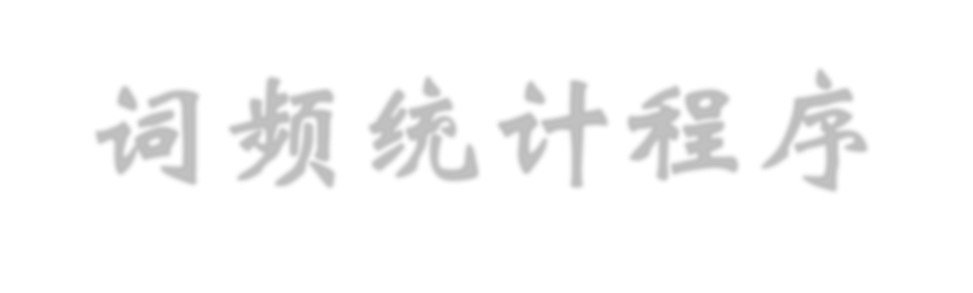


案例：



词频统计程序

张

华

WHU



需求

一篇文章，出现了哪些词？哪些词出现得最多？英文文本中文文本

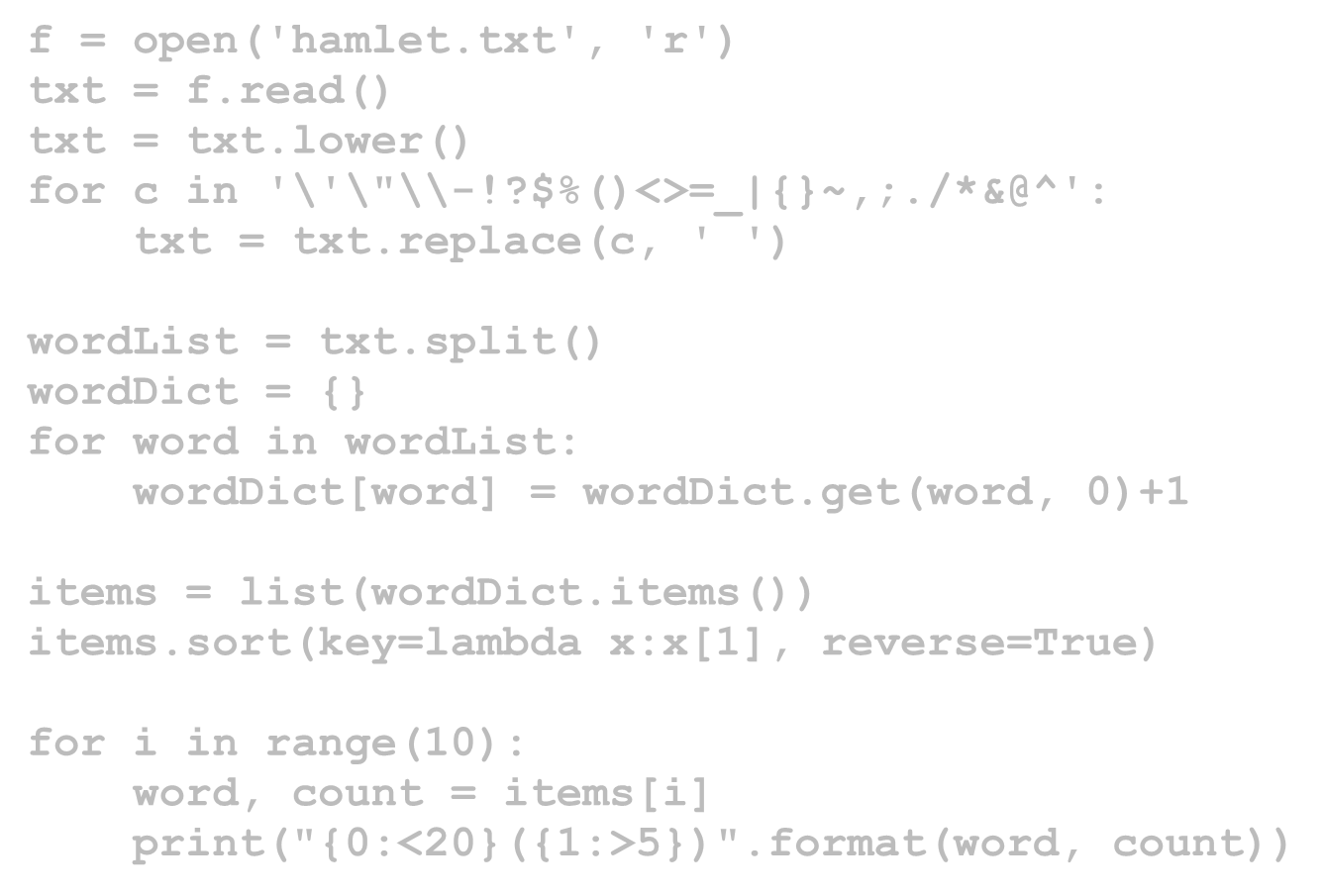
设计

文本是保存在文件中的一个长字符串从文本中分离出单词统计单词出现的次数用字典来保存单词及其出现次数实现**1** 统计《哈姆雷特》中出现次数最多的**10**个单词

切割单词

把特殊符号换成空格

打开文档



**f =**

**open(**

**'hamlet.txt**

**', 'r')**

**txt**

**=**

**f.read**

**()**

**txt =**

**txt.lower**

**()**

**for c in '**

**\**

**'**

**\**

**"**

**\**

**\**

**-**

**!**

**?$%()<>=\_|{}~,;./\*&@^'**

**:**

**txt =**

**txt.replace**

**(**

**c, ' '**

**)**

**wordList**

**=**

**txt.split**

**()**

**wordDict**

**=**

**{}**

**for word in**

**wordList**

**:**

**wordDict**

**[**

**word] =**

**wordDict.get**

**(**

**word,**

**0)+1**

**items = list(**

**wordDict.items**

**())**

**items.sort**

**(**

**key**

**=**

**lambda x:x[1]**

**, reverse=True)**

**for**

**i**

**in range(10):**

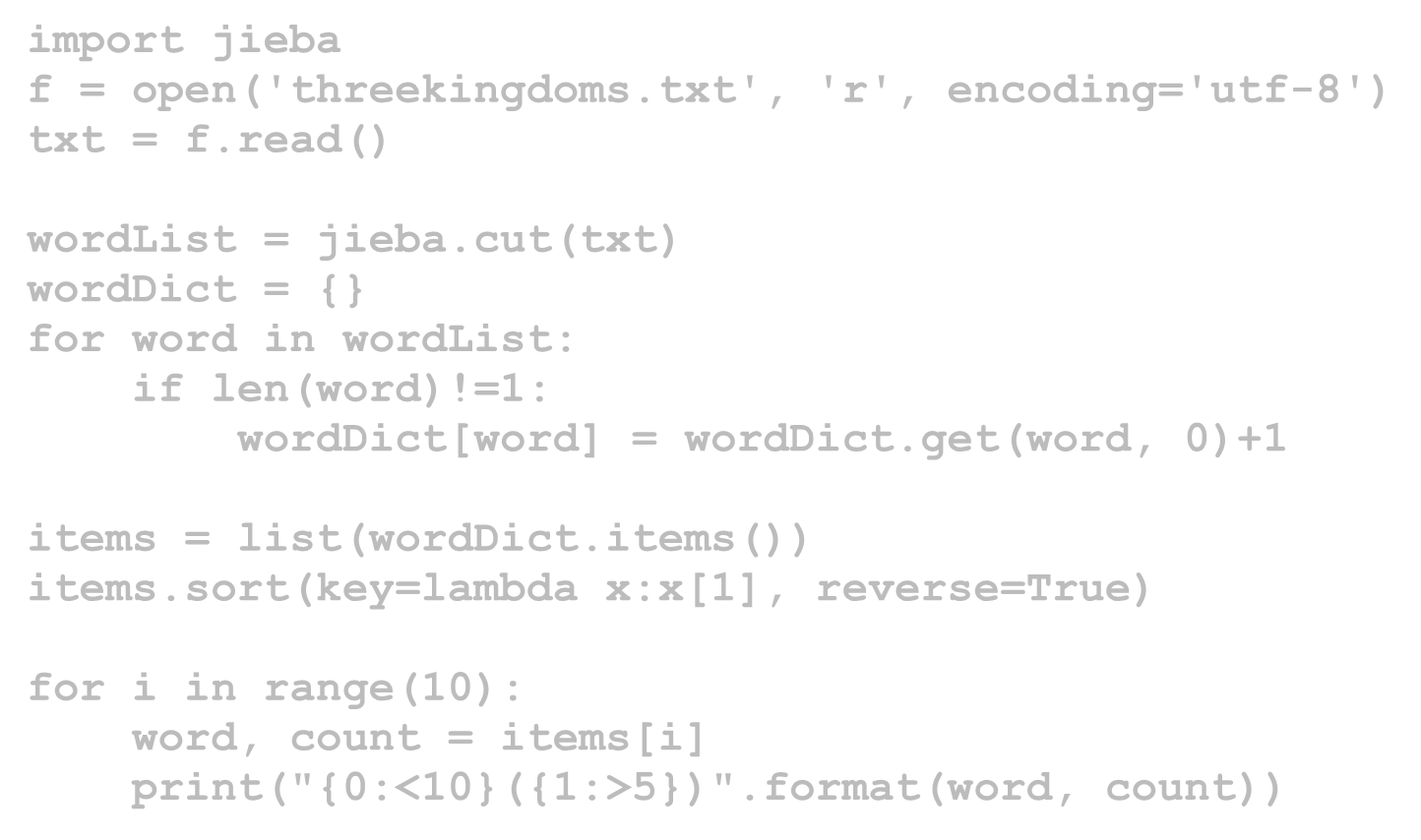
**word, count = items[**

**i**

**]**

**print("{0:<20}({1:>5})".format(word, count))**

实现**2** 统计《三国演义》中出场次数最多的**10**个人物



**import**

**jieba**

**f = open('threekingdoms.txt', 'r', encoding='utf**

**-**

**)**

**'**

**8**

**txt =**

**f.read**

**()**

**wordList**

**=**

**jieba.cut**

**(**

**txt**

**)**

**wordDict**

**=**

**{}**

**for word in**

**wordList**

**:**

**if**

**len**

**word**

**)!=1:**

**(**

**wordDict**

**word] =**

**[**

**wordDict.get**

**(**

**word,**

**0)+1**

**items = list(**

**wordDict.items**

**())**

**items.sort**

**(**

**key=lambda x:x[1], reverse=True**

**)**

**for**

**i**

**in range(10):**

**word, count = items[**

**i**

**]**

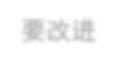
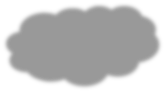
**print("{0**

**:<10**

**}({1:>5})**

**".format(word, count**

**))**



要改进