O podciągach i palindromach słów kilka

*Radosław ŻAK**

* Student, Wydział Matematyki i Informatyki, Uniwersytet Jagielloński

Autor prezentuje część wyników ze swojej pracy *Podciągi*, nagrodzonej złotym medalem (*ex aequo* z pracą *Szczególne podgrupy skończonego indeksu w grupie warkoczy* B_3 autorstwa Bartłomieja Bychawskiego) w 43. Konkursie Uczniowskich Prac z Matematyki im. Pawła Domańskiego. (przyp. red.)

Na I etapie XXVI Olimpiady Informatycznej pojawiło się zadanie o mniej więcej takiej treści:

Zadanie. Dla liczby naturalnej $N \leqslant 10^{18}$ wypisać słowo o dokładnie N różnych podciągach i nie więcej niż 1000 znakach.

Słowem jest tu dowolny ciąg znaków, zaś jego *podciąg* to dowolny ciąg, który możemy dostać, usuwając z naszego słowa pewne znaki, nie zmieniając przy tym kolejności pozostałych.

Dla wyrobienia sobie intuicji przed wyruszeniem w dalszą wędrówkę przedstawię kilka przykładów. Pojedyncza litera „a” ma dwa podciągi – siebie samą oraz słowo puste, czyli „” (będziemy je oznaczać przez $\mathsf{0}$). Słowo „baca” ma ich już 14 – w kolejności leksykograficznej są to: $\mathsf{0}$, a, aa, ac, aca, b, ba, baa, bac, baca, bc, bca, c, ca. Musimy uważać, by podciągów nie mylić z pojęciem, które nazywa się czasem *podsłowami* bądź po prostu podciągami spójnymi, dla których znaki brane do podciągu muszą tworzyć spójny fragment słowa. Zerkając na nasz przykład – chociażby „bc” jest podciągiem, ale już nie podsłowem, słowa „baca”.

Wróćmy do naszego zadania. Informatycy rozwiązali je informatycznie. Ja tak nie umiałem, więc wymyśliłem inny sposób – zająłem się słowami zawierającymi tylko dwa różne znaki A, B (i tylko takie będziemy rozważać w tym artykule).

Definicja. Dla dwóch względnie pierwszych liczb $a, b \geqslant 1$ niech słowo $\mathrm{gen}(a,b)$ będzie zdefiniowane w następujący, rekurencyjny sposób:

$$\mathrm{gen}(a,b) = \begin{cases} \mathsf{0}, & \text{jeśli } a = b = 1, \\ \mathsf{A} \circ \mathrm{gen}(a-b,b), & \text{jeśli } a > b, \\ \mathsf{B} \circ \mathrm{gen}(a,b-a), & \text{jeśli } b > a, \end{cases}$$

gdzie $\circ$ oznacza konkatenację (czyli zwykłe złączenie) słów.

W pewnym sensie słowo $\mathrm{gen}(a,b)$ jest zapisem algorytmu Euklidesa dla liczb a i b – w każdym momencie zapisujemy, od której z liczb jest odejmowana druga. Dla przykładu $\mathrm{gen}(11,7)$ to ABABB – dostajemy kolejno pary $(4,7)$, $(4,3)$, $(1,3)$, $(1,2)$ oraz $(1,1)$, na której proces się kończy. Warto w tym momencie zwrócić uwagę, że przedstawioną procedurę można odwrócić: mając dowolne słowo dwuliterowe, możemy, czytając je od końca, dojść do generującej to słowo pary liczb względnie pierwszych. Na przykład dla słowa BABA mamy

$$(1,1) \xrightarrow{\mathsf{A}} (2,1) \xrightarrow{\mathsf{B}} (2,3) \xrightarrow{\mathsf{A}} (5,3) \xrightarrow{\mathsf{B}} (5,8),$$

skąd $\mathrm{gen}(5,8) = \mathsf{BABA}$.

Może nie być jasne, do czego słowa $\mathrm{gen}(a,b)$ będą nam w ogóle przydatne, ale pokazuje to następująca własność:

Twierdzenie 1. *Słowo* $\mathrm{gen}(a,b)$ *ma dokładnie* $a+b-1$ *podciągów. Ponadto dokładnie* $a-1$ *z nich zaczyna się od* A*, zaś* $b-1$ *od* B*.*

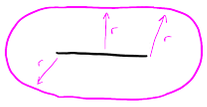
Dowód. Pierwsze zdanie wynika z drugiego, które znowuż jest symetryczne – wystarczy więc, że pokażemy część o liczbie podciągów zaczynających się od A. Dowodzimy indukcyjnie po długości $\mathrm{gen}(a,b)$, dla słowa pustego nietrudno to sprawdzić. Jeśli $b > a$, to $\mathrm{gen}(a,b)$ ma na początku B, i usunięcie tej litery nie zmieni liczby podciągów zaczynających się od A, pozostawi zaś słowo $\mathrm{gen}(a,b-a)$, które ma ich dokładnie $a-1$ z założenia indukcyjnego. Jeśli zaś $a > b$, to pierwszą literą $\mathrm{gen}(a,b)$ jest A. Każdy podciąg zaczynający się od A można zapisać tak, by składał się z tej właśnie pierwszej litery oraz jakiejś pozostałej części – która, jako iż $\mathrm{gen}(a,b) = \mathsf{A} \circ \mathrm{gen}(a-b,b)$, będzie podciągiem $\mathrm{gen}(a-b,b)$. Ich jest zaś, z założenia indukcyjnego, $(a-b)+b-1 = a-1$, czyli dokładnie tyle, ile chcemy. □

A zatem nasze zadanie ma bardzo proste rozwiązanie – dla danego N wybierzmy dowolne a względnie pierwsze z $N+1$ i zapiszmy $\mathrm{gen}(a, N+1-a)$. I to prawie

Rozwiązania zadań z artykułu *Czy dobrze rysuję obwarzanki?*

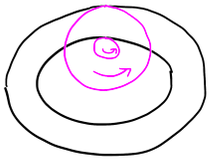
1. Dziury nie widzimy dokładnie wtedy, gdy r-otoczka elipsy o' wypełnia całe jej wnętrze, co z kolei ma miejsce, jeśli r jest nie mniejsze od krótszej półosi o'. Elipsa ta ma półosie długości R i $R\cos\alpha$, więc otrzymujemy warunek $r \geqslant R\cos\alpha$.

2. W takim przypadku płaszczyzna rzutu P jest prostopadła do płaszczyzny okręgu o, więc obserwowany obraz o' jest odcinkiem. A r-otoczka odcinka ma kształt parówki – prostokąta zakończonego półkolami.

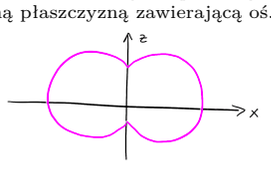


3. Punkty po przeciwnych stronach „wystającej części uśmiechu” leżą na rysunku blisko siebie, ale odpowiadające im punkty na właściwym torusie są daleko. Chodzi o to, że jedna część torusa niejako zasłania nam drugą.

4. Należy uzasadnić, że zewnętrzny zarys r-otoczki o' nie może być elipsą. Przyjmijmy, że elipsa o' ma półosie $a > b$. Wówczas od wewnętrznej strony toczy się po niej bez uskoków koło o promieniu $\frac{b^2}{a}$. Po jej r-otoczce toczy się więc koło o r większe, czyli o promieniu $r + \frac{b^2}{a}$ – najłatwiej to zauważyć, łącząc dwa takie koła na sztywno i tocząc razem (jedno po elipsie, drugie po otoczce, jak na rysunku). Z drugiej strony, gdyby tą r-otoczką była elipsa, to musiałaby mieć półosie $a+r$ i $b+r$, więc największe toczące się koło miałoby promień $\frac{(b+r)^2}{a+r}$. Ta wartość jest jednak ostro mniejsza od $r + \frac{b^2}{a}$.



5. Nie wynika – w dwóch punktach na osi „torusa” powstają wklęsłe dzióbki. Taki „torus” nadal jest bryłą obrotową, więc dzióbki łatwo zobaczyć w przekroju dowolną płaszczyzną zawierającą oś.



wszystko. Prawie, bo musimy zatroszczyć się jeszcze o długość słowa, co w teorii jest trudniejsze niż w praktyce – da się wykazać, że większość słów otrzymanych w ten sposób ma długość $O(\log N \log\log N)$, ale dowód jest bardzo trudny.

Teraz jednak rozejrzyjmy się po nowych możliwościach, jakie otwiera przed nami powyższe twierdzenie. Wprowadźmy garść oznaczeń: $\mathsf{P}(\mathfrak{s})$ będzie oznaczało po prostu liczbę podciągów słowa $\mathfrak{s}$. $\mathsf{P}^{\mathsf{A}}(\mathfrak{s})$ oraz $\mathsf{P}^{\mathsf{B}}(\mathfrak{s})$ będą oznaczały liczbę podciągów słowa $\mathfrak{s}$ zaczynających się odpowiednio od A i od B, zaś gdy będziemy zapisywali literę w indeksie dolnym: $\mathsf{P}_{\mathsf{A}}(\mathfrak{s})$ i $\mathsf{P}_{\mathsf{B}}(\mathfrak{s})$, będziemy zliczali podciągi kończące się na odpowiednią literę – we wszystkich tych przypadkach wliczamy słowo puste. Nasze twierdzenie możemy teraz zgrabnie zapisać jako $\mathsf{P}^{\mathsf{A}}(\mathrm{gen}(a,b)) = a$. W ten sposób dostajemy bijekcję między słowami o N podciągach a elementami $\mathbb{Z}^*_{N+1}$, czyli resztami z dzielenia przez $N+1$ względnie pierwszymi z $N+1$. W jedną stronę jest to $\mathsf{P}^{\mathsf{A}}(\cdot)$, w drugą zaś $a \mapsto \mathrm{gen}(a, N+1-a)$. Za darmo dostajemy więc, że słów o N podciągach zawierających co najwyżej dwa różne znaki jest dokładnie $\varphi(N+1)$.

Przykład dla $N = 8$:
$1 \leftrightarrow \mathrm{gen}(1,8) = \mathsf{BBBBBBBB}$
$2 \leftrightarrow \mathrm{gen}(2,7) = \mathsf{BBBA}$
$4 \leftrightarrow \mathrm{gen}(4,5) = \mathsf{BAAA}$
$5 \leftrightarrow \mathrm{gen}(5,4) = \mathsf{ABBB}$
$7 \leftrightarrow \mathrm{gen}(7,2) = \mathsf{AAAB}$
$8 \leftrightarrow \mathrm{gen}(7,2) = \mathsf{AAAAAAAA}$

Mając dane słowo $\mathfrak{s}$, możemy zapisać je od tyłu i dostać $\overline{\mathfrak{s}}$. Możemy także zamienić wszystkie litery A na B i *vice versa* – powstałe tak słowo oznaczymy przez $\mathfrak{s}^*$. Ostatnią opcją jest wykonanie obu tych operacji naraz; rezultat będziemy oznaczali przez $\widetilde{\mathfrak{s}}$. Wszystkie te działania, jak nietrudno zauważyć, zachowują liczbę podciągów słowa – te symetrie muszą zatem odpowiadać pewnym symetriom $\mathbb{Z}^*_{N+1}$. Nie jest trudno zauważyć, że jeśli $\mathfrak{s} = \mathrm{gen}(a,b)$, to $\mathfrak{s}^* = \mathrm{gen}(b,a)$, ta sytuacja nie jest więc aż tak skomplikowana. Okazuje się, że i w drugim przypadku odpowiednia symetria nie jest trudna do zapisania.

Twierdzenie 2.

$$\mathsf{P}^{\mathsf{A}}(\mathfrak{s})\mathsf{P}^{\mathsf{A}}(\overline{\mathfrak{s}}) \equiv 1 \pmod{N+1}.$$

Dla przykładu: dla naszego wcześniej wspomnianego słowa $\mathfrak{s} = \mathsf{ABABB} = \mathrm{gen}(11,7)$, zapisując je od tyłu, dostaniemy $\mathsf{BBABA} = \mathrm{gen}(5,13)$. Te słowa odpowiadają resztom, odpowiednio, 11 i 5. Mamy $11 \cdot 5 \equiv 1 \pmod{18}$, a zapisując inaczej: $11^{-1} \equiv 5 \pmod{18}$. Zatem odwracanie w znaczeniu potocznym i liczbowym jest, w pewnym (dziwnym) sensie, tym samym.

Palindrom to słowo, które wygląda tak samo od przodu jak i od tyłu. Poprzedni fakt daje nam dobry opis takich słów – odpowiadają resztom w $\mathbb{Z}^*_{N+1}$, które podniesione do kwadratu dają 1. W ten sposób na przykład dowiemy się, że każde słowo binarne o 23 podciągach jest palindromem, i nie tak trudno pokazać, że to największa liczba o tej własności. Prawdopodobnie wydawałoby nam się to zaskakujące, gdybyśmy usłyszeli o tym fakcie przed rozpoczęciem lektury niniejszego artykułu, ale teraz już domyślamy się, że jest to spowodowane tym, że każda odwracalna reszta modulo 24 podniesiona do kwadratu daje resztę 1.

Podobnie *antypalindromem* nazwiemy takie słowo $\mathfrak{s}$, że $\mathfrak{s} = \widetilde{\mathfrak{s}}$ – po odwróceniu go oraz zamianie liter A na B i *vice versa* dostaniemy wyjściowe słowo. Takimi słowami są na przykład AABB czy ABBAAB. Nietrudno zauważyć, że każdy antypalindrom ma parzystą liczbę liter – inaczej byłby problem ze środkową z nich – i że możemy go zapisać jako $\mathfrak{s} = \mathfrak{t} \circ \widetilde{\mathfrak{t}}$.

Potrzebujemy jeszcze jednej obserwacji: jak wygląda liczba podciągów złączenia dwóch słów? Odpowiedź na to pytanie wyraża poniższy wzór:

$$\mathsf{P}(\mathfrak{s} \circ \mathfrak{t}) + 1 = \mathsf{P}_{\mathsf{A}}(\mathfrak{s})\mathsf{P}^{\mathsf{B}}(\mathfrak{t}) + \mathsf{P}_{\mathsf{B}}(\mathfrak{s})\mathsf{P}^{\mathsf{A}}(\mathfrak{t}).$$

Dowód opiera się na przesuwaniu liter z $\mathfrak{t}$ do $\mathfrak{s}$ po kolei i pokazaniu, że prawa strona nie zmienia się przy tej operacji (co można zrobić, bo wiemy już, jak zmienia się liczba podciągów odpowiedniego typu po dodaniu jednej litery).

Czas na finał. Weźmy dowolną liczbę pierwszą p postaci $4k+1$. Wiemy już, że istnieje dokładnie $p-1$ słów binarnych o dokładnie $p-1$ podciągach. Możemy pogrupować je w czwórki $\{\mathfrak{s}, \mathfrak{s}^*, \overline{\mathfrak{s}}, \widetilde{\mathfrak{s}}\}$. Taka czwórka może się zdegenerować, jeśli któreś dwa z tych słów są takie same, ale $\mathfrak{s} \neq \mathfrak{s}^*$, czyli $\mathfrak{s}$ do tego celu musiałoby być palindromem albo antypalindromem. Palindromy mamy tylko dwa, tworzące jedną parę: $\mathsf{A}\ldots\mathsf{A}$ i $\mathsf{B}\ldots\mathsf{B}$, oba po $p-2$ liter. Istotnie, palindromy odpowiadają rozwiązaniom równania $x^2 \equiv 1 \pmod{p}$, czyli $p \mid (x-1)(x+1)$ – jedyne opcje to $x \equiv 1$ lub $x \equiv -1 \pmod{p}$.

W takim razie, ponieważ wszystkich słów jest $4k$, dla zachowania podzielności przez 4 musimy mieć jeszcze co najmniej jedną zdegenerowaną czwórkę, odpowiadającą $\mathfrak{s}$ będącemu antypalindromem. Taki antypalindrom możemy zapisać jako $\mathfrak{s} = \mathfrak{t} \circ \widetilde{\mathfrak{t}}$, i ze wzoru powyżej dostajemy

$$p = \mathsf{P}_{\mathsf{A}}(\mathfrak{t})\mathsf{P}^{\mathsf{B}}(\widetilde{\mathfrak{t}}) + \mathsf{P}_{\mathsf{B}}(\mathfrak{t})\mathsf{P}^{\mathsf{A}}(\widetilde{\mathfrak{t}}) = \mathsf{P}_{\mathsf{A}}(\mathfrak{t})^2 + \mathsf{P}_{\mathsf{B}}(\mathfrak{t})^2,$$

co oznacza, iż p jest sumą dwóch kwadratów.

Nie jest to może ani oryginalny, ani niespodziewany wynik, ale ciekawe, że można go otrzymać, wychodząc od czysto kombinatorycznych rozważań.