CS/CPE 590

Huffman Coding

Kazi Lutful Kabir

Spring 2023

Huffman Codes

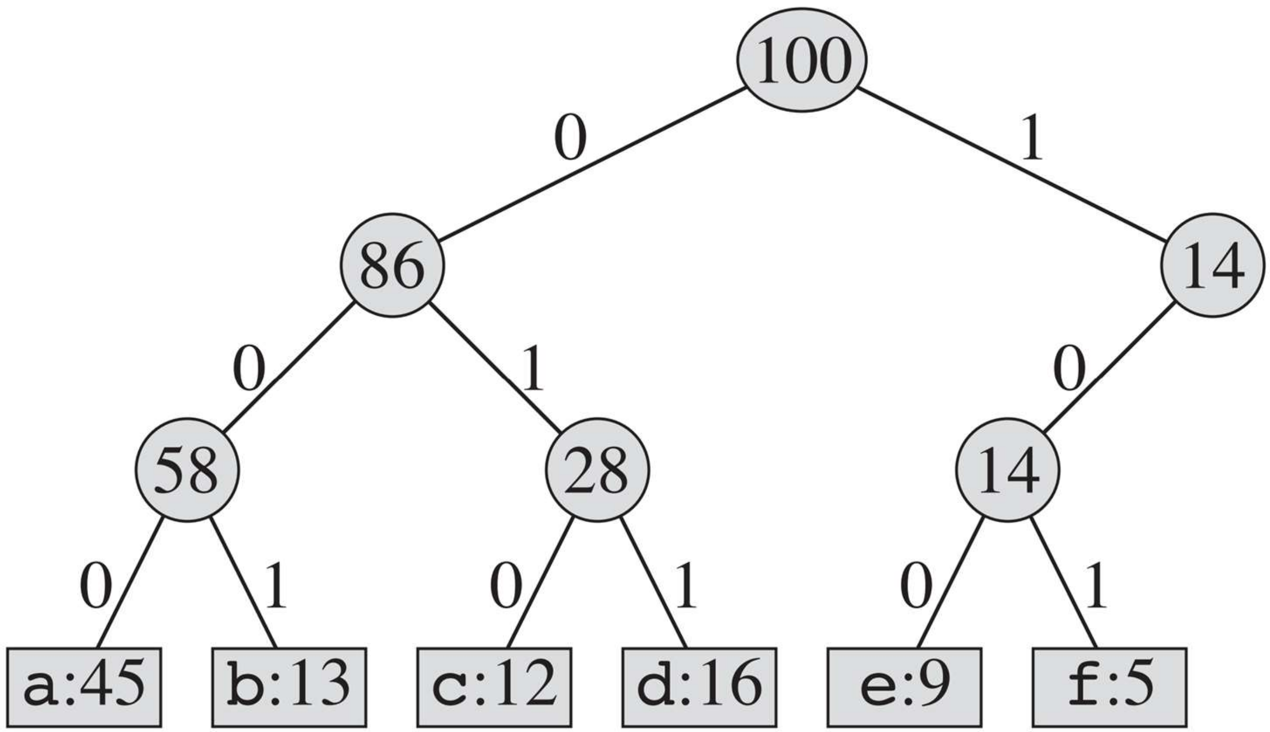
Huffman codes compress data very effectively: savings of 20% to 90% are typical, depending on the characteristics of the data being compressed. We consider the data to be a sequence of characters. Huffman's greedy algorithm uses a table giving how often each character occurs (i.e., its frequency) to build up an optimal way of representing each character as a binary string.

- Suppose we have a 100,000-character data file that we wish to store compactly.
- We observe that the characters in the file occur with the frequencies given by

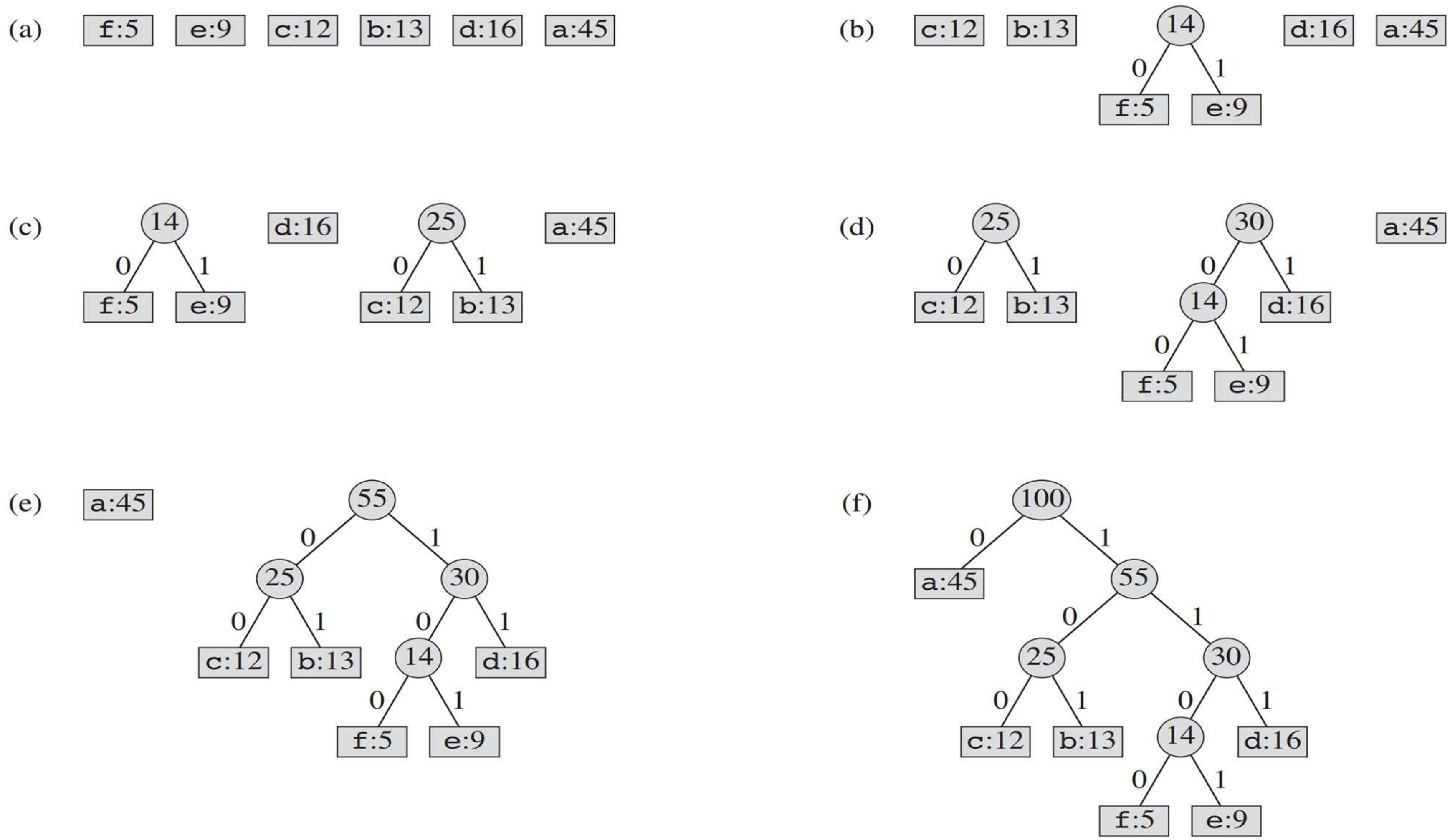
	a	b	c	d	e	f
Frequency (in thousands)	45	13	12	16	9	5

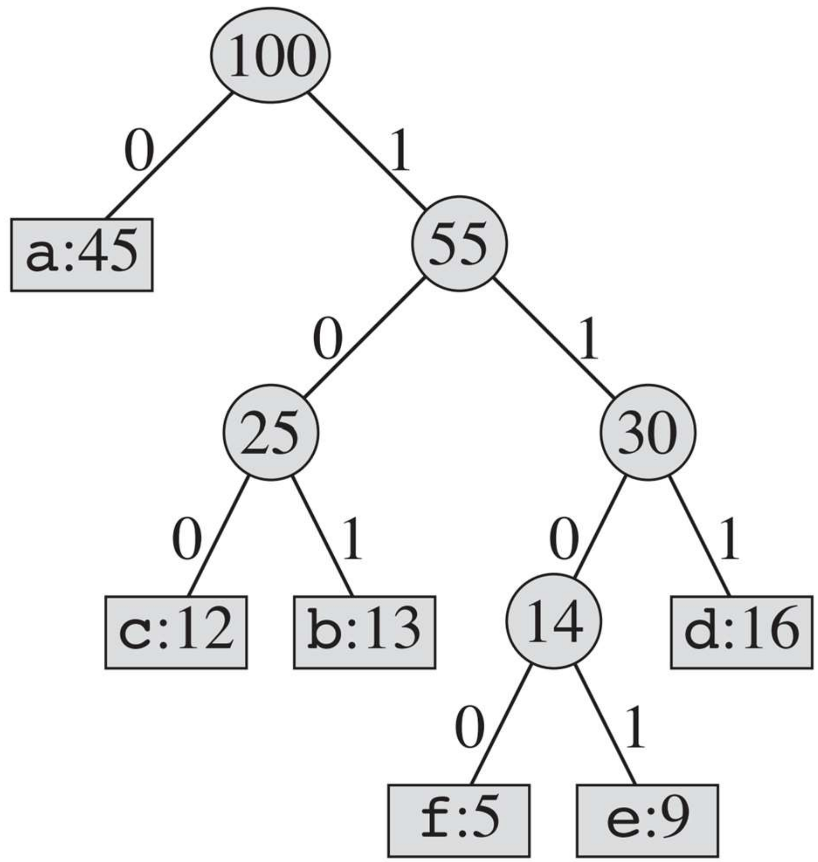
- That is, only 6 different characters appear, and the character "a" occurs 45,000 times.
- We have many options for how to represent such a file of information.
- We consider the problem of designing a binary character code

Fixed-length Coding



Variable-length Coding





String 001011101 parses uniquely as 0 0 101 1101, which decodes to "aabe"

Algorithm

Input: a collection C of objects containing a character and its frequency.
Output: the root of a Huffman tree

```
HUFFMAN(C)
n = |C|
Q = C
for i = 1 to n - 1
    allocate a new node z
    z.left = x = EXTRACT-MIN(Q)
    z.right = y = EXTRACT-MIN(Q)
    z.freq = x.freq + y.freq
    INSERT(Q, z)
return EXTRACT-MIN(Q)      // return the root of the tree
```

Running time: $(O(n \lg n))$

Compression Savings

- 100,000 characters each taking 1 byte = 8 bits
  -- 800,000 bits (without any compression)

- With fixed-length codes, 3 bits for each character
  -- 300,000 bits (62.5% savings)

- Variable-length codes (25.33% savings w.r.t fixed length codes & 72% savings w.r.t when stored as characters)

$$(45 \cdot 1 \; + \; 13 \cdot 3 \; + \; 12 \cdot 3 \; + \; 16 \cdot 3 \; + \; 9 \cdot 4 \; + \; 5 \cdot 4) \cdot 1{,}000 = 224{,}000 \text{ bits}$$

Contents of this presentation are based on Book Chapter- 16, Introduction to Algorithms by *Cormen, Leiserson, Rivest,* & *Stein*

STEVENS
INSTITUTE OF TECHNOLOGY
1870

THANK **YOU**

Stevens Institute of Technology