

## Cây quyết định

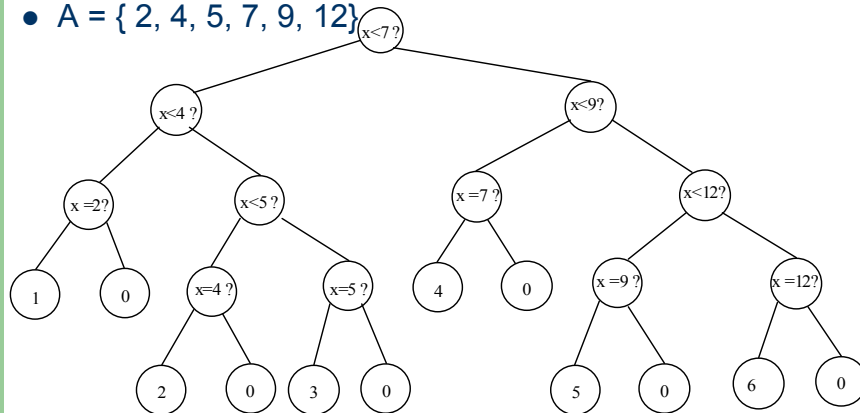
- Cây quyết định là cây:
  - Mỗi nút trong tương ứng với một câu hỏi trên dữ liệu
  - Mỗi nhánh của cây tương ứng với một khả năng trả lời cho một câu hỏi
  - Mỗi nút lá của cây tương ứng với một kết luận
- Khi tính toán trên cây quyết định
  - Xác định một kết luận bằng việc bắt đầu từ nút gốc, xác định một đường đi đến một nút lá
  - Tại các nút trung gian, việc chọn khả năng trả lời với dữ liệu sẽ dẫn ta đến một nút ở mức tiếp theo

## Cây quyết định

- Ví dụ : Bài toán tìm kiếm
  - Cho một mảng A gồm n số được sắp xếp. Xác định xem một giá trị X có xuất hiện trong A không, nếu có cho biết vị trí trong mảng A của X
- Cây quyết định để giải bài toán
  - Là cây nhị phân
  - Mỗi nút trong biểu diễn một câu hỏi dạng  $X < k$  ?
  - Mỗi nút sát lá biểu diễn câu hỏi dạng  $X = A[i]$  ?
    - Mỗi nút lá tương ứng là một giá trị về vị trí i hoặc giá trị 0 thể hiện sự không xuất hiện

## Cây quyết định – Bài toán tìm kiếm

- $A = \{2, 4, 5, 7, 9, 12\}$



## Mã hóa văn bản

- Bài toán mã hóa văn bản :
  - Cho một văn bản D trên bảng chữ cái C. Cần mã hóa văn bản đã cho
    - Mã hóa văn bản với ASCII
    - Mã hóa văn bản với độ dài cố định
    - Mã hóa văn bản với mã phi tiền tố

## Mã hóa với độ dài cố định

Thuộc dạng mã phi tiền tố, không phân biệt chữ hoa chữ thường  
mã hóa sử dụng mã nhị phân độ dài cố định

- Mã hóa một chữ cái sử dụng 5 bit, không phân biệt chữ hoa chữ thường
  - A → 00001
  - B → 00010
  - C → 00011
  - D → 00100
  - E → 00101
  - F → 00110
  - G → 00111
  - ...
- 000110000110100 ↔ CAT

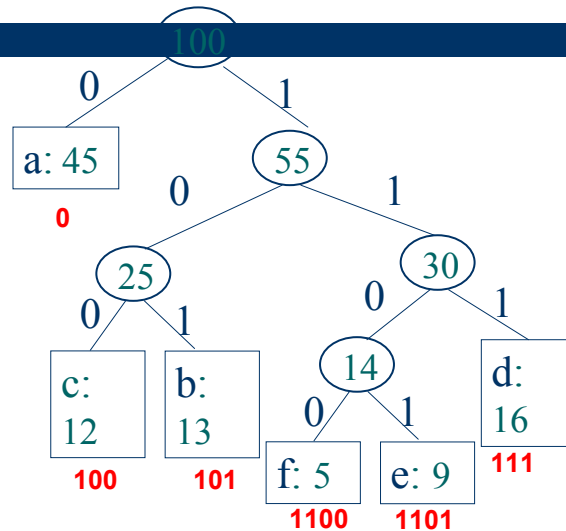
## Mã Huffman

- Thuộc dạng mã phi tiền tố
  - Mã hóa mỗi ký tự sao cho mã của một ký tự bất kỳ không phải là đoạn đầu của mã của một ký tự nào trong số các ký tự còn lại
- Mã hóa các ký tự sử dụng đoạn mã có độ dài khác nhau
  - Ký tự có tần suất xuất hiện lớn hơn thì được mã hóa sử dụng đoạn mã ngắn hơn

## Mã Huffman

- Một mã phi tiền tố được biểu diễn bằng một cây nhị phân
  - Các nút lá biểu diễn các ký tự và tần suất xuất hiện của ký tự trong văn bản
  - Các nút trong có 2 nhánh tương ứng với : 0 – nhánh bên trái và 1 – nhánh bên phải. Các nút này chứa tổng tần suất xuất hiện của các nút trong các nhánh con của nó
  - Một đường đi từ nút gốc đến nút lá chính là một mã nhị phân biểu diễn ký tự ở nút lá
  - Ký tự có tần suất lớn sẽ xuất hiện ở mức thấp hơn, ký tự có tần suất nhỏ sẽ xuất hiện ở mức cao hơn trong cây mã Huffman

## Mã Huffman

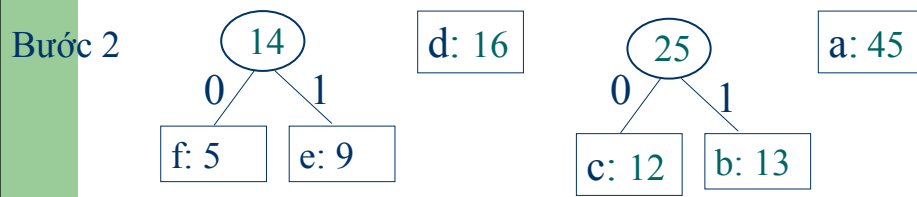
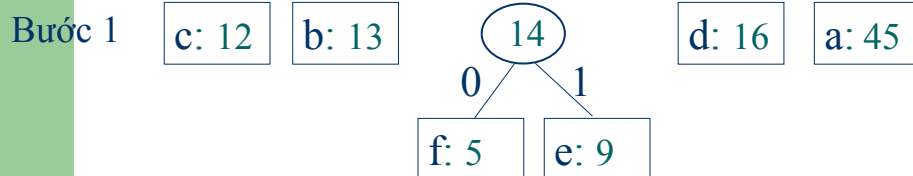


## Thiết lập mã Huffman

- Đầu vào: Bảng chữ cái C. Tần suất xuất hiện của các ký tự trong C
- Đầu ra: Cây mã hóa Huffman
- Ý tưởng:
  - Dựng cây từ dưới lên, xuất phát với các nút lá
  - Tạo lập một nút nhánh (nút trong) bằng cách nhóm hai nhánh đã có sẵn mà tần suất của hai nhánh đó là nhỏ nhất

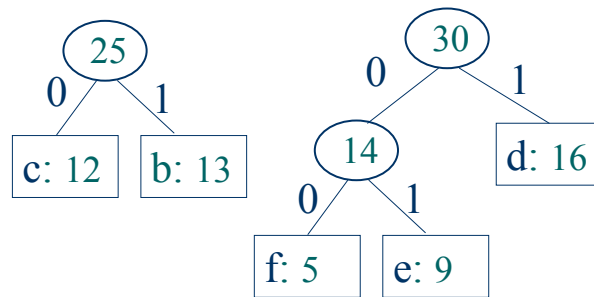
## Thiết lập mã Huffman – Ví dụ

Bắt đầu: f: 5 e: 9 c: 12 b: 13 d: 16 a: 45



## Thiết lập mã Huffman – Ví dụ

Bước 3

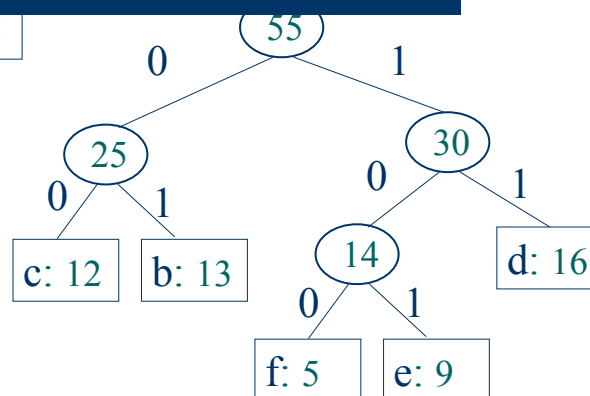


a: 45

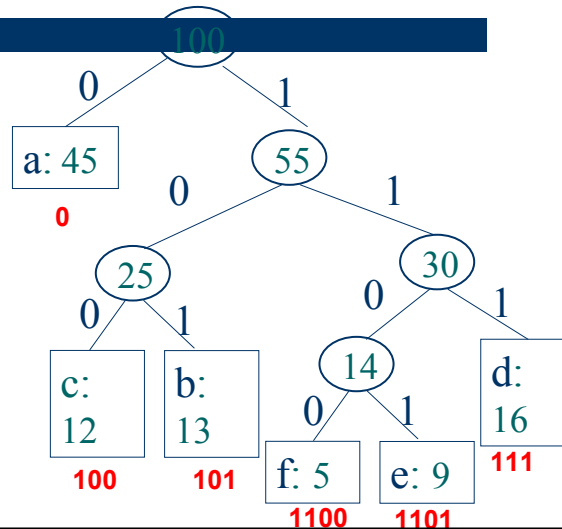
## Thiết lập mã Huffman – Ví dụ

Bước 4

a: 45



## Thiết lập mã Huffman – Ví dụ



## Giải mã Huffman

- Xuất phát từ gốc của cây mã hóa
- Đọc lần lượt từng ký tự trong chuỗi mã hóa
  - Nếu là 0 : đi sang trái
  - Nếu là 1: đi sang phải
- Nếu chạm tới một nút lá ghi ký tự chứa tại nút lá ra, sau đó quay lại nút gốc của cây mã hóa.

## Giải mã Huffman

