第一次接触php 使用yaf+yar做一个电影推荐系统。

主要系统思路:

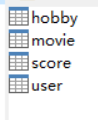
1. 分为两个服务：平台服务，推荐服务—中间使用yar rpc远程调用（微服务那味）
2. 登录使用token验证，redis存放token进行主动过期
3. 用户看完电影（假装你看过了，看电影的代码还没写）评分，然后传消息给kafka，消费端消费之后调用推荐平台进行电影推荐
4. 推荐算法：给你篇文章自己看吧 https://www.jianshu.com/p/27d1e8b84c64

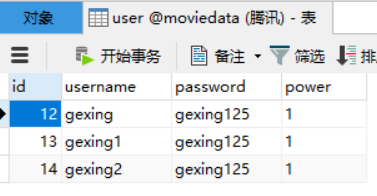
初步设计

Mysql设计

以下为系统平台需要的表

先以用户为主，管理员后台部分后期再加











Redis

存放token

使用List存放movie json

存放评分数量

建表完成--------------

爬虫找电影数据

代码：from urllib import request

from lxml import etree

from urllib.request import urlopen, Request

#构造函数，抓取第i页信息

def crow(i):

header={'User-Agent':'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.159 Safari/537.36'}

# 构造第i页的网址

url='https://movie.douban.com/top250?start='+str(25\*i)

# 发送请求，获得返回的html代码并保存在变量html中

ret = Request(url, headers=header)

res = urlopen(ret)

html=res.read().decode('utf-8')

#将返回的字符串格式的html代码转换成xpath能处理的对象

html=etree.HTML(html)

#先定位到li标签，datas是一个包含25个li标签的list，就是包含25部电影信息的list

datas = html.xpath('//ol[@class="grid\_view"]/li')

a=0

for data in datas:

data\_title=data.xpath('div/div[2]/div[@class="hd"]/a/span[1]/text()')

data\_info=data.xpath('div/div[2]/div[@class="bd"]/p[1]/text()')

data\_quote=data.xpath('div/div[2]/div[@class="bd"]/p[2]/span/text()')

data\_score=data.xpath('div/div[2]/div[@class="bd"]/div/span[@class="rating\_num"]/text()')

data\_num=data.xpath('div/div[2]/div[@class="bd"]/div/span[4]/text()')

data\_picurl=data.xpath('div/div[1]/a/img/@src')

print("No: "+str(i\*25+a+1))

print(data\_title)

#保存电影信息到txt文件，下载封面图片

with open('douban.txt','a',encoding='utf-8')as f:

#封面图片保存路径和文件名

f.write(str(i\*25+a+1)+',')

f.write(data\_title[0]+',')

f.write(data\_picurl[0]+',')

f.write(data\_score[0]+',')

f.write(str(data\_info[1]).strip() + '\n')

f.write('\n')

#下载封面图片到本地，路径为picname

#request.urlretrieve(data\_picurl[0],filename=picname)

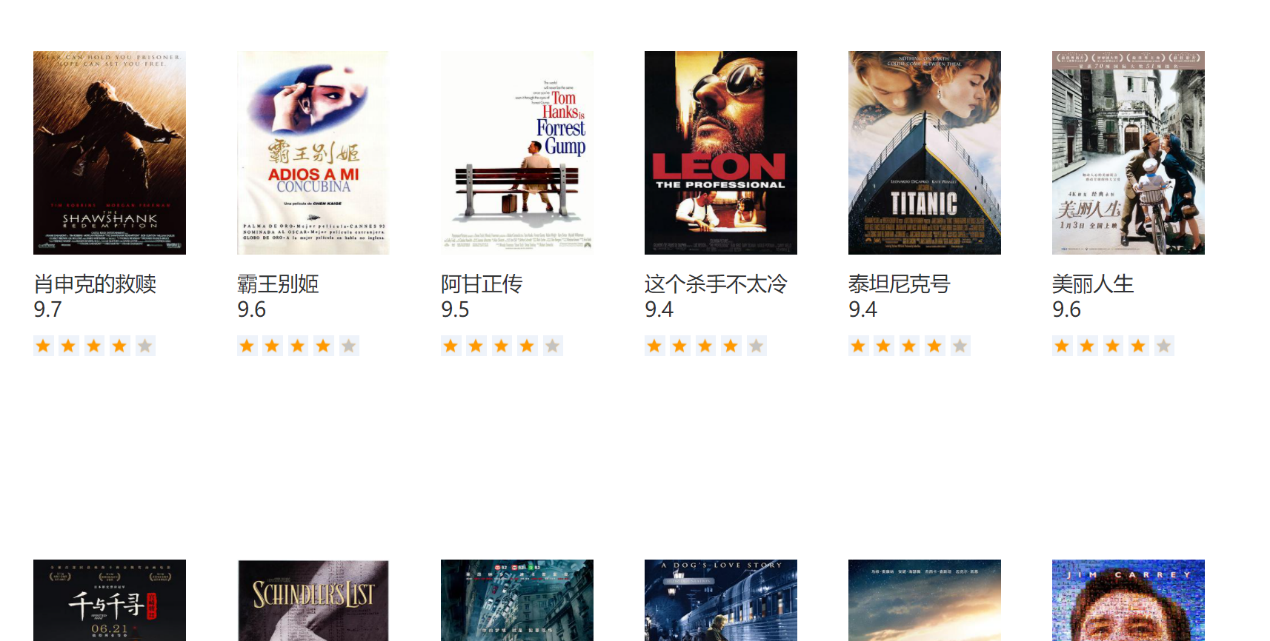
a+=1

for i in range(10):

crow(i)



到现在为止基本的评分平台已经搭建完成。比较简陋。。。前端使用的vue简单实现了一下。



买了一个阿里云的kafka服务做消息队列，使用yar将平台与推荐分析模块分离。

----------推荐模块是根据用户的协同过滤推荐，推荐5个电影



到此为止用时10天左右

下一步---加上es全文检索