Zadanie 4 z listy 6 - “Kompresja Danych”

Łukasz Klasiński

10 maja 2020

Zadanie 4

Pokaż, że dla słów postaci $(a_1a_2 \dots a_m)^*$ algorytm Sequitur wykona $\Omega(n)$ operacji dodania nowej produkcji i $\Omega(n)$ operacji usunięcia produkcji (gdzie n to długość słowa).

Rozwiązanie

1. $\Omega(n)$ operacji dodania nowej produkcji: Zauważmy, że na początku po wczytaniu pojedyńczego $(a_1a_2 \dots a_m)$, będziemy mieli pojedyńczą produkcję:

   $$S \to a_1a_2 \dots a_m$$

   Po wczytaniu kolejnego znaku a_1, nic się nie stanie. Z kolei po dodaniu symbolu a_2, dostaniemy powtórzenie:

   $$S \to a_1a_2 \dots a_m a_1 a_2$$

   Zatem wyrzucimy je i dodamy nową produkcję:

   $$S \to Aa_3 \dots a_m A$$

   $$A \to a_1a_2$$

   Następnie po dodaniu a_3 algorytm zauważy powtórzenie $Aa_3 \dots a_m Aa_3$, zatem stworzy nową produkcję:

   $$S \to Ba_4 \dots a_m B$$

   $$A \to a_1a_2$$

   $$B \to Aa_3$$

   Ale teraz mamy kolejne załamanie niezmiennika - produkcja A występuje tylko raz. Zatem usunie się, i zostaną produkcje

   $$S \to Ba_4 \dots a_m B$$

   $$B \to a_1a_2a_3$$

   Dodajemy a_4 - dostajemy produkcję

   $$S \to Ca_5 \dots a_m C$$

   $$B \to a_1a_2a_3$$

   $$C \to Ba_4$$

I znowu B jest użyte tylko raz, więc zostanie usunięte.

Sytuacja powtarza się aż do dojścia do zakodowania słowa $a_1 \dots a_m a_1 \dots a_m$.

Zobaczmy teraz co się stanie gdy mamy już produkcję S i dodajemy kolejne słowa z cyklu.

Wiemy że poza S mamy jeszcze produkcję $G \to a_1a_2 \dots a_m$. Zatem widać że po dodaniu do S a_1 nic się nie stanie, ponieważ aby coś zmieniać musimy mieć dopasowanie wielkości 2.

Dodajmy a_k. Mamy dwa przypadki:

1. Mamy tylko jedną produkcję $G \to a_1 a_2 \dots a_m$. Wtedy możemy zmachować $a_{k-1}a_k$ z G i stworzyć nową produkcję. Zatem wykonujemy jedną operację dodania produkcji.
2. Mamy produkcję $G \to a_1 \dots K a_k a_{k+1} \dots a_m$ oraz $K \to a_i \dots a_{k-1}$. Wtedy możemy stworzyć nową produkcję $X \to K a_k$, po czym zastąpić w Ka_k w G na X i usunąć K, zamieniając $X \to K a_k$ na to co miało K. Zatem usuwamy i dodajemy po jednej produkcji.
3. Dodajemy ostatnie słowo z cyklu - a_m. Wtedy $G \to K a_m$ oraz $K \to a_1 a_2 \dots a_{m-1}$, zatem po zastąpieniu Ka_m poprzez nową produkcję $X \to K a_m$, będziemy mogli usnąć produkcję K oraz X i otrzymamy $G \to a_1 a_2 \dots a_m$, czyli wrócimy do punktu 1). Łącznie dodajemy produkcję, po czym usuwamy dwie.

Widać zatem, że wykonujemy zamortyzowanie dokładnie jedną operację dodania predykatu oraz jedną operację usunięcia podczas dodawania wszystkich znaków z alfabetu poza a_1. Poza tymi operacjami mamy także dodatkowe operacje dodania predykatu w ramach zawijania produkcji $S \to XXXX$ na nowy predykat (jest *logn* takich operacji).Zatem ostatecznie w algorytmie wykonamy $O(n)$ yoperacji dodawania i usuwania, więc dla odpowiednio dużego n, $\Omega(n)$ jest prawdziwa.