# 第四课 简单爬取新浪新闻列表

新闻列表页网页结构：



输出h2：

res = requests.get('http://news.sina.com.cn/china/')

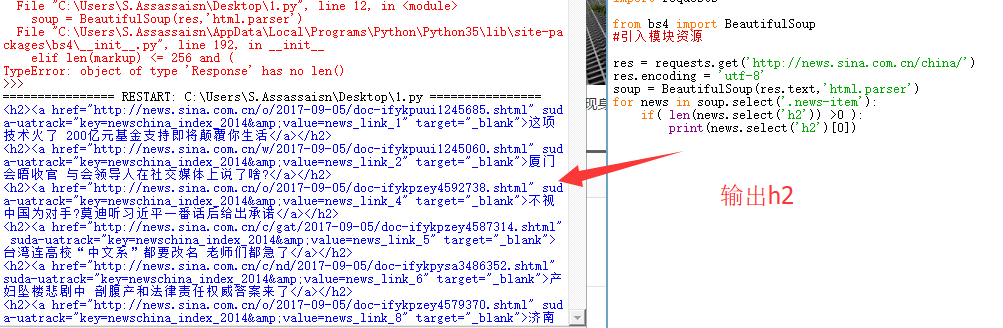
res.encoding = 'utf-8'

soup = BeautifulSoup(res.text,'html.parser')

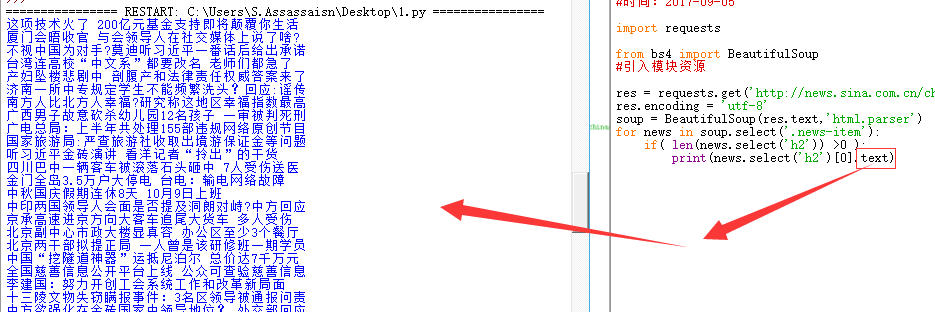
for news in soup.select('.news-item'):

if( len(news.select('h2')) >0 ):

print(news.select('h2')[0])



.text



如法炮制：

获取新闻时间 标题 超链接

res = requests.get('http://news.sina.com.cn/china/')

res.encoding = 'utf-8'

soup = BeautifulSoup(res.text,'html.parser')

for news in soup.select('.news-item'):

if( len(news.select('h2')) >0 ):

h2 = news.select('h2')[0].text

time = news.select('.time')[0].text

a = news.select('a')[0]['href']

print(time,h2,a)

