

LZW 文档说明

陈李锋 14 通信 2014081025

2017 年 4 月 30 日

操作系统和编译器说明

- 操作系统: ubuntu 16.04lts
- 编译器: gcc
- 输入文件文件类型: html

lzw 压缩算法简单说明

重要对象说明:

- 字符 (character): 最基础的数据元素, 在文本文件中就是一个字节
- 字符串 (string): 由几个连续的字符组成
- 码长 (codeLength): 取 12 位
- 字典容量 (dictionarySize): 取 2 的 12 次幂 4096 个
- 编码压缩 (compress): 编码依据是先将数据的个别单一字符 ascii 码先创建成一个字符串编码表, 再分别给予编号, 在随后的编 (解) 码过程, 字符串编码表会随着字符串键入而逐渐扩大
- 解码译码 (decompress): 解码依据是将压缩数据与原先字符串编码表对照, 并将对应的字符放于一个暂存队列中, 依序将压缩数据读入, 若为重复数据保存于队列中, 若不为重复数据, 则扩充一个新的码置于字符串编码表中。
- 文档格式: 需要输入的是 html 英文文档, 压缩后的为.lzw 格式文件

操作

编译: gcc lzw.c -o lzw

压缩文件: ./lzw c test.html

解压文件: ./lzw d test.pdf.lzw

结果演示

压缩：把 501k 大小的 test.html 压缩成 112k test.html.lzw

```

→ LZW-master gcc lzw.c -o lzw
→ LZW-master ./lzw c test.html
→ LZW-master ll
总用量 793K
-rwxrwxrwx 1 ferris ferris 3.2K 4月 30 17:03 algorithms.c
-rwxrwxrwx 1 ferris ferris 629 4月 30 17:03 array.c
-rwxrwxrwx 1 ferris ferris 2.5K 4月 30 17:14 dictionary.c
-rwxrwxrwx 1 ferris ferris 1.2K 4月 30 17:03 file.c
-rwxrwxrwx 1 ferris ferris 14K 4月 30 17:15 lzw
-rwxrwxrwx 1 ferris ferris 57 4月 30 17:00 lzw-2014081025.aux
-rwxrwxrwx 1 ferris ferris 51K 4月 30 17:00 lzw-2014081025.log
-rwxrwxrwx 1 ferris ferris 61K 4月 30 17:00 lzw-2014081025.pdf
-rwxrwxrwx 1 ferris ferris 4.6K 4月 30 17:00 lzw-2014081025.synctex.gz
-rwxrwxrwx 1 ferris ferris 1.8K 4月 30 17:00 lzw-2014081025.tex
-rwxrwxrwx 1 ferris ferris 1.9K 4月 30 17:03 lzw.c
-rwxrwxrwx 1 ferris ferris 134 4月 30 17:03 README
-rwxrwxrwx 1 ferris ferris 501K 4月 29 23:37 test.html
-rwxrwxrwx 1 ferris ferris 122K 4月 30 17:15 test.html.lzw

```

test.html 原始 501k
压缩后 122k

解压：由 112k test.html.lzw 解压成原来的 501k 大小的 test.html

```

→ LZW-master rm test.html
→ LZW-master ll
总用量 369K
-rwxrwxrwx 1 ferris ferris 3.2K 4月 30 17:03 algorithms.c
-rwxrwxrwx 1 ferris ferris 629 4月 30 17:03 array.c
-rwxrwxrwx 1 ferris ferris 2.5K 4月 30 17:14 dictionary.c
-rwxrwxrwx 1 ferris ferris 1.2K 4月 30 17:03 file.c
-rwxrwxrwx 1 ferris ferris 14K 4月 30 17:15 lzw
-rwxrwxrwx 1 ferris ferris 57 4月 30 17:00 lzw-2014081025.aux
-rwxrwxrwx 1 ferris ferris 51K 4月 30 17:00 lzw-2014081025.log
-rwxrwxrwx 1 ferris ferris 61K 4月 30 17:00 lzw-2014081025.pdf
-rwxrwxrwx 1 ferris ferris 4.6K 4月 30 17:00 lzw-2014081025.synctex.gz
-rwxrwxrwx 1 ferris ferris 1.8K 4月 30 17:00 lzw-2014081025.tex
-rwxrwxrwx 1 ferris ferris 1.9K 4月 30 17:03 lzw.c
-rwxrwxrwx 1 ferris ferris 134 4月 30 17:03 README
-rwxrwxrwx 1 ferris ferris 122K 4月 30 17:15 test.html.lzw
-rwxrwxrwx 1 ferris ferris 79K 4月 30 17:21 选区_004.png

```

删除原来html文件

```

→ LZW-master ./lzw d test.html.lzw
→ LZW-master ll
总用量 949K
-rwxrwxrwx 1 ferris ferris 3.2K 4月 30 17:03 algorithms.c
-rwxrwxrwx 1 ferris ferris 629 4月 30 17:03 array.c
-rwxrwxrwx 1 ferris ferris 2.5K 4月 30 17:14 dictionary.c
-rwxrwxrwx 1 ferris ferris 1.2K 4月 30 17:03 file.c
-rwxrwxrwx 1 ferris ferris 14K 4月 30 17:15 lzw
-rwxrwxrwx 1 ferris ferris 57 4月 30 17:00 lzw-2014081025.aux
-rwxrwxrwx 1 ferris ferris 51K 4月 30 17:00 lzw-2014081025.log
-rwxrwxrwx 1 ferris ferris 61K 4月 30 17:00 lzw-2014081025.pdf
-rwxrwxrwx 1 ferris ferris 4.6K 4月 30 17:00 lzw-2014081025.synctex.gz
-rwxrwxrwx 1 ferris ferris 1.8K 4月 30 17:00 lzw-2014081025.tex
-rwxrwxrwx 1 ferris ferris 1.9K 4月 30 17:03 lzw.c
-rwxrwxrwx 1 ferris ferris 134 4月 30 17:03 README
-rwxrwxrwx 1 ferris ferris 501K 4月 30 17:26 test.html
-rwxrwxrwx 1 ferris ferris 122K 4月 30 17:15 test.html.lzw
-rwxrwxrwx 1 ferris ferris 79K 4月 30 17:21 选区_004.png
-rwxrwxrwx 1 ferris ferris 74K 4月 30 17:25 选区_005.png

```

无损失解压原来文件
大小还是501k

由此可以看出，代码实现了压缩的效果，并且压缩效果也不错