

Kompresja danych

Czas: 2 × 45 min

praca indywidualna lub w parach

Cel: Zbadanie, jak format archiwum i poziom kompresji wpływają na stopień i czas kompresji różnych typów plików, oraz wyjaśnienie, dlaczego niektóre pliki prawie się nie kompresują.

Po zajęciach potrafisz:

- kompresuje pliki w formatach zip, 7z i gzip na różnych poziomach w 7-Zip (interfejs graficzny i wiersz poleceń)
- mierzy czas kompresji i oblicza stopień kompresji
- opracowuje wyniki w arkuszu w postaci tabeli i wykresu
- porównuje 7-Zip z innym programem do kompresji
- wyjaśnia, dlaczego pliki jpg, mp3 i mp4 słabo się kompresują

Sprzęt i materiały

- komputer stacjonarny (system szkolny lub Windows 11 z zadania 13)
- pendrive z Ventoy (na pliki testowe i wyniki)

Oprogramowanie

- **7-Zip** — formaty zip, 7z, gzip; wiersz poleceń `7z.exe`
7-zip.org/download.html
- **PeaZip** — porównanie: formaty ZPAQ, Zstandard (zst), Brotli; wersja portable
peazip.github.io
- **Arkusz kalkulacyjny (Excel lub LibreOffice Calc)** — import CSV, tabela i wykres

Przygotowanie przed zajęciami

- Przygotuj folder `pliki` z plikami testowymi o rozmiarze ok. 2 MB i ok. 20 MB (po jednym z każdego wiersza): grafika — jpg, png, bmp; dźwięk — mp3, wav, flac; wideo — mp4 (tylko 20 MB); tekst — txt, docx; inne — pdf, exe. Nauczyciel udostępnia zestaw na serwerze szkolnym — możesz go użyć.
- Pliki pobieraj tylko z legalnych źródeł (własne zdjęcia i nagrania, domena publiczna, np. teksty z wolnelektury.pl).

Bezpieczeństwo

- Nie umieszczaj w plikach testowych danych osobowych (zdjęć innych osób, dokumentów) — wyniki trafiają do sprawozdania.
- Nie rozpakowuj archiwów z nieznanych źródeł; pliki exe służą tylko do kompresji, nie uruchamiaj ich.
- Po zajęciach usuń z komputera folder z plikami testowymi i archiwami.

Przebieg zadania

1 Kompresja ręczna — jeden plik

Poznanie okna 7-Zip.

1. Kliknij prawym przyciskiem plik txt 20 MB → Pokaż więcej opcji → 7-Zip → Dodaj do archiwum...
2. Ustaw format 7z i poziom kompresji „Najszybsza”, zatwierdź. Odczytaj w oknie postępu czas, a w Eksploratorze rozmiar archiwum.
3. Powtórz dla poziomów „Normalna” i „Ultra”. Zapisz wyniki i zwróć uwagę na metodę kompresji i rozmiar słownika wybierane dla każdego poziomu.
4. Rozpakuj archiwum „Ultra” do innego folderu. Porównaj sumy SHA-256 pliku oryginalnego i rozpakowanego (`Get-FileHash plik.txt`) — potwierdź, że kompresja jest bezstratna.

2 Pomiary automatyczne

PowerShell, w folderze nadrzędnym względem folderu `pliki`. Poziomy: 1 — najszybszy, 5 — normalny, 9 — ultra.

1. Utwórz folder wyników: `mkdir wyniki` i ustaw ścieżkę programu: `$z = "C:\Program Files\7-Zip\7z.exe"`.
2. Uruchom pętlę:

```
foreach ($f in Get-ChildItem .\pliki -File) { foreach ($t in 'zip','7z','gzip') { foreach ($m in 1,5,9) { $o = "wyniki\$(($f.Name)).$m.$t"; $s = (Measure-Command { & $z a "-t$t" "-mx=$m" $o $f.FullName | Out-Null }).TotalSeconds; "$($f.Name);$($f.Length);$t;$m;$((Get-Item $o).Length);$s" >> wyniki.csv } } }
```
3. Poczekaj na zakończenie (kilka minut). W tym czasie wpisz do sprawozdania przewidywania: które typy plików skompresują się najlepiej, a które najgorzej.
4. Otwórz `wyniki.csv` w arkuszu (separator: średnik). Dodaj nagłówki: plik; rozmiar przed [B]; format; poziom; rozmiar po [B]; czas [s].

3 Opracowanie w arkuszu

1. Dodaj kolumnę „stopień kompresji [%]” = $(1 - \text{rozmiar po} / \text{rozmiar przed}) \times 100$.
2. Oblicz średni stopień kompresji dla każdej kategorii plików (grafika, dźwięk, wideo, tekst, inne).
3. Zbuduj tabelę przestawną lub zestawienie: typ pliku × format i poziom → stopień kompresji.
4. Narysuj wykres kolumnowy stopnia kompresji dla plików 20 MB (format 7z, poziom 9) oraz wykres czasu kompresji pliku txt 20 MB dla wszystkich formatów i poziomów.

4 Porównanie z innym programem

Jeden typ plików (np. tekstowe lub bmp).

1. W PeaZip skompresuj wybrane pliki formatem ZPAQ na najwyższym poziomie oraz Zstandard (zst) na najwyższym poziomie. Zapisz rozmiar i czas (stoper lub okno postępu).
2. Porównaj z najlepszym wynikiem 7-Zip dla tych samych plików. Wskaż, który program i format spakował dane najmocniej i jakim kosztem czasu.
3. Spróbuj ponownie skompresować gotowe archiwum 7z (poziom 9). Zapisz wynik i wyjaśnij go.

W sprawozdaniu umieść

- ☐ wyniki kompresji ręcznej pliku txt (trzy poziomy)
- ☐ tabela wyników z pliku CSV ze stopniem kompresji
- ☐ wykres stopnia kompresji według typu pliku i wykres czasu kompresji
- ☐ wynik porównania z PeaZip
- ☐ wyjaśnienie, które pliki kompresują się dobrze, a które słabo, i dlaczego
- ☐ odpowiedzi na pytania kontrolne

Pytania kontrolne

1. Dlaczego pliki jpg, mp3 i mp4 prawie nie zmniejszają się po kompresji, a bmp, wav i txt zmniejszają się wyraźnie?
2. Czym różni się kompresja bezstratna od stratnej? Podaj po dwa przykłady formatów.
3. Dlaczego format gzip pakuje tylko jeden plik i jak w Linuksie archiwizuje się cały folder (tar.gz)?
4. Kiedy warto wybrać najszybszy poziom kompresji zamiast ultra?

Kryteria oceny

Kryterium	Punkty
Kompletne pomiary (ręczne i automatyczne)	5
Tabela ze stopniem kompresji i poprawne wykresy	5
Porównanie z innym programem	3
Wyjaśnienie różnic między typami plików	3
Odpowiedzi na pytania kontrolne	4
Razem	20

Sprawozdanie przygotuj na wzorze dołączonym do zadania (plik .docx o tym samym numerze).