

Zadanie 5 z listy 6 - “Kompresja Danych”

Łukasz Kłasiński

10 maja 2020

Zadanie 5

Znajdź najdłuższe słowo, dla którego gramatyka składa się z co najmniej $n/4$ produkcji (n to długość słowa). Jeśli istnieją dowolnie długie słowa o tej własności, znajdź przykład rodziny takich słów.

Takie słowo:

$$\Sigma = a_1, a_2, a_3, \dots, a_n$$

$$w = a_1 a_1 a_1 a_1 : a_1 a_2 a_1 a_2 : a_1 a_3 a_1 a_3 \dots a_2 a_2 a_2 a_2 : a_2 a_1 a_2 a_1 \dots$$

Dlaczego mamy dokładnie $n/4$ produkcji? Budujemy słowo tak, że doklejamy po 2 unikatowe pary, które wcześniej nie występowały w prefiksie słowa, dzięki czemu tworzymy nową produkcję. Zatem na każde 4 symbole przypada dokładnie jedna produkcja. Po wyczerpaniu wszystkich możliwych kombinacji par zostajemy z produkcją:

$$S \rightarrow A_{11}A_{11} : A_{12}A_{12} : A_{13}A_{13} \dots A_{21}A_{21} : A_{22}A_{22} \dots A_{N0}A_{N0} \dots A_{NN}A_{NN}$$

Zauważmy teraz, że możemy kontynuować takie słowo - dodając od początku $aaaa : abab \dots$, ponieważ nasza para p_{ij} zostanie zamieniona na odpowiadającą mu produkcję A_{ij} , która otrzyma nową produkcję A'_{ij} (ponieważ A_{ij} będzie występować po 2 razy obok siebie). Widzimy że możemy w taki sposób skonstruować dowolnej długości słowo w , gdzie $|w|/4$ to ilość produkcji które powstaną po zaaplikowaniu do niego algorytmu Sequitur.