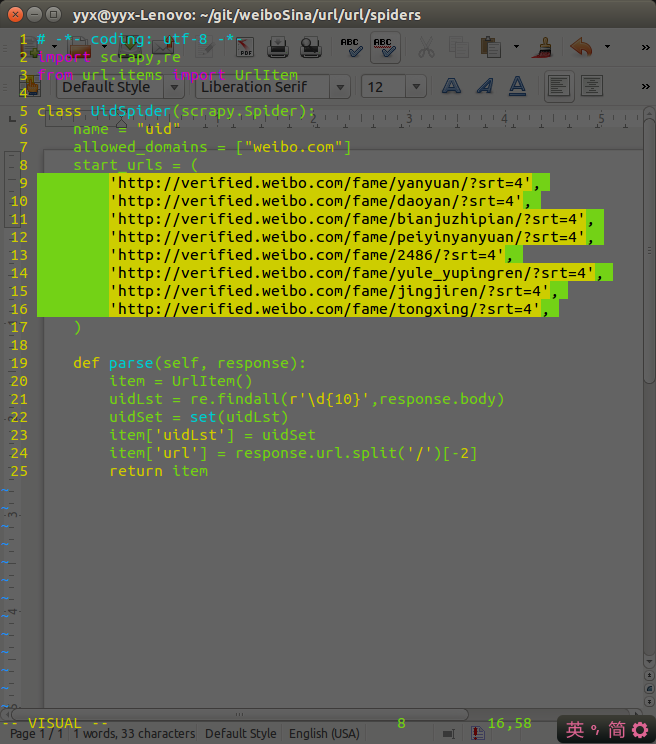
1.weiboSina/url/url/spider/uid.py

把start\_urls中的url改为对应类别的url。例如：



特征：结尾为/?srt=4

2.weiboSina/url/url/

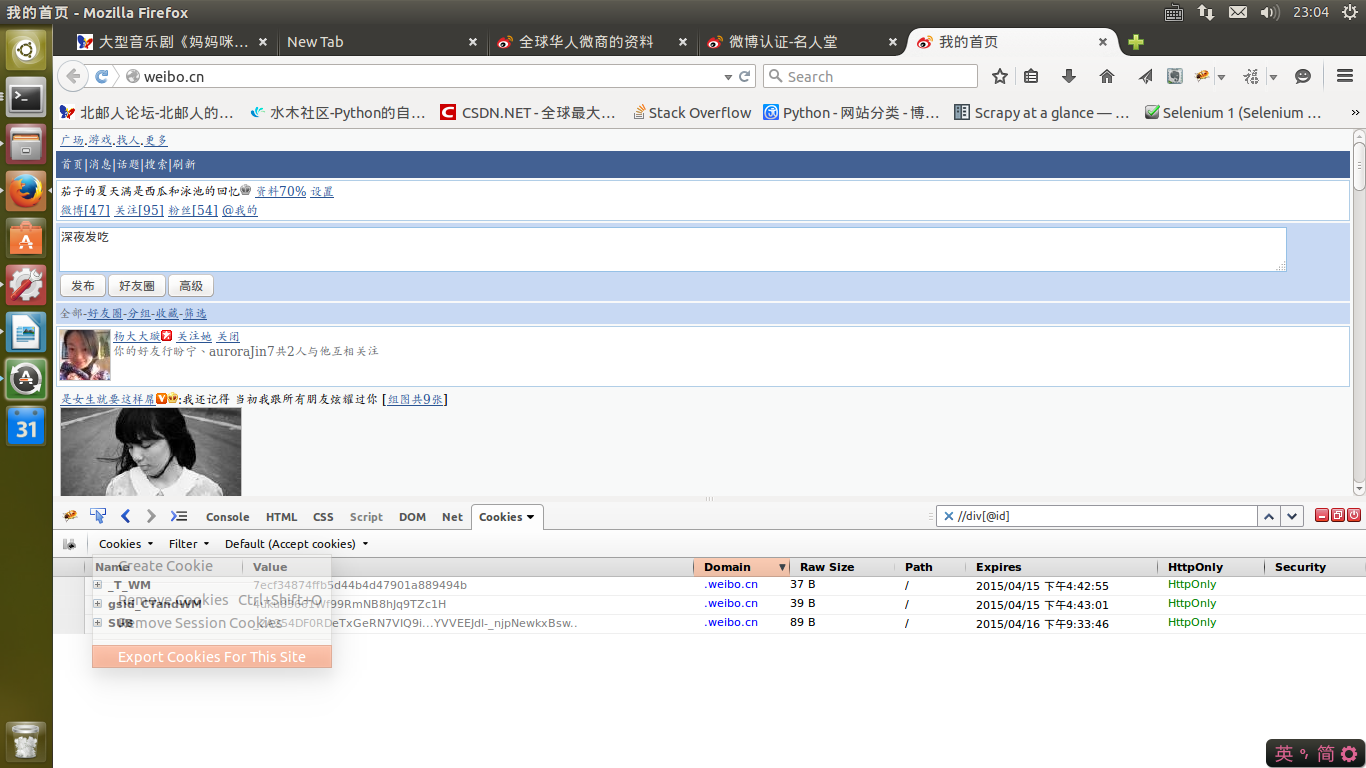
$scrapy crawl uid

3.把weiboSina/url/url/yule/文件夹移动到weiboSina/test123/test123/spiders/文件夹中

4.获取cookie文件

登陆www.weibo.cn

使用Firefox浏览器，使用firebug工具



记住路径

5.修改weiboSina/test123/test123/spiders/uinfo.py

第17行，cookiePath修改为第4步中设置的路径

第32行，把self.item['user\_label']修改为爬取的类别名字（替换‘娱乐’）

6.建立数据库

ubuntu系统下

$sudo aptitude install mysql-server

密码设为yuanyunxu

$mysql -u root -p mydb

>create table blogs (blog\_id varchar(30), user\_id int(10), blog\_flag int(10), blog\_content varchar(400), blog\_forward\_num int(10), blog\_comment\_num int(10), blog\_praise\_num int(10), blog\_pub\_time varchar(20), blog\_download\_time varchar(20),primary key (blog\_id))default charset=utf8;

> create table userInfo (user\_id int(10) not null, user\_label varchar(20), user\_nickname varchar(20), blog\_num int(10), following\_num int(10), follower\_num int(10), user\_tags varchar(200), user\_description varchar(200) ,user\_sexual varchar(10),user\_birth varchar(10),user\_location varchar(10),user\_cert varchar(30),user\_ori varchar(10),user\_status varchar varchar(10),primary key (user\_id))default charset=utf8;

7.weiboSina/test123/test123/spiders/

$scrapy crawl uinfo