

Zadania domowe. Zestaw 2.1

Maciej Poleski

25 marca 2013

1

$n - p$ jest długością prefixu powiedzmy a . Przedewszystkim a nie jest potęgą całkowitą. Gdyby był, to istniałby dłuższy prefikso sufiks od naszego maksymalnego. Zauważmy że ten wzorec powtarza się (jeżeli $n - p | n$ to $p \not\leq \frac{n}{2}$ bo najmniejszą liczbą pierwszą jest 2). Słowo a jest prefiksem maksymalnego prefikso sufiksu będącego prefiksem słowa w . Skoro tak, to jest również prefiksem maksymalnego prefikso sufiksu będącego sufiksem słowa w . Czyli istnieje prefiks aa . Po wykonaniu dostatecznie wielu kroków okazuje się że cały tekst jest konkatenacją wzorca a z samym sobą dostatecznie wiele razy.

Teraz w drugą stronę. Mamy tekst powiedzmy tej postaci: $aaaaa...aaaa$ (gdzie a oczywiście nie jest potęgą całkowitą (gdyby było, to bierzemy to mniejsze słowo które jest potęgowane)). Oczywiście widać że istnieje prefikso sufiks długości $n - |a|$ i jak widać jego długość dzieli n . Dodatkowo jest to maksymalny prefikso sufiks (bo a nie jest potęgowe).

2

Możliwe są dwie sytuacje. Nasze słowo jest albo prefiksem albo sufiksem wynikowego palindromu. Gdyby tak nie było, to można kasować skrajne znaki aż zredukujemy problem do jednego z dwóch powyższych. Nasze słowo z punktu widzenia wynikowego palindromu może wyglądać na dwa sposoby: $ab\hat{b}$ lub $b\hat{b}a$. Gdy dokleimy do niego jego samego odwróconego uzyskamy $ab\hat{b}b\hat{b}a$ lub $b\hat{b}a\hat{a}bb$. Maksymalizujemy długość prefikso sufiksu (czyli szukamy maksymalnego prefikso sufiksu). Wybieramy tę wersję w której maksymalny prefikso sufiks będzie dłuższy. Pozostaje nam wtedy dokleić do wyjściowego słowa z odpowiedniej strony to co nie zostało dopasowane (odwrócone). Na szczęście w przypadku palindromu nieparzystego nic się nie zmienia (tylko wskakuje dodatkowa litera (np c) i pojawiają się podśłowa $bc\hat{b}$ zamiast $b\hat{b}$).

3

4

5

Będziemy dopasowywać wzorec po kawałku. Umieszczamy tekst w tablicy rezerwując na jej początku dostatecznie dużo miejsca, aby móc dopisać

wzorzec. Robimy przejście po wzorcu dzieląc go na kawałki oddzielone znakami '*' (powtarzające się '***...***' logicznie sklejamy w jeden '*' - nic one nie wnoszą). Każdy tak uzyskany kawałek będziemy doklejać przed resztę tekstu która nam pozostała (odzielając oczywiście jakimś znakiem nie należącym do alfabetu). Mając tak przygotowany napis uruchamiamy standardowy KMP. Z chwilą gdy znajdziemy jakieś dopasowanie (czyli prefiks sufiks równy długości danego fragmentu wzorca) odcinamy ten prefiks tekstu który już przeanalizowaliśmy i do pozostałej doklejamy kolejny fragment wzorca i powtarzamy ten algorytm do oporu (nie ma potrzeby kończyć KMP - dalsze dopasowania nas nie interesują) (aż braknie fragmentów, lub skończy się nie mając dopasowania). Jeżeli zabrakło fragmentów - istnieje dopasowanie. Jeżeli zabrakło tekstu - nie istnieje.

Cały algorytm opiera się na idei zachłannej. Mogąc dopasować dany fragment dopasowuję go tak wcześnie jak to możliwe (nigdy nie jest to błąd). Z kolei fragmenty muszą być dopasowywane w ściśle określonej kolejności. Koszt całości to jedna iteracja po wzorcu + iteracja po wszystkich fragmentach (czyli ponownie jedna po wzorcu) i jedna iteracja po tekście.