

생물정보학 및 실습 2 실습 보고서 4

2021-20471 박명규

<https://github.com/PMKYU98/BioInfo2>

Q1 (ba5d)

Edge-weighted directed acyclic graph에서 최장 경로 찾기

이번 주차 과제 세 문제 중 가장 어렵고 시간이 많이 들었다.

처음에는 학부 시절 배운 adjacent list를 표현하는 class로 node를, 그들을 포함하는 class로 graph를 표현하려 했으나 node 단위에서 생각보다 많고 복잡한 method가 필요할 것 같지 않다고 느껴졌다. 그래서 solution을 크게 간소화하여 graph class만 선언하였다.

생각만큼 문제 해결이 잘 되지 않아 마지막에는 straightforward 하게 시작점에서부터 다른 정점에 도달하기까지 가장 긴 거리를 구하기 위해 재귀함수를 이용하였고, 마지막에는 문제의 타겟이 되는 정점에서부터 backtracking을 진행했다.

Q2 (ba5e)

BLOSUM62 Global alignment

3주차 과제 2번(격자 공간에서 가장 긴 경로 찾기)을 응용하였다. 단, 아래와 오른쪽으로 이동하는 경우는 통일된 gap penalty를 적용하고 substitution matrix를 따라 대각선으로 이동하는 경우를 추가하였다.

Substitution matrix는 주어진 BLOSUM62 matrix를 parsing하여 사용하였다. 그리고 생물정보학 및 실습 I에서 배웠던 Needleman-Wunsch algorithm을 코드화하여 쉽게 진행할 수 있었다.

Q3 (ba5f)

PAM250 Local alignment

Q2 ba5e에서 substitution matrix를 PAM250으로 변경하고, local alignment의 대표적인 알고리즘인 Smith-Waterman algorithm을 사용하기 위해서 크기를 비교하는 값에 0을 추가하였다.

또한 가장 높은 점수로부터 backtracking을 시작하였는데, local alignment가 끝난 후에도 양끝에 gap을 추가하는 상황을 방지하기 위해서 0점에 도달하면 즉시 backtracking을 멈추도록 했다.
