

# “基于用户画像的电影推荐系统”软件使用说明书

## 一、软件编制人员

本软件由作者编写完成，包括系统的分析、设计、实现以及测试等，项目前后耗时一个多月。

## 二、运行环境

### 1. 硬件环境：

- (1) 操作系统：Windows 10 64 位 、CentOS
- (2) 处理器：Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz 2.71 GHz
- (3) 内存：8 GB
- (4) 主硬盘：SAMSUNG\_MZVLV256 ( 256 GB / 固态硬盘 )

### 2. 软件环境

#### (1) Windows 10 64位：

- Python 3.7
- MySql 5.6
- Redis 5.0.7
- 以及对应所需的Python库

#### (2) Linux：

- Java 1.8
- Spark 3.1.1
- Hadoop 2.7

## 三、软件安装方法

### 1. 使用前提

运行本系统的前提是需要将本系统运行的软件环境安装配置好。

#### (1) Windows软件安装：

- Python 3.7

安装包下载地址：<https://www.python.org/downloads/windows/>。

找到3.7版本的 Download Windows x86-64 executable installer 下载。

下载完后直接运行，可全部默认配置安装。

- MySql 5.6

安装包下载地址：<https://downloads.mysql.com/archives/get/p/25/file/mysql-installer-community-5.6.45.0.msi>。

下载完后直接运行安装。

安装完成后设置mysql的用户名和密码。

- Redis 5.0.7

安装包下载地址：<https://github.com/tporadowski/redis/releases>。

找到对应版本的.zip或.msi文件下载然后安装。

打开一个 cmd 窗口，使用 cd 命令切换到Redis目录运行：

redis-server.exe redis.windows.conf 启动Redis。

另起一个cmd窗口，切换到 redis 目录下运行：

redis-cli.exe -h 127.0.0.1 -p 6379 测试Redis是否连接成功。

- 安装运行本系统所需的python库

打开一个 cmd 窗口，切换到本系统程序的根目录下，运行：

```
pip install --index https://pypi.tuna.tsinghua.edu.cn/simple/ -r requirements.txt
```

即可安装本程序所有需要的库。

## (2) Linux软件安装:

- 安装JDK

安装包下载地址：<https://www.oracle.com/java/technologies/javase/javase-jdk8-downloads.html>

找到名为：jdk-8u291-linux-x64.tar.gz 点击下载

将文件传到Linux上，解压文件：

命令：tar -zxvf jdk-8u211-linux-x64.tar.gz (jdk-8..这个是刚下载的jdk名称)

配置环境变量：

命令：vim /etc/profile

在末尾添加如下几行代码：（JAVA\_HOME是刚才上传解压到Linux的目录）

```
export JAVA_HOME=/usr/java/jdk1.8.0_201-amd64
```

```
export PATH=$PATH:$JAVA_HOME/bin
```

刷新环境变量命令:source /etc/profile

查看是否安装成功：查询jdk版本命令：java -version

- 安装Hadoop 2.7

首先关闭系统防火墙：

```
systemctl stop firewalld
```

```
systemctl disable firewalld
```

再配置本机DNS解析：（\$ip为本机ip地址，可使用ifconfig查看）

```
echo "$ip master" >> /etc/hosts
```

再配置主机名：

```
echo "HOSTNAME=master" >> /etc/sysconfig/network
```

创建hadoop存放目录：

```
mkdir /root/Hadoop
```

去hadoop官网下载hadoop2的版本都可，然后上传到Linux里，并解压到刚刚差UN宫颈癌in的hadoop目录，将其重命名为hadoop。

```
mv /root/hadoop/hadoop-3.2.0 /root/hadoop/Hadoop
```

配置免密登陆：

```
(ssh-keygen -t rsa;cp ~/.ssh/id_rsa.pub ~/.ssh/authorized_keys;chmod 644
authorized_keys;sudo service sshd restart)
```

配置hadoop环境变量:

```
echo "" >> /etc/profile
echo "export HADOOP_HOME=/root/hadoop/hadoop" >> /etc/profile
echo "export HADOOP_MAPRED_HOME=\$HADOOP_HOME" >> /etc/profile
echo "export HADOOP_COMMON_HOME=\$HADOOP_HOME" >> /etc/profile
echo "export HADOOP_HDFS_HOME=\$HADOOP_HOME" >> /etc/profile
echo "export YARN_HOME=\$HADOOP_HOME" >> /etc/profile
echo "export HADOOP_COMMON_LIB_NATIVE_DIR=\$HADOOP_HOME/lib/native" >>
/etc/profile
echo "export PATH=\$PATH:\$HADOOP_HOME/sbin:\$HADOOP_HOME/bin" >>
/etc/profile
echo "export HADOOP_INSTALL=\$HADOOP_HOME" >> /etc/profile
echo "export HADOOP_OPTS=\"-
Djava.library.path=\$HADOOP_HOME/lib:\$HADOOP_COMMON_LIB_NATIVE_DIR\"" >>
/etc/profile
echo "" >> /etc/profile
source /etc/profile
```

修改hadoop配置:

```
sed -i "s/\${JAVA_HOME}/\/root\/hadoop\/jdk/g"
/root/hadoop/hadoop/etc/hadoop/hadoop-env.sh
```

```
sed -i "s/<\/configuration>/ /g" /root/hadoop/hadoop/etc/hadoop/core-
site.xml
```

```
echo -e
"<property><name>hadoop.tmp.dir</name><value>/root/hadoop/hadoop/tmp</value></
property>\n<property><name>fs.defaultFS</name><value>hdfs://master:9000</value
></property>
```

```
<property><name>hadoop.proxyuser.root.hosts</name><value>*</value></proper
ty><property><name>hadoop.proxyuser.root.groups</name><value>*</value></proper
ty>\n</configuration>" >> /root/hadoop/hadoop/etc/hadoop/core-site.xml
```

```
sed -i "s/<\/configuration>/ /g" /root/hadoop/hadoop/etc/hadoop/hdfs-
site.xml
```

```
echo -e "<property><name>dfs.replication</name><value>1</value></property>
<property><name>dfs.webhdfs.enabled</name><value>true</value></property>
<property><name>dfs.permissions</name><value>false</value></property>\n</c
onfiguration>" >> /root/hadoop/hadoop/etc/hadoop/hdfs-site.xml
```

```
cp /root/hadoop/hadoop/etc/hadoop/mapred-site.xml.template
/root/hadoop/hadoop/etc/hadoop/mapred-site.xml
```

```
sed -i "s/<\n</configuration>/ /g" /root/hadoop/hadoop/etc/hadoop/mapred-
site.xml
```

```
echo -e
"<property><name>mapreduce.framework.name</name><value>yarn</value></property>
\n</configuration>" >> /root/hadoop/hadoop/etc/hadoop/mapred-site.xml
```

```
sed -i "s/<\n</configuration>/ /g" /root/hadoop/hadoop/etc/hadoop/yarn-
site.xml
```

```
echo -e
"<property><name>yarn.resourcemanager.hostname</name><value>Master</value></pr
operty>\n<property><name>yarn.nodemanager.aux-
services</name><value>mapreduce_shuffle</value></property>\n</configuration>"
>> /root/hadoop/hadoop/etc/hadoop/yarn-site.xml
```

```
echo "master" > /root/hadoop/hadoop/etc/hadoop/slaves
```

初始化hadoop:

```
sh /root/hadoop/hadoop/bin/hdfs namenode -format
```

当初始化结束后若出现successfully则表示hadoop初始化成功，可使用start-all.sh启动hadoop集群。

集群启动完成后可使用jps命令查看hadoop组件启动状态。

- 安装Spark 3.1.1

安装Spark之前需要先安装Scala，百度搜索Spark和Scala，分别进入其官网下载安装包，同时上传到Linux上。

更改配置: vim /etc/profile

将Spark和Scala的目录配置到系统环境变量中:

```
export SCALA_HOME=/root/hadoop/scala
export PATH=$PATH:$SCALA_HOME/bin
export SPARK_HOME=/root/hadoop/spark-2.4.1-bin-hadoop2.7
export PATH=$PATH:$SPARK_HOME/sbin
```

使配置生效: `source /etc/profile`

验证scala安装: `scala -version` 或者直接 `scala`

更改spark配置文件:

```
vim spark-env.sh (从template中复制更改)
```

添加如下环境变量:

```
export JAVA_HOME=/root/hadoop/jdk
export SCALA_HOME=/root/hadoop/scala
export SPARK_MASTER_IP=master
export MASTER=spark://master:8080
#export SPARK_MASTER_PORT=7077
#export SPARK_WORKER_CORES=1
#export SPARK_WORKER_INSTANCES=1
export SPARK_WORKER_MEMORY=900M
export HADOOP_INSTALL=/root/hadoop/hadoop
export HADOOP_CONF_DIR=$HADOOP_INSTALL/etc/hadoop
export LD_LIBRARY_PATH=/root/hadoop/hadoop/lib/native
```

添加节点: `vim slaves` (从template中复制更改)

Master

由于hadoop和spark的启动和关闭命令相同, 所以启动spark时必须到spark/sbin目录中用start-all.sh来启动spark。(或者去sbin目录中将start-all.sh和stop-all.sh改成start.sh和stop.sh)

测试spark, 使用命令: `spark-shell` 或者访问 master:8080、master:4040 (需进入spark-shell命令式时才能访问)

jps会出现进程: SparkSubmit (需进入spark-shell命令式)、Worker、Master

- 更改配置

上述Linux服务搭建并启动好后, 还需在Windows系统中解压spark以及hadoop并配置环境变量, 如:

```
HADOOP_HOME = D:\app\eclipse-jee-photon-R-win32-x86_64\hadoop-3.2.0\hadoop-3.2.0
HADOOP_USER_NAME = root
SPARK_HOME = D:\tmp\spark-3.1.1-bin-hadoop2.7
以及在Path中添加: %HADOOP_HOME%\bin; %SPARK_HOME%\bin
```

### (3) 修改配置文件:

当上述一切配置完成后,可将本系统根目录中config/conf.ini文件打开,然后修改其中的配置。首先是系统的默认配置服务、MySQL数据库配置、Redis数据库配置、邮件验证系统配置和Hadoop等配置。

将项目目录下spark/jars文件夹中的mysql-connector-java-8.0.24.jar文件复制到spark的jars目录下。

修改spark目录下的spark.py文件中的spark配置和mysql配置,将其上传到Linux下,加入Linux定时任务每日执行。

在项目目录运行命令,创建表结构,同时创建Django管理系统的管理员帐号:

```
python manage.py makemigrations
```

```
python manage.py migrate
```

```
python manage.py createsuperuser
```

下载基础数据:

链接: <https://pan.baidu.com/s/1c83HPCp5MJey0s3CHk3OrQ> 提取码: 38ms

导入基础数据,登陆进入mysql控制台,并通过source 命令调用sql文件,如:(此处的sql文件为上面下载的三个基础数据文件)

```
mysql -uroot -p123456 -Dtest
```

```
mysql> source sys.sql
```

上述所有工作都准备完成后使用命令启动该系统服务: `python manage.py runserver 0.0.0.0:8001`

(8001为端口号,可根据需求更换,0.0.0.0表示所有主机都可访问,若设置为127.0.0.1,则表示只可本地访问)

## 四、软件功能说明

用户首次访问该网站时为访客身份,只能查询电影信息,以及搜索电影。此时用户可以选择点击注册账号,然后填写相关信息,完成注册后登陆,即可以普通用户身份使用本系统。

普通用户可对电影进行搜索、查看、收藏、评分和评论等功能，且系统会记录用户的每一次操作，系统会根据用户的操作对用户进行打标签，修改标签权重，形成用户画像，然后根据用户标签进行精准的电影推送。

发表评论

评论标题

你还可以输入500 个字符

确认提交 / Ctrl+Enter

在此处输入评论内容...

全部评论 (2)

全国计算机设计大赛

大赛

Admin 2021-04-17 21:09:33 删除

评论

好，非常好!

Admin 2021-04-17 21:11:11

在历史记录显示区域里，会显示用户的浏览记录，收藏记录以及还会推荐电影。

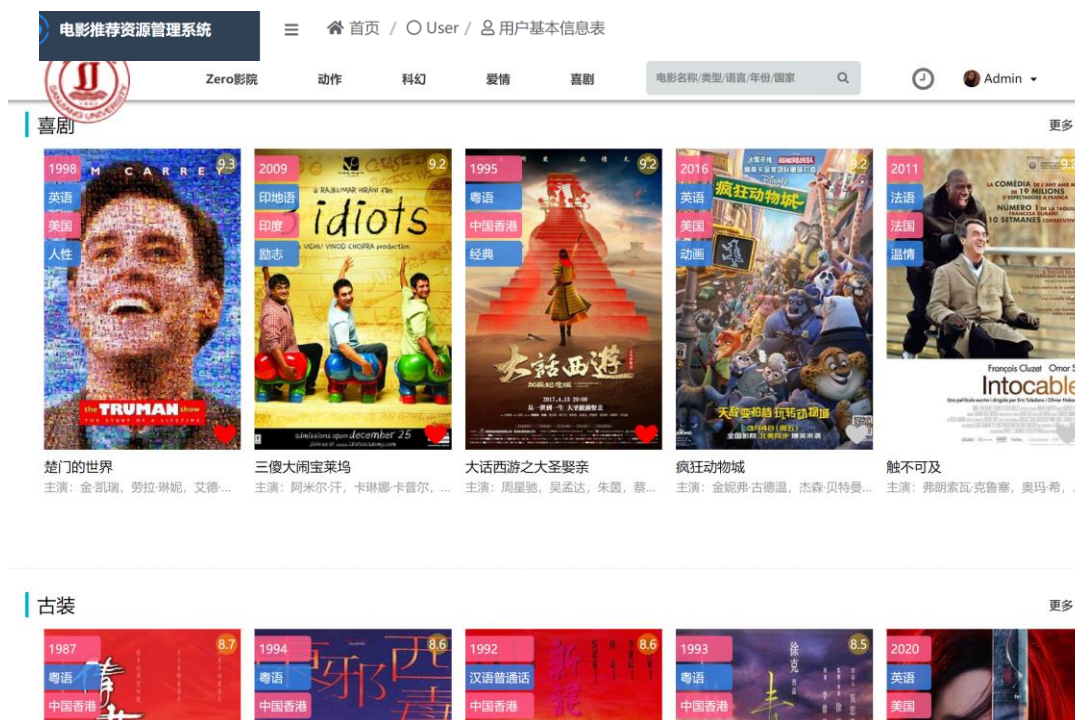
The screenshot displays a web interface for a movie recommendation system. At the top, there is a navigation bar with categories like 'Zero影院', '超', '动作', '科幻', '爱情', '喜剧', and '压'. A search bar and a user profile 'Admin' are also present. The main content area is divided into two sections: '历史记录' (History) and '收藏记录' (Collection). The '历史记录' section shows a list of movies with their posters, titles, and details. The '收藏记录' section shows a list of movies that the user has collected. The movies shown include '楚门的世界' (The Truman Show), '隐形人' (The Invisible Man), '分歧者3：上升之势' (Divergent: The Ascending Path), '我，机器人2' (I, Robot 2), '重回地球' (Return to Earth), and '大话西游之大圣娶亲' (The Chinese Odyssey Part Three: The Wedding of the Holy Son).

历史记录	收藏记录
<p>1998 英语 美国 人性 9.3</p> <p>楚门的世界</p>	<p>1995 粤语 中国香港 经典 9.2</p> <p>大话西游之大圣娶亲</p>
<p>2020 英语 澳大利亚 惊悚 7.3</p> <p>隐形人</p>	
<p>2020 英语 美国 科幻</p> <p>分歧者3：上升之势</p>	
<p>2020 英语 美国 科幻</p> <p>我，机器人2</p>	
<p>2020 汉语普通话 中国大陆 喜剧</p> <p>重回地球</p>	

在首页系统还会根据用户所选的喜欢的电影类型来显示4个类别的电影推荐，在电影详情页也会推荐类似该电影的电影。

同时用户可以在个人中心修改个人的账号信息、密码、邮箱，以及管理个人的收藏信息、评论信息等。用户使用结束后也可进行注销操作。

若成为管理员，只需在登录后台管理系统<http://127.0.0.1:8001/admin/#/admin/user/usersbase/>后，修改相关用户信息角色字段，将其修改为2即可成为管理员。若需创建后台管理员账号则需使用命令：`python manage.py createsuperuser`。



在管理系统中，管理员可对所有信息进行更改，可管理电影信息、用户信息、首页显示信息等。

