

在线教育平台用户手册

1 引言	3
1.1 编写目的.....	3
1.2 背景.....	3
1.3 定义.....	3
1.4 参考资料.....	4
2 用途	4
2.1 功能.....	4
2.2 性能.....	6
2.2.1 精度.....	6
2.2.2 时间特性.....	6
2.2.3 灵活性.....	7
2.3 安全保密.....	7
3 运行环境	7
3.1 硬设备.....	7
3.2 支持软件.....	8
3.3 数据结构.....	8
4 使用过程	11
4.1 安装与初始化.....	11
4.2 输入.....	14
4.2.1 输入数据的现实背景.....	14
4.2.2 输入格式.....	15
4.2.3 输入举例.....	15
4.3 输出对每项输出作出说明.....	16
4.3.1 输出数据的现实背景.....	16
4.3.2 输出格式.....	16
4.3.3 输出举例.....	16
4.4 文卷查询.....	16
4.5 出错处理和恢复.....	17
4.6 终端操作.....	18

用户手册（GB8567——88）

1 引言

1.1 编写目的

为了帮助用户更好地了解和使用该软件,提高用户与软件的亲和度。用户手册讲述怎样安装、配置和使用基于协同过滤的在线教育平台系统,以及该软件使用过程中应注意的一些问题。

系统提供了注册、登录、评分、评论、收藏、记录浏览历史、个性化推荐、文本分类等功能。该文档还将介绍系统预期目标,性能和其他操作方面的内容。

1.2 背景

- a. 待开发的软件系统的名称:基于协同过滤的在线教育平台;
- b. 本项目任务由方涛提出,开发者主要由胡恒昌、姜美羨担任,目标用户是贫困山区儿童、普通在校大学生和在职人员。
- c. 该软件与其他教育平台、知识科普网络存在数据上的交互,且依赖大量网络教育资源。

1.3 定义

(1) 在线教育:

或称远程教育、在线学习,现行概念中一般指的是指一种基于网络的学习行为,与网络培训概念相似。

(2) 文本分类:

文本分类用电脑对文本集(或其他实体或物件)按照一定的分类体系或标准进行自动分类标记。

(3) 网络爬虫 (Reptilia):

是一种自动获取网页内容的程序。是搜索引擎的重要组成部分,因此搜索引擎优化很大程度上就是针对爬虫而做出的优化。

(4) 协同过滤推荐 (Collaborative Filtering recommendation):

协同过滤分析用户兴趣,在用户群中找到指定用户的相似(兴趣)用户,综合这些相似用户对某一信息的评价,形成系统对该指定用户对此信息的喜好程度预测。

(5) 朴素贝叶斯算法 (Naive Bayesian Model):

朴素贝叶斯法是基于贝叶斯定理与特征条件独立假设的分类方法。

1.4 参考资料

- [1]刘露, 彭涛, 左万利, 戴耀康. 一种基于聚类的 PU 主动文本分类方法[J]. 软件学报, 2013, 11:2571-2583.
- [2]平源. 基于支持向量机的聚类及文本分类研究[D]. 北京邮电大学, 2012.
- [3]杨杰明. 文本分类中文本表示模型和特征选择算法研究[D]. 吉林大学, 2013.
- [4]李荣陆. 文本分类及其相关技术研究[D]. 复旦大学, 2005.
- [5]王煜. 基于决策树和 K 最近邻算法的文本分类研究[D]. 天津大学, 2006.
- [6] 苏金树, 张博锋, 徐昕. 基于机器学习的文本分类技术研究进展[J]. 软件学报, 2006, 09:1848-1859.
- [7]周平红. 我国高等教育信息化水平测评与发展预测研究[D]. 华中师范大学, 2012.
- [8]范坤. 推进我国教育信息化建设进程的对策研究[D]. 华中师范大学, 2004.
- [9]牛龙飞. 基于 Web 的我国教育信息化公共服务平台的设计与实现[D]. 华中师范大学, 2013.
- [10] 艾瑞咨询 2015 中国在线教育行业发展报告

2 用途

2.1 功能

本项目旨在开发一款基于包含爬虫、数据挖掘、个性教育的在线教育系统。用以将老师从繁重的查询资料、备课任务中解放。通过爬虫获取优质的互联网知识，进行文本挖掘的处理后，可以通过个性化推荐模块在既定的教学大纲内向学生推送最合适的内容，供其挑选与学习、以达到因材施教的效果。预期将大大减轻教师备课授课的压力，也能极大程度的激发学生的学习兴趣，提高学习效率。

2.1.1 用户注册功能

新用户可通过注册账号来登录系统从而进一步查看和使用本系统中的服务。已注册的用户将被允许正常登录本系统。新用户必须需要填写用户名、邮箱、密码和再次输入密码，感兴趣的话题是可选。系统会验证用户信息的完整性和邮箱格式。如果通过验证，系统将用户信息正确写入数据库，并跳转至主界面。系统同时能够并发处理多个用户同时提出的注册请求。

2.1.2 用户登录功能

允许用户从客户端登陆本系统，从而进一步查看和使用本系统中的服务。本模块与数据库交互，允许用户提出登陆请求并接收用户发送的用户邮箱、密码并校验数据符合数据库中已存在用户条目后用户将被允许正常登录本系统。如果是新用户，可跳转至注

册界面。本程序模块常驻内存，能够并发处理多个用户同时提出的登陆请求。

2.1.3 书籍搜索功能

用户可在书籍列表页面的搜索框中输入书籍全名或书名关键词以定位到指定书籍。提供点击搜索按钮，将表单提交到后台。数据库接收文本框的值，利用 sql 的 like 模糊查询在书籍列表里查找到指定书籍，并返回给前端，刷新界面，显示在书籍列表中。系统支持中英文和关键词搜索，若没有找到匹配的书籍，则返回空列表。

2.1.4 书籍收藏功能

此功能允许用户在遇到感兴趣的书籍时，将它们保存下来，以便二次查阅。用户通过点击书籍列表上的“+”将书籍添加到收藏列表。在个人页面中可查看所收藏的所有书籍，按照收藏时间升序排列。点击书目，跳转至书籍详细信息页面。用户也可以在个人页面管理收藏的书籍，通过点击右上角的“×”将书籍移出列表。系统可以利用用户收藏信息来进行行为分析，实现个性化推荐。

2.1.5 历史浏览功能

用户近三个月访问过的书籍都会被自动记录在个人主页，按照访问时间升序排列，这些记录通常被用来帮助用户进行“重访问”。用户也可以在个人主页管理收藏的书籍，通过点击“×”来删除历史记录。

同时，这些记录蕴含了丰富的用户特征和偏好，因此还可以用来帮助系统分析用户行为，获取推荐列表。

2.1.6 用户评论功能

在书籍详细信息页设有评论列表，用户可以查看别人的评论，也可以发表自己的见解。系统允许中英文输入，字数限制在五百字以内。

评论功能可以增加与用户之间的互动性。许多用户阅读过后可能会讨论书籍的内容和感受。如果没有评论功能，那么用户阅读文章后就无法发表自己的读后感。这是评论最重要的意义。

2.1.7 用户评分功能

用户为书本打分，基于统计与分析，能够获取到书本的优质信息、受欢迎程度，每一个用户的喜好、偏好倾向。满分为 5 分，分数为正整数。采用取平均的方式，将书本的评分显示给用户看，并且每一个 ip 地址也可以自主为每一本书打分。

2.1.8 文本分类功能

此功能的实现基于将网络爬虫所获取的信息依据其内容在给定的标签集合内进行文本分类。本模块需要预先使用样本文本对分类器进行训练。本程序模块无需常驻内存，并且最终生成标签列表不可重复、避免相似标签，同时能够并发处理多个文本分类请求。

2.1.9 个性推荐功能

个性化推荐建立在海量数据挖掘基础上，向顾客提供个性化的信息服务。本程序功能目的在于为特定用户推送可能喜好的书籍列表。使用本功能前提是用户在本系统中注册过，且有过一定数目的浏览记录。使用本功能时用户向服务端发送获取推荐书单列表请求，服务端根据发送请求的用户邮箱，调用本功能生成相似用户列表并返回推荐书籍列表。

2.2 性能

2.2.1 精度

1. 用户输入邮箱格式包含@且域名存在，发送邮件验证用户存在性
2. 用户名长度 1~20 个字符，不允许重名
3. 用户密码必须为英文字母加数字组合形式，且长度大于 8 个字符，并且加密
4. 用户注册确认密码与第一次密码必须相同
5. 用户第一次注册时可选择多个自己喜欢的方向（书籍类型）
6. 用户浏览记录只保存三个月
7. 用户浏览记录和收藏夹按时间升序排列
8. 用户评论支持中英文，不超过 500 字
9. 书籍简介文本信息不超过 3000 个字

2.2.2 时间特性

1. 系统接收客户端用户注册请求的延迟时间不超过 2.0 秒；
系统处理用户注册请求、写入数据库并且返回注册成功信息至客户端时间上限 2.0 秒。
2. 系统接收客户端用户登陆请求的延迟时间不超过 1.0 秒；
系统处理用户登陆请求并且返回登陆成功或失败信息至客户端时间上限 1.0 秒。
3. 爬虫程序对一篇国外文章的抓取约为 0.3s，主要由网站响应时间决定。相比起来国内网站的响应时间较快，能够达到 1s 抓取 6~7 篇文章。
4. 系统对一篇 500 字左右文章进行分类的工作时间不超过 5.0 秒
5. 系统返回生成的推荐数目列表时间不超过 3.0 秒
6. 评分模块的响应时间主要在于于数据库之间的响应速度，一般评分时间在 0.2~0.4s。

2.2.3 灵活性

- a. 用户操作方式：用户既可以用键盘直接输入信息，也可以直接在列表框中选择输入信息，比如：用户名、密码、邮箱等
- b. 运行环境：只要接入互联网的客户端都可以使用该系统；
- c. 时间特性：运行环境的改变不影响该软件的响应时间以及数据传输效率

2.3 安全保密

- a. 用户之间不能相互修改密码，即各用户只有自己登录系统才能修改自己的登录密码；各人员的密码皆不可见，即使是管理人员也不能查询各用户的密码。
- b. 各用户只能在自己的权限范围内操作，无法实现超越自己权限的任何操作，实现了数据的安全和保密。
- c. 本系统对用户密码采用 MD5 方式加密，所有用户密码经 MD5 加密后存储于服务端数据库中；此外所有用户相关信息的查询、修改等向服务端发送的请求均需要用户 COOKIE 或用户密码验证。

3 运行环境

3.1 硬设备

（1）客户端设备要求

CPU	1GHZ
内存	512MB
带宽	1MB/s
硬盘	无要求

（2）服务端设备要求

CPU	1GHZ
内存	1GB
带宽	2MB/s
硬盘	50GB

3.2 支持软件

支持软件列表：

服务端	
操作系统	CentOS 7
编译语言支持	Python2.7
支持库	Django 1.10.2
	MysqlClient
	Numpy
	Jieba（结巴分词）
	Scikit-learn
	scipy
其他	JQuary,Bootstrap
数据库	Mysql
客户端	
操作系统	Windows7， 8， 10
	Android
	Mac OS， IOS
浏览器	火狐， Chrome
	IE(可能显示不正常)

3.3 数据结构

3.3.1 逻辑结构要点

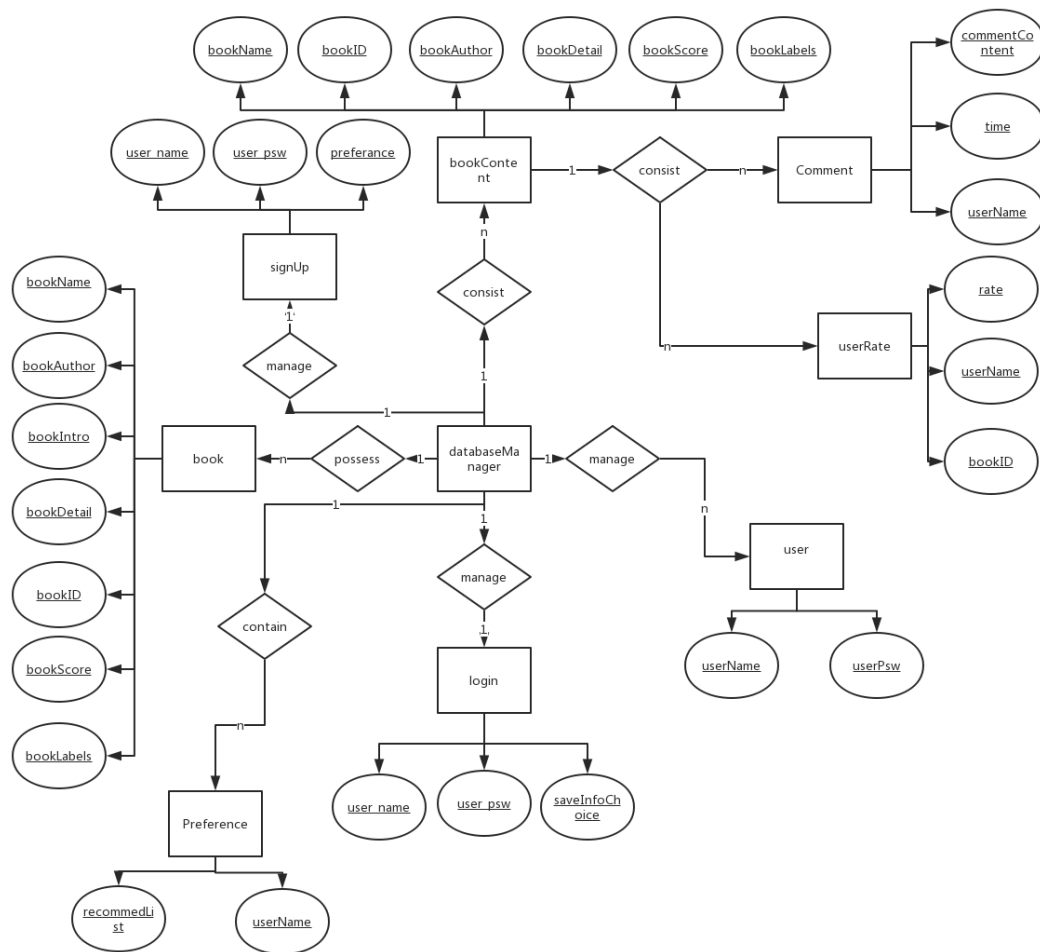


图 1

本程序逻辑结构设计要点在于数据结构模块化划分的清晰，一级逻辑数据模块包括：login、preference、book、signup、bookContent、Comment、UserRate、user，这些模块彼此相对功能独立，但是也有千丝万缕的联系。

Login 是本平台的登入的第一个模块,其包括 `user_name`(用户名)、`user_psw`(用户密码)、`saveInfoChoice` (保存个人化信息选项),这些二层数据对接下来饿模块也具有影响,比如对用户的 `user` 数据库操作的时候,需要将 `user_name` 保存下来然后对服务器进行相关信息的提交,做个人化信息、书籍的推出的时候,也需要用到 `saveInfoChoice`。

Book 包括 bookName (书籍名字)、bookAuthor (书籍作者)、bookIntro (书籍简介)、bookDetail (书籍内容)、bookID (书籍编码)、bookScore (书籍评分)、bookLabel (书籍标

签), 这些书籍也与 bookContent、userRate 里的而成逻辑数据有着紧密的联系。

3.3.2 物理结构要点

存取方法: 调用外部依赖库 mysqlclient, 使用 Django 的 model 模版创建表结构, 并且在 view 里进行相关操作。

存取单元结构: 根据逻辑结构的指标以及 DBMS 支持的数据类型, 所确定的数据项的存储类型和长度以及元组的存储结构等, 即: 数据文件及其数据项在介质上的具体存储结构。

存放位置: 指根数据库文件和索引文件等在介质上的具体存储位置。

存储介质: 用于存储文件的物理存储设备包括磁盘、磁带、光盘、磁盘阵列、磁带库、光盘阵列, 具体包括: 介质容量的大小、存取速度与费用

3.3.3 数据结构与程序的关系

服务器程序在对用户进行增删改、书籍进行增删改、评分进行增删改、喜好推荐表进行修改的过程中需要对数据库数据结构也就是数据表进行查询和修改, 在查询用户、书籍等操作中都需要对数据库中的所有表, 进行联合查询、修改。

物理数据结构主要用于各模版之间函数的信息传递, 接口传递的信息将是以数据结构封装了的数据, 以参数传递或返回值的形式在各模版间传输。出错信息将送入显示模版中, 一些数据的测试模块则送入准备模块中准备打印格式。

我们程序的数据结构在程序上的操作时线性的操作, 从表的一段开始, 向另外一端逐个按给定值与关键码进行比较, 若找到, 则是查找成功, 并给出数据元素在表中的位置, 若整个表监测完, 未找到相同的关键码, 则查找失败, 给出失败信息。从数据结构的逻辑关系层面考虑, 顺序查找的方向是可以从左到右, 也可以是从右到左。但是如果进一步考虑存储结构, 该结论就不一定正确, 比如单链表只能从左到右, 如果决定使用链表, 又要考虑从右到左的查找, 显然必须启用双向链表, 为了操作方便性而付出空间代价。

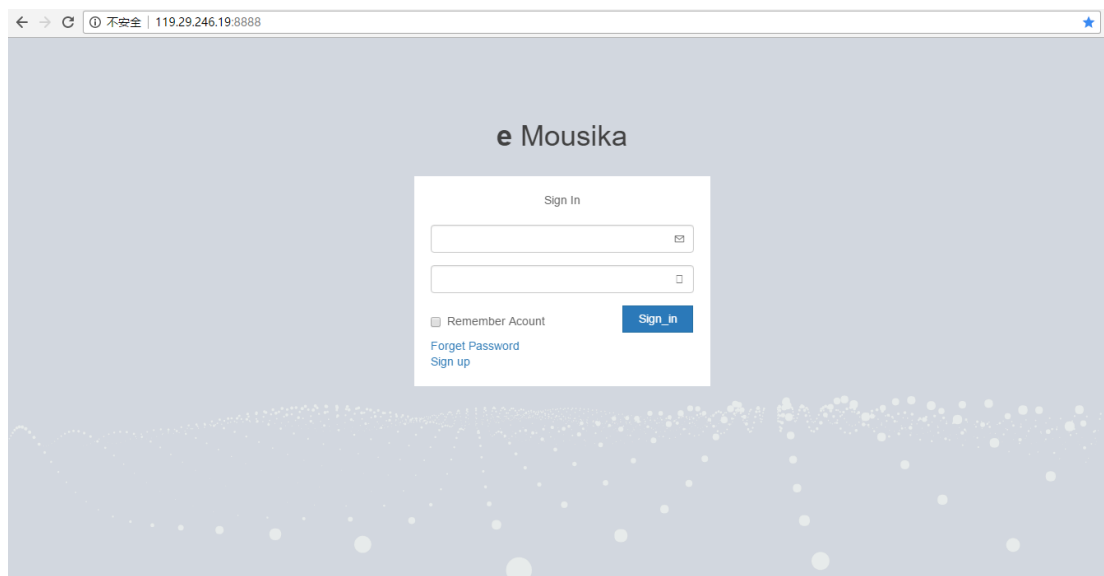
。

4 使用过程

4.1 安装与初始化

4.1.1 客户端初始化

用户客户端连接上网络，在浏览器中输入网址：<http://119.29.246.19:8888/>，访问系统登录界面。输入用户名和密码，跳转至系统主界面。



若新用户没有账号，可点击“sign up”跳转至注册界面。输入完整的个人信息，注册新账号。若用户忘记密码，可点击“forget password”跳转至注册界面，重新注册账号。



登录系统后，可在主界面进行一些列操作。



4.1.2 服务端安装与配置

a. 安装 Python

打开 WEB 浏览器访问 <http://www.python.org/download/>
选择适用于 Unix/Linux 的源码压缩包。

下载及解压压缩包

执行 `./configure` 脚本

`make`

`make install`

b. 环境变量配置

在 `csh` shell: 输入

`setenv PATH "$PATH:/usr/local/bin/python"`，按下"Enter"。

在 `bash` shell (Linux): 输入

`export PATH="$PATH:/usr/local/bin/python"`，按下"Enter"。

在 `sh` 或者 `ksh` shell: 输入

`PATh="$PATH:/usr/local/bin/python"`，按下"Enter"。

c. 安装 MySQL

1、下载 `mysql` 的 `linux` 版本的二进制安装包：

2、解压安装包：

在需要安装的位置解压缩安装包：执行命令如下：

`cd /opt`

`tar -xPf tinglyun-mysql-5.6.22.tar.gz`

3、创建 `mysql` 安装目录，并修改目录权限：

执行命令如下：

创建 `mysql` 目录：`mkdir mysql`

打开 `mysql` 目录：`cd mysql`

创建 `data` 目录：`mkdir data`

4、安装 `mysql`

在 `mysql` 目录下执行安装命令 `mysql_install_db`

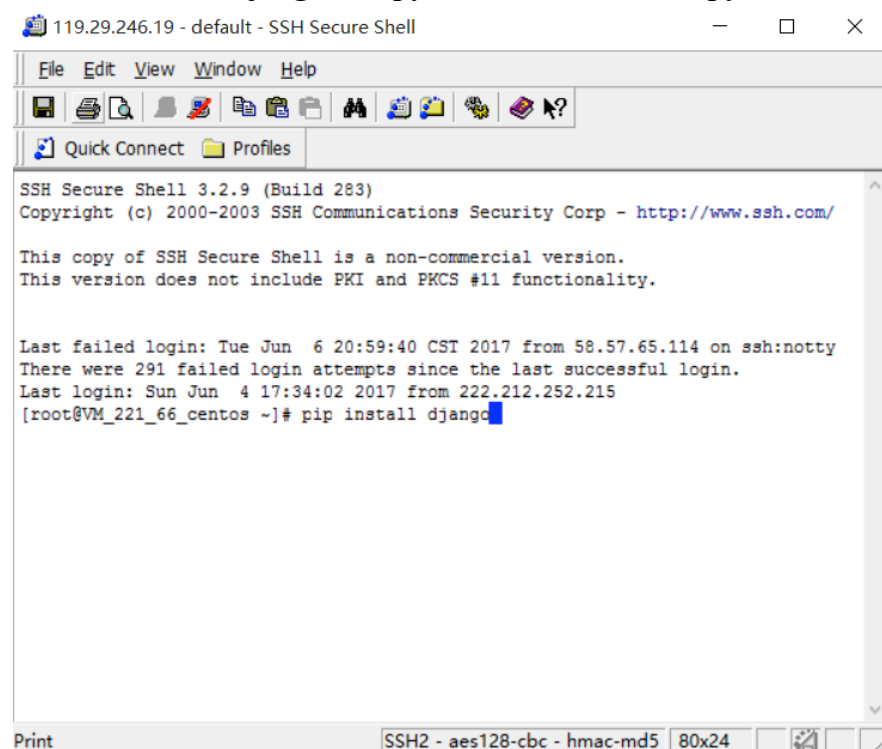
5、启动 `mysql`

进入 `init.d` 目录执行命令：

```
cd /etc/init.d/
```

```
mysqld restart
```

d.安装依赖库 (django, scipy, scikit-learn, numpy)



SSH Secure Shell 3.2.9 (Build 283)
Copyright (c) 2000-2003 SSH Communications Security Corp - <http://www.ssh.com/>

This copy of SSH Secure Shell is a non-commercial version.
This version does not include PKI and PKCS #11 functionality.

Last failed login: Tue Jun 6 20:59:40 CST 2017 from 58.57.65.114 on ssh:notty
There were 291 failed login attempts since the last successful login.
Last login: Sun Jun 4 17:34:02 2017 from 222.212.252.215
[root@VM_221_66_centos ~]# pip install django

e.解压缩本源码，放置在合适位置

/root				
名字	大小	已改变	权限	操作
↑		2017/6/6 21:18:50	r-xr-xr-x	r
goudan		2017/6/3 15:23:29	rw-r--r--	r
helloworld		2017/3/16 19:58:32	rw-r--r--	r
it		2017/6/6 14:28:59	rw-r--r--	r
it-en		2017/5/7 0:37:52	rw-r--r--	r
protectTooth		2017/6/3 15:44:16	rw-r--r--	r
soft		2016/10/5 12:58:01	rw-r--r--	r
test2		2017/5/6 15:13:55	rw-r--r--	r
testdj		2017/3/16 10:39:00	rw-r--r--	r
九洲商务		2016/10/5 10:13:44	rw-r--r--	r
anaconda-ks.cfg	3 KB	2016/4/21 15:10:43	rw-----	r
booksources.py	8 KB	2017/3/27 21:48:11	rw-r--r--	r
new.txt	80 KB	2017/6/6 21:19:01	rw-r--r--	r
nohup.out	1 KB	2017/4/9 18:12:38	rw-----	r
test.py	1 KB	2017/5/26 13:34:48	rw-r--r--	r

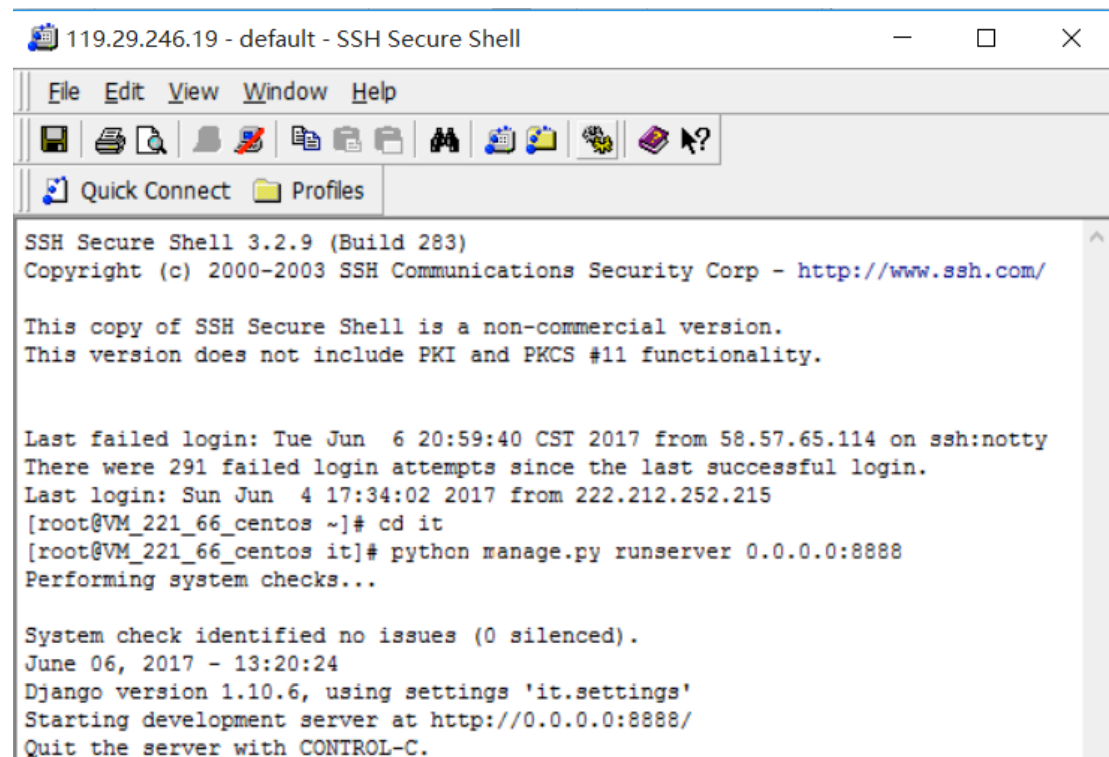
It 文件夹下为中文服务端，it-en 为英文服务端

f.运行服务端

输入指令：

cd it

python manage.py runserver 0.0.0.0:8888



The screenshot shows an SSH Secure Shell window titled "119.29.246.19 - default - SSH Secure Shell". The terminal output is as follows:

```
SSH Secure Shell 3.2.9 (Build 283)
Copyright (c) 2000-2003 SSH Communications Security Corp - http://www.ssh.com/

This copy of SSH Secure Shell is a non-commercial version.
This version does not include PKI and PKCS #11 functionality.

Last failed login: Tue Jun  6 20:59:40 CST 2017 from 58.57.65.114 on ssh:notty
There were 291 failed login attempts since the last successful login.
Last login: Sun Jun  4 17:34:02 2017 from 222.212.252.215
[root@VM_221_66_centos ~]# cd it
[root@VM_221_66_centos it]# python manage.py runserver 0.0.0.0:8888
Performing system checks...

System check identified no issues (0 silenced).
June 06, 2017 - 13:20:24
Django version 1.10.6, using settings 'it.settings'
Starting development server at http://0.0.0.0:8888/
Quit the server with CONTROL-C.
```

4.2 输入

4.2.1 输入数据的现实背景

- a. 情况：
 - 用户向服务端提交注册、登陆请求；
 - 用户发送消息请求服务器返回书籍等信息。
- b. 情况来源：
 - 用户请求；
 - 测试人员测试数据。
- c. 输入媒体：
 - 键盘，网络通信。
- d. 限制：
 - 所有输入数据均为字符串形式；
 - 用户密码必须经 MD5 加密才能输入数据库；
 - 用户密码、名称必须符合格式。
- e. 质量管理：
 - 建立出错情况日志。
- f. 支配：

- 输入格式符合要求的用户注册信息能被保存于数据库。

4.2.2 输入格式

模块	输入名称	数据类型	有效范围	能否为空	默认值	输入方式	是否加密
注册模块	用户名	String	1~20 字符	N	‘’	Inputbox	否
	用户密码	String	8~20 字符	N	‘’	Inputbox	是
	用户邮箱	String	10~20 字符	N	‘’	Inputbox	否
	用户性别	Bool	True/False	N	True	Radio button	否
登录模块	用户名	String	1~20 字符	N	‘’	Inputbox	否
	用户密码	String	8~20 字符	N	‘’	Inputbox	是
爬虫模块	爬取数量	Integer	Int 有效值内	N	0	人为规定	否
分类模块	书本 id	Int	1~8 字符	N	‘’	参数传值	否
	文本内容	String	8~20 字符	Y	‘’	参数传值	否
个性推荐	用户邮箱	String	10~20 字符	N	‘’	Inputbox	否
评分模块	评分分数	Integer	Int 有效值内	N	0	用户选择	否

4.2.3 输入举例

功能名称	输入举例
网站登录	“duolala@qq.com” “123456”
用户注册	“duolala@qq.com” “123456” “Science”
书本搜索	“人工智能”
用户评分	5
收藏书本	“电影理论”
浏览记录	“2015/3/4”
书籍推荐	服务端在个人页面中自动定期推送由系统推荐的文章
标签系统	“人工智能在当下社会成为热点。”

用户评论	“这本书措辞优美。”
------	------------

4.3 输出对每项输出作出说明

4.3.1 输出数据的现实背景

- a. 使用：
 - 提示用户请求操作是否成功；
 - 返回用户请求数据。
- b. 媒体：
 - 网络通信传输，显示屏。
- c. 质量管理：
 - 建立输出错误日志。
- d. 支配：
 - 用户历史纪录和收藏会被保存于数据库中。

4.3.2 输出格式

模块	输出名称	数据类型	有效范围	是否加密
注册模块	Check	Bool	True/false	否
登录模块	Check	Bool	True	否
爬虫模块	Labels	String[]	-	否
	BookName	String	-	否
	BookContent	String	-	否
	Author	String	-	否
分类模块	Check	Bool	True/False	否
推荐模块	RecommendedBooks	String[]	-	否
评分模块	Average_Score	Double	0.00~5.00	否

4.3.3 输出举例

功能名称	输出举例
网站登录	True/False

用户注册	True/False
书本搜索	["人工智能理论"]
用户评分	True
收藏书本	True
浏览记录	["电影理论 "]
书籍推荐	["人工智能理论 ", "电影理论"]
标签系统	"人工智能"
用户评论	True

4.4 文卷查询

当用户选择以 sqlite3 数据库存储数据:

- (1) 进入项目目录 (cd it);
- (2) 进入数据库 shell:
Sqlite3 db.sqlite3
- (3) 查询用户信息:
Select * from model_userinfo
- (4) 查询书籍信息:
Select * from model_bookinfo
- (5) 查询书籍目录信息:
Select * from model_labels

4.5 出错处理和恢复

4.5.1 出错处理

功能模块	异常说明	输出信息	处理方法
登陆模块	网络状况异常, 信息通讯交换不稳定	无法获得正常返回信息	对程序中网络信息传输的稳定加保护
	输入值非法: 长度溢出	可能造成 IndexOutOfBoundsException 溢出	对输入值的长度进行限制
	输入值非法: SQL 语言注入	非常规登入	加入登录防注入

查阅书籍 模块	带有 href 属性的元素的值未在 Django 注册表里登记	无法寻找到链接	对每一个外链 url 进行定义
	查询的字符串非法	未能搜索成功	对字符串合法监测
	网络状况异常，信息通讯交换不稳定	无法获得正常返回信息	对程序中网络信息传输的稳定加保护
评论模块	输入值非法	未能评论成功	对字符串合法监测
	Decode 方法未正确定义	评论显示出问题	Decode 正确定义

4.5.2 恢复

我们的系统除了在数据库上，在本地也有备份，可以放置数据库的系列错误导致信息不可恢复的结果。在操作数据的过程中，如果因为软件错误或失误，则会进行人为就诊和修改错误代码。并且会及时更新系统。

后备技术说明准备采用的后备技术，当原始系统数据万一丢失时启用的副本的建立和启动的技术，例如周期性地把磁盘信息记录到磁带上就是对于磁盘媒体的一种后备技术；

降效技术说明准备采用的后备技术，使用另一个效率稍低的系统或方法来求得所需结果的某些部分，例如一个自动系统的降效技术可以是手工操作和数据的人工记录；

恢复及再启动技术说明将使用的恢复再启动技术，使软件从故障点恢复执行或使软件从头开始重新运行的方法。

4.6 终端操作

a. 终端连接方式

(1) 连接互联网，启用浏览器（chrome）

(2) 输入项目地址并跳转：

服务器地址 + 端口号（8888）

b. 终端操作能力

(1) 用户数据上传：

包括用户注册信息上传，和用户登录信息上传。

(2) 用户信息修改：

包括用户对自身用户名、用户密码、用户偏好的修改。

(3) 书籍信息搜索:

用户可以通过终端实现对服务端数据库中书籍的搜索操作。

(4) 书籍检索:

用户可以根据终端提供的目录快速检索某类别书籍。