1. 返回的结构都是一个可迭代的generator，可以使用for循环来获得。返回如下：

<generator object Tokenizer.cut at 0x0000000004BE9410>

1. 读取、写入文件的编码错误：

file = open(filename,'w',encoding='utf-8') # 写入时加入encoding='utf-8'

file = open(filename,'rb') # 读取时，使用'rb'而不使用'r'

text = file.read() # 如果报错，修改IDE的设置File-setting-file encoding—全部设为utf-8

decode('utf-8', 'ignore') 用于html爬取下网页，某些评论使用了不可以解析的字符

1. 文件保存方法  
   1) 自带的open()，然后write

2) NumPy模块， 保存：np.savetxt(‘filename’, array)  
 读取：np.loadtxt(‘data.csv’,delimiter=’,’, usecols=(1, 3), skiprows=1) 读取csv文件，用逗号分隔｛貌似读不了数据类型不一致的csv｝,取1,3列，跳过第一行表头(因为数据类型不同)

3）Pandas模块 pandas.read\_csv(r"c:\csv\test.csv")

1. Numpy处理时间序列