

반복(Repeat)

[문제] {a, t, g, c} 4개의 염기로 이루어진 DNA 문자열이 있다. 반복(repeat)이란 연속하여 반복되는 부분자열이다. 만일 문자열이 아래와 같다면 caggtc는 길이가 6인 반복서열이다.

agtcaggtccaggtcgcgatacgaatac

단 반복서열은 반드시 연결되어 있어야 하므로 아래에서 caggtc는 반복이 아니다. 길이가 1인 반복 서열이다.

agtcaggtcgcgataccaggtcaatac

여러분은 주어진 DNA 문자열에서 **가장 긴 반복(repeat)**을 찾아서 그 길이를 정수로 출력해야 한다. 아래의 예를 보면 cag도 반복이지만 그보다 더 큰 반복인 cagcag가 존재하므로 이 입력의 경우 답은 3이 아니라 6이 되어야 한다.

agtcagcagcagcaggcgatacgaatac

[입출력] 입력파일 이름은 **repeat.inp** 이다. 파일의 첫 줄에 DNA 문자열(string)이 주어진다. 문자열은 문자 { a, g, t, c }로만 구성되어 있으며 길이 N은 $3 \leq N \leq 100$ 범위다. 여러분은 이 문자열에서 **가장 긴 반복(repeat)**의 길이를 **repeat.out**에 정수로 출력해야 한다.

[예제]

repeat.inp	repeat.out
agtccgagat <u>gagag</u> gagagt	5
agtc <u>ggggggg</u> attgacgatagat	3
agtctgcatgac	0

[조건] 제출 프로그램의 이름은 repeat.py이다. 마감 시간은 4월 16일 금요일 저녁 23시 59분이다. 허용 제출횟수는 20번이며 검사는 10개의 데이터로 시행한다.

[참고] 조환규 교수님의 2019년 입문 과제물을 참고하여 수업 진도에 맞춰 수정하였습니다.