KOMBINATORYKA NA SŁOWACH

JAROSŁAW GRYTCZUK

1. Ciągi bez repetycji

Niech $S = s_1 s_2 \dots s_n$ będzie ciągiem długości n. *Segmentem* w S nazywamy podciąg składający się z kolejnych wyrazów ciągu S. Ciąg S nazywamy *repetycją* jeżeli składa się dwóch identycznych segmentów, czyli $n = 2k$ i $s_i = s_i + k$ dla każdego $i = 1, 2, \dots, k$. Mówimy, że ciąg S nie zawiera repetycji jeżeli żaden jego segment nie jest repetycją. Poniższe twierdzenie odkrył i udowodnił Axel Thue [7] (por. [6]).

Twierdzenie 1 (Thue, 1906)**.** *Istnieją dowolnie długie ciągi bez repetycji o wyrazach ze zbioru* 3*-elementowego.*

Dowód twierdzenia Thuego wykorzystuje następującą konstrukcję. Niech T będzie nieskończonym ciągiem generowanym przez podstawienie $a \to ab$, $b \to ba$:

$$T = abbabaabbaababbabaababbaabbabaab\dots.$$

Niech T_1 oznacza podciąg ciągu T o indeksach nieparzystych, zaś T_2 o indeksach parzystych. Łatwo zauważyć, że $T_1 = T$ zaś $T_2 = T'$, gdzie T' jest ciągiem, który powstał z T przez zamianę liter a i b.

Nakładką nazywamy ciąg składający się z dwóch identycznych, nachodzących na siebie segmentów, czyli segmentów, których nośniki (indeksy wyrazów) mają niepusty przekrój. Nietrudno przekonać się, że każda nakładka ma postać $xBxBx$, gdzie x jest pojedynczą literą a B dowolnym ciągiem.

Lemat 1 (Thue, 1906)**.** *Ciąg T nie zawiera nakładek.*

Dowód. Przypuśćmy na przekór, że T zawiera nakładkę i niech $N = xBxBx$ będzie najkrótszą nakładką w T. Niech $B = b_1 b_2 \dots b_n$. Jeżeli $n = 2k+1$, to ciąg $xB'xB'x$, gdzie $B' = b_2 b_4 \dots b_{2k}$, jest nakładką w $T_1 = T$ lub w $T_2 = T'$, krótszą od nakładki N. Sprzeczność.

Przypuśćmy zatem, że $n = 2k$ i że pierwszy x nakładki N należy do T_1. Wówczas środkowy x musi należeć do T_2, zaś ostatni x ponownie do T_1. Przyjmijmy (bez straty ogólności), że $x = a$. Wówczas b_1 musi leżeć w T_2 i wobec tego $b_1 = b$. Ale w drugim segmencie B wyraz b_1 należy do T_1, a zatem $b_2 = a$. Podobnie, wyraz b_2 w pierwszym segmencie B należy do T_1, co pociąga $b_3 = b$. Rozumując tak dalej dostajemy, że $B = baba\dots ba$, przy czym ostatni wyraz pierwszego segmentu nakładki należy do T_1. Zatem kolejnym wyrazem powinno być b, a mamy tam środkowy $x = a$. Sprzeczność. Jeżeli pierwszy wyraz nakładki należy do T_2, to rozumowanie jest podobne. Dowód jest więc zakończony. □

Dla dowodu Twierdzenia 1 skonstruujemy następujący ciąg R, którego wyrazami będą liczby $\{0, 1, 2\}$. Niech r_i oznacza liczbę wystąpień litery b pomiędzy i-tym a $(i+1)$-szym wystąpieniem litery a w ciągu T:

$$R = 210201210120210201202101210201\dots.$$

Nietrudno zobaczyć, że jeżeli ciąg R zawierałby repetycję, to ciąg T musiałby mieć nakładkę, co jest sprzeczne z Lematem 1.

Niedawno Tetsuo Kurosaki [5] znalazł inną, bezpośrednią konstrukcję ciągów bez repetycji. Niech K będzie dowolnym ciągiem o wyrazach ze zbioru $\{0, 1, 2\}$. Niech $f(K)$ oznacza ciąg powstały z K przez zamianę symboli 0 i 1. Niech $g(K)$ oznacza ciąg powstały z K przez zamianę symboli 1 i 2. Niech teraz $K_1 = 1$ oraz $K_{n+1} = f(K_n)K_n g(K_n)$ dla każdego $n \geqslant 1$. Początkowe ciągi wyglądają następująco:

$$K_1 = 1$$

$$K_2 = 012$$

$$K_3 = 102012021$$

$$K_4 = 012102120102012021201021012$$

Można wykazać, że żaden ciąg K_n nie zawiera repetycji.

Twierdzenie 2 (Kurosaki, 2008)**.** *Żaden ciąg K_n nie zawiera repetycji.*

Ciągi bez repetycji na zbiorze trzech symboli można tworzyć na wiele sposobów. Wiadomo, że istnieje continuum nieskończonych takich ciągów. Z drugiej strony udowodniono, że każdy ciąg skończony bez repetycji jest prefiksem ciągu *maksymalnego*, czyli takiego, którego nie da się już wydłużyć przez dodanie wyrazu na końcu.

Rozważmy ogólniejszy sposób rozszerzania ciągów polegający na wstawieniu nowego wyrazu w dowolne miejsce danego ciągu. Na przykład, ciąg $S = 102012$ możemy rozszerzyć wstawiając nowy wyraz $x \in \{0, 1, 2\}$ w dowolne miejsce ciągu S, na początku $x102012$, na końcu $102012x$ lub gdziekolwiek w środku, np. $1020x12$. Naturalne pytanie brzmi, czy każdy ciąg bez repetycji na trzech symbolach można w ten sposób rozszerzyć tak, aby nowy ciąg też nie zawierał repetycji? Odpowiedź jest negatywna, ale być może jest to prawdą dla dostatecznie długich ciągów bez repetycji.

Hipoteza 1. *Niech $S = s_1 s_2 \ldots s_n$ będzie ciągiem bez repetycji o wyrazach ze zbioru $\{0, 1, 2\}$. Jeżeli $n \geqslant 1000$, to istnieje takie $i \in \{0, 1, \ldots, n\}$ oraz taki $x \in \{0, 1, 2\}$, że ciąg $S' = s_1 \ldots s_i x s_{i+1} \ldots s_n$ nadal nie zawiera repetycji.*

2. Ciągi bez repetycji z list

Udowodnimy teraz twierdzenie, którego inspiracją jest *listowe kolorowanie* grafów. W klasycznym problemie kolorowania grafu chcemy tak pokolorować wierzchołki aby sąsiedzi otrzymali różne kolory. Korzystamy przy tym z jednego ustalonego zbioru kolorów. W kolorowaniu listowym każdy wierzchołek ma swój z góry przypisany zbiór kolorów. Jest to więc zagadnienie ogólniejsze i znane są przykłady grafów o liczbie chromatycznej dwa, których nie da się listowo pokolorować przy szczególnym przypisaniu wierzchołkom nawet dowolnie dużych zbiorów kolorów.

W listowej wersji problemu Thuego założenie jest podobne: każda pozycja w ciągu ma swój zbiór i tylko z tego zbioru może pochodzić wyraz ciągu na tej pozycji. Pokażemy, że jeżeli te zbiory są rozmiaru tylko nieco większego niż 3, to da się utworzyć ciąg bez repetycji z tych zbiorów. Pierwszy dowód pojawił się w pracy [3]. Poniżej przedstawimy ulepszony dowód pochodzący z pracy [2].

Twierdzenie 3 (Grytczuk, Przybyło, Zhu, 2011)**.** *Niech n będzie dowolną liczbą naturalną i niech $A_1, A_2, \ldots, A_n$ będą dowolnymi zbiorami 4-elementowymi. Wówczas istnieje ciąg bez repetycji $S = s_1 s_2 \ldots s_n$ taki, że $s_i \in A_i$ dla każdego $i = 1, 2, \ldots, n$.*

Dowód. Niech n i N będą ustalonymi liczbami naturalnymi, przy czym N jest dużo większe od n. Niech $L = l_1 l_2 \dots l_N$ będzie ciągiem "losowym" o wyrazach ze zbioru $\{1, 2, 3, 4\}$. Załóżmy ponadto, że elementy zbiorów A_i zostały uporządkowane.

Rozważmy następującą procedurę tworzenia ciągu S. W pierwszym kroku wybieramy jako s_1 element zbioru A_1 o numerze l_1 i zapisujemy znak $+$ w ciągu znaków Z. W drugim kroku wybieramy element o numerze l_2 ze zbioru A_2 i dopisujemy go jako s_2 jednocześnie dodając kolejny znak $+$ do ciągu Z. Kontynuujemy to postępowanie aż do momentu, w którym pojawi się w konstruowanym ciągu S repetycja. Powiedzmy, że nastąpiło to w kroku k. W tym momencie mamy $S = s_1 s_2 \dots s_j BB$, gdzie B jest segmentem długości $(k-j)/2$. Wtedy zmazujemy powtórzony segment B dopisując jednocześnie $(k-j)/2$ znaków $-$ do ciągu Z. W kolejnym $(k+1)$-szym kroku dodajemy do ciągu $S = s_1 s_2 \dots s_j B$ element zbioru $A_{|S|+1}$ o numerze l_{k+1}. I tak dalej. Na przykład, jeżeli $L = 121231123231$ oraz $A_i = \{a, b, c\}$ dla każdego i, to procedura utworzy ciąg znaków $Z = ++++--+++-++----++--+$ i ciąg wynikowy $S = abca$.

Przypuśćmy, że dla dowolnego losowego ciągu L opisana procedura nie doprowadziła do uzyskania ciągu bez repetycji S długości n. Oczywiście każdy taki ciąg wyznacza jednoznacznie parę (S, Z), której składnikami są wynikowy ciąg S oraz ciąg znaków Z rejestrujący przebieg wydłużania i skracania ciągu S. Pokażemy teraz, że na odwrót, każda taka wynikowa para (S, Z) determinuje ciąg L jednoznacznie. W istocie, niech $S = s_1 s_2 \dots s_m$ będzie ostatnim ciagiem i niech Z będzie wynikowym ciągiem znaków. Jeżeli ostatnim znakiem w Z jest $+$, to oczywiście mamy $l_N = s_m$. Jeżeli zaś ciąg Z ma na końcu serię h znaków $-$, to znaczy, że w ostatnim kroku zmazaliśmy powtórzony segment długości h. Możemy go jednak odtworzyć z ciągu S co daje nam poprzednią postać S przed zmazaniem h wyrazów. Możemy teraz usunąć serię h minusów z końcówki ciągu Z aż do pierwszego plusa i powtarzać całą tę czynność aż do odtworzenia całego ciągu L.

W konsekwencji widzimy, że ciągów L jest tyle samo co wynikowych par (S, Z), czyli 4^N. Z drugiej strony, par (S, Z) nie może być więcej niż $4^n C_N$, gdzie C_N jest N-tą liczbą Catalana. W istocie, ciągi znaków Z mają w każdym prefiksie co najwyżej tyle minusów co plusów, a liczba takich ciągów długości $2N$ z dokładnie N plusami to właśnie liczba Catalana C_N. Jak wiadomo, $C_N = \frac{1}{N+1}\binom{2N}{N} = o(4^N)$. To daje sprzeczność dla dostatecznie dużych N przy ustalonym n. Dowód jest więc zakończony. $\square$

Pytanie czy powyższe twierdzenie jest optymalne pozostaje otwarte.

Hipoteza 2 (Grytczuk, 2008). *Niech n będzie dowolną liczbą naturalną i niech $A_1, A_2, \dots, A_n$ będą dowolnymi zbiorami 3-elementowymi. Wówczas istnieje ciąg bez repetycji $S = s_1 s_2 \dots s_n$ spełniający warunek $s_i \in A_i$ dla każdego $i = 1, 2, \dots, n$.*

3. Niepowtarzalne kolorowanie grafów

Przedstawimy teraz uogólnienie twierdzenia Thuego na kolorowanie grafów. Kolorowanie wierzchołków grafu G nazywamy *niepowtarzalnym* jeżeli ciąg kolorów na żadnej ścieżce nie zawiera repetycji. Niech $\pi(G)$ oznacza najmniejszą liczbę kolorów potrzebną w takim kolorowaniu grafu G. Twierdzenie Thuego mówi zatem, że $\pi(P_n) = 3$ dla każdej ścieżki P_n, gdzie $n \geqslant 4$. Poniższy rezultat, pochodzący z pracy [1], stanowi uogólnienie tego twierdzenia na grafy o ograniczonym stopniu.

Twierdzenie 4 (Alon, Grytczuk, Hałuszczak, Riordan, 2002). *Każdy graf G o maksymalnym stopniu Δ spełnia nierówność $\pi(G) \leqslant 100\Delta^2$.*

Dowód tego twierdzenia korzysta z metody probabilistycznej. W tej samej pracy dowiedziono, że istnieją grafy G dla których oszacowanie z twierdzenia jest asymptotycznie optymalne z dokładnością do czynnika logarytmicznego.

Hipoteza 3. *Każdy graf G o maksymalnym stopniu Δ spełnia nierówność $\pi(G) \leqslant \Delta^2$.*

Pokażemy teraz, że liczba $\pi(T)$ jest również ograniczona dla drzew, pomimo tego, że drzewa mogą mieć dowolnie duży stopień.

Twierdzenie 5. *Każde drzewo T spełnia nierówność $\pi(T) \leqslant 4$.*

Dowód. Niech T będzie drzewem z korzeniem r i niech $k \geqslant 1$ oznacza wysokość T (maksymalną długość ścieżki zaczynającej się w korzeniu). Niech L_i, $i = 0, 1, \ldots, k$, oznacza zbiór wierzchołków w odległości i od korzenia r. Rozważmy ciąg $S = s_0 s_1 \ldots s_k$, $s_i \in \{0, 1, 2, 3\}$, który jednocześnie nie zawiera repetycji i palindromów (*palindrom* to ciąg, który wygląda tak samo czytany od tyłu, np. 01210). Taki ciąg możemy otrzymać biorąc dowolny ciąg bez repetycji na zbiorze $\{0, 1, 2\}$ i wstawiając czwarty symbol 3 pomiędzy kolejne pary sąsiednich wyrazów. Na przykład z ciągu 0121021201210 uzyskamy w ten sposób ciąg 0132130231230132130.

Rozważmy teraz kolorowanie $c : V(T) \to \{0, 1, 2, 3\}$ zdefiniowane przepisem $c(v) = s_i$ dla $v \in L_i$. Pokażemy, że to kolorowanie jest niepowtarzalne. W istocie, przypuśćmy, że istnieje ścieżka $P = v_1 v_2 \ldots v_{2n}$ taka, że ciąg kolorów $R = c(v_1)c(v_2) \ldots c(v_{2n})$ jest repetycją. Ponieważ ciąg S nie zawiera repetycji, ścieżka P nie może być monotoniczna i musi mieć najwyższy wierzchołek, powiedzmy v_h, którego wierzchołki sąsiednie v_{h-1} i v_{h+1} muszą leżeć na tym samym poziomie L_i. Bez straty ogólności możemy przyjąć, że $2 \leqslant h \leqslant n$ i że v_h jest korzeniem drzewa T. Wówczas ciąg R będzie miał postać:

$$R = s_{h-1} s_{h-2} \ldots s_1 s_0 s_1 \ldots s_{h-1} s_h \ldots s_{2n-h}.$$

Jeżeli $h < n$, to palindrom $s_1 s_0 s_1$ leży w całości w pierwszej części repetycji R, a stąd i w drugiej, a zatem także w S. Jeżeli natomiast $h = n$, to mamy

$$R = (s_{n-1} s_{n-2} \ldots s_1 s_0)(s_1 \ldots s_{n-1} s_n).$$

Ponieważ R jest repetycją, mamy $s_i = s_{n-i}$ dla wszystkich $i = 0, 1, \ldots n - 1$. Zatem ciąg $s_0 s_1 \ldots s_n$ jest palindromem. W obu przypadkach dostajemy zatem sprzeczność, która kończy dowód. □

Powyższe twierdzenie można uogólnić na k-drzewa.

Twierdzenie 6. *Każde k-drzewo G spełnia nierówność $\pi(G) \leqslant 4^k$.*

Niech P będzie dowolną ścieżką. Niech $c : V(P) \to \{0, 1, 2, 3\}$ będzie kolorowaniem, w którym nie ma ani repetycji ani palindromów. Wreszcie niech P' oznacza ścieżkę P z dodaną pętlą w każdym wierzchołku. *Spacerem* w grafie nazywamy dowolny ciąg wierzchołków, w którym dwa kolejne wyrazy są połączone krawędzią.

Lemat 2. *Jeżeli S jest spacerem na ścieżce P' takim, że ciąg kolorów $c(S)$ jest repetycją, to S też jest repetycją.*

Dowód. Niech $S = XY$ będzie spacerem na ścieżce P', gdzie $X = x_1 x_2 \ldots x_n$ i $Y = y_1 y_2 \ldots y_n$. Niech $c(S) = AB$, gdzie $A = a_1 a_2 \ldots a_n$ i $B = b_1 b_2 \ldots b_n$, będzie ciągiem kolorów wyznaczonym przez S, to znaczy, że $c(x_i) = a_i$ oraz $c(y_i) = b_i$ dla każdego $i = 1, 2, \ldots, n$. Przypuśćmy, że $c(S)$ jest repetycją, czyli $a_i = b_i$ dla wszystkich i. Wykażemy, że wówczas S też jest repetycją, czyli, że $x_i = y_i$ dla każdego $i = 1, 2, \ldots, n$.

Zastosujemy indukcję względem n. Najpierw zauważmy, że wierzchołki ścieżki P leżące w odległości 1 i w odległości 2 są w różnych kolorach (co wynika z braku palindromów na ścieżce P). Stąd teza twierdzenia wynika dla $n = 1$ i $n = 2$. Przypuśćmy teraz, że w ciągu kolorów $c(S)$ mamy gdzieś $a_i = a_{i+1}$, dla pewnego $1 \leqslant i < n$. Oczywiście wówczas mamy również $b_i = b_{i+1}$. Ponieważ bezpośrednie powtórzenie koloru jest możliwe tylko przez pętlę, mamy również $x_i = x_{i+1}$ i $y_i = y_{i+1}$. Możemy wobec tego usunąć wierzchołki x_{i+1} i y_{i+1} ze spaceru S otrzymując krótszy spacer S', którego ciąg kolorów $c(S')$ również jest repetycją. Z założenia indukcyjnego S', a w konsekwencji S, jest również repetycją. Jeżeli $a_n = b_1$, to rozumowanie jest podobne. Mamy wówczas $x_n = y_1$, ale także $a_1 = b_n$. Usuwając wierzchołki x_n i y_n ze spaceru S dostajemy krótszy spacer i znów stosujemy indukcję.

Przypuśćmy teraz, że dla pewnego $i \leqslant n - 2$ zachodzi $a_i = a_{i+2}$. Mamy wówczas także $b_i = b_{i+2}$. Zachodzą ponadto odpowiednie równości dla wierzchołków spaceru S, mianowicie $x_i = x_{i+2}$ oraz $y_i = y_{i+2}$. Możemy wobec tego usunąć z S wierzchołki $x_{i+1}, x_{i+2}, y_{i+1}, y_{i+2}$ otrzymując krótszy spacer S', którego ciąg kolorów $c(S')$ jest repetycją. Stosując indukcję dostajemy równość obu części spaceru S'. Ponadto nietrudno zobaczyć, że wówczas musi być także $x_{i+1} = y_{i+1}$, co kończy rozumowanie w tym przypadku.

Jeżeli natomiast nie zachodzi żaden z powyższych przypadków, to spacer S musi składać się z dwóch podścieżek w P, co szybko daje sprzeczność podobnie jak w dowodzie wyniku dla drzew. □

Niech G będzie dowolnym grafem i niech $L_0, L_1, \ldots L_m$ będzie podziałem zbioru wierzchołków na *poziomy*, to znaczy, że każda krawędź grafu G albo łączy wierzchołki sąsiednich poziomów L_i i L_{i+1}, albo leży całkowicie w pewnym poziomie L_i. Niech Rozważmy pokolorowanie c wierzchołków grafu G takie jak w dowodzie twierdzenia dla drzew: wierzchołki na ustalonym poziomie mają ten sam kolor, a ciąg kolorów po poziomach nie ma repetycji i palindromów. Nazwijmy to kolorowanie *poziomującym*. Powyższy lemat daje natychmiast następujący wniosek.

Wniosek 1. *Niech c będzie kolorowaniem poziomującym grafu G i niech P będzie ścieżką w G. Jeżeli ciąg kolorów $c(P)$ jest repetycją, to $P = P_1P_2$ i ścieżki P_1 i P_2 mają ten sam ciąg poziomów.*

Niech $L_0, L_1, \ldots, L_m$ będzie podziałem zbioru wierzchołków grafu G na poziomy. Dla ustalonego j, oznaczmy przez G_j podgraf indukowany przez L_j. Podobnie, przez $G_{<j}$ i $G_{>j}$ oznaczmy odpowiednio podgrafy grafu G indukowane przez $L_0 \cup L_1 \cup \cdots \cup L_{j-1}$ oraz $L_{j+1} \cup L_{j+2} \cup \cdots \cup L_m$. Dla każdego zbioru wierzchołków A oznaczmy przez $N(A)$ zbiór wszystkich wierzchołków mających sąsiada w A. Niech $N_k(A) = N(A) \cap L_k$. Powiem, że podział na poziomy grafu G jest *klawy* jeżeli dla każdej składowej C podgrafu $G_{>j}$ zbiór $N_j(C)$ indukuje klikę.

Twierdzenie 7 (Kündgen, Pelsmajer, 2008)**.** *Niech G będzie grafem posiadającym klawy podział na poziomy $L_0, L_1, \ldots, L_m$. Niech $M = \max\{\pi(L_j) : j = 0, 1, \ldots, m\}$. Wówczas $\pi(G) \leqslant 4M$.*

Dowód. Niech c_1 będzie kolorowaniem poziomów bez repetycji i palindromów. Niech c_2 będzie niepowtarzalnym M-kolorowaniem grafów G_j indukowanych przez poziomy L_j. Pokażemy, że kolorowanie produktowe $c(v) = (c_1(v), c_2(v)$ jest niepowtarzalnym kolorowaniem całego grafu G.

Niech P będzie ścieżką, na której ciąg kolorów $c(P)$ jest repetycją. Wówczas, na mocy Wniosku 1 dostajemy, że $P = AB$ i ścieżki A i B maja ten sam ciąg poziomów. Niech j

będzie najmniejszą liczbą taką, że P ma jakieś wierzchołki na poziomie L_j. Niech $A' = A \cap L_j$ i $B' = B \cap L_j$. Ponieważ $c(A) = c(B)$ dostajemy, że również $c(A') = c(B')$. Pokażemy, że $A'B'$ jest ścieżką w G_j. W istocie, dwa kolejne wierzchołki ciągu $A'B'$ są albo połączone krawędzią w P, albo łączy je ścieżka całkowicie zawarta w $G_{>j}$. Wtedy na mocy założenia o klawym podziale są one również połączone krawędzią w G_j. To jest sprzeczne z niepowtarzalnością kolorowania c_2. □

Twierdzenie 8 (Barát, Wood, 2009)**.** *Dla każdego grafu G istnieje równoważny topologicznie graf H taki, że $\pi(H) \leqslant 4$.*

Literatura

[1] N. Alon, J. Grytczuk, M. Hałuszczak, and O. Riordan, Nonrepetitive colorings of graphs, Random Structures and Algorithms 21 (2002) 336–346.

[2] J. Grytczuk, J. Kozik, P. Micek, New approach to nonrepetitive sequences, Random Structures and Algorithms 42 (2013) 214–225.

[3] J. Grytczuk, J. Przybyło, and X. Zhu, Nonrepetitive list colorings of paths, Random Structures and Algorithms 38 (2011) 162–173.

[4] A Kündgen, M.J. Pelsmajer. Nonrepetitive colorings of graphs of bounded tree-width. Discrete Math., 308 (2008) 4473–4478.

[5] T. Kurosaki, Direct definition of a ternary infinite square-free sequence, Information Processing Letters 106 (2008) 175–179.

[6] M. Lothaire, Combinatorics on Words, Addison-Wesley, Reading, MA, 1983.

[7] A. Thue, Über unendliche Zeichenreichen, Norske Vid. Selsk. Skr., I Mat. Nat. Kl., Christiania 7 (1906) 1–22.

Wydział Matematyki i Nauk Informacyjnych, Politechnika Warszawska, Plac Politechniki 1, 00-662 Warszawa

E-mail address: `j.grytczuk@mini.pw.edu.pl`