

CSE 101

数据结构和算法介绍编程作业8

在这个项目中，你将重新创建pa7的字典ADT，但现在是基于红黑树。红黑树在文本的第13章中有所涉及，并将在课堂上详细讨论。所有关于红黑树（和BST）的相关算法都发布在网页的例子/伪代码下。除了用RBT作为其底层数据结构外，你的Dictionary ADT的界面将只有轻微的变化。本项目推荐的方法是直接从pa7中复制Dictionary.cpp并进行必要的修改，但如果你觉得有必要，也可以从头开始。头文件Dictionary.h被贴在Examples/pa8中。它与pa7的头文件最显著的区别是一个新的int类型的Node字段，叫做color。除此之外，唯一的区别是新增了一个RBT辅助函数的部分。尽管这些函数被列为可选的，你可以随心所欲地进行修改，但你应该认为它们对这个项目来说是绝对必要的。

你将在本作业中创建两个顶级客户。第一个将被称为Order.c，其规格与pa7中相同。没有必要对该项目进行任何修改。例子/pa8中同样给出了五对输入-输出文件，以及一个随机输入文件生成器。请注意，输入文件与pa7的文件相同，但成对的输出却不同。特别是，输出文件中给出预排序遍历中的所有键的部分是不同的，因为现在树是由RBT算法平衡的。

第二个顶级客户端将被称为WordFrequency.cpp。它将读入文件的每一行，解析每一行中的单个单词，将每个单词转换为所有小写字母，然后将其（作为键）放入一个字典。但是，输入文件中的个别单词可能会重复出现。一个给定的词被遇到的次数（其频率）也将被储存在字典中（作为值）。因此，当你的程序读入单词时，它应该首先使用contains()检查该单词（key）是否已经存在。如果它是一个新词，就用setValue()把它加进去。如果它已经存在，通过调用getValue()来增加相应的值。回顾一下，getValue()函数返回一个值的引用，然后可以用来改变这个值。使用发布在/Examples/C++/FileIO中的例子FileIO.cpp作为WordFrequency.cpp的起点，因为你需要的大部分东西已经在那里了。FileIO.cpp程序包含一个名为delim的字符串变量，它被初始化为一个空格。

```
string delim = " "。
```

这是字符串函数find_first_of()和find_first_not_of()所使用的分隔符，用于确定哪些字符属于标记，哪些不属于标记。因此，FileIO.cpp在空格周围对文件进行标记。你的程序WordFrequency.cpp将围绕一个更大的字符集进行标记。我们文件中的词将是只包含字母字符的子串。为了达到这个目的，你可以按以下方式重置delim。

```
string delim = "\\t\\\"\\'<.>/? ;:[{}]|`~!@#$%^&*()-_+=0123456789" 。
```

因此，为了解析输入文件，要删除所有的空白、标点符号和特殊字符。剩下的就是要放在字典里的词，以及它们的频率。

一旦输入文件中的所有单词连同它们的频率被放入字典，你的程序WordFrequency.cpp将把字典打印到输出文件中。文件夹/Examples/pa8/包含两个非常大的文本文件，名为Shakespeare（包含Willi

am Shakespeare的全部作品) 和Gutenberg (包含Project Gutenberg提供的几个英语文本)。还包括相应的输出文件Shakespeare-out和Gutenberg-out。用这些来测试WordFrequency.cpp。

另外，和以前一样，在 `Examples/pa8` 中发布了一个名为 `DictionaryClient.cpp` 的测试客户端。这个程序与 `pa7` 版本相似，但它有不同的输出，你可以在 `DictionaryClient-out` 中找到它。你仍然应该把它看作是对 `Dictionary` ADT 的一个弱测试，并一如既往地设计你自己的测试。

总的来说，这应该是一个简单的任务，特别是如果 `pa7` 对你来说很顺利。在宽限期结束前，以常规方式提交以下7个文件。

`README` Written by you, a catalog of submitted files and any notes to grader
`Makefile` Provided, alter as you see fit.
`Dictionary.h` 提供的, 你可以改变 "辅助函数" 部分, 但不包括其他内容
`Dictionary.cpp` 由你来写, 这个项目的大部分工作都是由你来完成的。
`DictionaryTest.cpp` 由你编写, 是你的字典测试客户端
`ADT Order.`
`cpp` 由你编写, 是本项目的客户端, 与 `pa7` 没有变化
`WordFrequency.cpp` 由你编写, 是本项目的客户端

像往常一样，不要上交任何可执行文件、二进制文件或上面没有列出的任何东西。祝你好运。