



# Báo cáo Đồ án

## CẤU TRÚC DỮ LIỆU VÀ GIẢI THUẬT 18CTT5C

Kết hợp các kiến thức kỹ thuật lập trình và cấu trúc dữ liệu để xây dựng một ứng dụng nén dữ liệu.

# I. Các thành viên

Nguyễn Xuân Tuấn – 18120635

- Phân công: Nén file, làm menu.

Cao Huy Tùng – 18120639

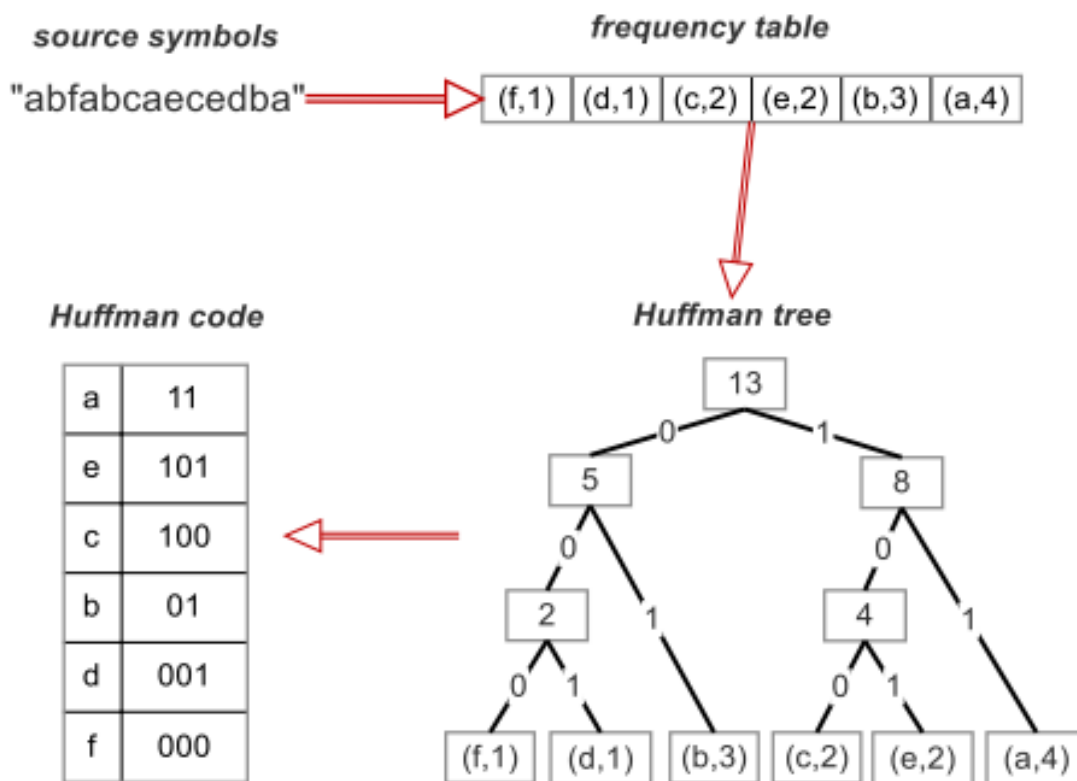
- Phân công: Nén folder, viết báo cáo

## II. Tiến trình đồ án

- Nén 1 tập tin chuỗi
- Nén 1 thư mục chứa nhiều tập tin chuỗi
- Nén 1 tập tin có những kiểu dữ liệu còn lại
- Nén 1 thư mục có chứa nhiều tập

### III. Thuật toán nén Huffman tĩnh

Nén Huffman là phương pháp mã hóa bằng mã có độ dài thay đổi (variable length encoding) trong đó chỉ sử dụng vài bit để biểu diễn 1 ký tự và độ dài mã bit cho các ký tự không giống nhau (ký tự xuất hiện nhiều lần được biểu diễn bằng mã ngắn và ngược lại).



Thuật toán nén Huffman bao gồm 5 bước:

- B1: Thống kê tần số xuất hiện của các ký tự có trong file.
- B2: Xây dựng cây Huffman dựa vào bảng thống kê trên.
- B3: Phát sinh bảng mã bit cho từng ký tự tương ứng.
- B4: Duyệt file, thay thế các ký tự trong tập tin bằng mã bit tương ứng.
- B5: Lưu lại thông tin của cây Huffman cho giải nén.

Để thực hiện việc dựng cây và tạo mã bit, 2 struct được xây dựng như sau:

```
struct NODE{  
    unsigned char c;    //ký tự của node  
    int freq;           //số lần xuất hiện  
    bool used;          //flag để đánh dấu node đã thuộc cây hay chưa  
    int nLeft;          //node trái  
    int nRight;         //node phải  
};
```

Cây được lưu như sau : NODE hufftree[MAX\_NODE]

```
struct MABIT{  
    char* bits;         //mã bit tương ứng của kí tự  
    int soBit;           //độ dài mã bit  
};
```

Bảng mã bit của 256 kí tự ASCII được lưu như sau: MABIT bangMaBit[256]

### 1. Bước 1:

Các kí tự sẽ được duyệt qua và ghi nhận tần số xuất hiện vào mảng hufftree với vị trí tương ứng của nó trong bảng mã ASCII, ví dụ kí tự 'a' có giá trị thập phân là 97 sẽ được lưu tần số tại vị trí hufftree[97]. Vì vậy cần phải tạo ra 256 node để phục vụ việc truy xuất nhanh chóng, đặc biệt là đối với các file có độ dài lớn.

### 2. Bước 2:

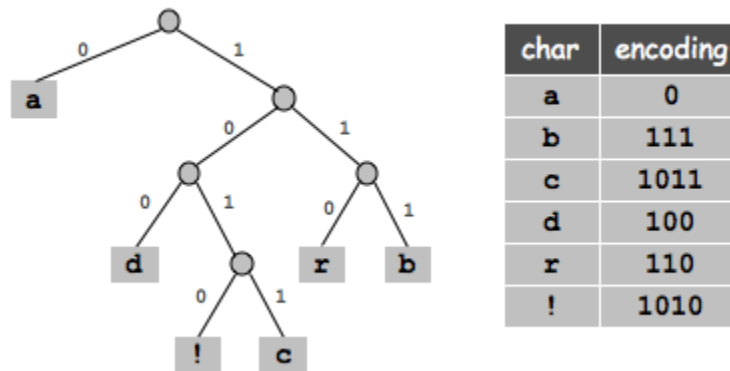
Các bước phát sinh cây:

- B1: Chọn trong bảng thống kê 2 phần tử có tần suất thấp nhất.
- B2: Tạo 2 node của cây cùng với node cha z có trọng số bằng tổng trọng số 2 nút con.
- B3: Loại 2 phần tử x, y khỏi bảng thống kê (bật biến used = true trong node).
- B4: Thêm phần tử z vào bảng thống kê.
- B5: lặp lại bước 1 đến bước 4 cho đến khi còn 1 phần tử trong bảng thống kê.

### 3. Bước 3:

- Node gốc chính là node cuối cùng có được từ bước 2. Duyệt đệ quy từ node gốc, thêm bit 0 khi đi qua nhánh trái và bit 1 khi đi qua nhánh phải, khi chạm node lá (tức nLeft = -1 và nRight = -1) sẽ phát sinh được mã bit của kí tự tại node lá đó.

- Vì đoạn mã kết quả có thể có số bit chưa hoàn chỉnh để tạo thành 1 byte (không chia hết cho 8) nên cần phải bổ sung số bit 0 tương ứng



#### 4. Bước 4 & Bước 5:

- Đầu tiên, cần lưu các thông tin giải mã vào đầu file như: số kí tự khác nhau, bảng tần số, độ dài đoạn mã, số bit được thêm bổ sung (đã nêu ở bước 3).
- Sử dụng 1 biến unsigned char để lưu các mã bit mã hóa của các kí tự tương ứng trong file. Xuất biến đó ra file.
- Đối với nén folder, cần lưu thêm các thông tin như cấu trúc thư mục, số file trong thư mục, vị trí của file trong thư mục.

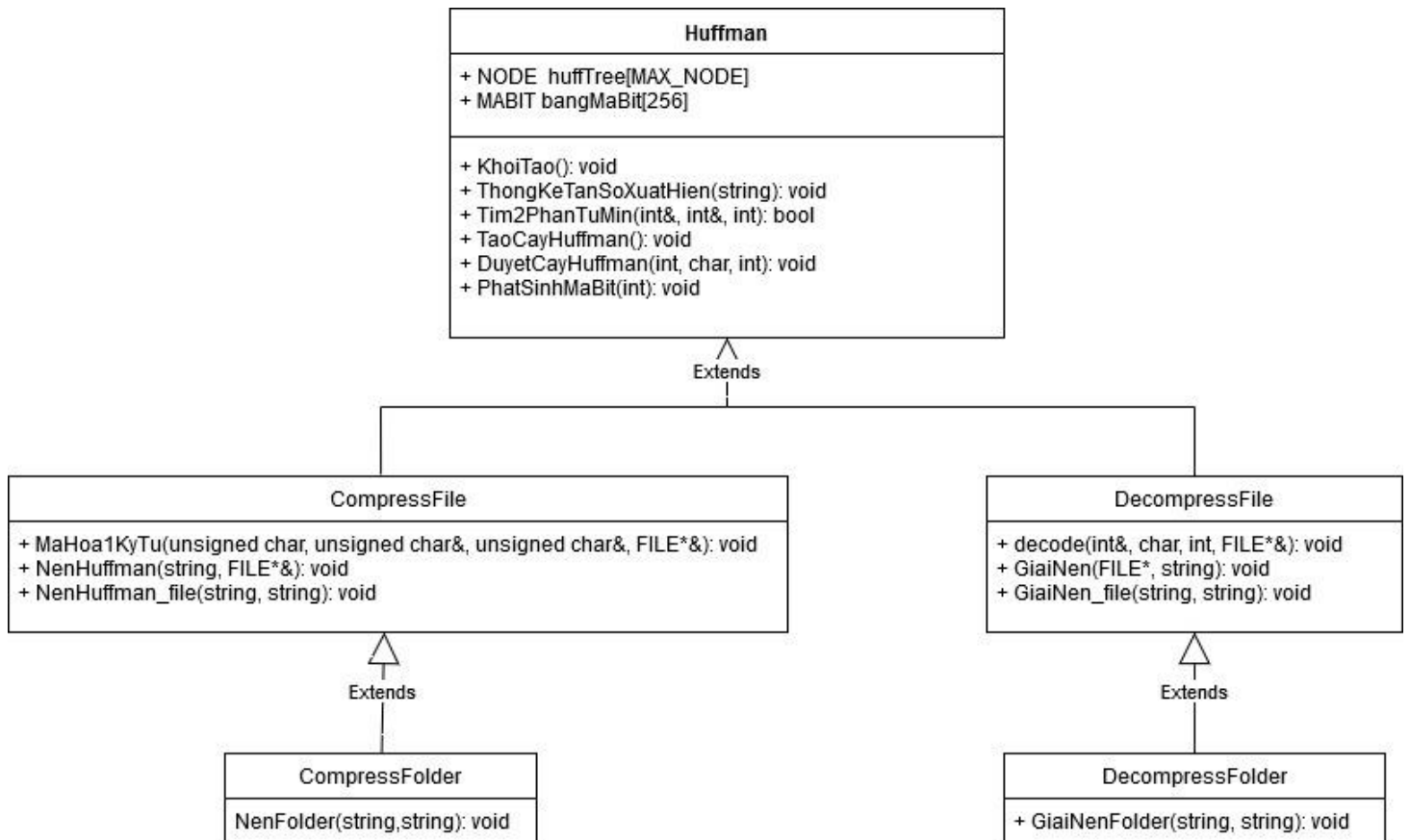
#### 5. Giải nén:

- Đọc các thông tin giải mã và dựng lại cây huffman.
- Bắt đầu node gốc, tiến hành duyệt đoạn mã, khi gặp bit 0 sẽ rẽ qua nhánh phải, qua nhánh trái nếu gặp bit 1.
- Khi chạm node lá sẽ xuất ra kí tự tương ứng tại node đó, đặt lại node gốc để giải mã kí tự tiếp theo.

#### 6. Đánh giá

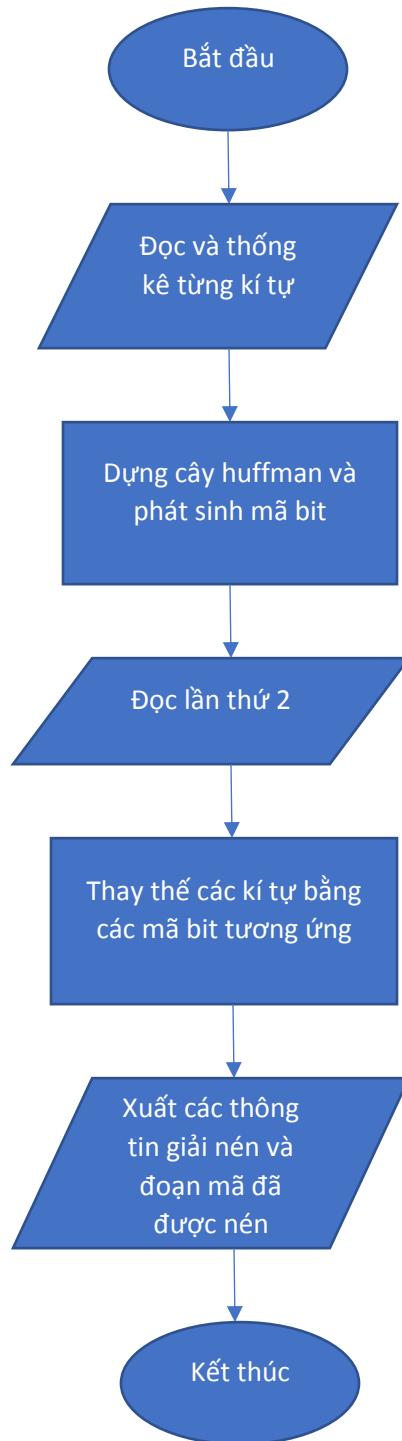
- *Ưu:*
  - Có thể nén được những file, folder lớn
  - Thuật toán dễ thực hiện
  - Đòi hỏi ít bộ nhớ
- *Nhược:*
  - Phải duyệt file 2 lần khi nén
  - Việc lưu trữ thông tin giải mã vào file nén trong một số trường hợp nhất định sẽ làm file nén lớn hơn file gốc
  - Phải dựng lại cây khi giải nén

## IV. Sơ đồ lớp



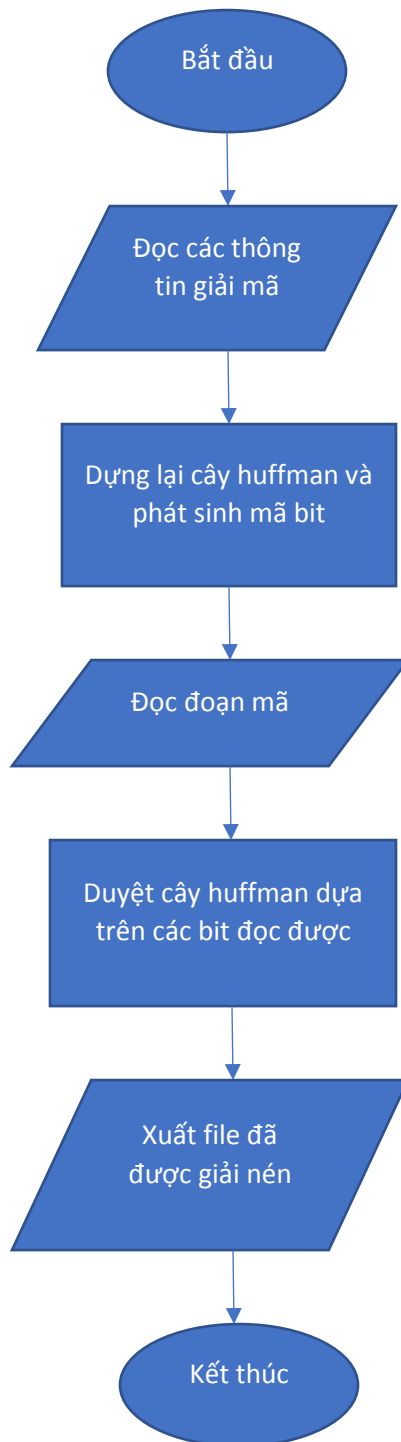
# V. Lưu đồ thuật toán

## 1. Nén Huffman tĩnh:





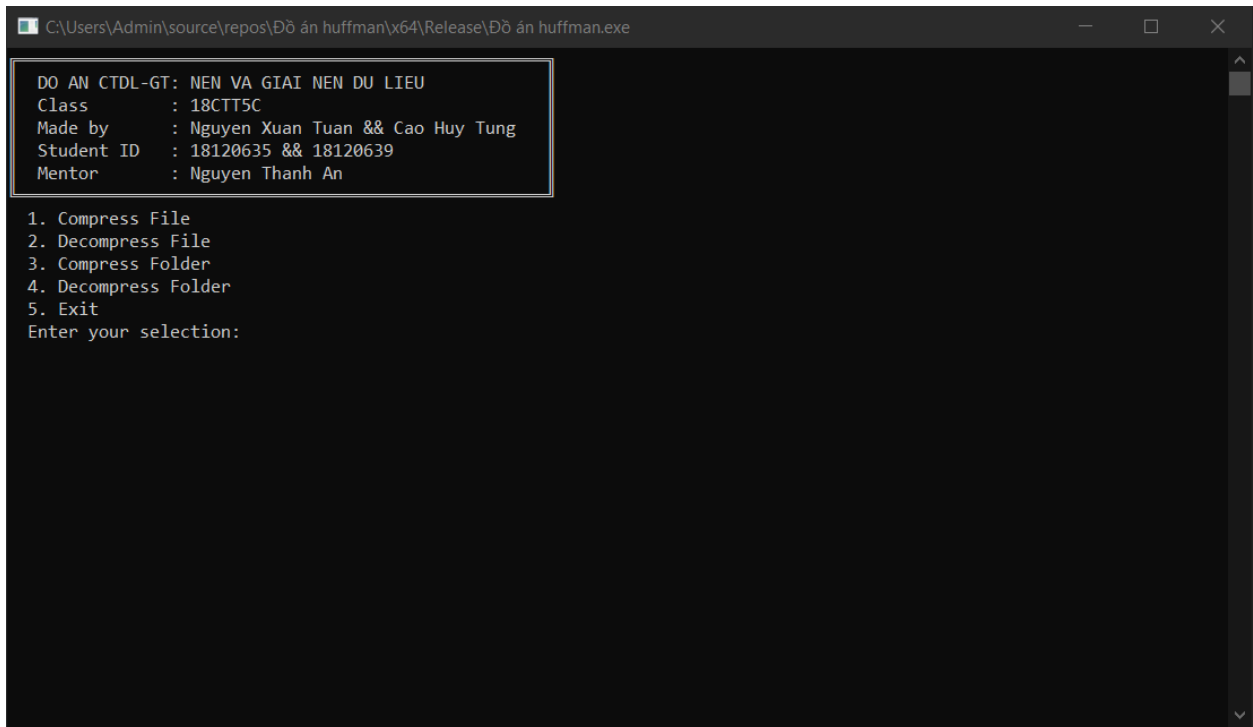
## 2. Giải nén:



## VI. Hướng dẫn sử dụng

Xem video hướng dẫn:

<https://www.youtube.com/watch?v=CIUUBSyHplI&feature=youtu.be>

A screenshot of a Windows command prompt window titled "C:\Users\Admin\source\repos\Đồ án huffman\x64\Release\Đồ án huffman.exe". The window displays the program's header information in a box: "DO AN CTDL-GT: NEN VA GIAI NEN DU LIEU", "Class : 18CTT5C", "Made by : Nguyen Xuan Tuan & Cao Huy Tung", "Student ID : 18120635 & 18120639", and "Mentor : Nguyen Thanh An". Below this, a menu is shown: "1. Compress File", "2. Decompress File", "3. Compress Folder", "4. Decompress Folder", and "5. Exit". The prompt "Enter your selection:" is at the bottom of the menu.

```
C:\Users\Admin\source\repos\Đồ án huffman\x64\Release\Đồ án huffman.exe

DO AN CTDL-GT: NEN VA GIAI NEN DU LIEU
Class      : 18CTT5C
Made by    : Nguyen Xuan Tuan & Cao Huy Tung
Student ID : 18120635 & 18120639
Mentor     : Nguyen Thanh An

1. Compress File
2. Decompress File
3. Compress Folder
4. Decompress Folder
5. Exit
Enter your selection:
```

- Khởi động file exe để tiến hành nén.
- Giao diện cơ bản gồm các chức năng:
  - Nén file
  - Giải nén file
  - Nén folder
  - Giải nén folder
  - Thoát chương trình
- Khi chọn chức năng nén, chương trình sẽ yêu cầu nhập các đường dẫn cho file, folder cần nén và đường dẫn xuất file nén. Đối với giải nén thì nhập file .hf cần giải nén và đường dẫn cho output.

The image displays two screenshots of a Windows command prompt window. The title bar of the window is "C:\Users\Admin\source\repos\Đồ án huffman\x64\Release\Đồ án huffman.exe".

**Top Screenshot:**

```
DO AN CTDL-GT: NEN VA GIAI NEN DU LIEU
Class      : 18CTT5C
Made by    : Nguyen Xuan Tuan && Cao Huy Tung
Student ID : 18120635 && 18120639
Mentor     : Nguyen Thanh An

1. Compress File
2. Decompress File
3. Compress Folder
4. Decompress Folder
5. Exit
Enter your selection: 2
Enter the input file path: D:\\test.hf
Enter the folder path containing the decompressed file: D:

Thời gian giải nén là: 0.03
File ban đầu có dung lượng: 78157 byte
File sau khi giải nén có dung lượng: 129955 byte
Press any key to continue . . .
```

**Bottom Screenshot:**

```
DO AN CTDL-GT: NEN VA GIAI NEN DU LIEU
Class      : 18CTT5C
Made by    : Nguyen Xuan Tuan && Cao Huy Tung
Student ID : 18120635 && 18120639
Mentor     : Nguyen Thanh An

1. Compress File
2. Decompress File
3. Compress Folder
4. Decompress Folder
5. Exit
Enter your selection: 1
Enter the input file path: E:\\test\\test.txt
Enter the folder path containing the compressed file: D:

Thời gian nén là: 0.01
File ban đầu có dung lượng: 129955 byte
File sau khi nén có dung lượng: 77824 byte
Press any key to continue . . .
```

- Khi kết thúc, chương trình cũng sẽ cho biết thời gian nén/giải nén, dung lượng input và output.

## VII. Tham khảo

- Tài liệu Nén Huffman tỉnh HCMUS
- Thư viện dirent.h: <https://github.com/tronkko/dirent>