

python中的urlencode与urldecode

当url地址含有中文，或者参数有中文的时候，这个算是很正常了，但是把这样的url作为参数传递的时候（最常见的callback），需要把一些中文甚至'/'做一下编码转换。

一、urlencode

urllib库里面有个urlencode函数，可以把key-value这样的键值对转换成我们想要的格式，返回的是a=1&b=2这样的字符串，比如：

```
>>> from urllib import urlencode
>>> data = {
...     'a': 'test',
...     'name': '魔兽'
... }
>>> print urlencode(data)
a=test&name=%C4%A7%CA%DE
```

如果只想对一个字符串进行urlencode转换，怎么办？urllib提供另外一个函数：quote()

```
>>> from urllib import quote
>>> quote('魔兽')
'%C4%A7%CA%DE'
```

二、urldecode

当urlencode之后的字符串传递过来之后，接受完毕就要解码了——urldecode。urllib提供了unquote()这个函数，可没有urldecode()！

```
>>> from urllib import unquote
>>> unquote('%C4%A7%CA%DE')
'\xc4\xa7\xca\xde'
>>> print unquote('%C4%A7%CA%DE')
魔兽
```

三、讨论

在做urldecode的时候，看unquote()这个函数的输出，是对应中文在gbk下的编码，在对比一下quote()的结果不难发现，所谓的urlencode就是把字符串转成gbk编码，然后把\x替换成%。如果你的终端是utf8编码的，那么要把结果再转成utf8输出，否则就乱码。

可以根据实际情况，自定义或者重写urlencode()、urldecode()等函数。