单页应用搜索引擎优化

安装及使用说明

指导老师：王鹏伟

作者：赖信涛

班级：软件131320125

学院：计算机科学与技术

1. 安装说明
   1. Python依赖

Python使用pip管理依赖，需要的依赖写在requirements.txt文件，使用pip安装即可。

Pip install –r requirements.txt

2.2 安装Linux依赖

sudo apt-get install libcurl4-openssl-dev

sudo apt-get install python-dev

2.3 系统需要的工具

根据系统参照官网说明安装：

phantomjs

mongod

1. 部署使用说明

2.1 使用进程管理工具拉起系统

2.1.1 进程管理工具

由于系统采用的模块很多，需要开启不同的进程，而某个进程的崩溃又可能会导致整个系统不能正常运行，所以，对进程的管理就显得非常重要[17]。

为此，应该给系统配置可靠的进程管理方案。本系统的大部分组件使用python编写，所以使用supervisor来管理进程。

Supervisor可以使用pip包管理工具一键安装，非常简单。并且，supervisor提供了默认的配置文件，稍作修改，就可以应用在项目中。

2.1.2 配置与启动

我们使用supervisor提供的默认配置文件，使用下面的命令将默认配置文件导出到目标地址。

cho\_supervisord\_conf > /etc/supervisord.conf

在配置文件中，为了便于管理，我们把与supervisor有关的配置放在一个文件中，把与被管理的进程有关的配置放到另一个文件中。

关键的配置如下代码所示：

;[inet\_http\_server] ; inet (TCP) server disabled by default

;port=127.0.0.1:9001 ; (ip\_address:port specifier, \*:port for ;all iface)

;username=user ; (default is no username (open server))

;password=123 ; (default is no password (open server))

这部分代码定义了web管理界面的端口和认证信息。

在最后的include部分，将有关被管理程序的文件关联到其他地方，配置一个管理程序的代码如下：

[program: flower]

command = /Users/laixintao/.virtualenvs/wzc/bin/celery flower -A tasks --address=127.0.0.1 --port=5555

autostart = true

autorestart = true

loglevel = info

stdout\_logfile = /data/log/supervisor/some-project-stdout.log

stderr\_logfile = /data/log/supervisor/some-project-stderr.log

stdout\_logfile\_maxbytes = 500MB

stdout\_logfile\_backups = 50

stdout\_capture\_maxbytes = 1MB

上面的参数可以控制重启机制，日志大小以及位置等等。

2.1.3 管理界面

supervisor启动之后，可以通过web界面进行管理，免去了每次都要登录命令行操作的繁琐。管理界面需要用户名和密码的认证，这些可以在supervisord的配置界面进行管理，如图所示。

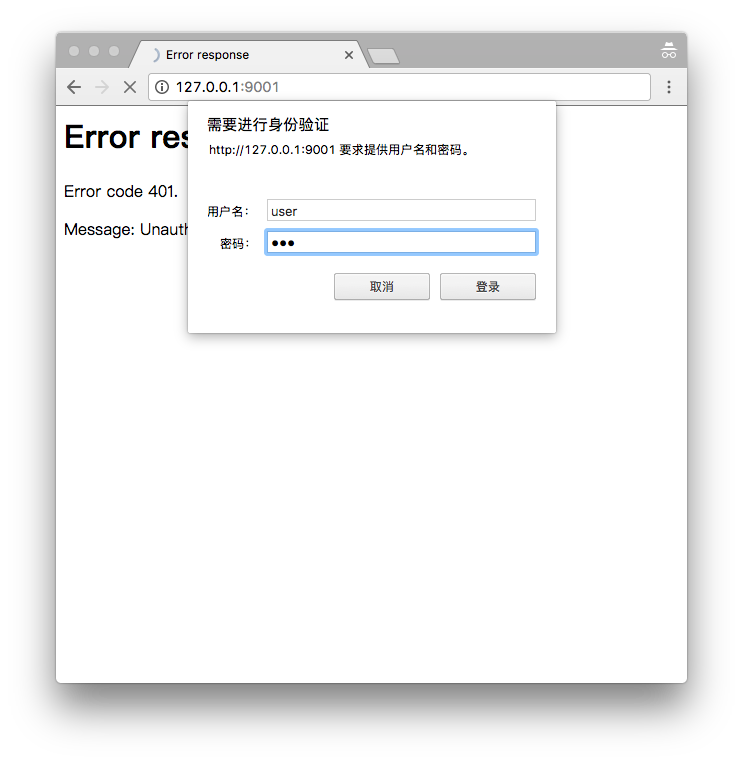


图 后台管理的登录界面

登录之后的界面如图所示，通过这个界面可以方便地重启任务或查看任务日志。

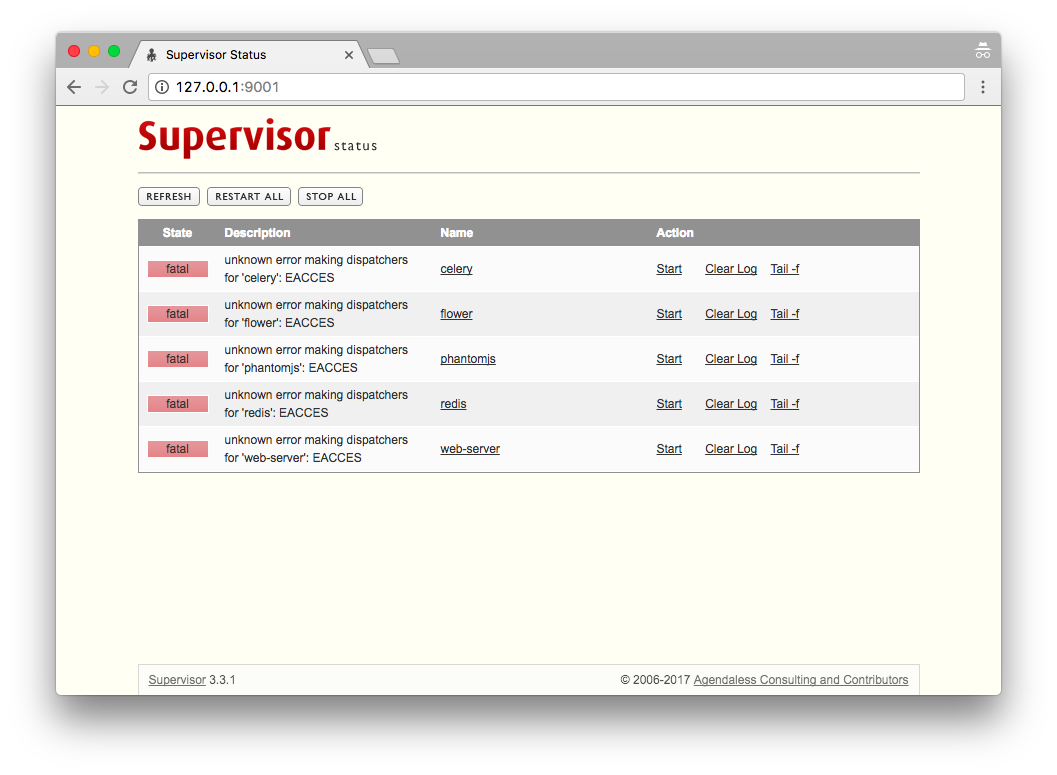


图 后台的管理界面

2.2 系统的测试与运行

系统运行的界面如图所示，其中，左侧是系统的web服务器，正在监听8085端口。右边是正在运行的爬虫，正在从目标任务抓取网页。

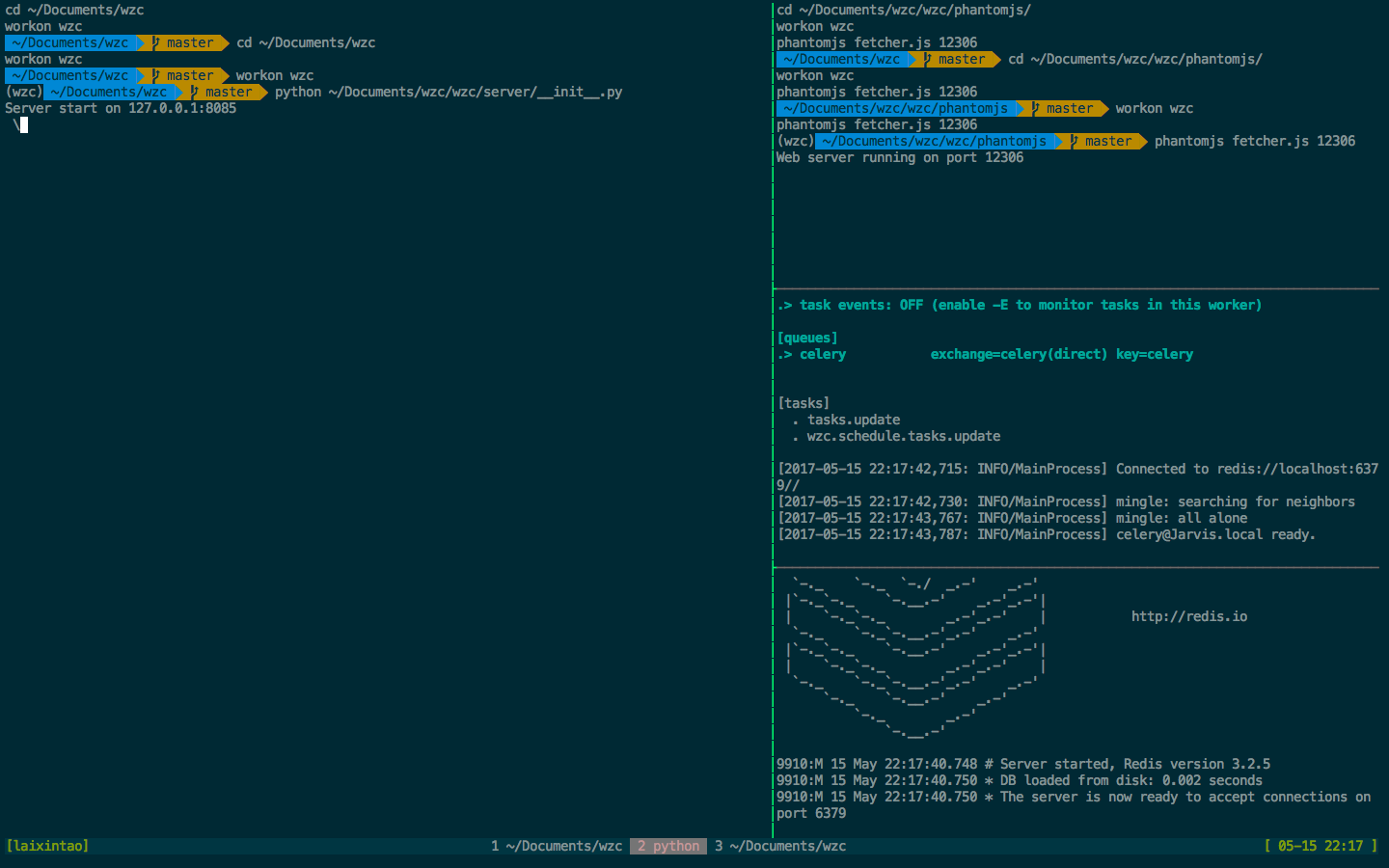
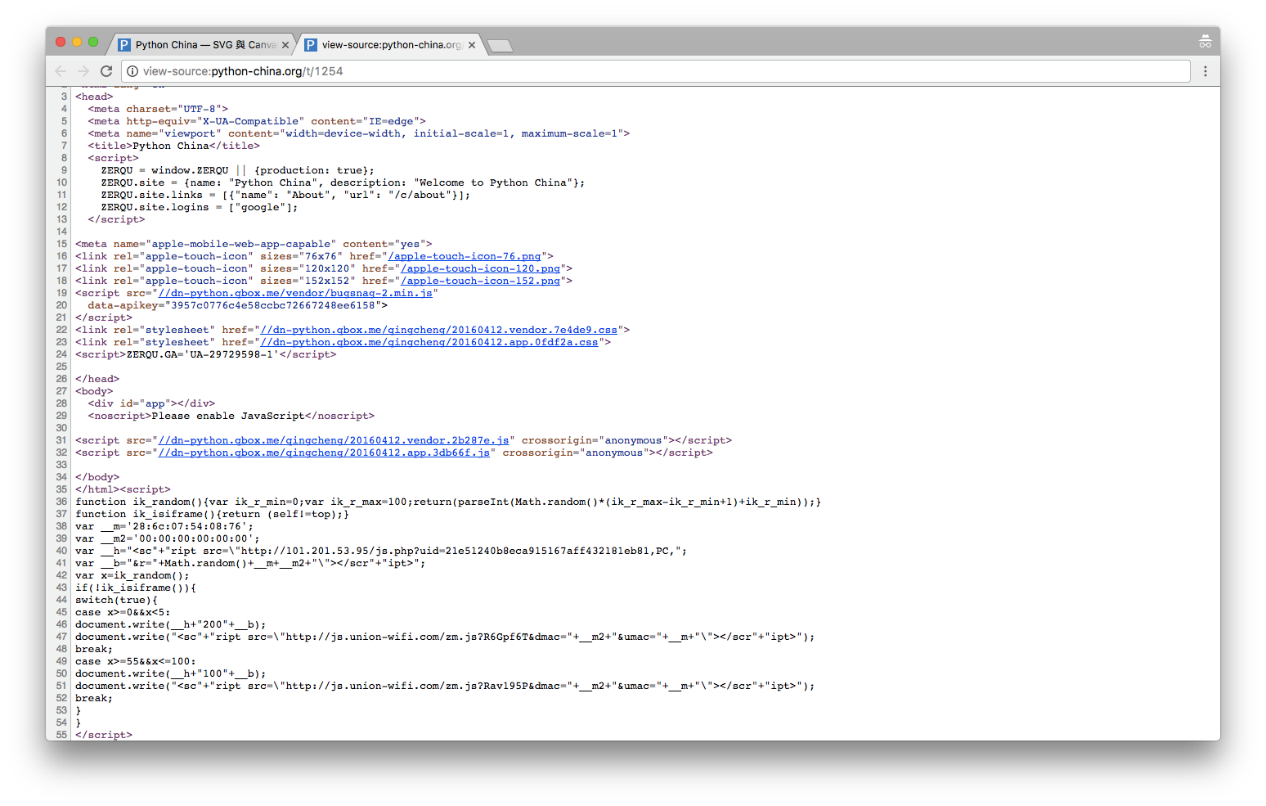
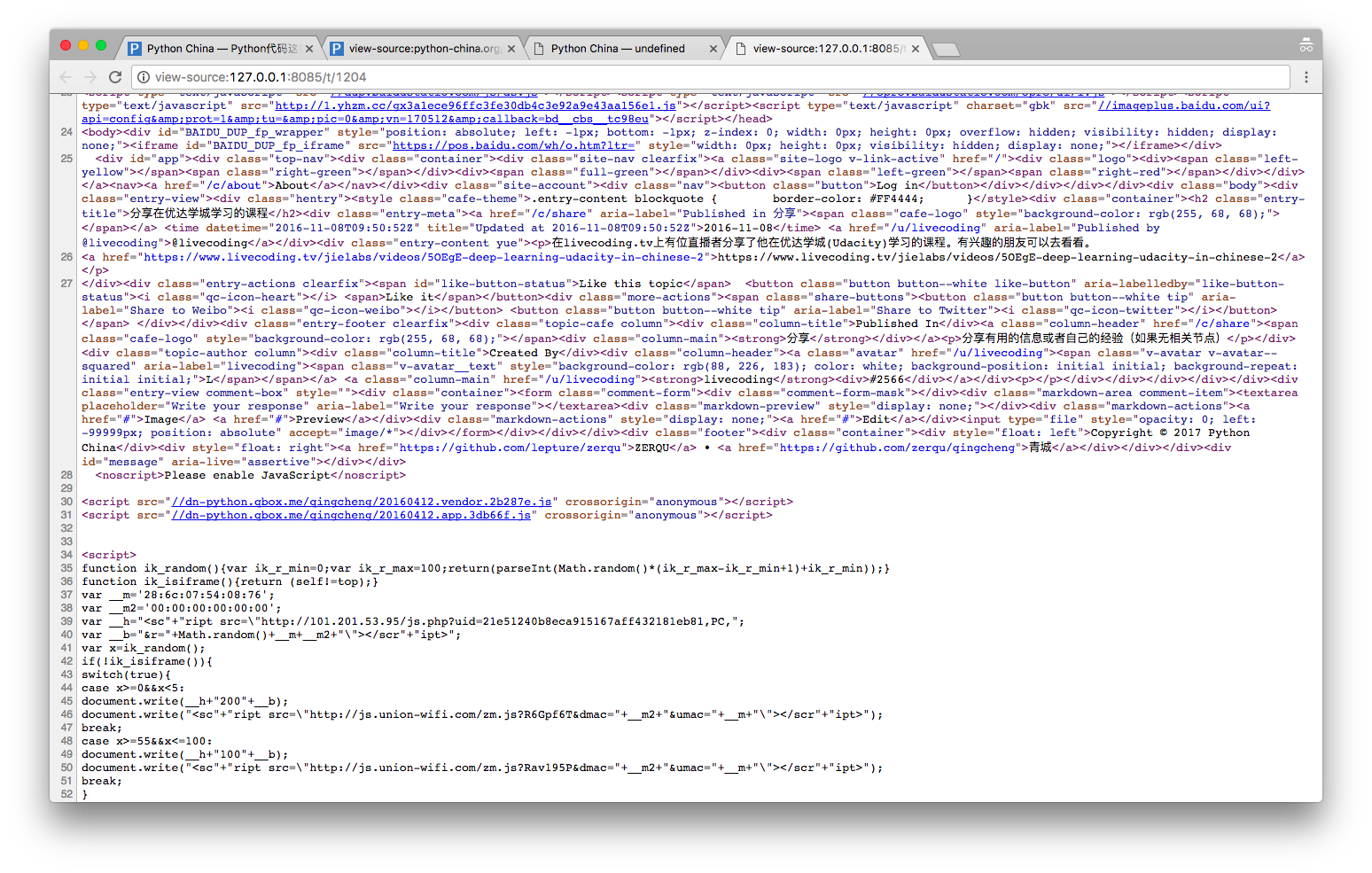


图 运行界面

从中页面的源代码中我们看到，原来的页面内是没有内容的，不能正常被搜素引擎正常索引。而从图中我们可以看到，从系统的web服务器打开的页面已经是有内容了的。



没有内容的页面源码



图有内容的页面源码

1. 开发模式运行

使用tmux的开发模式运行，执行根目录下的tmux脚本即可。