

Huffman Coding

주어진 문자열에 대해 허프만 코드를 적용한 후의 문자열 길이를 출력하시오.

입력/출력

aaaabbbb
7
해설 : a와 b가 각각 1bit 코드로 변환되며, a가 4글자, b가 3글자이므로 전체 문자열은 7의 길이를 갖게 됩니다.
aaaaaabbbc
14
해설 : a는 1bit, b와 c는 각각 2bit 코드로 변환되며, 전체 문자열은 14bit가 됩니다.
abracadabra
23
visionzquestionzonionzcaptionzgraduationzeducation
185

조건

- 자료구조 라이브러리는 사용할 수 없으며, 단순 배열은 사용해도 무방합니다. 또한 직접 구현한 자료구조는 사용할 수 있습니다.
- 문자열 관련 라이브러리 함수와 클래스, 객체 함수는 사용해도 무방합니다.
- 문자열은 'a'부터 'z' 까지 26개의 알파벳 소문자로만 이루어져 있습니다.
- 공백 문자가 없으므로 문자열 입력을 받을 때 다음과 같은 예제를 따르면 정상 작동합니다.

C
char array[100]; scanf("%s", array);
C++
char array[100]; cin >> array;
Java
Scanner s; String str = s.nextLine(); str.charAt(n); // (n번째 위치의 문자를 가져오는 함수)

- 허프만 인코딩에 대해서는 다음 페이지에 설명되어 있습니다.

허프만 트리 구축 과정(1)

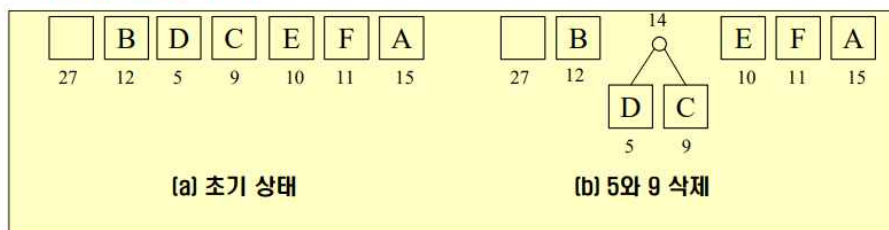
❖ 최소 힙

- 허프만 트리의 자료구조에 가장 적합

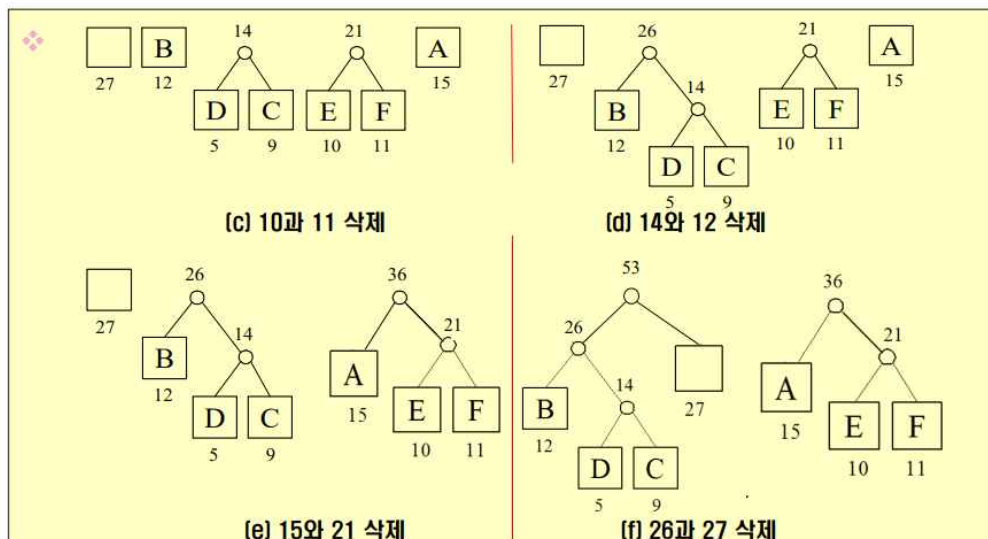
❖ 주어진 텍스트의 빈도수를 계산

		A	B	C	D	E	F
k	0	1	2	3	4	5	6
count[k]	27	15	12	9	5	10	11

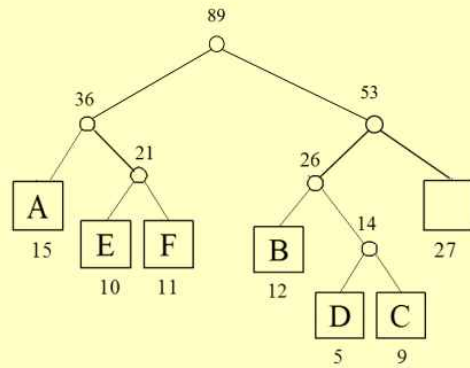
❖ 허프만 트리 구축



허프만 트리 구축 과정(2)



허프만 트리 구축 과정(3)



(g) 36과 53 삭제

		A	B	C	D	E	F
k	0	1	2	3	4	5	6
	11	00	100	1011	1010	010	011