问题：

urllib.request.urlopen() 方法经常会被用来打开一个网页的源代码，然后会去分析这个页面源代码，但是对于有的网站使用这种方法时会抛出"HTTP Error 403: Forbidden"异常。

例如： 执行下面的语句时

[python] view plain copy

```
1. urllib.request.urlopen("http://blog.csdn.net/eric_sunah/article/details/11099295")
```

会出现以下异常:

[python] view plain copy

```
1. File "D:\Python32\lib\urllib\request.py", line 475, in open
2.     response = meth(req, response)
3.   File "D:\Python32\lib\urllib\request.py", line 587, in http_response
4.     'http', request, response, code, msg, hdrs)
5.   File "D:\Python32\lib\urllib\request.py", line 513, in error
6.     return self._call_chain(*args)
7.   File "D:\Python32\lib\urllib\request.py", line 447, in _call_chain
8.     result = func(*args)
9.   File "D:\Python32\lib\urllib\request.py", line 595, in http_error_default
10.     raise HTTPError(req.full_url, code, msg, hdrs, fp)
11. urllib.error.HTTPError: HTTP Error 403: Forbidden
```

分析:

之所以出现上面的异常，是因为如果用 urllib.request.urlopen() 方式打开一个URL，服务器端只会收到一个单纯的对于该页面访问的请求，但是服务器并不知道发送这个请求使用的浏览器，操作系统，硬件平台等信息，而缺失这些信息的请求往往都是非正常的访问，例如爬虫。

有些网站为了防止这种非正常的访问，会验证请求信息中的UserAgent（它的信息包括硬件平台、系统软件、应用软件和用户个人偏好），如果UserAgent存在异常或者是不存在，那么这次请求将会被拒绝（如上错误信息所示）
所以可以尝试在请求中加入UserAgent的信息。

方案：

对于Python 3.x来说，在请求中添加UserAgent的信息非常简单，代码如下：

[python] view plain copy

```
1. headers = {'User-
Agent':'Mozilla/5.0 (Windows NT 6.1; WOW64; rv:23.0) Gecko/20100101 Firefo
x/23.0'}
2. req = urllib.request.Request(url=chaper_url, headers=headers)
3. urllib.request.urlopen(req).read()
```