**上機習題 #7 (binary tree之應用)**

題目：以Huffman 演算法製作一個檔案壓縮與解壓縮視窗軟體

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，其方法與程式碼詳見於本課程課本。

**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(可以引用課本的程式碼，但不可引用網路或他人的程式碼)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。

Step 5：計算壓縮率（compression ratio）。並將原始檔案位元組數、壓縮後檔案位元組數（含檔頭部份）、壓縮率儲存於檔頭。

Huffman code特別規定：

為使每人所製作之Huffman code 均相同，特別規定如下：

規則1：製作Huffman code時，左子樹為0、右子樹為1。

規則2：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 (比較他們最小的符號，即可知道順序)。

規則3：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。

解壓縮時，利用檔頭的編碼表，逐一將每個字元還原即可。

**此壓縮與解壓縮軟體應設計成一般視窗操作方式。**

**作業繳交方式：**

第一次：以文字模式，完成檔案壓縮、檔案解壓縮兩種完整功能(包含上述所有功能)。須具備正確開檔及讀檔功能，並將Huffman table輸出至文字檔案。

第二次：將上述功能視窗化。需親自至助教處進行現場Demo(時間將另行公佈)。

**測試範例（左為輸入，右為Huffman table輸出）：**

|  |  |
| --- | --- |
| BACCABACABAA | A=0  B=10  C=11 |
| DGHFDFGCJBCHAJDGAGCGBIHJCACDHFJCCIJHDGBCJDBJEHIEFADGBAGDCAGBFFEHGFGJDHEFCFEHEEDAHIGIIEDBEJCAGIJCGGEJFAEIHBHCHCCGBAGBFJEJAJBHHBBFJHHGHDGFGDJEIBCJDJAIIFAJGDIFGBBFJIEIBADAEEEEHGDBHFJGCBHIFHEBIFJHFDIIDBIJGEDDDIGAEIIIJHHGEDADAJCFEAFJEGJCCEHHFEIBCIJDGIACBAFBBAIFAGEGCFIGCIEHCEAGCJJAIBDBBADEADJBJGJDDIAFGGEAAEGCGHFGJIHCICJJGACHGBDCBFJJBEJBAHFIHAEIAECJGBAECCCAFFCJACIDIAEAJBJGCFEEJJDGAFACJEDFBHEDBEGJECCGEBCIIJCIIIGIDIDDDIAEHGIJAGIHJADDDHDCGFCGFIHJGAEBAEIHAIGCEHJDJCIDABHIJBFEJCFHEJJEFJDFHAIBJJHICFDEJACABJGCBCAHDBBJAFGH | A=010  B=1101  C=000  D=1111  E=001  F=1100  G=100  H=1110  I=011  J=101 |

* 測試時會有非文字檔案 (Ex: 圖片、影片)，為使程式能適用於任何格式檔案，請以binary模式開檔。
* 除上述範例外，demo時助教將會另外準備測試檔案。