

Python - 编码转换

标签: GBK gbk python Python UNICODE Unicode unicode UTF-8 utf-8

2013-01-14 22:09 6307人阅读 评论(0) 收藏 举报

版权声明：本文为博主原创文章，未经博主允许不得转载。

```
[python]    
01. # coding: utf-8  
02.  
03. s = 'abc'  
04. print type(s) # str(utf-8)  
05. print len(s) # 3  
06.  
07. s = unicode(s) # str -> unicode, 其中str的每个字符值必须小于128  
08. print type(s) # unicode  
09. print len(s) # 3  
10.  
11. s = u'abc'  
12. print type(s) # unicode  
13. print len(s) # 3  
14.  
15. s = s.encode('utf-8') # unicode -> str(utf-8) | 载:  
16. print type(s) # str  
17. print len(s) # 3  
18.  
19. s = s.decode('utf-8') # str(utf-8) -> unicode, 这里str的每个字符值任意  
20. print type(s) # unicode  
21. print len(s) # 3  
22.  
23. s = '中国' # 由于整个文件以utf-8编码  
24. print type(s) # str(utf-8)  
25. print len(s) # 6  
26.  
27. s = u'中国'  
28. print type(s) # unicode  
29. print len(s) # 2  
30.  
31. s = s.encode('utf-8')  
32. print type(s) # str(utf-8)  
33. print len(s) # 6  
34.  
35. s = s.decode('utf-8')  
36. print type(s) # unicode  
37. print len(s) # 2  
38.  
39. s = raw_input(u'输入: ') # windows下貌似中文按gbk编码, 每个中文占2个字节  
40. print type(s) # str(gbk)  
41. print len(s) # 4  
42.  
43. s = s.decode('gbk') # 要想gbk编码转为utf-8编码, 先将gbk编码转为unicode  
44. print type(s) # unicode
```

```
45. print len(s) # 2
46.
47. s = s.encode('gbk')
48. print type(s) # str(gbk)
49. print len(s) # 4
50.
51. # 根据以上的验证，得出结论
52. # 各种编码都可以通过unicode来转化，unicode可以假想为一张各种字符的对照表，在这个表中可以找到世界范围内的任何一种字符
53. # 当然，也包括中文，每个字符都对应一个序号，如'a' -> 0x61, '中' -> 0x4e2d
54. # unicode -> utf-8 unicode.encode('utf-8')
55. # utf-8 -> unicode str.decode('utf-8')
56. # gbk -> unicode str.decode('gbk')
57. # unicode -> gbk unicode.encode('gbk')
```