**豆瓣电影信息搜索**

目录:

一、需求概述

对电影信息提供搜索服务，可以分类搜索,

二、功能概述

从电影网站获取影片信息，提供分类查询

三、需求模块

1. 抓取模块
2. 存储模块
3. 索引模块
4. 查询模块

四、需求功能详细说明

1.抓取模块

任务:抓取电影信息

描述:需要抓取的内容有：29个分类下每个分类下按照排名前200个电影的评分分数，评分人数，剧情简介，导演，编剧，主演，类型，制片国家/地区，语言，上映日期，片长，又名，十条热门影评，豆瓣成员常用的标签

2.存储模块

任务:对抓取到的数据进行持久化

描述:保存到数据库中

数据库表

|  |  |  |  |
| --- | --- | --- | --- |
| MySQL数据库表结构 表名:t\_douban250 | | | |
| 字段名 | 数据类型 | 非空 | 说明 |
| **movieID(PK)** | int(11) | TRUE | 电影ID |
| **movieName** | varchar(20) | TRUE | 电影名 |
| **movieScore** | float(2,1) | FALSE | 电影分数 |
| **movieJudges** | int(11) | FALSE | 打分人数 |
| **movieScenario** | text | TRUE | 剧情简介 |
| **movieDirector** | varchar(50) | TRUE | 电影导演 |
| **movieScriptwriter** | varchar(50) | TRUE | 电影编剧 |
| **movieActor** | varchar(20) | FALSE | 主要演员 |
| **movieType** | varchar(20) | FALSE | 电影类型 |
| **movieLocation** | varchar(20) | FALSE | 电影国家 |
| **movieLanguage** | varchar(50) | FALSE | 电影语言 |
| **movieDate** | datatime | TRUE | 上映日期 |
| **movieTime** | smallint | TRUE | 电影时长 |
| **movieNickname** | varchar(50) | FALSE | 电影别名 |
| **movieTag** | varchar(50) | FALSE | 电影标签 |

|  |  |  |
| --- | --- | --- |
| 热门影评表 表名:t\_HotTalk | | |
|  | **数据类型** | **说明** |
| **Hid(PK)** | int(11) |  |
| **movieID** | int(11) |  |
| **title** | varchar(30) | 标题 |
| **content** | text | 正文 |

|  |  |  |
| --- | --- | --- |
| 电影类型表 表名:t\_movieStyle | | |
|  | **数据类型** | **说明** |
| **Sid(PK)** | int(11) |  |
| **movieID** | int(11) |  |
| **style** | smallint(4) | 电影类型编号 |

3.索引模块

任务:对内容建立索引进行排序

描述:上传数据到Elastic Search上

4.查询模块

任务:提供用户查询接口

1. 流程图

