OMÓWIENIE ZADANIA CZATBBB (I ETAP XXXI OI)

Podzadanie 1 - $n \leq 100, b \leq 1000$:

Limity w tym podzadaniu pozwalają na bezpośrednią symulację treści. Dla każdego k-literowego podsłowa słowa S zliczamy litery występujące po nim. Dzięki temu dla określonego k-literowego sufiksu wiemy, jaka litera zostanie po nim dopisana. Po dopisaniu jakiejś litery do naszego słowa, jesteśmy w stanie zaktualizować informacje o licznościach liter w czasie $O(1)$. Do takiego zliczania możemy wykorzystać drzewo trie lub kontener `unordered_map` w języku C++.

Całkowita złożoność wyniesie $O(n \cdot A + b \cdot k)$, przy czym A to rozmiar alfabetu.

Podzadanie 2 - $n \leq 10^6, b \leq 10^8$:

Zamiast indeksować strukturę całymi słowami, to możemy indeksować ich haszami wielomianowymi (patrz np. opis w języku angielskim lub opis w książce "Przygody Bajtazara. 25 lat Olimpiady Informatycznej. Wybór zadań", PWN 2018). Musimy uważać na paradoks dnia urodzin, więc trzeba zastosować moduł rzędu 10^{18} lub dwa rzędu 10^9. Dzięki temu złożoność wyniesie $O(n \cdot A + b)$.

Podzadania 3 oraz 4 - wcześniejsze wystąpienie sufiksu R zawsze będzie istnieć i za każdym wystąpieniem będzie znajdować się ta sama litera:

Skoro wcześniejsze wystąpienia zawsze będą istnieć oraz zawsze będzie występować po nich ta sama litera, to jeżeli k-literowym sufiksem jest słowo P, a po nim k-literowym sufiksem staje się słowo Q, to zawsze gdy następnym razem k-literowym sufiksem będzie słowo P, kolejnym k-literowym sufiksem będzie ponownie Q. Rozważmy graf skierowany, którego wierzchołkami są k-literowe słowa oraz istnieje krawędź ze słowa P do słowa Q wtedy i tylko wtedy, gdy po sufiksie P zawsze następuje sufiks Q. Graf ten składa się z wierzchołków o stopniu wyjściowym 1 oraz odizolowanych wierzchołków o stopniu wejściowym i wyjściowym 0.

Nasze zadanie można sprowadzić do następującego problemu: Dany jest graf skierowany oraz wierzchołek v o stopniu wyjściowym 1; należy stwierdzić, w jakim wierzchołku się znajdziemy, jeżeli t razy przejdziemy krawędzią, zaczynając od wierzchołka v. Każda taka ścieżka od pewnego momentu (niekoniecznie po t krawędziach) się zacykli. Po wyznaczeniu ścieżki do cyklu oraz cyklu (oba rozmiaru $O(n)$) możemy wyznaczyć wierzchołek końcowy, a następnie cofnąć się odpowiednią liczbę razy, wyznaczając przy tym odpowiednie litery (od końca).

W zależności od użytej metody do budowy grafu rozwiązanie to ma złożoność obliczeniową $O(n \cdot A + (b - a))$ lub $O(n \cdot A + n^2 + (b - a))$.

Podzadanie 5 - $k \leq 20, b \leq 10^{10}, n \leq 10^6$**, użyte są tylko litery** a **i** b**:**

Okazuje się, że aby stwierdzić, jaka litera występuje po k-literowym słowie wychodzącym już poza słowo S, wystarczy sprawdzić, jaka litera najczęściej występuje po tym podsłowie wśród wszystkich wystąpień tego podsłowa ściśle wewnątrz słowa S. Za każdym nowym wystąpieniem tego podsłowa, częstość wystąpienia najliczniejszej litery wzrasta o jeden, więc nadal jest największa, czyli przy kolejnym wystąpieniu tego k-literowego podsłowa kolejna litera będzie taka sama. Oznacza to, że podejście grafowe z poprzedniego podzadania można zastosować i tutaj. Ponieważ alfabet jest dwuliterowy oraz $k \leq 20$, to możliwych słów jest co najwyżej 2^{20}. Jest to dostatecznie mało, aby skonstruować cały graf.

Podzadanie 6 - $n \leq 10^6, k < n, b \leq 10^{18}$

Zbiór wierzchołków grafu osiągalnych z wierzchołka startowego tak naprawdę zawsze jest rozmiaru $O(n)$. Faktycznie, jeżeli k-literowy sufiks P nie występował wcześniej jako podsłowo słowa S, to dopisujemy literę a. To może się wydarzyć pod rząd co najwyżej k razy - po $k+1$-wszym dodaniu litery a już zawsze będziemy dopisywać a. Jeżeli natomiast k-literowy sufiks P występował w całości w słowie S, to po dodaniu nowej litery, k-literowy sufiks Q również będzie występować w całości w słowie S - wiemy, że najczęściej występująca litera po P jest również najczęściej występującą literą po P wśród wszystkich wystąpień w całości w słowie S i ma krotność przynajmniej 1, więc musi istnieć słowo Q w całości występujące w słowie S.

Wynika z tego, że jest tylko $n + k$ różnych słów długości k do rozpatrzenia i są to podsłowa słowa Sa^k. Możemy sobie zatem pozwolić na konstrukcję interesującej nas części grafu i wyznaczenie w nim ścieżki do cyklu oraz cyklu, począwszy od zadanego wierzchołka startowego. Podsłowa reprezentujemy za pomocą identyfikatorów (hasze lub np. deterministyczny słownik podsłów bazowych KMR) i utrzymywać je w haszmapie (w przypadku haszy) lub w tablicy (w przypadku KMR). Ostateczne rozwiązanie ma złożoność $O(n \cdot A + (b - a))$ lub $O(n \cdot A + n \log n + (b - a))$ w przypadku deterministycznym.

Olimpiada Informatyczna, OEIiZK
ul. Nowogrodzka 73 :: 02-006 Warszawa :: tel. 22 50 140 14 lub 22 50 140 15 :: olimpiada@oi.edu.pl

Olimpiada finansowana jest ze środków Ministerstwa Edukacji i Nauki w ramach zadania publicznego „Organizacja i przeprowadzenie olimpiad przedmiotowych i interdyscyplinarnych w latach szkolnych 2022/2023, 2023/2024, 2024/2025”.