Kamil Borkowski WCY22IY1S1

TIiK cw 3 Kodowanie słownikowe metodą LZ77

Dane do kodowania:

DEEEAADDEEEEDEEEDDAAAAAABA

Kompresja danych:

Metoda kompresji LZ77 korzysta z bufora słownika o długości n znaków i bufora wejścia o długości m znaków. Bufor słownika przechowuje ostatnio zakodowane symbole. Bufor wejścia przechowuje symbole do zakodowania.

Wielkosci n i m mogą być różne, ale dla różnych wartości można uzyskać różne wyniki kodowania. Spróbuję wyznaczyć takie wielkości n i m, żeby otrzymać najkrótsze dane wyjściowe kodowania z możliwie najmniejszymi wielkościami n i m.

Kompresję danych przeprowadzę na podstawię algorytmu:

* Bufor wejściowy wypełniam m pierwszymi symbolami danych
* Bufor słownika wypełniam początkowym symbolem bufora wejsciowego
* Póki w buforze wejściowym są dane:
  + W buforze słownika szukam najdłuższego podciągu równego początkowi bufora wejściowego
  + Wynikiem jest indeks P z zakresu od 0 do n -1 początku wyszukanego podciągu oraz jego długość C, ograniczona z góry przez m – 1.
  + Jeśli nie znajdę podciągu to P i C będzie równe 0.
  + Na wyjście kodowania wyprowadzam trójkę (P,C,S)
  + Oba bufory są przesuwane w lewo o C + 1 pozycji
  + Jeśli są jeszcze jakies znaki danych to są dopisywane na koniec bufora wejściowego C + 1 danych

Próba 1:

Najpierw zakoduję dane przy użyciu największych wartości n i m, czyli n = m = 26.

[…]

Wynikiem tego kodowania jest osiem trójek i symbol początkowy. W kolejnej próbie użyję znacznie mniejszych wartości n i m, czyli n = m = 4.

Próba 2:

[..]

W tej próbie wyniki są mniej zadowalające, ponieważ otrzymałem o dwie trójki więcej niż w poprzedniej próbie. Spróbuję wyznaczyć najmniejsze n = m, dla których w wyniku będzie się znajdować dziewięć trójek.

Próba 3:

n = m = 16

Próba 4:

N = m = 8

Próba 5:

N = m = 7

Próba 6:

N = m = 6

Próba 7:

N = m = 5

Dla tej próby wartości n i m dają dłuższy wynik kodowania niż przy wcześniejszej próbie. Sprawdzę czy można zmniejszyć wartość bufora wejścia bez wydłużania wyniku kodowania.

Próba 8:

N = 6 , m = 5

[…]

Wynik jest taki sam jak w poprzedniej próbie, co wskazuje, że optymalnymi wartościami dla n i m jest 6.

Dekompresja danych:

Dekompresję danych przeprowadzę na podstawię algorytmu:

* Bufor słownika wypełniam początkowym symbolem
* Dla wszystkich trójek (P,C,S):
  + Kopiuje z bufora słownikowego do bufora wyjściowego symbole z zakresu od P do P + C – 1
  + Dopisuję na koniec bufora wyjściowego symbol S
  + Na wyjście wyprowadzam C + 1 początkowych symboli z bufora wyjściowego
  + Przesuwam zawartość bufora o C + 1 pozycji w lewo