

查找后继词

markov.pas (exe)

【题目叙述】

我们现在需要一些随机的可以读的英语文本，但是计算机运用语言的能力跟一个婴儿差不了多少，由它随机生成的文本是毫无意义的。当然，我们也可以给它一本字典，让它从中随机挑选单词，但这也无济于事，因为计算机不懂语法。实际上，我们可以用一种相当简单的方法，让计算机写出一段比较像样的文章来。这种技术被称为“马尔可夫链算法”。它利用一个样本来产生一段结构与样本相似的文章。

我们可以把一段英语文本看成是一系列相互重叠的短语的组合，该算法把每个短语分成两部分：由一个以上的词构成的前趋和只有一个词的后继。这个算法用一句话就能说明问题：根据特定的前趋在样本中随机地选取后继。实践证明，采用 3 个词的短语效果最好。算法具体过程如下：

- 1、随机地在样本中选取连续的两个词 w1、w2 作为文本的开头。
- 2、输出 w1、w2。
- 3、在样本中随机选取一个 w1、w2 的后继 w3，如果不存在 w3，则结束算法。
- 4、输出 w3。
- 5、令 w1=w2，w2=w3。
- 6、转 3。

这个算法也可以根据你的需要随时停止。为了帮助理解算法，这里有一个例子：

Show your flowcharts and conceal your tables and I will be mystified. Show your tables and your flowcharts will be obvious.

下面是一些输入的前趋和它们的后继：

前趋	后继
Show your	flowcharts tables
your flowcharts	and will
flowcharts and	conceal
flowcharts will	be
your tables	and and
will be	mystified. obvious.
be mystified.	Show

如果选择了 Show your 作为文本的开头，那么它的后继有两个：flowcharts 和 tables，从中随机选取一个。假设选取了 flowcharts 作为后继，那么前趋就变为 your flowcharts，这样不断重复，直到没有任何后继或是产生了足够多的词。

你的任务是完成这个算法的最核心部分，即对于一系列给定的前趋，分别输出它们所有可能的后继。

【输入文件】

样本文件名为 sample.txt，它包括了不超过 10000 个的单词。这些单词将是随机生成的或是一段随机选取的文章，也就是说你可以确保单词选取的随机性。词与词

之间用一个空格分开。注意：标点符号也被看成是单词的一部分。样本中将不包括除了空格之外的任何分隔符。文件大小不超过 200k。

输入文件名为 `markov.in`，第一行包括了一个整数 `n`，表示给定的前趋数目， $1 \leq n \leq 2,000,000$ 。接下来的 `n` 行每行包括了两个单词，表示给定的前趋，词与词之间用一个空格分开。每个词语的长度不超过 255。

【输出文件】

输出文件名为 `markov.out`，它共有 `n` 行，第 `i` 行包括了第 `i` 个前趋的所有后继，所有后继之间用一个空格分开。如果该前趋在样本中不存在或找不到任何后继，则输出一个空行。如果某前趋的一个后继在文章中出现了不止一次，则输出的次数等于出现的次数。（这样在随机选取的时候可以使该词被选中的频率较高）

【输入输出样例】

Sample.txt

and my heart dances to the distant music of the ancient xylophone
--

Markov.in

1 the ancient

Markov.out

xylophone
