Lecture 05: Huffman Coding

Swakkhar Shatabda

CSI 227: Algorithms, Summer 2014
Department of Computer Science and Engineering
United International University

United International University
QUEST FOR EXCELLENCE

Huffman Coding

1. Compress data very efficiently
2. Data is typically a text file
3. Saves from 20% to 90%
4. Suppose we are given a text file with 100,000 characters
5. And the alphabet, $\sum = \{a, b, c, d, e, f\}$

	a	b	c	d	e	f
frequency	45k	13k	12k	16k	9k	5k
ASCII	01100001	01100010	01100011	01100100	01100101	01100110
	97	98	99	100	101	102

Total bits needed = 100k*8=800k bits, can we do better?

Fized Length Coding

	a	b	c	d	e	f
frequency	45k	13k	12k	16k	9k	5k
ASCII	01100001	01100010	01100011	01100100	01100101	01100110
	97	98	99	100	101	102
Fixed	000	001	010	011	100	101

For six different characters we need only 3 bits.

Now total bits = 3*100k = 300k bits, we can do even better.

How? variable length coding.

Variable length coding

	a	b	c	d	e	f
frequency	45k	13k	12k	16k	9k	5k
ASCII	01100001	01100010	01100011	01100100	01100101	01100110
	97	98	99	100	101	102
Fixed	000	001	010	011	100	101
Variable	0	101	100	111	1101	1100

Now the total number of bits =

45k*1+13k*3+12k*3+16k*3+9k*4+5k*4 = 224k

It saves! why? Smaller code for most frequent characters

Coding

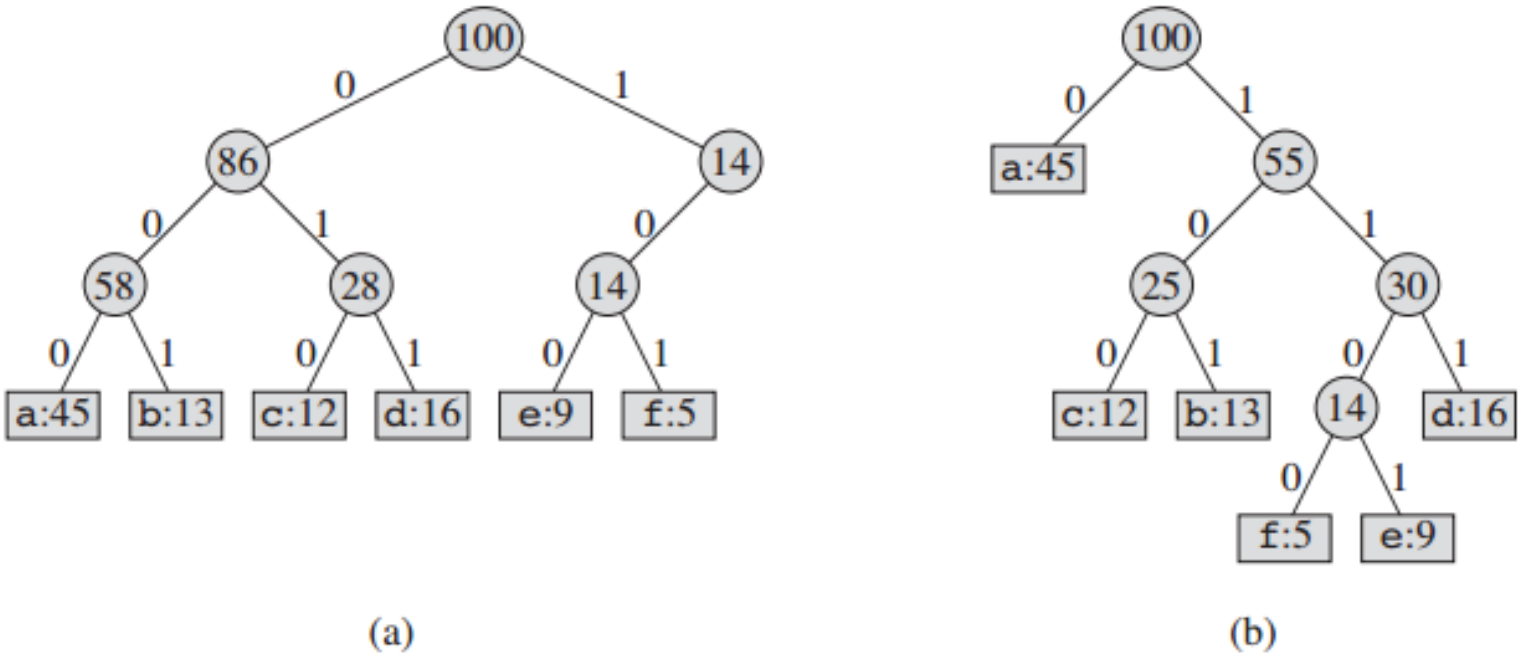
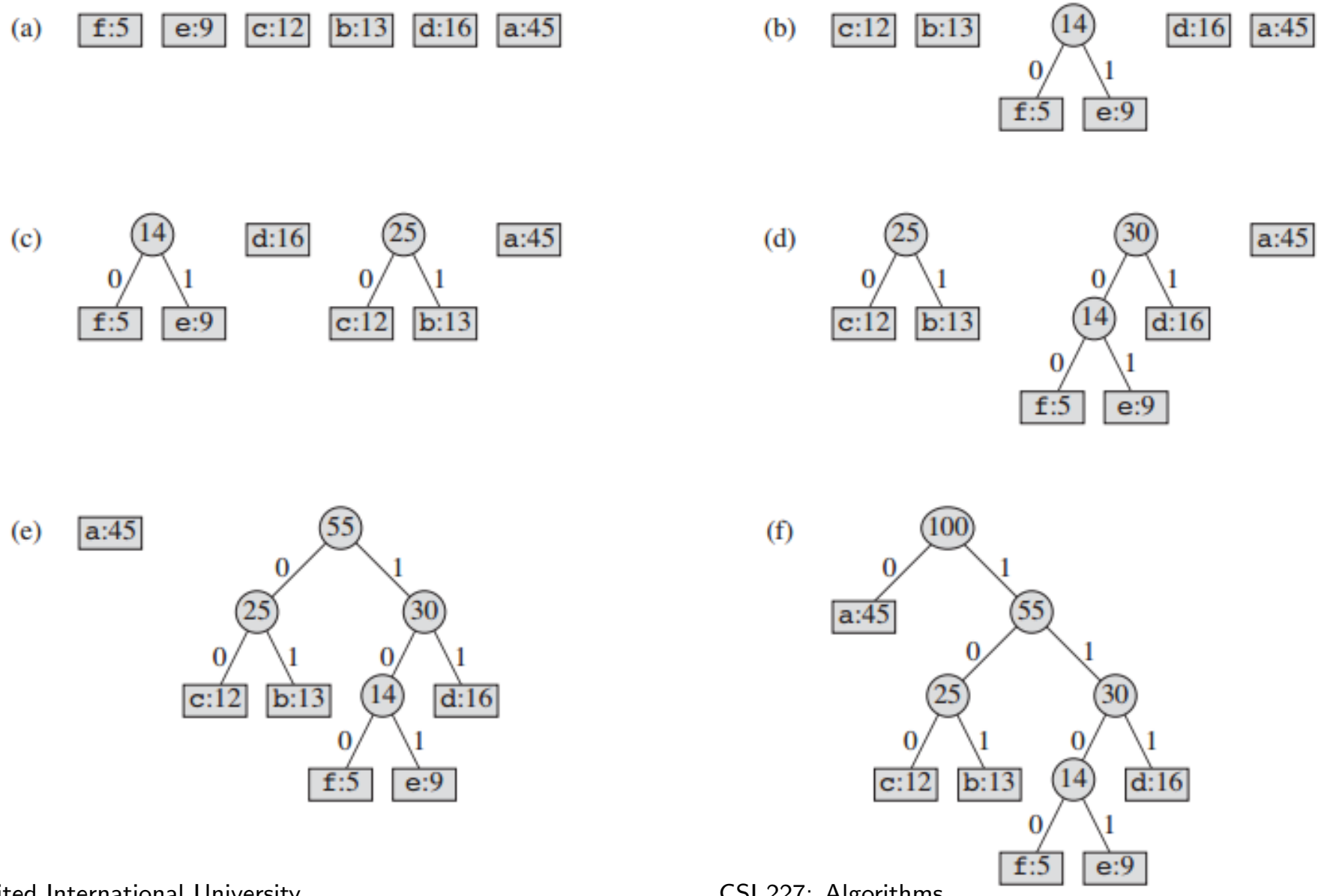


Figure 16.4 Trees corresponding to the coding schemes in Figure 16.3. Each leaf is labeled with a character and its frequency of occurrence. Each internal node is labeled with the sum of the frequencies of the leaves in its subtree. **(a)** The tree corresponding to the fixed-length code a = 000, . . . , f = 101. **(b)** The tree corresponding to the optimal prefix code a = 0, b = 101, . . . , f = 1100.

Huffman Coding

```
HUFFMAN(C)
 1  n = |C|
 2  Q = C
 3  for i = 1 to n − 1
 4        allocate a new node z
 5        x = EXTRACT-MIN(Q)
 6        y = EXTRACT-MIN(Q)
 7        z.left = x
 8        z.right = y
 9        z.freq = x.freq + y.freq
10        INSERT(Q, z)
```

An Example


(a)
f:5
e:9
c:12
b:13
d:16
a:45
(b)
c:12
b:13
14
0
1
f:5
e:9
d:16
a:45
(c)
14
0
1
f:5
e:9
d:16
25
0
1
c:12
b:13
a:45
(d)
25
0
1
c:12
b:13
30
0
1
14
d:16
0
1
f:5
e:9
a:45
(e)
a:45
55
0
1
25
30
0
1
0
1
c:12
b:13
14
d:16
0
1
f:5
e:9
(f)
100
0
1
a:45
55
0
1
25
30
0
1
0
1
c:12
b:13
14
d:16
0
1
f:5
e:9

Reading

Chapter 16 (Pages 428-433)

Thats it!

Thank you