# CSE3081 (2반): 알고리즘 설계와 분석
# [숙제 2]

**담당 교수:** 임 인 성

2020년 11월 27일

**마감:** 12월 7일 토요일 오후 8시 정각 (이후 제출자에 대해서 LATE 감점 적용)
**제출물 및 제출 방법:** 조교가 사이버 캠퍼스 공지 사항 참조.

**목표:** (1) 알고리즘 설계 기법 중의 하나인 Dynamic Programming 방법에 대한 이해도를 높이도록 한다. (2) 주어진 문제로부터 재귀적인 구조를 유추하고, 이를 테이블을 사용하여 계산하는 과정에 대하여 연습하여 본다.

다음 글을 읽은 후 아래에 기술하는 gapped alignment 문제를 풀어보자 (출처: Wikipedia).

> In bioinformatics, a sequence alignment is a way of arranging the sequences of DNA, RNA, or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences. Aligned sequences of nucleotide or amino acid residues are typically represented as rows within a matrix. ***Gaps are inserted between the residues so that identical or similar characters are aligned in successive columns***.

두 개의 문자열 (string)에 대해 임의로 indel이라고 하는 gap −을 적절히 삽입할 수 있으나, 그러한 삽입을 가급적 방지하기 위하여 매번 -$p$점 만큼의 감점을 부과한다. 만약 indel이 아닌 두 대응되는 문자가 일치하면 $s$점 만큼의 점수를, 아닐 경우 -$f$점 만큼의 감점을 부과한다. 만약 $s = 2$, $f = 1$, $p = 2$라고 가정할 경우, `X = ATCGGATCT`와 `Y = ACGGACT`에 대해 `X = ATCGGAT-CT`와 `Y = A-C-GG-ACT`와 같이 gap이 삽입되었다면 전체 유사성 점수는 1점, 그리고 `X = ATCGGATCT`와 `Y = A-CGG-ACT`와 같이 gap이 삽입되었다면 전체 유사성 점수는 7점이 된다.

이제 임의로 주어진 문자열 $X = x_1x_2 \cdots x_m$과 $Y = y_1y_2 \cdots y_n$에 대하여, $O(mn)$ 시간 복잡도를 가지는 dynamic programming 알고리즘을 사용하여, 전체 유사성 점수를 최대로 해주는 gap 삽입 방법을 출력해주는 프로그램을 작성하여 보자. 여러분의 프로그램은 다음과 같은 입출력 요구사항을 만족해야 한다.

**입력 형식**

프로그램이 수행되면 이름이 `input.txt`인 텍스트 파일에서 다음과 같은 형식으로 저장되어 있는 정보를 읽어 들여야 한다.

```
twostrings.bin
s f p
```

이 파일의 첫 번째 줄에는 두 문자열 데이터를 저장하고 있는 텍스트 파일의 이름이 저장되어 있다. 이 파일에는 이진 형식 (binary format)으로 데이터가 저장되어 있는데, 첫 4 바이트에는 $X$의 길이 $m$, 그리고 다음 4 바이트에는 $Y$의 길이 $n$이 각각 `int` 타입으로 저장이 되어 있고, 다음 $m$ 바이트에는 $X$가, 그리고 그 다음 $n$ 바이트에는 $Y$가 `char` 타입으로 저장이 되어 있다. 다음 세 개의 양의 정수 `s`, `f`, `p`가 (의미는 분명) `int` 타입으로 저장되어 있다.

**출력 형식**

프로그램 수행 후 이름이 `output.txt`인 텍스트 파일에 다음과 같은 방식으로 계산 결과를 출력하라.

```
1
10
1
8
3
2
4
7
```

여기서 첫 줄에는 자신이 구한 최대 전체 유사성 점수, 두 번째 줄에는 gap을 포함한 전체 문자열의 길이가 저장이 되어야 한다. 다음 $X$에 삽입된 gap의 개수와 그 개수 만큼 gap이 삽입된 위치를 저장되어야 하며, 그 다음 마찬가지로 $Y$에 대한 gap 정보가 저장되어야 한다.

- **[주의]**
  1. 조교가 숙제 제출 방법 및 기타 공지 사항을 사이버 캠퍼스에 올릴 예정이니 항상 확인하기 바람.
  2. 조교는 자신의 명령어 파일과 입력 데이터를 사용하여 여러분의 프로그램이 정확한 값을 계산하는지 확인할 예정이니, 자신이 생성한 데이터 (필요 시 적절히 압축하여)와 원시 코드를 조교가 수행하기 편리한 형태로 제출할 것.
  3. 제출한 원시 코드에서 대해서는 copy-check를 수행할 예정이며, 다른 사람의 코드 또는 보고서를 복사할 경우 **관련된 사람 모두에 대하여 (즉 복사한 사람과 복사 당한 사람 모두)** 과목 최종 성적의 50%를 감점함.