‘‘일이 이어져야 좋다’’ 문제 풀이

작성자: 이재웅

부분문제 1

부분문제 1은 N이 2의 거듭제곱 꼴인 경우이다. 이 경우, S_N은 1과 0이 한 개씩 교대로 나타내는 형태를 띄게 된다.

예를 들어, $S_1 = 1, S_2 = 101, S_4 = 1010101$과 같은 꼴이 되는 것을 확인할 수 있다.

따라서 질의 (i, j, k)의 답은 i나 j가 홀수일 경우 $min(2k+1, j-i+1)$이고, i, j가 짝수일 경우 $min(2k, j-i+1)$를 만족하게 된다.

이와 같이 특수한 형태의 N에 대한 문자열의 성질을 관찰하면 $O(Q)$ 시간에 문제를 해결할 수 있다.

부분문제 2

부분문제 2는 N이 1000 이하로 매우 작은 경우이다. 이때, 문자열 S_N의 길이는 $2N - 1$ 이하임을 관찰함으로써 문자열의 길이 역시 매우 작음을 알아낼 수 있다.

분할 정복 등의 알고리즘을 통해 문자열의 길이에 대한 선형 시간에 문자열의 각 원소를 모두 알아낼 수 있고, 문자열의 길이가 $2N - 1$ 이하이므로 $O(N)$ 시간에 S_N의 모든 원소를 구할 수 있다.

이제 문자열 S_N의 원소를 알아낸 것을 바탕으로 어떻게 질의를 해결할 수 있을지 생각해보자.

우선, 시작점이 s로 고정되어 있고 0을 최대 k개까지 포함하는 $S_N[i..j]$의 가장 긴 부분문자열에 주목하자. 문자열의 모든 원소를 이미 알고 있는 상태라면, $S_N[s..e]$에 포함되는 0의 개수가 k개를 초과하거나 e가 j보다 커지기 직전까지 부분문자열의 끝점 e를 1씩 증가시켜 감으로써 조건을 만족하는 부분문자열을 구할 수 있다.

또한, 시작점이 $i, i+1, \cdots, j$일 경우에 대해 조건을 만족하는 가장 긴 부분문자열의 길이를 구한 뒤, 그것들의 최댓값을 계산함으로써 질의 (i, j, k)를 해결할 수 있다.

이때, 시작점이 s일 경우 0을 최대 k개까지 포함하는 가장 긴 부분문자열의 끝점의 위치가 s가 커짐에 따라 단조 증가한다는 성질을 바탕으로, 투 포인터 기법을 통해 각각의 질의를 $O(j - i + 1)$만에 해결할 수 있다.

이 경우 S_N의 모든 원소를 계산하는 데 $O(N)$이 필요하고, Q개의 질의를 각각 $O(N)$만에 해결하므로 총 $O(QN)$ 시간에 문제를 해결할 수 있다.

부분문제 3

부분문제 3은 $j - i + 1$의 합이 $100,000$ 이하인 경우이다. 따라서 $j - i + 1$의 합이 N에 비해 작다는 점을 이용해 $S_N[i..j]$의 원소를 알아내는 방법에 대해 생각해보자.

S_N의 x번째 원소 $S_N[x]$에 대해, $S_{\lfloor N/2 \rfloor}$의 길이가 L이고 S_N의 길이가 $2L + 1$ 이라면 아래와 같은 식이 성립한다.

$$S_N[x] = \begin{cases} S_{\lfloor N/2 \rfloor}[x], & \text{if } x < L + 1 \\ N\%2, & \text{if } x = L + 1 \\ S_{\lfloor N/2 \rfloor}[x - L - 1], & \text{if } x > L + 1 \end{cases}$$

위 식을 이용하면 $S_{\lfloor N/2 \rfloor}[x]$를 $O(\log N)$에 계산할 수 있다.

$S_N[i..j]$의 모든 원소를 알고 있다면, 부분문제 2에서와 같이 해당 질의를 $O(j-i+1)$만에 해결할 수 있고, 주어진 문자열의 모든 원소를 구하는 데 $O((j-i+1)\log N)$이 필요하므로, 총 $O(\sum_{q=1}^{Q}(j_q-i_q+1)\log N)$ 시간에 문제를 해결할 수 있다.

부분문제 4

부분문제 4는 모든 질의에 대해 $k=0$인 경우이다.

먼저, 질의 (s,e,k)에 대해, 가장 왼쪽부터 시작해서 연속한 1의 최대 길이, 가장 오른쪽부터 시작해서 연속한 1의 최대 길이, 구간 (s,e) 전체에서 연속한 1의 최대 길이를 각각 $L(s,e), R(s,e), A(s,e)$로 정의하자.

$$L(s,e)=\begin{cases} L(s,m-1), & \text{if } L(s,m-1)<m-s \\ (m-s)+L(m,e), & \text{if } L(s,m-1)=m-s \end{cases}$$

$$R(s,e)=\begin{cases} R(m,e), & \text{if } R(m,e)<e-m+1 \\ (e-m+1)+R(s,m-1), & \text{if } R(m,e)=e-m+1 \end{cases}$$

$$A(s,e)=max(A(s,m-1),A(m,e),R(s,m-1)+L(m,e))$$

정의에 따라 위와 같은 식이 성립하므로, 우리는 각 질의에 대한 답 $A(s,e)$를 분할정복을 통해 찾을 수 있다.

단순히 위 식만 이용해 계산할 경우 $T(n)=2T(n/2)+O(1)$로부터 각 질의당 $O(N)$의 시간이 걸림을 알 수 있지만, 이때 $S_N[s..e]$는 수많은 $S_{\lfloor N/2^i \rfloor}$ 꼴 문자열의 합으로 표현할 수 있고, 이를 바탕으로 모든 $S_{\lfloor N/2^i \rfloor}$에 대해 각각의 L,R,A 값을 미리 계산해 둠으로써 분할정복에 걸리는 시간을 $O(\log N)$으로 단축시킬 수 있다.

따라서 전체 질의를 총 $O(Q\log N)$ 시간에 해결할 수 있다.

부분문제 5

부분문제 5는 부분문제 4와 달리 k가 0보다 큰 값을 가질 수 있으므로, L,R,A의 정의에도 변화가 필요하다.

구간 (s,e)의 가장 왼쪽부터 시작해서 0을 최대 k개 포함하는 가장 긴 문자열의 길이를 $L(s,e,k)$

구간 (s,e)의 가장 오른쪽부터 시작해서 0을 최대 k개 포함하는 가장 긴 문자열의 길이를 $R(s,e,k)$

구간 (s,e)의 부분 문자열 중 0을 최대 k개 포함하는 가장 긴 문자열의 길이를 $A(s,e,k)$로 정의하자.

$$L(s,e,k)=\begin{cases} L(s,m-1,k), & \text{if } L(s,m-1,k)<m-s \\ (m-s)+L(m,e,k-k'), & \text{if } L(s,m-1,k')=m-s, \end{cases}$$

(k'은 $L(s,m-1,x)=m-s$를 만족하는 가장 작은 x)

$$R(s,e)=\begin{cases} R(m,e,k), & \text{if } R(m,e,k)<e-m+1 \\ (e-m+1)+R(s,m-1,k-k'), & \text{if } R(m,e,k')=e-m+1 \end{cases}$$

(k'은 $R(m,e,x)=e-m+1$을 만족하는 가장 작은 x)

$$A(s,e)=max(A(s,m-1),A(m,e),max_{k'=0}^{k}(R(s,m-1,k')+L(m,e,k-k')))$$

정의에 따라 위와 같은 식이 성립하므로, 우리는 각 질의에 대한 답 $A(s,e,k)$를 분할정복을 통해 찾을 수 있다.

부분문제 4와 같이 모든 $S_{\lfloor N/2^i \rfloor}$에 대해 각각의 L, R, A 값을 미리 계산해 둠으로써 분할정복에 걸리는 시간을 $O(\log N)$으로 단축시킬 있고, 이러한 전처리에는 각각의 질의마다 $O(k \log N)$의 시간이 필요하므로 총 $O((\sum_{q=1}^{Q} k + Q) \log N)$ 시간에 문제를 해결할 수 있다.