

Notice

1. 먼저 학번과이름, 분반, 프로그램번호로 폴더를 만든다. 예를 들어, 60분반에 학번 201612111인 홍길동 학생이 No.1 프로그램을 작성할 때는 60_201712111홍길동_PA1 이라고 폴더를 만든다. 그리고 이 폴더에서 프로그램 작업을 한다.
2. 반드시 레포트와 요구하는 문서를 작성해서 같이 제출한다.(문서 작성도 채점에 포함됨)
3. 모든 작업이 끝나면 폴더를 압축하여 수업게시판 과제제출란에 올리기 바랍니다.
4. **Due Date : 4월 13일**

1 Dynamic Programming

PA02 : 염기서열에서 보존 영역 찾기

일반적으로 생물들은 많은 부분의 유전자가 서로 같다고 한다. 이를 바탕으로 서로 얼마나 가까운 종인지 아니면 외관상 닮았지만 유전자 서열을 살펴보면 생각보다 먼 종인지를 구분할 수 있다고 한다. 주어진 두 개의 서열(DNA, RNA, Protein)에서 서로 같은 부분이 얼마나 있는지 확인하는 프로그램(local alignment 방법)을 작성하고자 한다. 해결방법은 dynamic programming 기법으로 구현한다

문제 : 주어진 두 개의 서열에서 부분적으로 서로 같다고 판단되는 영역을 찾아서 출력하라. Alignment score는 다음과 같은 규칙을 가진다.

- Matched(일치) : 3점
- Mismatched(불일치) : =2점
- Gap penalty : -5점(첫번째, -2점(연속 두번째부터))

입력

- DNA 서열의 양식(fasta양식 :70자 한 줄. 편의상 한줄로 표시한 것임. 연속 서열임)은 다음과 같다.(파일은 PLMS 수업게시판에서 다운로드)

```
>gi|21426071|ref|NC_004002.1| Sheeppox virus 17077-99, complete genome
AATTTAAATCTAATTATAAAAAAGCCAATTAAACCTGTAAATGGATACTTTTTTCATTCAATCTTTTAAG
TCTAATTGAAGCCAATTTAAATCTAATTATAAAAGCCAATTAAACCTGTAAATGGATACTTTTTTCATTTC
AATCTTTTAAGTCTAATTATTAGTGCTGATTTCACTAGCGAAATATCGTTGTCAGAAACGAGGTCTCGAA
...
```

출력

- 서열의 점수를 부여하여 점수의 상위 5개의 부분서열을 출력하라. 5개 미만이면 순서대로 출력하라. 점수는 부분서열의 최고점수를 기준으로 한다. 서열의 끝은 최고점수를 지나고 연속해서 5개가 불일치 혹은 gap이 나오면 부분서열의 끝으로 간주한다.

예)

```
부분 일치 영역 1(score)
..A-CAGAAGGA..
..AG-AGTCC|T..
부분 일치 영역 2(score)
..A-CA-AAGGA..
..AG-AGTCG|T..
```

- 출력 양식 : 출력파일명은 output_filename1_filename2.algin이다. 그 양식은 다음과 같다.
 - 서열1 : 이름, 서열개수
 - 서열2 : 이름, 서열개수
 - 부분서열 : 점수(1)

```
서열1: 시작위치 ... 종료위치
seq1 ..AGGAGTCC|T..
seq2 ..AGG-GTCCGT..
```

플이조건

- 사용하는 알고리즘은 Dynamic programming 기법이다.
- 입력(파일이름)은 프로그램 실행때 인자로 주어져야 한다. 즉 command line 으로 처리하여야 한다.

```
%pa02 file_name1 file_name2
```

- 출력은 그 결과가 화면에 나타남과 동시에 output_filename1_filename2.align 파일로 생성한다. 출력파일 이름은 문자열 처리를 통해 자동으로 생성한다.
- 플이방식은 보고서(PA02-report)에 명시되어야 한다.
- 사용한 알고리즘의 시간복잡도(time complexity)을 분석하여 보고서에 포함시켜야 한다.

2 점검환경 및 제출

1. 구현언어는 C++를 추천하며 C, Java, python도 가능하다. 그리고 실행환경은 linux이다. C, C++용 컴파일러는 gcc이다.

2. 주점검 사항은 (1) 프로그램이 linux상에서 정상적으로 컴파일되는가? (2) 프로그램이 실행되는가? (3) 프로그램이 실행이 옳게 되는가? (4) Dynamic programming이 제대로 구현되었는가?(코드로 확인함) (5) 프로그램 코드 정리가 제대로 되었는가? (변수의 가독성, 주석달기, 코드의 가독성) 이다.
3. 작성한 보고서(PA02-report)와 프로그램 소스(PA02)는 압축하여 하나의 파일로 제출한다.
4. 제출은 PLMS Programming Assignment 게시판에 올리면 됩니다.
5. 마감:2019년4월13일