

“공통 괄호 문자열 사전” 문제 풀이

작성자: 최은수

부분문제 1

S 의 부분 문자열인 괄호열을 S 의 부분 괄호열이라 하자. 모든 괄호열은 여는 괄호 하나에서 시작해 닫는 괄호 하나에서 끝나기 때문에 A 의 부분 괄호열의 개수는 많아야 $\frac{|A|^2}{4}$ 개이다. 마찬가지로, B 의 부분 괄호열의 개수는 많아야 $\frac{|B|^2}{4}$ 개이다.

A 의 부분 괄호열 중, B 의 부분 괄호열인 것만을 모아, 이를 정렬하여 K 번째의 문자열을 출력하면 $O(|A|^5 + |B|^5)$ 에 문제를 해결할 수 있다.

부분문제 2

A 의 부분 괄호열과 B 의 부분 괄호열 쌍 모두를 비교하는 것이 아니라, 각각을 정렬하여 비교하면 $O(|A|^3 + |B|^3)$ 에 문제를 해결할 수 있다.

부분문제 3

A 의 부분 괄호열 S, T 가 있어, S 가 T 의 접두사가 된다면, T 는 S 보다 사전순으로 크므로 답이 아니다. 따라서 A 의 각 접미사에 대해, 괄호열인 접두사 중 가장 짧은 문자열만 구하면 된다.

각 문자부터 시작하여, 괄호 문자열 판별을 적용하면 되므로 $O(|A|^2)$ 의 시간복잡도를 가진다.

부분문제 4

A 의 모든 접미사에 대한 트라이를 만들고, 이 중 올바른 괄호열인 문자열에 해당하는 노드에 표시를 하자.

B 의 모든 접미사를 트라이에서 검색하며, 검색 중 참조하는 모든 노드에 표시를 하자.

구성된 트라이를 dfs로 순회하자. 순회 중, 여는 괄호에 해당하는 자식을 닫는 괄호에 해당하는 자식보다 먼저 방문하면 이는 사전순이 된다. 따라서 순회 중 K 번째로 만나는 두 번 표시된 노드에 해당하는 문자열이 답이 된다.

트라이의 구성 및 검색에는 문자열의 길이의 합만큼의 시간이 소요되므로, 시간복잡도는 $O(|A|^2 + |B|^2)$ 이다.

부분문제 5, 6, 7, 8

두 문자열 A 와 B 를 합친 $A * B$ 의 접미사 배열을 구하자. 그 후, 이 접미사 배열의 LCP 배열을 구하자.

A 의 접미사 $A[i..|A|]$ 에 대해, 이 문자열의 B 의 부분 문자열인 가장 긴 접두사는 다음의 두 가지 경우이다.

- 사전순으로 $A[i..|A|]$ 보다 앞인 B 의 접미사 중 가장 뒤에 있는 접미사와의 최장 공통 접두사
- 사전순으로 $A[i..|A|]$ 보다 뒤인 B 의 접미사 중 가장 앞에 있는 접미사와의 최장 공통 접두사

두 경우의 길이 모두 스윙핑을 통하여 모든 접미사에 대해 $O(|A| + |B|)$ 에 구할 수 있으며, 둘 중 최댓값의 배열을 $l_1, l_2, \dots, l_{|A|}$ 라 하자.

또한, 어떤 문자열 S 가 있을 때, S 의 i 번째 문자까지 있는 여는 괄호의 개수에서 닫는 괄호의 개수를 빼 값을 문자열 S 의 i 번째 높이 h_i 로 정의하자. 편의상 $h_0 = 0$ 으로 정의하자.

$A[i..j]$ 가 올바른 괄호열인 조건과, $h_{i-1} = h_j$ 이며, $i \leq k < j$ 를 만족하는 모든 k 에 대해 $h_k \geq h_j$ 인 조건은 동치이다.

$1 \leq i \leq |A|$ 를 만족하는 i 에 대해, 다음 방법을 통해 $A[i..j]$ 가 올바른 괄호열인 가장 큰 j 를 구할 수 있다.

먼저, 이분탐색을 통해 $i \leq k$ 이면서, $h_{i-1} - 1 = h_k$ 인 가장 작은 k 를 찾는다. 없다면 $k = |A|$ 로 둔다. $h_{i-1} = h_j$ 이며 $i \leq j \leq k$ 인 것은 $A[i..j]$ 가 올바른 괄호열인 것과 동치이다.

$A[i..j]$ 가 B 의 부분 문자열인 것은, $i \leq j < i + l_i$ 를 만족하는 것과 동치이다. 따라서 i 번째 문자에서 시작하는 A 의 부분 문자열 중 문제의 조건을 만족하는 것의 개수는, $i \leq j \leq \min(k, i + l_i - 1)$ 이며 $h_{i-1} = h_j$ 를 만족하는 j 의 개수이다.

A 의 접미사 배열의 순서대로 조건을 만족하는 부분 문자열의 개수를 세면서, K 개가 넘어갈 때 답을 구해주면 된다.

주의해야 할 점은, 답은 중복을 허용하지 않는 집합에서의 K 번째 원소라는 점이다. 이 문제를 해결하는 방법은, 접미사 $A[i..|A|]$ 에 대한 계산을 할 때 이미 계산된 부분을 빼 주는 것이다. LCP 배열을 lcp_i 라 한다면, $i \leq j < i + lcp_i$ 에 해당하는 부분을 빼 주면 된다.

부분문제 5, 6, 7의 경우, 문제를 푸는 과정에서 일부의 처리를 건너뛰어도 해결할 수 있다.

- $A = B$ 인 경우, $l_i = |A| - i + 1$ 로 두어 해결할 수 있다.
- $K = 1$ 인 경우, 조건을 만족하는 j 중 가장 작은 하나만 보아 해결할 수 있다.

시간복잡도는 $O(|A| \log |A| + |B| \log |B|)$ 이다.