Zliczamy nadciągi

*Tomasz IDZIASZEK**

*Informatyk, prowadzi stronę internetową algonotes.com/

Na Obozie Naukowo-Treningowym im. A. Kreczmara w roku 2019 pojawiło się zadanie pt. *Nadciągi*, które brzmiało następująco: dane jest n-literowe słowo s nad alfabetem A-literowym i chcemy znaleźć liczbę słów m-literowych nad tym samym alfabetem, które zawierają s jako podciąg (czyli niekoniecznie spójny ciąg liter). Wynik podajemy jako resztę z dzielenia przez $P = 10^9 + 7$. Ograniczenia były dość duże, bo $n \leqslant 10^6$ oraz $m, A \leqslant 10^9$.

W zadaniu szukamy zatem nadciągów słowa s, czyli słów, które możemy uzyskać, dopisując nowe litery w dowolnych miejscach słowa s. Na przykład dla alfabetu $\{\mathtt{A},\mathtt{B}\}$ oraz słowa $s = \mathtt{ABA}$ mamy 5 czteroliterowych nadciągów:

$$\underline{\mathtt{A}}\mathtt{A}\underline{\mathtt{BA}},\quad \underline{\mathtt{ABA}}\mathtt{A},\quad \underline{\mathtt{ABA}}\mathtt{B},\quad \underline{\mathtt{AB}}\mathtt{B}\underline{\mathtt{A}},\quad \mathtt{B}\underline{\mathtt{ABA}}.$$

Zaczniemy od prostego rozwiązania wykorzystującego programowanie dynamiczne: przez $dp[i, j]$ oznaczymy liczbę słów j-literowych, które zawierają i-literowy prefiks s jako podciąg, ale nie zawierają prefiksu długości $i + 1$. Przypadki bazowe to $dp[i, j] = 0$ dla $i < 0$ lub $i > j$ oraz $dp[0, 0] = 1$, zaś wzór rekurencyjny ma postać:

$$dp[i, j] = dp[i, j - 1] \cdot (A - [i \neq n]) + dp[i - 1, j - 1].$$

Wynika on z faktu, że i-literowy podciąg może znajdować się w $(j - 1)$-literowym prefiksie słowa (a wtedy ostatnią literą może być dowolna litera inna niż $s[i + 1]$, chyba że $i = n$, to wtedy dowolna) albo nie, i wtedy ostatnią literą musi być $s[i]$, a o literę krótszy podciąg musi znajdować się w $(j - 1)$-literowym prefiksie słowa.

To rozwiązanie wyznacza odpowiedź $dp[n, m]$ w złożoności czasowej $O(nm)$, więc zdecydowanie zbyt wolno. Pozwala nam ono jednak zaobserwować ciekawą rzecz: z powyższego wzoru wynika, że odpowiedź *nie zależy* od konkretnych liter słowa s, a jedynie od jego długości n. Tak więc równie dobrze możemy wyznaczyć liczbę słów, które zawierają n takich samych liter x jako podciąg. Ustalmy pozycje $1 \leqslant i_1 < i_2 < \ldots < i_n \leqslant m$, na których stoją te litery (przy czym wybieramy pozycje z pierwszymi n wystąpieniami litery x, licząc od lewej). Tak więc wszystkie litery przed pozycją i_1 nie mogą być literą x, ale mogą być dowolną inną; analogicznie wszystkie litery pomiędzy pozycjami i_k oraz i_{k+1}. Natomiast litery za pozycją i_n mogą być już dowolne. Tak więc w zależności od wyboru pozycji i_n mamy wzór:

$$Ans = \sum_{i_n=n}^{m} \binom{i_n - 1}{n - 1} (A - 1)^{i_n - n} \cdot A^{m - i_n}.$$

Ta suma nie jest jednak zbyt przyjemna. Wygodniej byłoby rozważyć wszystkie pozycje, na których występuje litera x; załóżmy, że jest ich $N \geqslant n$. Wtedy na pozostałych $m - N$ pozycjach mogą wystąpić dowolne inne litery. To upraszcza sumę do:

$$Ans = \sum_{N=n}^{m} \binom{m}{N} (A - 1)^{m-N}.$$

Ta suma ma aż $O(m)$ składników. Zauważmy, że jest to $m + 1 - n$ końcowych składników z rozwinięcia w dwumian Newtona $((A - 1) + 1)^m$, zatem możemy ją zamienić na n początkowych składników:

$$Ans = A^m - \sum_{i=0}^{n-1} \binom{m}{i} (A - 1)^{m-i}.$$

Co ciekawe, możemy kombinatorycznie uzasadnić poprawność powyższego wzoru. Skoro składnik A^m oznacza liczbę wszystkich słów m-literowych, to występująca za nim suma ze znakiem minus powinna oznaczać liczbę słów, które *nie zawierają* s jako podciągu. Tak jest w istocie: jeśli słowo nie zawiera s, ale zawiera jego i-literowy prefiks (dla $i < n$), to znowu, idąc zachłannie od lewej, przed każdą pozycją możemy wybrać z $A - 1$ liter, ale też za ostatnią pozycją mamy $A - 1$ możliwości, bo te litery muszą być inne niż $s[i + 1]$.

Jak szybko wyznaczyć Ans? Aby obliczyć początkowe A^m, użyjemy szybkiego potęgowania w czasie $O(\log m)$, podobnie zrobimy z pierwszym składnikiem sumy (dla $i = 0$), który jest równy $(A - 1)^m$. Każdy kolejny składnik sumy wyznaczymy w czasie stałym na podstawie poprzedniego:

$$\binom{m}{i + 1} (A - 1)^{m-(i+1)} = \binom{m}{i} (A - 1)^{m-i} \cdot \frac{m - i}{(i + 1)(A - 1)}.$$

Potrzebujemy do tego odwrotności modulo P dla liczby $A - 1$ oraz liczb $1 \leqslant i \leqslant n$; w uwadze na marginesie przypominamy, jak wyznaczyć je efektywnie. Ostatecznie uzyskujemy algorytm o złożoności czasowej $O(n + \log P + \log m)$.

Odwrotność modulo P jednej liczby x możemy wyznaczyć w czasie $O(\log P)$ za pomocą rozszerzonego algorytmu Euklidesa lub małego twierdzenia Fermata (obliczając $x^{P-2} \bmod P$).

Odwrotności liczb $1 \leqslant i \leqslant n$ możemy wyznaczyć w czasie $O(n + \log P)$ następująco. Na początku obliczamy tablicę $fac[i] = i! \bmod P$, a następnie $finv$ – odwrotność $fac[n]$ modulo P. W końcu robimy pętlę po $i = n, \ldots, 1$:

$$inv[i] = finv \cdot fac[i - 1] \bmod P;$$
$$finv = finv \cdot i \bmod P.$$

Rozwiązanie zadania M 1749. Najpierw udowodnimy, że punkty osiowe mają następującą własność:

Niech Π będzie dowolną płaszczyzną przechodzącą przez punkt osiowy pewnego wielościanu wypukłego. Wtedy po obu stronach tej płaszczyzny znajdzie się tyle samo wierzchołków tego wielościanu.

Istotnie, weźmy dowolny wierzchołek A po jednej stronie płaszczyzny Π. Wtedy na prostej AP leży inny wierzchołek B tego wielościanu. Ponieważ wielościan jest wypukły, to punkt P leży wewnątrz odcinka AB. Oznacza to, że punkty A i B leżą po przeciwnych stronach płaszczyzny Π. Powtarzając to rozumowanie dla każdego wierzchołka, dostajemy tezę.

Załóżmy teraz, że istnieją dwa punkty osiowe P i Q. Poprowadźmy płaszczyznę Σ_1 przechodzącą przez P oraz pewien wierzchołek danego wielościanu. Następnie poprowadźmy płaszczyznę Σ_2 równoległą do Σ_1, przechodzącą przez Q. Oznaczmy przez x i x' liczbę wierzchołków wielościanu leżących na płaszczyznach, odpowiednio, Σ_1 i Σ_2. Oczywiście $x \geqslant 1$. Niech teraz a będzie liczbą wierzchołków wielościanu znajdujących się po tej stronie płaszczyzny Σ_1, po której nie ma płaszczyzny Σ_2; b liczbą wierzchołków między płaszczyznami Σ_1 i Σ_2; a c liczbą wierzchołków po tej stronie płaszczyzny Σ_2, po której nie ma płaszczyzny Σ_1. Z własności udowodnionej na początku mamy

$$a = b + x' + c \quad \text{oraz} \quad c = a + x + b.$$

Sumując obie równości stronami, dostajemy

$$2b + x + x' = 0,$$

co jest niemożliwe, więc płaszczyzny Σ_1 i Σ_2 się pokrywają. Wobec dowolności wyboru wierzchołka, przez który przechodzi Σ_2, otrzymujemy zatem, że wielościan wypukły może mieć co najwyżej 1 punkt osiowy. Nietrudno wskazać również wielościan z punktem osiowym (np. sześcian).