

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN  
KHOA CÔNG NGHỆ THÔNG TIN



CẤU TRÚC DỮ LIỆU VÀ GIẢI THUẬT:  
**BÁO CÁO BÀI TẬP P2**

GIẢNG VIÊN: NGUYỄN TRI TUẤN

**LỚP 16CTT3**

1612426      Hoàng Nghĩa

*Tp. Hồ Chí Minh, ngày 17 tháng 12 năm 2017*

1) Mô tả, giải thích rõ ý nghĩa của cấu trúc file nén đã cài đặt

- Đây là cấu trúc dùng để lưu trữ thông tin của 1 file con nằm trong folder

```
}struct fileNen{                                // Cấu trúc lưu trữ dữ liệu của 1 file nén
    unsigned char doDaiTenFile;
    string tenFile;
    int sizeTruocNen;                          // dung lượng file trước nén
    int sizeSauNen;                            // dung lượng file sau khi nén
    int bitBatDau;                             // vị trí bắt đầu đọc bitEnCoding để giải nén
    unsigned char soBitCuoiCoNghia;           // số bits có nghĩa trong bytes cuối
    string DataNen_file;                       // chuỗi string bits dùng để giải nén
};
```

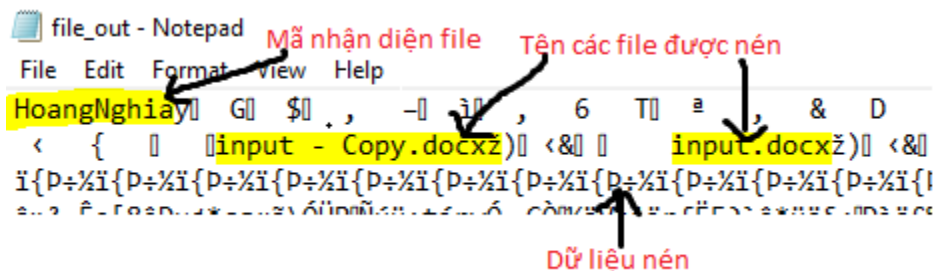
- Đây là cấu trúc lưu trữ thông tin nén của 1 folder

```

struct headerFolder{
    string maNhanDangFile; // em chọn mã nhận dạng file là chuỗi "HoangNghia"
    int bangTanSo[256]; // lưu tần số xuất hiện của 256 kí tự => tạo cây huffman
    int soLuongFile; // số lượng file có trong folder
    /* tiếp theo là từng cấu trúc nén của file con, gồm có : */
    unsigned char doDaiTenFile; // độ dài tên file
    string tenFile; // tên file được nén
    int sizeTruocNen; // dung lượng file trước khi nén
    int sizeSauNen; // dung lượng file sau khi nén;
    unsigned char soBitCuoiCoNghia; // phục vụ cho việc giải nén
    /* tiếp theo là những vùng nhớ tương tự cho những file kế tiếp */
    //.....
    string DataNen_Folder; // được tạo thành bằng cách cộng dồn các dataNen_File
};

```

- Hình ảnh minh họa 1 file nén folder



## 2) Mô tả các cấu trúc dữ liệu quan trọng khác trong chương trình, bao gồm nhưng không giới hạn bởi các cấu trúc dữ liệu của thuật toán Huffman.

- Các cấu trúc dữ liệu quan trọng là :

```
struct NODE {
    unsigned char  c;      // ký tự
    int    freq;    // số lần xuất hiện
    bool   used;    // đánh dấu node đã xử lý chưa
    int    nLeft;   // chỉ số nút con bên trái
    int    nRight;  // chỉ số nút con bên phải
};

struct MABIT {
    char*  bits;
    int    soBit;
};

const int MAX_NODE = 256 * 9;
const int MAX_BIT_LEN = 10000;
NODE    huffTree[MAX_NODE];
MABIT   bangMaBit[256];
```

- + Cây Huffman lưu trữ dưới dạng mảng tĩnh.
- + Bảng mã bit của 256 ký tự

## 3) Mô tả các hàm chính của chương trình

// Tên hàm, ý nghĩa, mô tả ngắn về ý tưởng thuật toán  
// Mô tả input  
// Mô tả output

// nén

- Lập bảng thống kê tần số xuất hiện của các ký tự.
- Xây dựng cây Huffman dựa vào bảng thống kê tần số xuất hiện.
- Phát sinh bảng mã bit cho từng ký tự tương ứng.
- Duyệt tập tin, thay thế các ký tự trong tập tin bằng mã bit tương ứng.
- Lưu lại thông tin của cây Huffman cho giải nén (lưu lại bảng tần số xuất hiện của các ký tự)


// giải nén

- Load dữ liệu từ file nén vào bộ nhớ trong
- Từ bảng tần số, tạo lại cây, phát sinh mã bit
- Từ dữ liệu nén, bắt đầu duyệt cây phát sinh kí tự tương ứng
- Cộng dồn các kí tự rồi xuất ra tập tin giải nén

\_ Các hàm chính như:

- `void` ThongKeTanSoXuatHien(`string` tenFile) {}
- `int` TaoCayHuffman()
- `void` PhatSinhMaBit(`int` nRoot)
- `void` MaHoa1KyTu(`unsigned char` c, `unsigned char` &out, `unsigned char` &viTriBit, `string` &DuLieuNen)
- `void` Nen\_Folder(`string` folder\_in, `string` file\_out)
- `bool` GiaiNen\_Folder(`string` file\_out, `string` folder\_out)
- `bool` XemThongTinFileNen(`string` file\_nen)
- `bool` GiaiNen\_TheoFile(`string` file\_nen, `string` folder\_out)

\_ inPut và outPut:

 D:\OneDrive - VNU-HCMUS\P2\1612426\_Project\_2\Release\1612426\_Project\_2.exe

```

  *==*==*==*==*==*==*==*==*==*==*==*==* Huffman Compression *==*==*==*==*==*==*==*==*==*==*==*==*
Menu:
1. Nen cac file trong folder.
2. Xem noi dung file da nen.
3. Giai nen tat ca cac file trong tap tin nen.
4. Chon va giai nen cac file riêng le.
5. Thoat!
-----
Lua chon cua ban :
```

- ✓ Với yêu cầu “ Nén các file trong folder” :
  - \_ input sẽ là đường dẫn folder cần nén (**folder\_in**\\), đường dẫn + tên file nén (**file\_out**)
  - \_output sẽ là file nén (**file\_out**) chứa tất cả dữ liệu nén của các file con nằm trong folder

```

D:\OneDrive - VNU-HCMUS\P2\1612426_Project_2\Release\1612426_Project_2.exe
*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~* Huffman Compression *~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*
Menu:
1. Nen cac file trong folder.
2. Xem noi dung file da nen.
3. Giai nen tat ca cac file trong tap tin nen.
4. Chon va giai nen cac file rieng le.
5. Thoat!
-----
Lua chon cua ban : 1

Nhap vao folder ban muon nen (folder_in\\) : folder_in\\

Nhap vao duong dan + ten file nen (file_out) : file_out

1 : input - Copy.docx
2 : input.docx
3 : test.txt
4 : Test2.txt

Nen thanh cong !
Press any key to continue .

```

Folder/File	Date/Time	Type	Size
folder_in	12/17/2017 4:30 PM	File folder	
Release	12/17/2017 8:53 PM	File folder	
1612426_Project_2	12/17/2017 12:36 ...	VC++ Project	5 KB
1612426_Project_2.vcxproj	12/17/2017 12:36 ...	VC++ Project Filte...	2 KB
file_out	12/17/2017 9:03 PM	File	149 KB
main.cpp	12/17/2017 3:28 PM	C++ Source	3 KB
Menu	12/17/2017 3:14 PM	C/C++ Header	1 KB
resource	12/17/2017 8:53 PM	C/C++ Header	25 KB

✓ Tương tự như yêu cầu 1, các yêu cầu khác cũng phải nhập vào các đường dẫn folder và file.

```

*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~* Huffman Compression *~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*
Menu:
1. Nen cac file trong folder.
2. Xem noi dung file da nen.
3. Giai nen tat ca cac file trong tap tin nen.
4. Chon va giai nen cac file rieng le.
5. Thoat!
-----
Lua chon cua ban : 4

Nhap vao file nen(file_out) : file_out

Nhap vao folder chua cac file nen : folder_out\\

Dang load data nen ! ....vui long doi giay lat.
STT      Ten File                      Size truoc nen      Size sau nen
1        input - Copy.docx             76190               75403
2        input.docx                    76190               75403
3        test.txt                       34                  29
4        Test2.txt                      0                   0

Moi nhap thu tu cac file can giai nen :
giai nen file : 4
Ban co muon nhap tiep (y/n) : y

giai nen file : 2
Ban co muon nhap tiep (y/n) : n

Dang trong qua trinh giai nen! ....Vui long doi :)

```

```

    **==**==**==**==**==**==**==**==**==**==**==**==**==**==**==**==**==**==**==**==**==**
Menu:
1.  Nén các file trong folder.
2.  Xem nội dung file đã nén.
3.  Giải nén tất cả các file trong tập tin nén.
4.  Chọn và giải nén các file riêng lẻ.
5.  Thoát!
-----
  Lựa chọn của bạn : 3

Nhập vào tập tin cần giải nén(file_out) : file_out

Nhập vào folder chứa các file giải nén(folder_out\\) : folder_out1\\

Đang trong quá trình giải nén ! .... Vui lòng đợi :) .

Ten File                Dung luong truoc nen  Dung luong sau khi giai nen
input - Copy.docx       76190                76190                Checksum complete with no error!
input.docx              76190                76190                Checksum complete with no error!
test.txt                34                  34                  Checksum complete with no error!
Test2.txt               0                   0                   Checksum complete with no error!

Giải nén thành công !

```

❖ *Lưu ý: Đường dẫn folder phải có 2 dấu “\\” ở sau.*