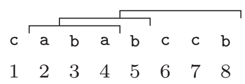
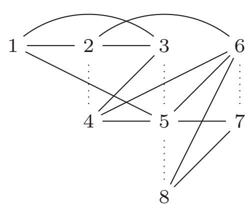
Informatyczny kącik olimpijski (100): Równoważność palindromiczna

W jubileuszowym odcinku kącika omówimy zadanie *Równoważność palindromiczna*, które pojawiło się na Obozie Naukowo-Treningowym im. A. Kreczmara w 2010 roku. Dwa słowa w i v o długości n nazwiemy *równoważnymi palindromicznie*, jeśli dla każdej pary liczb i oraz j, takich że $1 \leqslant i \leqslant j \leqslant n$, podsłowo $w[i..j]$ złożone z liter na pozycjach od i-tej do j-tej jest palindromem wtedy i tylko wtedy, gdy palindromem jest podsłowo $v[i..j]$ złożone z liter na tych samych pozycjach. Mając dane słowo w, należy wyznaczyć liczbę słów równoważnych mu palindromicznie, zawierających litery z ustalonego A-literowego alfabetu.



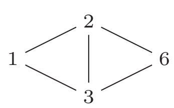
Rys. 1. W 8-literowym słowie $w =$ `cababccb` na pozycji $i = 3$ występuje palindrom nieparzysty `aba` o długości $2r + 1 = 3$, gdyż $w[2] = w[4]$ oraz $w[1] \neq w[5]$.

Rozważmy pewien palindrom $w[i - r\,..\,i + r]$ o nieparzystej długości $2r + 1$, którego środek jest na pozycji i w słowie w, oraz który nie może być rozszerzony (tzn. nie istnieje dłuższy palindrom o tym samym środku). Zatem w słowie v także musi istnieć analogiczny palindrom, zatem muszą być spełnione równości liter $v[i - k] = v[i + k]$ dla $1 \leqslant k \leqslant r$, oraz nierówność $v[i - r - 1] \neq v[i + r + 1]$. Nietrudno się przekonać, że zbiór takich (nie)równości dla wszystkich pozycji i (wraz z analogicznym zbiorem dla palindromów o parzystej długości) w pełni opisuje warunki, jakie musi spełniać słowo v, aby być palindromicznie równoważnym słowu w (patrz rys. 1).



Rys. 2. Graf G odpowiadający słowu w. Wierzchołki 2 i 4 są w tej samej składowej grafu G, zaś wierzchołki 1 i 5 są połączone krawędzią ciągłą.

Zbiór ten możemy przedstawić jako graf G o wierzchołkach $\{1, 2, \ldots, n\}$ odpowiadających pozycjom w słowie v. Krawędź łącząca dwa wierzchołki będzie przerywana, jeśli odpowiadające im pozycje w słowie muszą mieć tę samą literę, lub ciągła, jeśli muszą mieć różne litery. W ten sposób sprowadzimy nasze zadanie do problemu kolorowania grafu: liczba słów palindromicznie równoważnych słowu w jest bowiem równa liczbie kolorowań grafu G przy pomocy A kolorów (przy zachowaniu ograniczeń na wynikających z istnienia krawędzi ciągłych i przerywanych; rys. 2).



Rys. 3. Graf G po zastąpieniu składowych pojedynczymi wierzchołkami. Liczba kolorowań tego grafu to $A(A-1)(A-2)^2$.

Niestety, nie znamy wielomianowego algorytmu dla ogólnego problemu zliczania kolorowań grafów, będziemy musieli więc nieco bliżej przyjrzeć się szczególnej postaci grafu z zadania. Na początek pozbędziemy się krawędzi przerywanych, scalając łączone przez nie wierzchołki. Powiemy, że dwa wierzchołki i oraz j należą do tej samej *składowej* w grafie G, jeśli są połączone przerywaną ścieżką. Każdą składową zastępujemy pojedynczym wierzchołkiem (o numerze będącym *najmniejszym* numerem wierzchołka z tej składowej), do którego wchodzą wszystkie krawędzie ciągłe poprzednio wchodzące do wierzchołków tej składowej.

W tak uzyskanym grafie będziemy zachłannie kolorować wierzchołki w kolejności rosnących numerów. Załóżmy, że chcemy pokolorować wierzchołek i oraz że jest on połączony krawędziami ciągłymi z wierzchołkami o mniejszych numerach $j_1, j_2, \ldots, j_k$. Kluczową własnością grafu, która umożliwi nam kolorowanie zachłanne jest to, że każda para wierzchołków spośród $j_1, j_2, \ldots, j_k$ jest również połączona krawędzią ciągłą, czyli wszystkie one *muszą mieć różne kolory*, zatem wierzchołek i możemy pokolorować na $A - k$ sposobów, niezależnie od kolorowania wierzchołków o mniejszych numerach (rys. 3).

Dowód kluczowej własności jest nieco techniczny i wymaga przyjrzeniu się, jak mogą być położone palindromy w słowie w. Powiemy, że dwa wierzchołki i oraz j są *sąsiednie*, jeśli należą do tej samej składowej, $i < j$ oraz nie istnieje k, że $i < k < j$ i k też należy do tej samej składowej. Dowód można przeprowadzić, udowadniając następujące własności:

- Jeśli i, j są sąsiednie, to $w[i..j]$ jest palindromem.
- Jeśli i, j są sąsiednie, $x < j$ oraz x i j są połączone krawędzią ciągłą, to x i i też są połączone krawędzią ciągłą (wynika z tego, że dla danej składowej wystarczy rozważać krawędzie ciągłe wchodzące do jednego wierzchołka z tej składowej – tego o najmniejszym numerze).
- Jeśli $x < y < j$ i j jest wierzchołkiem o najmniejszym numerze ze swojej składowej oraz x i j są połączone krawędzią ciągłą oraz y i j są połączone krawędzią ciągłą, to istnieje taki wierzchołek y' ze składowej wierzchołka y, że x i y' są też połączone krawędzią ciągłą.

Pozostaje oszacować efektywność powyższego algorytmu. Zbiór (nie)równości konstruowanych na podstawie słowa w, a które odpowiadają krawędziom grafu G może mieć rozmiar $O(n^2)$ i taka też będzie złożoność czasowa i pamięciowa powyższego algorytmu. Zauważmy jednak, że nierówności (czyli krawędzi ciągłych) jest jedynie $O(n)$, pozostałe to równości, czyli krawędzie przerywane, służące do wyznaczenie składowych grafu. Czytelnikom Dociekliwym polecamy zastanowić się, jak można wykorzystać algorytm Manachera znajdowania palindromów w słowie, do zmniejszenia liczby rozważanych krawędzi przerywanych do liniowej.

Tomasz IDZIASZEK