Suffix Array II

[문제] 텍스트 문자열 T이 주어진다. 여러분은 DC3 알고리즘을 이용하여 Suffix Array를 만드는 프로그램을 작성하고 그 과정을 상세히 보여야 한다.

[입력] 파일은 한 줄로 이루어진 텍스트 문자열 T이다. T는 {$,a,c,g,t}로 구성되어 있으며, 종료문자 $는 T의 마지막 위치에만 존재한다.

[구현] 최종적으로 구현하고자 하는 함수의 명세는 다음과 같다.

name	build_suffix_array(T)
argument	T: 텍스트 문자열 (Array of integers)
return	SA: 접미사 배열 (Array of integers)

※ 함수의 입출력 형식은 각자 사용하는 언어에 맞게 구현해도 좋다. 예를 들어 C의 경우 array의 크기를 추가로 입력받는다거나, C++에서 vector등의 STL로 사용해도 된다.

구현은 다음 단계를 거쳐서 이루어져야 한다.

[1단계] 문자열을 읽어와서 각 문자를 0~4까지 매핑하여 integer sequence로 만든다.
[2단계] 주어진 텍스트 문자열로부터 3개 문자씩 묶어서 하나의 문자로 만들어 triplet sequence S_0, S_1, S_2를 생성한다.
[3단계] S_1, S_2를 연결해 $S_{12} = S_1 \circ S_2$를 만들고 각 문자를 0부터 크기순으로 매핑한다.
[4단계] S_{12}에 대한 접미사 배열이 만들었다고 가정하고 이를 이용해 S_0를 정렬한다.
(이 부분을 재귀로 구현해도 되지만 처음부터 재귀로 작성하면 디버깅이 힘들기 때문에 merge 작업이 개발되기 전까지는 기초적인 방법으로 S_{12}에 대한 접미사 배열을 생성하여 이용하도록 한다)
[5단계] S_0와 S_1, S_2를 병합하여 전체 sequence에 대한 접미사배열을 생성한다.
[6단계] S_{12}의 접미사 배열을 생성하는 부분을 재귀로 대치한다.

[제출] 제출은 **6월 13일 토요일 저녁 10시**까지이다. 제출은 ESPA 과제물 게시판이며 제출물은 다음과 같다. **`SA_DC3.{c,cpp,py}`**에는 자신이 구현한 코드를, **`NAME_report.pdf`**에는 각 단계별 실험결과를 통해 올바르게 동작하고 있는지를 보이도록 한다.