
허프만 코드

Huffman Code

만든이 서정현
작성일 2017.10.08

01. 허프만 코드

- 각 글자의 빈도 수를 기반으로 메시지를 압축을 위해 고안된 특수화된 코드 형태

- 빈도가 높은 정보- 적은 비트 수
- 빈도가 낮은 정보- 많은 비트 수
- 전체 데이터의 표현에 필요한 비트의 양을 줄임
- 가변 길이 코드, 영상 압축에 많이 사용
- 팩스, 모뎀, 컴퓨터 네트워크, 고해상도 텔레비전

문자	코드
a	000
b	001
c	010
d	011
e	100
f	101
g	110
h	111

표 2) 압축을 반영하지 않은 코드

문자 비트 수가 줄어든다.



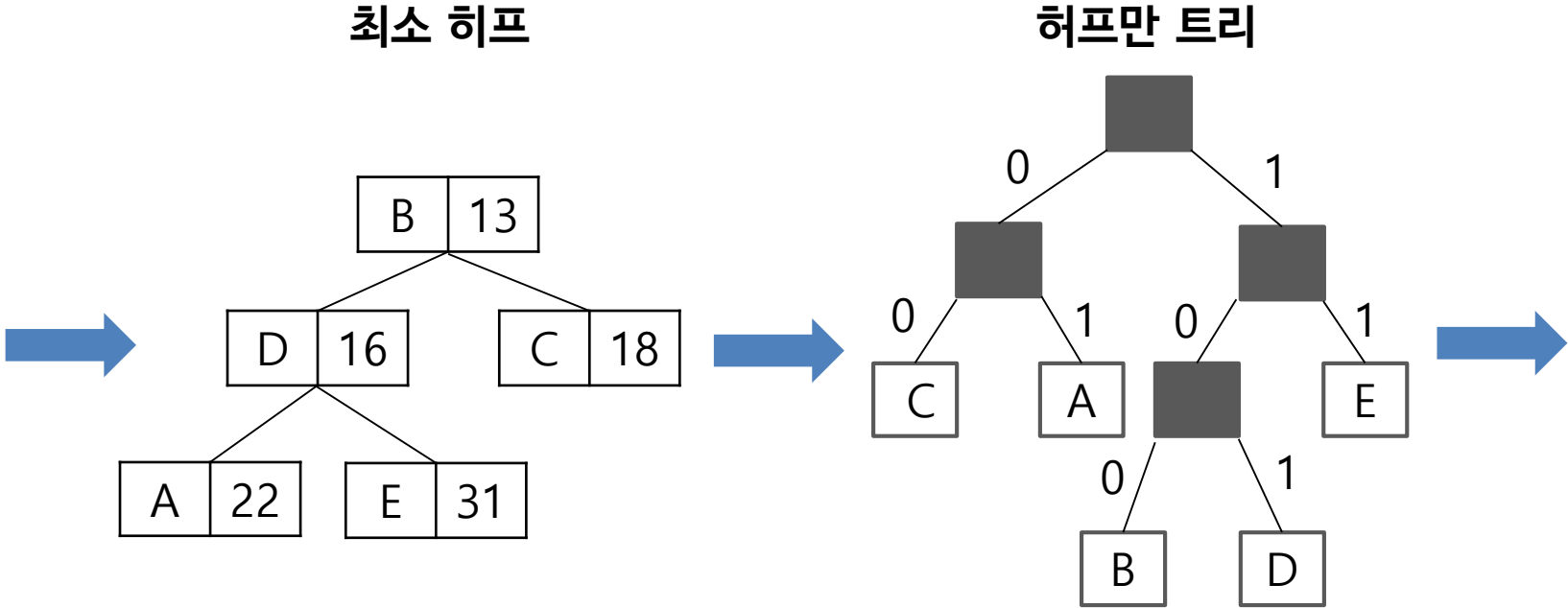
문자	코드
a	11
b	00
c	101
d	010
e	1000
f	1001
g	0110
h	0111

표 3) 허프만 코드

04. 허프만 코드

- 허프만 코드 생성 대략적인 과정

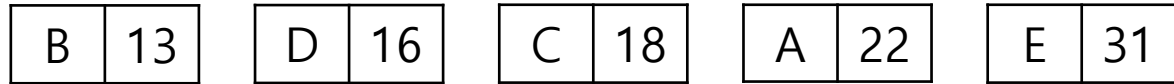
문자	빈도수
A	22%
B	13%
C	18%
D	16%
E	31%



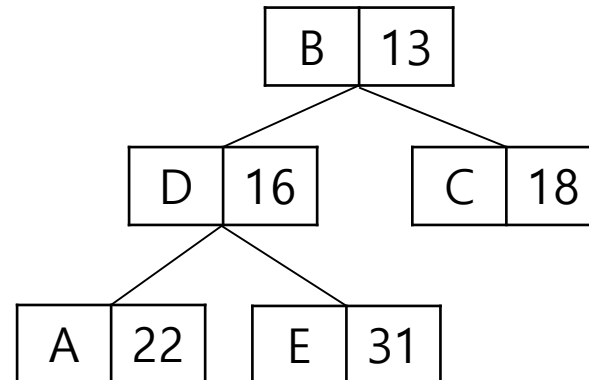
04. 허프만 코드

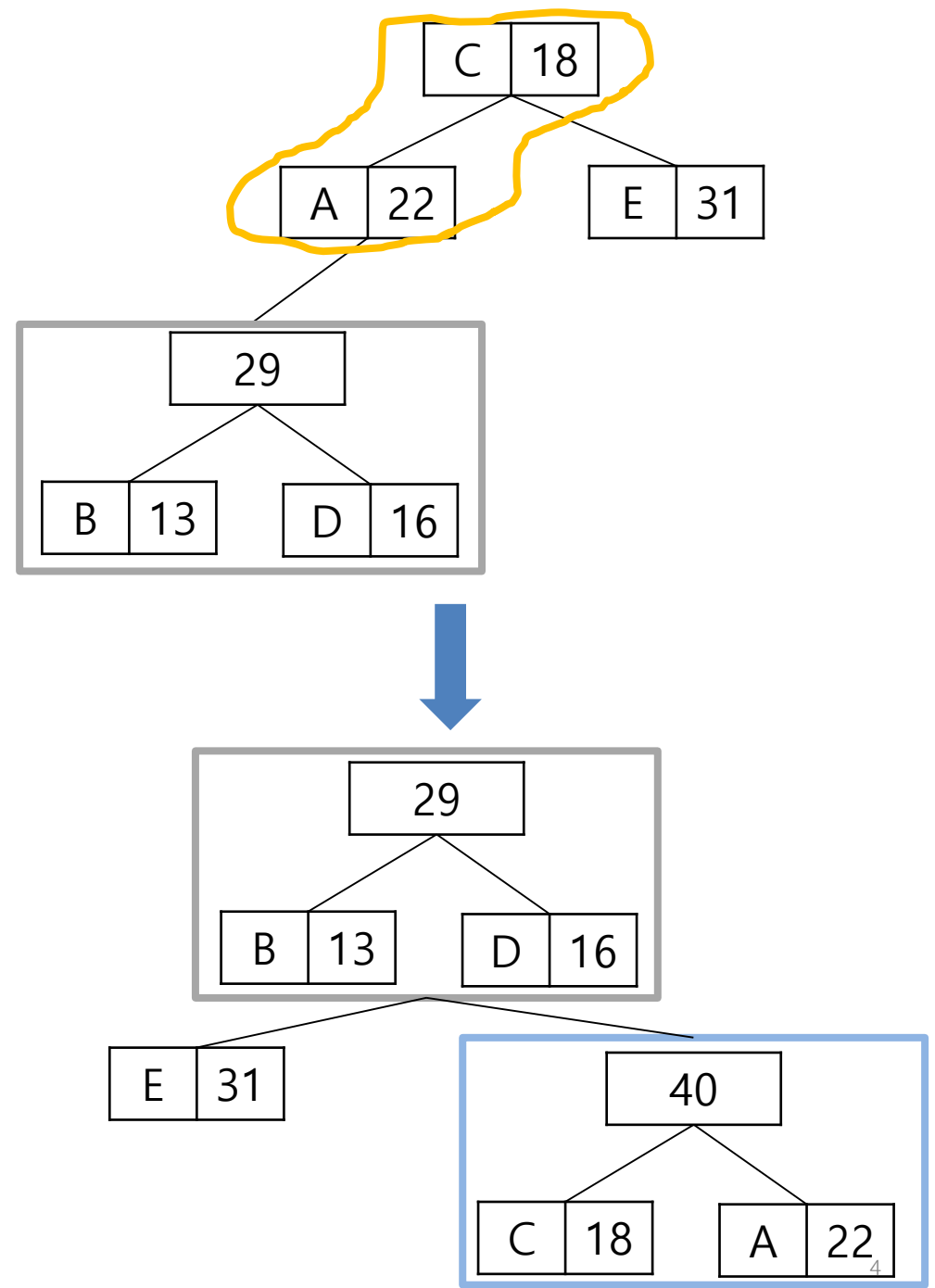
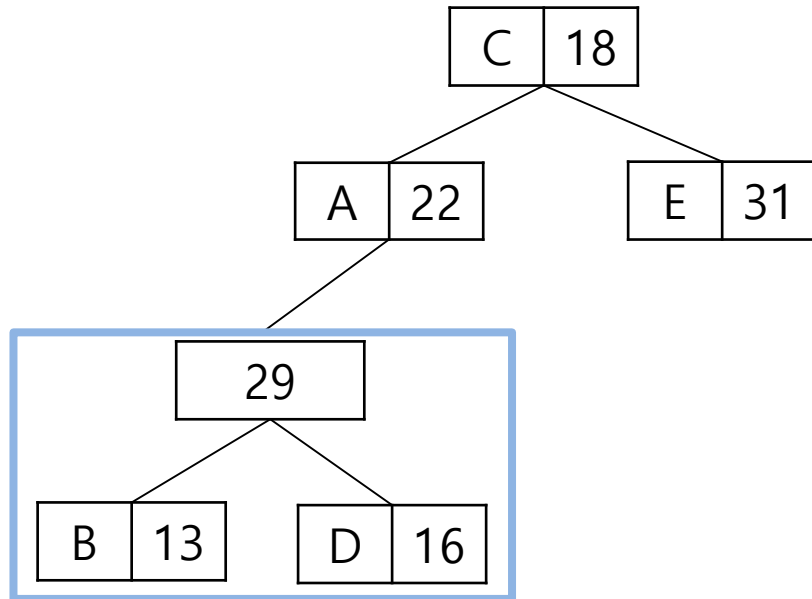
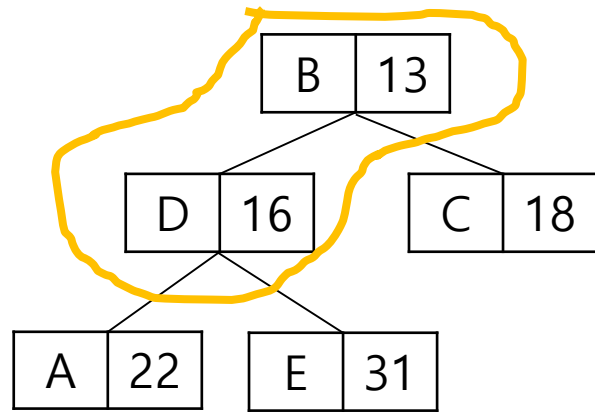
- 허프만 트리 만들기

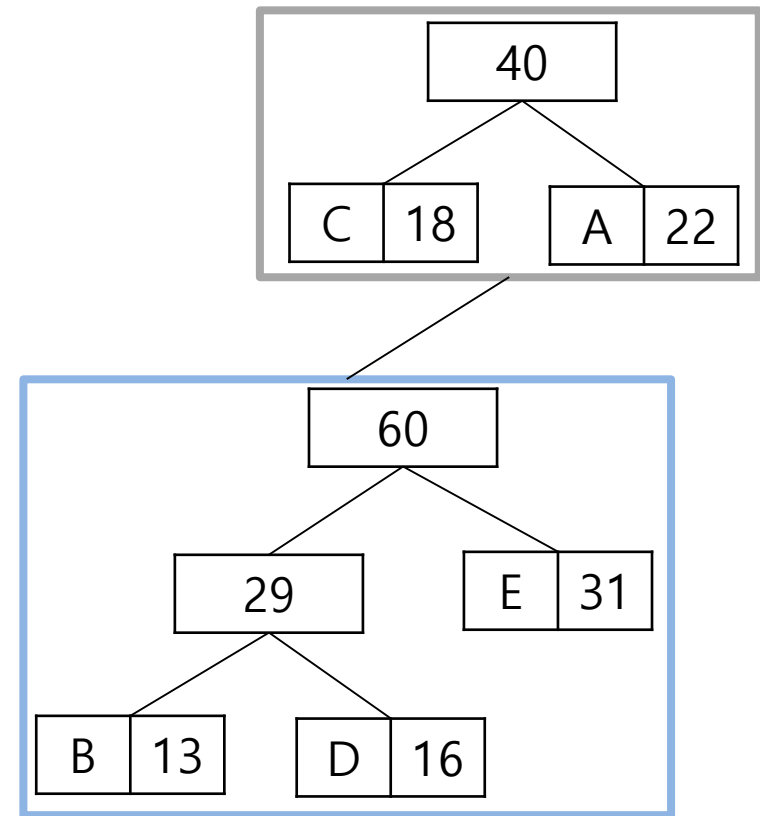
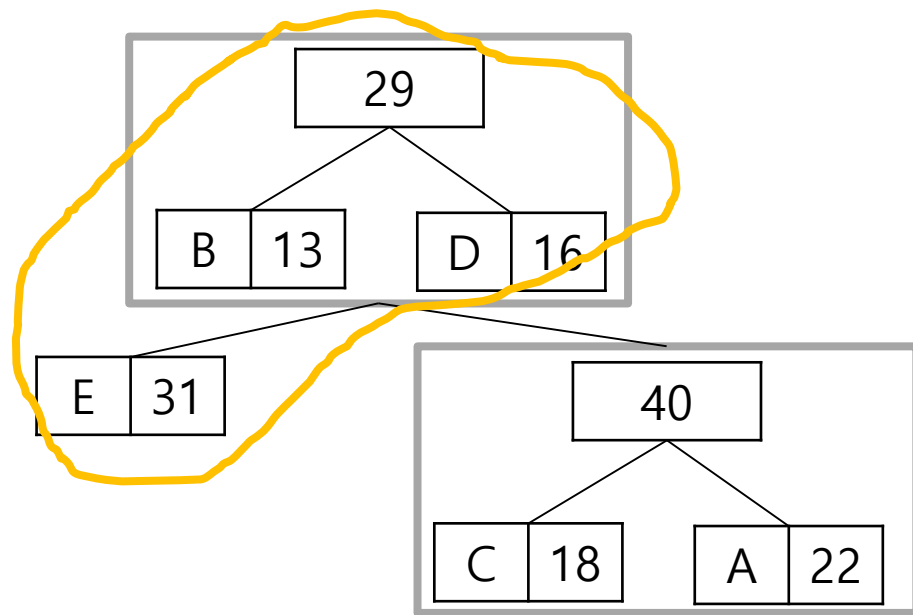
문자	빈도수
A	22%
B	13%
C	18%
D	16%
E	31%

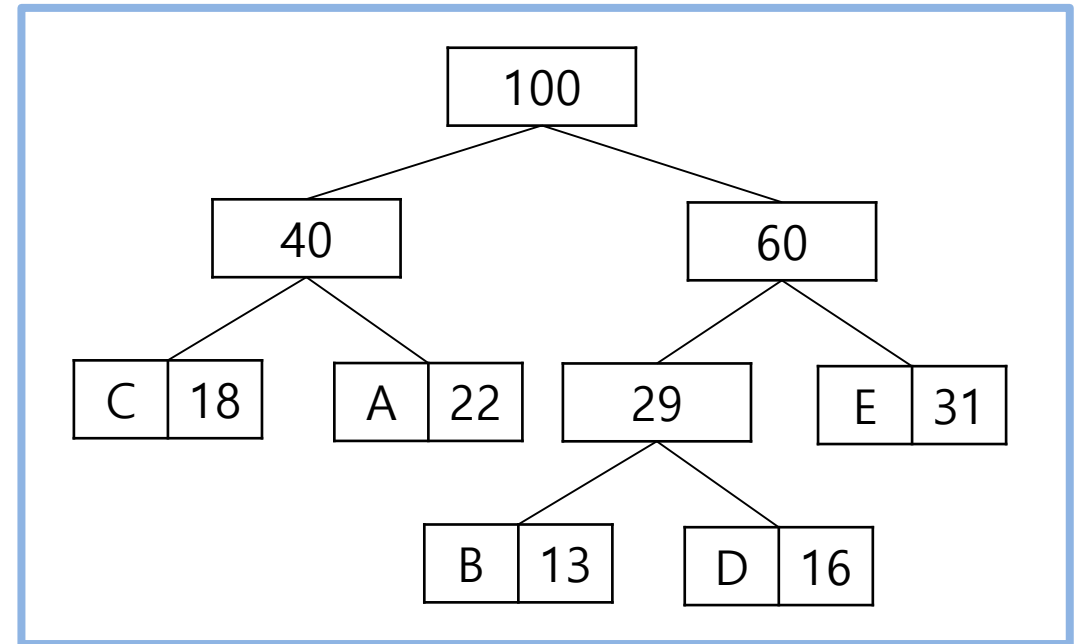
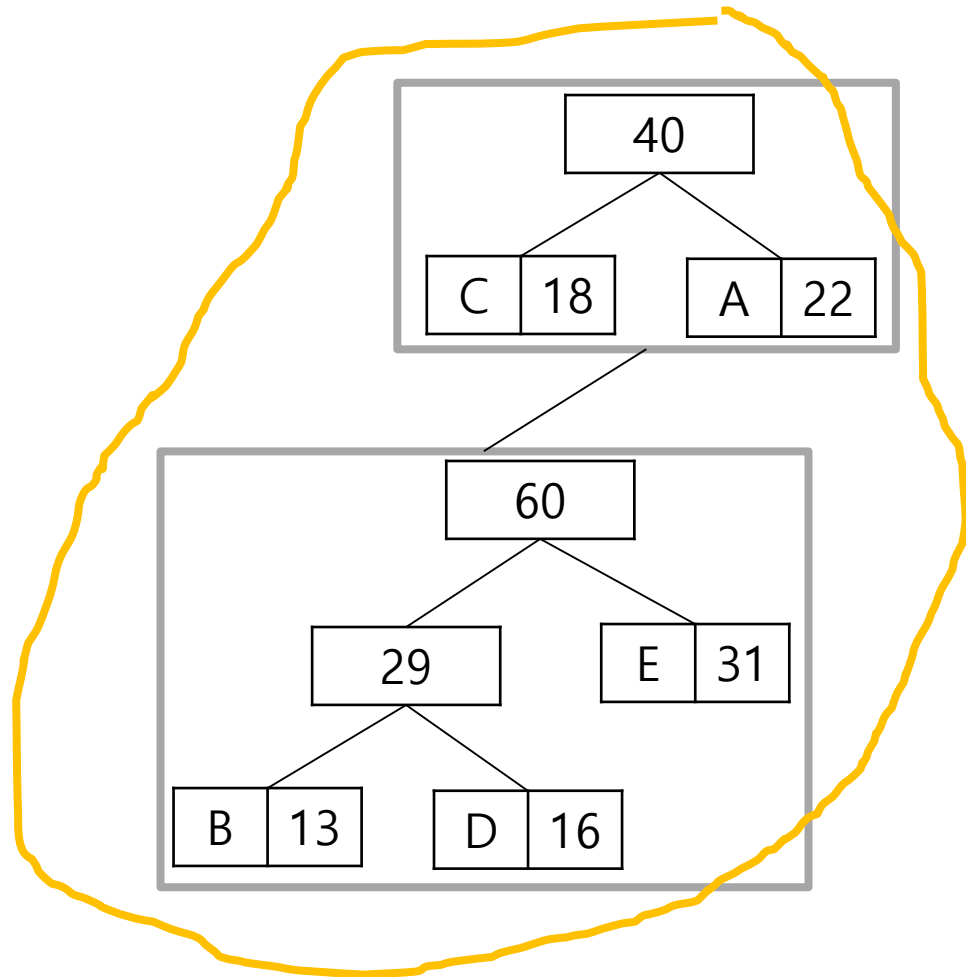


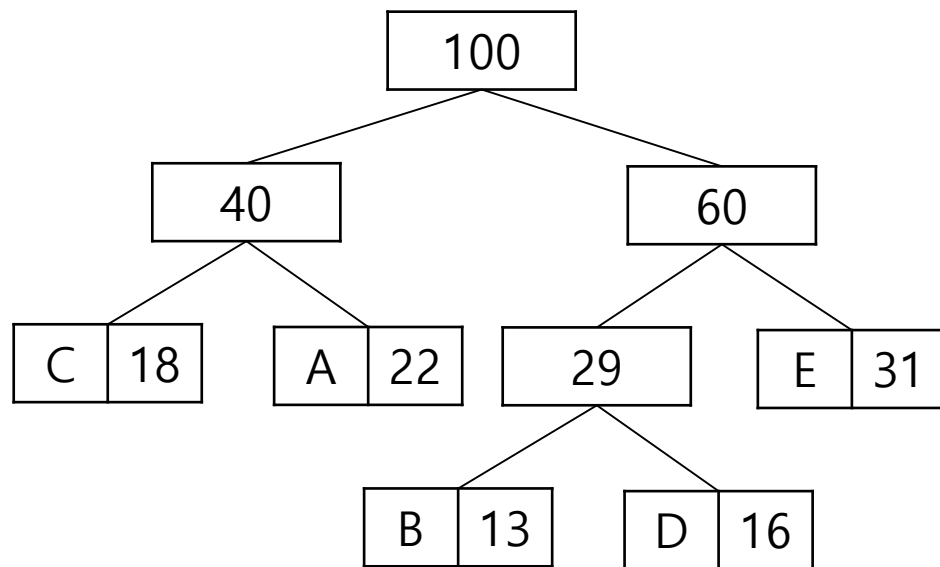
최소 힙



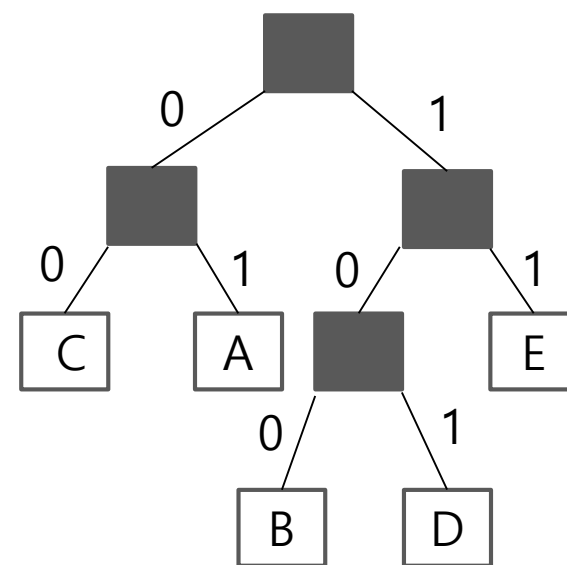






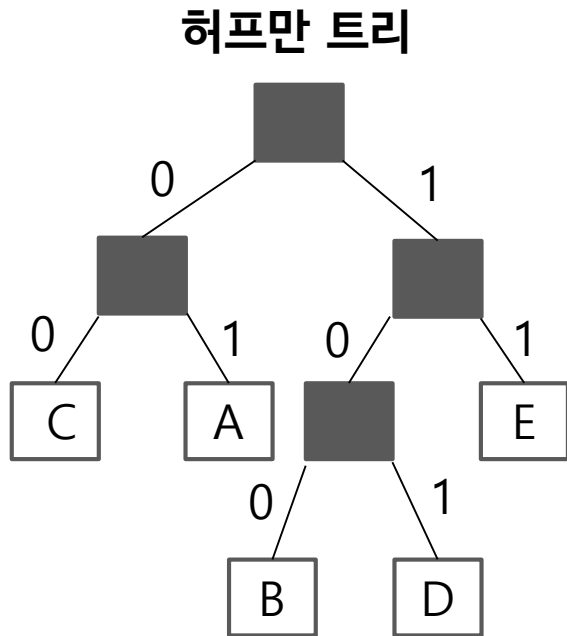


허프만 트리



01. 허프만 코드

- 허프만 코드 생성



허프만 코드

문자	코드
A	01
B	100
C	00
D	101
E	11

하나의 문자 코드가 다른 어떤 문자 코드의 접두부(prefix)와도 겹치지 않는다

→ 한 문자의 코드는 다른 코드의 처음 부분과 일치하지 않는다.

“코딩의 유일성 보장”

Ex) A의 코드는 01인데,
이를 다른 코드들의 처음 2 비트와 비교해보면
01로 시작되는 코드가 없다.

01. 허프만 코드

- 디코드(Decoding)

허프만 코드

문자	코드
A	01
B	100
C	00
D	101
E	11

디코드 과정

- 1.가장 짧은 코드 길이의 비트를 읽는다.
- 2. 코드표에 일치하는 문자가 있는 검사
- 3-1. 일치하는 문자가 없으면, 한 비트를 더 읽어온다.
- 3-2. 일치하는 문자가 있으면, 문자를 출력

Ex) **011010011100**

인코딩	디코딩
01/1010011100	A/1010011100
01/101/0011100	AD/0011100
01/101/00/11100	ADC/11100
01/101/00/11/100	ADCE/100
01/101/00/11/100/	ADCEB