | 无 |
| 后置条件 | 返回当天电影信息的列表 |
| JuheService.getCitycinemas | 语法 | public List<Cinema>  getCitycinemas(String cityname) |
| 前置条件 | 无 |
| 后置条件 | 返回检索城市的影院及电影信息的列表 |

1. MovieService

提供的服务

|  |  |  |
| --- | --- | --- |
| MovieService.getMovie | 语法 | public List<Movie> getMovie() |
|  | 前置条件 | 无 |
|  | 后置条件 | 返回已整合的电影信息的列表 |

1. 实体和数据访问层
2. Cinema

|  |  |  |
| --- | --- | --- |
| 字段名 | 类型 | 含义 |
| Id | int | 影院编号Id |
| cityName | String | 影院所在城市名 |
| Name | String | 影院名称 |
| Address | String | 影院地址 |
| telephone | String | 影院联系方式 |
| Movies | List<Movie> | 电影信息 |

1. Movie

|  |  |  |
| --- | --- | --- |
| Id | int | 电影编号Id |
| Name | String | 电影名 |
| Rating | double | 评分 |
| description | List<String> | 电影描述 |
| cinemaAddress | String | 上映影院的地址 |
| Price | double | 票价 |
| Remain | double | 余票数量 |
| pic\_url | String | 图片地址 |

1. 工具类
2. FileUtil

|  |  |  |
| --- | --- | --- |
| 提供的服务 | 语法 | 实现功能 |
| FileUtil.getDoubanStr | public static String getDoubanStr() | 读取json格式的豆瓣电影数据 |
| FileUtil.main | public static void main(String[] args) | 运行  FileUtil.getDoubanStr方法 |

1. Json2ObjUtil

|  |  |  |
| --- | --- | --- |
| 提供的服务 | 语法 | 实现功能 |
| Json2ObjUtil.listMovieFromDouban | public static List<Movie> listMovieFromDouban() | 将json格式的豆瓣电影数据转换为movie对象的列表 |

1. PythonRunnerUtil

|  |  |  |
| --- | --- | --- |
| 提供的服务 | 语法 | 实现功能 |
| PythonRunnerUtil.run | public static void run(String pythonFilePath, String[] params) | 运行Python脚本 |
| PythonRunnerUtil.main | public static void main(String[] args) | 启动Python的运行，获取豆瓣电影平台的所需信息 |

1. XMLUtil

|  |  |  |
| --- | --- | --- |
| 提供的服务 | 语法 | 实现功能 |
| XMLUtil.getMovieXML | public static String getMovieXML(String movieStr) | 获得XML格式的电影信息 |
| XMLUtil.formatXMLFromDouban | public static String formatXMLFromDouban(List<String> movieList) | 将豆瓣电影平台获得的数据转化为XML格式 |
| XMLUtil.createXml | public static Document createXml() | 获取一个XMLDocument对象 |
| XMLUtil.addElement | public static void addElement(Document document,String name,Map<String,Object> property) | 给Document添加节点 |
| XMLUtil.writeXml | public static void writeXml(Document document,String name) | 将XML写入本地 |
| XMLUtil.xml2json | public static String xml2json(String filePath) | xml转json字符串 |

4． Python提供的服务pyImpl

1. douban\_film.py

|  |  |  |
| --- | --- | --- |
| 提供的服务 | 语法 | 实现功能 |
| douban\_film.get\_top\_film | def get\_top\_film() | 获取top250的豆瓣电影数据 |

2）img\_thief.py

|  |  |  |
| --- | --- | --- |
| 提供的服务 | 语法 | 实现功能 |
| img\_thief.get\_douban\_img | def get\_douban\_img(url) | 爬取对应url的豆瓣电影海报 |

2）json\_util.py

|  |  |  |
| --- | --- | --- |
| 提供的服务 | 语法 | 实现功能 |
| json\_util.write\_file | def write\_file (file\_name, data) | 写文件 |
| json\_util.load\_obj(file\_path) | def load\_obj(file\_obj) | 将json转为dict并读入缓存 |

## 效果

根据后端接口的调用结果，前端展示整合后的电影信息，实现异构数据的集成与显示。