

# Temat 3

---

## To można odtworzyć! – *Kompresja tekstu*

### Streszczenie

Pamięć, na której przechowywane są informacje w komputerach, ma ograniczoną pojemność, dlatego dąży się do tego, by informacje zapisywane były w sposób możliwie efektywny. W tym celu stosuje się kompresję (upakowanie) informacji. Przed zapisaniem danych w pamięci czy przesłaniem przez sieć są one kompresowane (pakowane), a podczas ich odczytywania lub pobierania – dekompresowane (rozpakowywane). Dzięki temu w pamięci komputera można zapisać większą ilość danych czy przesłać dane przez sieć w krótszym czasie.

### Wiek

- ✓ 9 i więcej
- ✓

### Materiały

- ✓ Folia do rzutnika pisma (grafoskopu): To można odtworzyć! (s. 25)

Każde dziecko będzie potrzebować następujących kart pracy:

- ✓ To można odtworzyć! (s. 26)
- ✓ To też można odtworzyć! (s. 27)
- ✓ Krótki, ale przyjemny (s. 28)
- ✓ Dla ambitnych (s. 29)

# Można to odtworzyć!

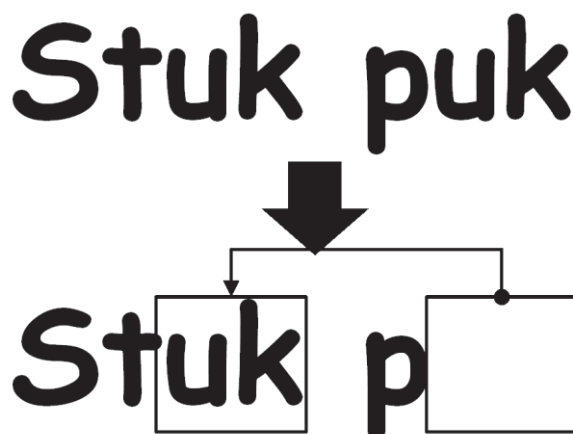
---

## Wprowadzenie

W systemach komputerowych przechowuje się i przekazuje przez sieć wielką liczbę danych. Aby ograniczyć ilość używanej pamięci lub skrócić czas przesyłania przez sieć, komputery kompresują (pakują) informacje. Sposób kompresji informacji tekstowych podobny jest do ukazanego poniżej.

## Prezentacja tematu i dyskusja

Spójrz na folię „Deszcz jesienny” (s. 25), zawierający fragment wiersza Leopolda Staffa. Przyjrzyj się grupom liter w tym wierszu. Czy można odnaleźć grupy dwóch lub więcej liter, które powtarzają się, a może nawet całe wyrazy czy frazy? (Zaznacz takie powtórzenia w sposób ukazany na poniższym diagramie.)



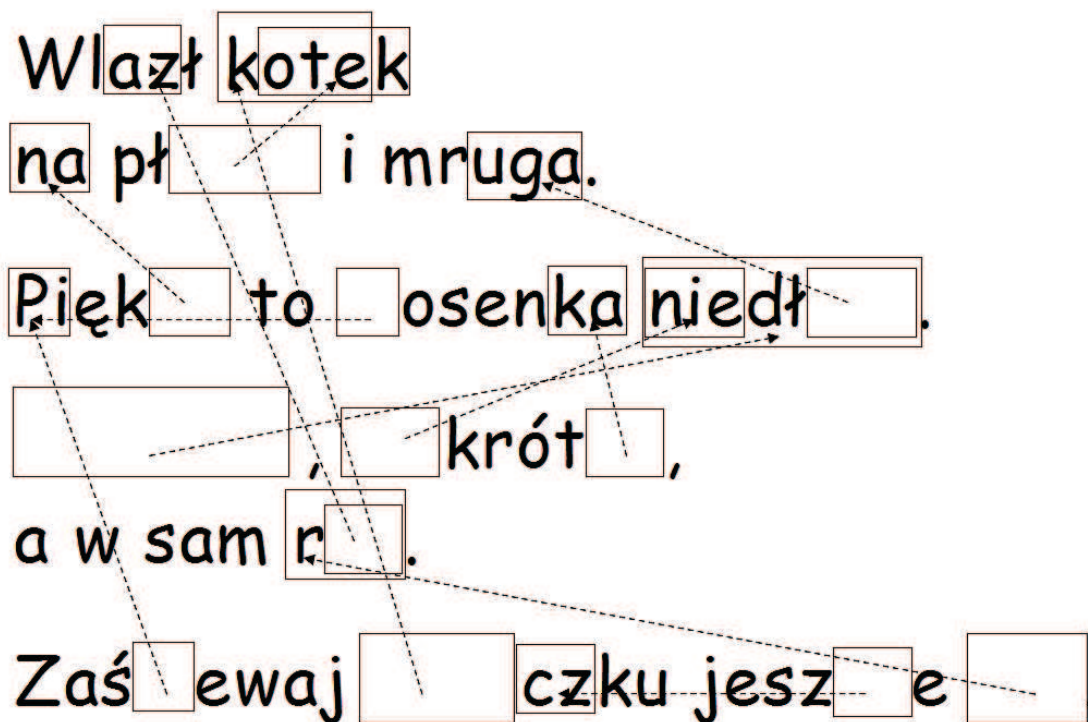
## Deszcz jesienny

O szyby deszcz dzwoni,  
deszcz dzwoni jesienny,

I pluszcze jednaki,  
miarowy, niezmienny,

## Karta pracy: Można to odtworzyć!

Z wiersza usunięto wiele słów i grup liter. Czy jest możliwe odtworzenie tekstu wiersza? Rozwiąż łamigłówkę (Za pomocą strzałek zaznaczono lokalizację tzw. wzorców.)



Wybierz jakiś prosty wiersz lub rymowankę i zaprojektuj swoją własną łamigłówkę. Pamiętaj o tym, że strzałki muszą zawsze wskazywać na wcześniejszy fragment tekstu. Tekst wiersza powinien być możliwy do odtworzenia zgodnie z kierunkiem czytania (od lewej do prawej i z góry na dół).

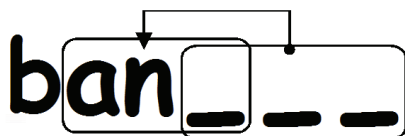
Sprawdź, jak wiele słów zostało usuniętych z oryginalnego tekstu!

**Wskazówka:** Unikaj nagromadzenia zbyt dużej liczby strzałek w jednym miejscu. Przygotuj odpowiednio dużo miejsca na zapisywanie słów wiersza (zwłaszcza na wypadek umieszczenia ramki wzorca wewnątrz innej ramki.)

Znacznie łatwiej przygotować taką łamigłówkę, jeśli najpierw zapiszesz cały tekst wiersza i na nim zaznaczysz odpowiednie prostokąty.

## Karta pracy: To też można odtworzyć!

Jakie jest rozwiązanie poniższej łamigłówki?



Czasami końcowa część wzorca stanowi początkową część usuniętego tekstu. W takim przypadku tekst można odtworzyć, jeśli litery odtwarzane są w kierunku od lewej do prawej. Odpowiednia część wzorca jest wtedy dostępna w momencie, kiedy jest potrzebna. Jest to użyteczne w przypadku, kiedy tekst zawiera dłuższy ciąg powtarzających się grup liter.

Spróbuj narysować własny przykład.

W komputerach rolę prostokątów i strzałek pełnią liczby. Na przykład słowo

**banana**

(tzn. rzeczownik *banan* w dopełniaczu) może być zapisane jako **ban(2, 3)**.

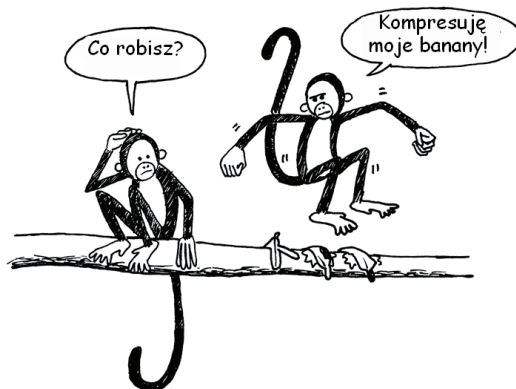
Liczba "2" wskazuje liczbę znaków, o jaką należy cofnąć się, by dotrzeć do pierwszego znaku wzorca

**ban\_ \_ \_**

a liczba "3" wskazuje liczbę znaków wzorca:

**ban\_ \_**

**banana**



Ponieważ do zakodowania (kompresji) takiego słowa używa się dwóch liczb, zwykle tylko grupy dwóch lub większej liczby liter poddaje się takiemu kodowaniu. W przeciwnym wypadku nie ma mowy o zmianie objętości danych. W skrajnym przypadku, wielkość zbioru danych powiększa się, jeśli do zakodowania jednej litery używa się dwóch liczb.

Zapisz kilka kodów słów z użyciem par liczb. Sprawdź, czy inni potrafią je odkodować.

## Karta pracy: Krótki, ale przyjemny

### Jakiej minimalnej liczby słów można użyć?

Wyobraź sobie siebie w roli komputera, którego zadaniem jest zapisanie na dysku tak dużej liczby danych, jaka jest to tylko możliwa. Usuń z tekstu wszystkie grupy dwóch lub większej liczby liter, które w tekście pojawiły się już wcześniej. Zastąp je odpowiednim wskaźnikiem wzorca. Usuń tak dużo liter, jak jest to możliwe.

Na jednym drzewie orzeł gdy z sową nocował,  
Że tylko w nocy widzi, bardzo jej żałował.  
Dziękowała mu sowa za politowanie.  
Wtem, uprzedzając jeszcze zorze i świtanie,  
Wkradł się strzelec pod drzewo; sowa to postrzegła  
I do orła natychmiast z przestroga pobiegła.  
Uszli śmierci; a wtenczas rzekł orzeł do sowy:  
„Gdybyś nie była ślepa, nie byłbym ja zdrowy”.

## Karta pracy: Dla ambitnych

### Czy jesteś przygotowany na naprawdę mocną kompresję?

Poniższy tekst w języku angielskim był kompresowany przez program komputerowy. W efekcie, z tekstu zostały usunięte 1633 litery.

A Ty ile potrafisz usunąć? Powodzenia!

Once upon a time, long, long ago, three little pigs set out to make their fortunes. The first little pig wasn't very clever, and decided to build his house out of straw, because it was cheap. The second little pig wasn't very clever either, and decided to build his house out of sticks, for the "natural" look that was so very much in fashion, even in those days. The third little pig was much smarter than his two brothers, and bought a load of bricks in a nearby town, with which to construct a sturdy but comfortable country home.

Not long after his housewarming party, the first little pig was curled up in a chair reading a book, when there came a knock at the door. It was the big bad wolf, naturally.

"Little pig, little pig, let me come in!" cried the wolf.

"Not by the hair on my chinny-chin-chin!" squealed the first little pig.

"Then I'll huff, and I'll puff, and I'll blow your house down!" roared the wolf, and he *did* huff, and he *did* puff, and the house soon collapsed. The first little pig ran as fast as he could to the house of sticks, and was soon safe inside. But it wasn't long before the wolf came calling again.

"Little pig, little pig, let me come in!" cried the wolf.

"Not by the hair on my chinny-chin-chin!" squealed the second little pig.

"Then I'll huff, and I'll puff, and I'll blow your house down!" roared the wolf, and he *did* huff, and he *did* puff, and the house was soon so much firewood. The two terrified little pigs ran all the way to their brother's brick house, but the wolf was hot on their heels, and soon he was on the doorstep.

"Little pig, little pig, let me come in!" cried the wolf.

"Not by the hair on my chinny-chin-chin!" squealed the third little pig.

"Then I'll huff, and I'll puff, and I'll blow your house down!" roared the wolf, and he huffed, and he puffed, and he huffed some more, but of course, the house was built of brick, and the wolf was soon out of breath. Then he had an idea. The chimney! He clambered up a handy oak tree onto the roof, only to find that there *was* no chimney, because the third little pig, being conscious of the environment, had installed electric heating. In his frustration, the wolf slipped and fell off the roof, breaking his left leg, and severely injuring his pride. As he limped away, the pigs laughed, and remarked how much more sensible it was to live in the city, where the only wolves were in the zoo. And so that is what they did, and of course they all lived happily ever after.

# O co w tym wszystkim chodzi?

---

Pojemność pamięci używanych przez komputery rośnie w niewiarygodnym tempie – w ciągu minionych 30 lat wielkość dysku komputerowego zwiększyła się milion razy, mimo to ciągle wydaje się ona w praktyce niewystarczająca. Na dyskach zapisywane są dziś już nie tylko dane tekstowe, ale np. muzyka i filmy. Wielkość plików stanowi również problem w przypadku przesyłania ich przez sieć Internet. Poza tym dochodzi problem miniaturyzacji – nawet w przypadku telefonu komputerowego oczekuje się, że będzie na nim można zapisać odpowiednio dużo informacji, również multimedialnych.

Istnieje rozwiązanie tego problemu. Zamiast kupować większe dyski lub szybsze modemy, możemy dane skompresować, tak aby zajmowały mniej pamięci. Proces kompresji i dekompresji jest wykonywany automatycznie przez komputer. To przetwarzanie danych dodatkowo obciąża komputer, ale dzięki temu strony WWW szybciej się ładują, a dysk twardy może więcej pomieścić.

Wymyślono wiele sposobów kompresji. Metoda, która używana była w czasie tych zajęć (oparta o pomysł użycia liczb wskazujących, kiedy wcześniej wystąpił dany ciąg znaków i z ilu bajtów się składał) jest nazywana kodowaniem słownikowym „Ziv-Lempel” lub kodowaniem „LZ” i wymyślona została przez dwóch izraelskich profesorów w latach 70-tych XX w. Może być używana dla teksów zapisanych w dowolnym języku i zwykle prowadzi do zmniejszenia objętości kodowanych danych co najmniej o połowę. Używana jest m.in. przy popularnej kompresji w formacie .zip, przy zapisie grafiki w formacie GIF oraz w modemach.

Inne metody bazują na pomycie związania długości kodów odpowiednich liter i częstości ich występowania: litery częściej używane mają krótsze kody niż inne. Metoda ta była już wcześniej używana w tzw. alfabecie Morse’a.

## Rozwiązania i wskazówki

---

**Można to odtworzyć!** (s. 26)

**Wlazł kotek  
na płotek i mruga.  
Piękna to piosenka niedługa.  
Niedługa, niekrótka,  
A w sam raz.  
Zaśpiewaj koteczku jeszcze raz.**