

## 0. 정의

문자열  $s$ , 정수  $a, b, x$ 에 대해 다음을 정의한다.

- $|s|$  := 문자열  $s$ 의 길이 (문자열  $s$ 에 들어있는 문자의 개수)
- $s[a]$  또는  $s_a$  := 문자열  $s$ 에 대해 인덱스  $a$ 에 위치한 문자  
문자열  $s$ 에 대해 인덱스의 최대치는  $|s|-1$  이다.
- $s[a:b]$  또는  $s_{a:b}$  := 문자열  $s$ 에 대해 인덱스  $a$  이상  $b$  미만에 해당하는 부분 문자열 ( $a < b$ )  
 $|s_{a:b}| = b - a$
- 접두사 := 첫 문자를 포함하는 부분 문자열,  $s_{0:x}$ ,  $0 < x < |s|$
- 접미사 := 마지막 문자를 포함하는 부분 문자열,  $s_{x:|s|}$

## 1. 문자열 알고리즘 개요

A	ORONDONTISS
B	NTI

문자열 B가 A에 포함되는지 여부를 판단한다.

ORONDONTISS  
NTI  $\Rightarrow$  매칭이 될 때까지  
미끄러진다.

위와 같이 미끄러지는 방법으로 문자열 알고리즘의 문제를 푸는 경우 최악의 경우  $O(|A||B|)$ 의 시간 복잡도가 된다.

최악의 예시

A	AAAAAAAAAA
B	AAAAAB

A | AAAAAAAAAA  
B | AAAAAB

$\rightarrow$  부분 문자열을 비교할 때  
마지막 문자까지 비교해야 한다.

KMP 알고리즘은 이 문제를  $O(|A|+|B|)$  시간 복잡도에 해결할 수 있게 해 준다.

## 2. 실패함수

- 실패함수  $F(x)$  := 문자열  $s[0:x+1]$ 에서 자기 자신을 제외하고 접두사와 접미사가 일치하는 최대 길이  
문자열  $s$  자기 자신을 제외해야 하므로  $F(x)$ 의 치역은  $0$  이상  $x$  이하의 정수이다.  
e.g. 문자열  $s = "ABABA"$ 에 대해  $F(4)$ 를 계산해 본다. 문자열  $s$ 에 대해 접두사와 접미사를 구하면 다음과 같다.

길이	점두사	점미사
0	(an empty string)	(an empty string)
1	A	A
2	AB	BA
3	ABA	ABA
4	ABAB	BABA
5	ABABA	ABABA

점두사와 점미사 중 일부는 서로 동일하다. 위 표에 의하면 길이0, 길이1, 길이3, 길이5에 대해 점두사와 점미사가 서로 동일하다. 이 중 문자열  $s$  그 자체와 동일한 길이5를 제외하고 가장 긴 길이 3이  $F(4)$ 가 된다.

이 실패함수의 값을 구하기 위해 길이  $|S|-1$  부터 점점 짧아지게 점두사와 점미사를 구하는 방법을 계산하면 시간이 너무 오래 걸리게 된다. => DP를 사용해 보자.

$F(5)$ 의 값을 바탕으로

$F(6)$ 의 값을 구할 수 있는가?

$s =$  A B A B C A B A B A

$F(5) = 1$

임을 알고 있다...