SE117, Spring 2016

Lab 3: Huffman Encoding

Assigned: May 30

Due: July 3, 23:59

Introduction

Huffman encoding is an example of a lossless compression algorithm that works particularly well on text and, in fact, can be applied to any type of file. It can reduce the storage required by a third or half or even more in some situations.

You are to write a program that allows the user to compress and decompress files using the standard Huffman algorithm for encoding and decoding.

Your program should compress a file by executing

```
    ./Compression  inputfilename  outputfilename
```

and decompress a file by executing

```
    ./Compression  -d  inputfilename  outputfilename
```

where `Compression` is your executable file produced by your code.

Requirement & Hint

1. You are required to complete this lab on Linux. Here is a simple guideline to make you get familiar with Linux and get this lab down. (为了说得更清楚开始用中文了 orz) Ubuntu 是对初学者比较友好的 Linux 发行版，如果你是初次接触 Linux 系统，建议使用 Ubuntu。你们可以安装双系统，或者使用虚拟机，课程 QQ 群提供一个已经安装好 Ubuntu 的虚拟机下载。

    Ubuntu 开机后，首先你需要将 lab3 工作目录拷贝到 Ubuntu 系统某位置，如果你对 shell

命令不熟悉，可以使用 Ubuntu 自带的文件管理器将 lab3 目录拷贝到你所期望的位置，操作方式类似于 Windows 资源管理器。

第二步，你需要打开一个 shell，在 Ubuntu 搜索页面中搜索 terminator 或终端以打开 shell。当你打开 shell 时，一般默认状况下你处在/home/username 目录，其中 username 是你当前登录用户名。使用 *cd directoryname* 命令进入名字为 *directoryname* 的目录，其中 *directoryname* 可以为绝对路径或者相对路径，使用 *cd ..* 返回上一级目录。现在你需要进入 lab3 目录，如果你不确定 lab3 文件夹目录位置，你在文件管理器中右键点击 lab3 目录，点击属性/properties 查看该文件夹路径，例如路径为/home/username/lab3，在 shell 中输入 cd /home/username/lab3 以进入该目录（文件名前有/表示绝对路径），进入该目录后，你可以使用 *ls* 命令查看当前目录下的文件和目录，输入 cd Compression 进入 Compression 目录（此处为相对路径），进入后输入 ls，你会发现一个名为 Compression.cc 的文件，在此次 lab 中，你只需要修改该文件即可。

第三步，在 Compression.cc 中进行代码编写。首先你需要一个文本编辑器，建议使用 vim，输入 sudo apt-get install vim 并键入密码进行安装。输入 vim Compression.cc 以通过 vim 编辑 Compression.cc，进入 vim 后按 i 键进入插入模式，现在你可以进行代码编写了，编写完成后，按 Esc 键，输入:w 进行保存，输入:q 退出，或者:wq 保存并退出。

第四步，在输入 Compression 目录输入 make 进行编译，请注意错误提示，如果编译成功，改目录下会生成名为 Compression 的可执行文件，输入./Compression + 参数进行压缩解压操作。

Finally，将目录切换至 Test，输入 make 编译测试文件，再输入./score 运行测试脚本，如果你是满分的话，congratulations!

2. Compressing a file will require reading through the file twice: first to count the characters, and then again when processing each character as part of writing the compressed output.

When writing the bit patterns to the compressed file, note that you do not write the ASCII characters '0' and '1' (that wouldn't do much for compression!), instead the bits in the compressed form are written one-by-one.

You are responsible for freeing memory. Make sure not to leak any tree nodes, and if you allocate any extra memory, be sure to deallocate it!

Grading

Your implementation will be evaluated using score in Test folder. You can evaluate your implementation by yourself. Try "./score " to evaluate your program. For trace 04 ~ 09, your compression ratio should be lower than 100%.

Hand-in

You should work in `Compression` folder. You may add or modify files in this folder. Only keep your `Makefile` can produce executable file named `Compression` correctly when we type `make` in this folder. You should compress your `Compression` folder marked with your student ID via following command:

```
tar czvf studentID+name.tar Compression
```

For example, if your student ID is 51323333 and your name is 张三, the command should be:

```
tar czvf 51323333+张三.tar Compression
```

Hand in `studentID+name.tar` to our course website.

TA-魏晓乐 and TA-李文泰 are responsible for this assignment. Enjoy your lab!