

# 허프만 코딩

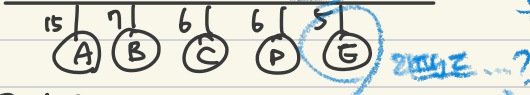
입력파일 문자 빈도수를 가지고 빈도수를 이용하여 파일을 압축하는 과정

허프만 트리를 만들어 압축하는 방법론의 원리

① 압축할 파일을 스캔 → 빈도수 계산

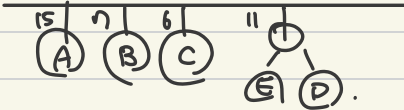
② 빈도수를 우선순위로 최소 힙을 구성

가장 작은 빈도가  
우선순위 높음

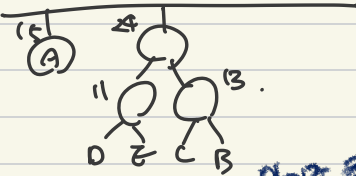
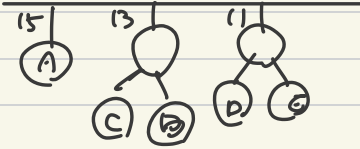


③ 빈도수가 제일 작은 두 노드를 삭제

④ 삭제한 두 노드 중 작은 것을 왼쪽 노드, 큰 것을 오른쪽



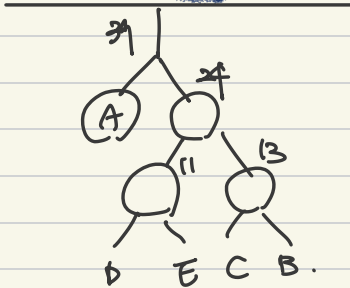
⑤ 노드가 하나 남을 때까지 반복



마지막 통합이 끝남

⑥ 마지막 노드가 루트 노드가 됨

빈도수가 높은 문자에는 짧은 이진 코드 (허프만 코드)를 부여하고 빈도수가 낮은 문자에는 긴 이진 코드를 부여

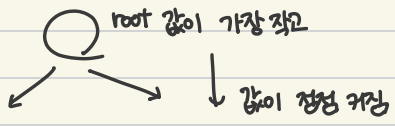


\* 최소 힙이란? min heap

부모노드 자식노드가 되는 평면 이진 트리 형태의 자료구조

마지막 레벨을 제외한 모든 레벨이 완전히 채워져 있어야함

마지막 레벨의 노드는 왼쪽부터 순서대로 채워져있어야함

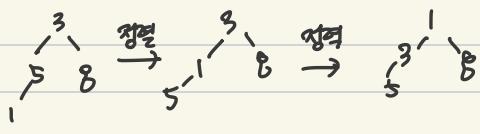
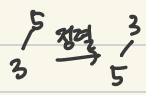


- 삽입: 새로운 값을 가장 아래에 삽입한 뒤 올라가며 정렬
- 삭제: 루트 (최소값)를 제거 가장 마지막노드를 루트로 옮긴뒤 아래로 내려가며 정렬

ex) ① 5

④ 1 삽입

② 3 삽입



③ 8 삽입



최소힙 ↔ 최대힙.