Python 生成句子

作者：mubai

原文链接：https://ld246.com/article/1484646877386

来源网站：链滴

```
from urllib.request import urlopen
from random import randint

def wordListSum(wordList):
    sum = 0
    for word, value in wordList.items():
        sum += value
    return sum

def retrieveRandomWord(wordList):
    randIndex = randint(1, wordListSum(wordList))
    for word, value in wordList.items():
        randIndex -= value
        if randIndex <= 0:
            return word

def buildWordDict(text):
    # 剔除换行符和引号
    text = text.replace("\n", "")
    text = text.replace("\"", "")

    # 保证每个标点符号都和前面的单词在一起
    # 这样不会被剔除，保留在马尔可夫链中
    punctuation = [',', '.', ';', ':']
    for symbol in punctuation:
        text = text.replace(symbol, " " + symbol + " ")

    words = text.split(" ")
    # 过滤空单词
    words = [word for word in words if word != ""]

    wordDict = {}
    for i in range(1, len(words)):
        if words[i-1] not in wordDict:
            # 为单词新建一个词典
            wordDict[words[i-1]] = {}
        if words[i] not in wordDict[words[i-1]]:
            wordDict[words[i-1]][words[i]] = 0
        wordDict[words[i-1]][words[i]] = wordDict[words[i-1]][words[i]] + 1
    return wordDict

text = str(urlopen("http://pythonscraping.com/files/inaugurationSpeech.txt").read(), 'utf-8')
wordDict = buildWordDict(text)

# 生成链长为100的马尔可夫链
length = 100
chain = ""
currentWord = "I"
for i in range(0, length):
    chain += currentWord + " "
    currentWord = retrieveRandomWord(wordDict[currentWord])

print(chain)
```

原文链接：Python 生成句子

参考：《Python网络数据采集》

原文链接：Python 生成句子