

上機習題 #8 製作 Huffman 壓縮軟體

題目：以 Huffman 演算法製作一個檔案壓縮與解壓縮之軟體

說明：

資料壓縮 (data compression) 是減少資料儲存空間或資料傳輸量的一種方式，其應用相當廣泛。資料壓縮可分成兩大種類，有失真壓縮與無失真壓縮。無失真壓縮方法所壓縮的資料經過解壓縮 (uncompress) 後，還原的資料與壓縮前完全一樣，常用於檔案壓縮。有失真壓縮方法的資料解壓縮 (uncompress) 後，還原的資料與壓縮前不完全一樣，有些許差異，但影響不大，常用於多媒體資料。

Huffman 演算法為無失真壓縮方法，其方法詳見於本課程課本。本題可以直接引用課本的 Huffman 程式、min heap 程式，不算抄襲。

進行檔案壓縮的步驟：

- Step 1：以檔案中，一個位元組(byte，即一個字元)為一個單位，計算各個字元出現的頻率。
- Step 2：利用 Huffman 演算法求得編碼表(encoding table)。
- Step 3：將編碼表儲存於檔案頭部(file header)。
- Step 4：利用編碼表將原始檔案每個字元加以編碼，並儲存於檔案中(放置於 file header 之後)。
- Step 5：計算壓縮率(compression ratio，減少的空間比率)。並將原始檔案位元組數、壓縮後檔案位元組數(含 file header)、壓縮率儲存於 file header。

Huffman tree 特別規定：

為使每人所製作之 Huffman tree 與 Huffman code 均相同，特別規定如下：

- 規則 1：製作 Huffman tree 時，左子樹為 0、右子樹為 1。
- 規則 2：每次合併兩個 node 時，請將字典順序較小的 node 設為左子樹，字典順序較大者設為右子樹(比較他們最小的符號，即可知道順序)。
- 規則 3：若超過兩個 node 出現相同頻率時，則先合併字典順序上最小的兩個 node。

解壓縮時，利用檔頭的編碼表，逐一將每個字元還原即可。本次習題以文字模式操作，需同時包含檔案壓縮、檔案解壓縮兩種功能。須具備正確開檔、讀檔、寫檔功能，並將 Huffman encoding table 輸出至文字檔案。

操作範例：

huffman -c -i infile -o outfile

(進行壓縮，infile 為輸入檔案，outfile 為輸出檔案)

huffman -u -i infile -o outfile

(進行解壓縮，infile 為輸入檔案，outfile 為輸出檔案)

注意事項：

1. 程式執行時，需將檔頭資料列印於標準輸出，包含原始檔案位元組數、壓縮後檔案位元組數、壓縮率、編碼表。
2. 本次習題，需要由命令列讀入資料。可經由 main(int argc, char* argv[])設定參數來讀入命令列字串。
3. 可能會有非文字檔案(例如圖片、影片等檔案)輸入，為使程式能適用於任何格式檔案，請以 binary 模式開檔。

測試範例(左為輸入檔案，右為編碼表)：

ABCABCADC	A=00 B=010 C=1 D=011
DGHFDFGCJBCHAJDGAGCGBI HJCACDHFJCCIJHDGBCJDBJE HIEFADGBAGDCAGBFFEHEGF GJDHEFCFEHEEDAHIIGIIEDE JCAGIJCGGEJFAEIHBBHCHCCG BAGBFJEJAJBHHBBFJHHGHD GFGDJEIBCJDJAIIFAJGDIFGBB FJIEIBADAEEEEHGDDBHFGJCB HIFHEBIFJHFDIIDBIJGEDDDIG AEIIIJHHGEDADAJCFEAFJEGJ CCEHHFEIBCIJDGIACBAFBBA IFAGEGCFIGCIEHCEAGCJJAIB	A=000 B=1000 C=010 D=1010 E=011 F=1001 G=110 H=1011 I=001 J=111

DBBADEADJBJGJDDIAFGGEA AEGCGHFGJIHCICJJGACHGBD CBFJJBEJBAHFIHAEIAECJGBA ECCCAFFCJACIDIAEAJBJGCFE EJJDGAFA CJEDFBHEDBEGJEC CGEBCIIJCIIIGIDIDDDIAEHGIJ AGIHJADDDHDCGFCGFIHJGA EBAEIHAIGCEHJDJCIDABHIJB FEJCFHEJJEJDFHAIBJJHICFD EJACABJGCB CAHDBBJAFGH	
---	--

作業繳交方式：

1. 至網站上傳程式相關內容。
2. 親自至助教處進行現場 demo。
3. 除上述範例外，demo 時助教將會另外準備測試檔案。
4. 若有做成視窗操作方式，將可獲得額外加分。