

Laboratorium 5 — Kompresja bezstratna (RLE, ByteRun)

Celem jest implementacja i porównanie kompresji bezstratnej RLE i ByteRun dla tablic NumPy (obrazy/wekory), z zachowaniem informacji o kształcie wewnętrz jednego wektora danych skompresowanych.

Uwagi dot. pamięci

Nagłówek skompresowanej postaci zawiera liczbę wymiarów i wymiary oryginalnej tablicy: [ndim, dim0, dim1, ...]. Dalej następuje strumień danych skompresowanych (RLE: pary [count, value]; ByteRun: naprzemienne znaczniki i wartości zgodnie z opisem).

Wyniki — zbiory testowe (tablice)

Test: rep_short

Algorytm	rozmiar [B]	kompr. [B]	CR	%
RLE	112	96	1.167	85.7
ByteRun	112	112	1.000	100.0

Test: cycle

Algorytm	rozmiar [B]	kompr. [B]	CR	%
RLE	72	160	0.450	222.2
ByteRun	72	96	0.750	133.3

Test: mix

Algorytm	rozmiar [B]	kompr. [B]	CR	%
RLE	104	160	0.650	153.8
ByteRun	104	160	0.650	153.8

Test: neg_vals

Algorytm	rozmiar [B]	kompr. [B]	CR	%
RLE	96	144	0.667	150.0
ByteRun	96	120	0.800	125.0

Test: zeros520

Algorytm	rozmiar [B]	kompr. [B]	CR	%
RLE	4160	40	104.000	1.0
ByteRun	4160	104	40.000	2.5

Test: range521

Algorytm	rozmiar [B]	kompr. [B]	CR	%
RLE	4168	8352	0.499	200.4
ByteRun	4168	4224	0.987	101.3

Test: eye7

Algorytm	rozmiar [B]	kompr. [B]	CR	%
RLE	392	232	1.690	59.2
ByteRun	392	280	1.400	71.4

Test: stack_eye

Algorytm	rozmiar [B]	kompr. [B]	CR	%
RLE	1176	240	4.900	20.4
ByteRun	1176	240	4.900	20.4

Test: ones10

Algorytm	rozmiar [B]	kompr. [B]	CR	%
RLE	80	80	1.000	100.0
ByteRun	80	80	1.000	100.0

Wyniki — obrazy (jeśli wczytane)

Obraz: dock.jpg

Algorytm	rozmiar [B]	kompr. [B]	CR	%
RLE	96000000	115610832	0.830	120.4

ByteRun	96000000	62336696	1.540	64.9
---------	----------	----------	-------	------

Obraz: kwiaty.jpg

Algorytm	rozmiar [B]	kompr. [B]	CR	%
RLE	439092360	864224032	0.508	196.8
ByteRun	439092360	447439000	0.981	101.9

Obraz: techniczny.jpg

Algorytm	rozmiar [B]	kompr. [B]	CR	%
RLE	96000000	2968768	32.337	3.1
ByteRun	96000000	4216016	22.770	4.4

Wnioski

RLE skutecznie kompresuje długie serie identycznych wartości (np. tła, formularze, rysunki), natomiast ByteRun radzi sobie lepiej także z fragmentami bez powtórzeń dzięki trybowi literalnemu. Efektywność zależy od zawartości: fotografie często kompresują się słabiej. Roundtrip zachowuje identyczność danych.