

# Problem 11: Łamacz szyfrów

Punkty: 30

Autor: Brett Reynolds, Annapolis Junction, Maryland, Stany Zjednoczone

## Wprowadzenie

Szyfr podstawieniowy to szyfr, który zastępuje jedne litery (lub ewentualnie grupy liter) innymi literami (lub ich grupami), szyfrując wiadomość i uniemożliwiając jej odczytanie. Przykładowo, szyfr Cezara, który był prawdopodobnie pierwszym stworzonym szyfrem, wymaga przesunięcia liter alfabetu. Każda litera w pierwotnej, niezaszyfrowanej wiadomości jest zastępowana inną literą, określaną przez przesunięcie o pewną liczbę liter do końca alfabetu; w ten sposób tworzy się tekst zaszyfrowany. W przypadku alfabetu angielskiego przesunięcie o 3 spowoduje, że A zastępujemy literą D, B staje się E itd. (X, Y i Z są z kolei zastępowane odpowiednio literami A, B i C, ze względu na zawinięcie do początku alfabetu).

Tekst nieszyfrowany:                    **ABCDEFGHIJKLMNOPQRSTUVWXYZ**

Przesunięcie Cezariańskie 3:        **DEFGHIJKLMNOPQRSTUVWXYZABC**

Szyfry podstawieniowe mają jednak istotną słabość; niektóre litery w alfabecie występują częściej niż inne. W szczególności samogłoski (w alfabecie angielskim: A, E, I, O i U) są często występującymi literami w każdym języku, ponieważ jest ich mało, a jednak są niezbędne, by słowa były możliwe do wymówienia. Pewne spółgłoski również występują częściej od innych (w angielskim R, S i T są powszechniejsze niż Q, X i Z). Jeśli kryptolog policzy, ile razy w danej wiadomości zakodowanej szyfrem podstawieniowym występują pewne litery, to zwykle jest w stanie złamać szyfr. Tworząc wykres przedstawiający względną częstotliwość występowania każdej litery w tekście zaszyfrowanym i porównując ją z średnią częstotliwością występowania liter w podejrzanym języku wiadomości kryptolog może zgadnąć, jakiego podstawienia użyto i stopniowo odszyfrować pierwotny przekaz.

## Opis problemu

W nowym zespole Lockheed Martin ds. cyberbezpieczeństwa pracujecie dla Narodowej Agencji Bezpieczeństwa w celu wykonania nowej bazy danych do analizy częstotliwościowej. Musicie napisać program, który jest w stanie odczytywać duże ilości tekstu i policzyć częstość występowania każdej litery angielskiego alfabetu. Następnie ma podać względną częstotliwość każdej liczby na liście, co zostanie później użyte do łamania szyfrów.

Względną częstotliwość danej litery ustala się korzystając z poniższego równania i wyraża jako wartość procentową:

$$\text{Względna częstotliwość} = \left( \frac{\text{Wystąpienia liter}}{\text{Całkowita liczba liter}} \right) \times 100\%$$

Przykładowo, w zdaniu „The cow is brown” mamy 16 liter. Litera „o” występuje dwukrotnie, co daje względną częstotliwość  $(2 \div 13) \times 100\% = 15.38\%$ . Przy obliczaniu względnych częstotliwości nie uwzględnia się znaków przestankowych i spacji. Nie są też rozróżniane wielkie i małe litery.

## Przykładowe dane wejściowe

Pierwszy wiersz danych wejściowych programu, otrzymanych przez standardowy kanał wejściowy, będzie zawierał dodatnią liczbę całkowitą oznaczającą liczbę przypadków testowych. Każdy przypadek testowy będzie zawierał:

- Wiersz zawierający dodatnią liczbę całkowitą, A, odpowiadającą liczbie zadanych wierszy tekstu.
- X wierszy, które będą zawierać analizowany tekst. Wiersze mogą mieć maksymalnie 2.000 znaków, a każdy z nich może zawierać dowolne znaki US-ASCII.

1

3

The quick red fox jumps over the lazy brown dog.

The above sentence contains every letter in the English language.

Don't forget to ignore punctuation and numbers; they're not relevant!

## Przykładowe dane wyjściowe

W każdym przypadku testowym wasz program musi wyświetlić jeden wiersz dla każdej litery angielskiego alfabetu, w kolejności alfabetycznej, w poniższym formacie:

- Litera (wielka)
- Dwukropek (:)
- Spacja
- Względna częstotliwość tej litery w tekście podanym w przypadku testowym, zaokrąglona do dwóch miejsc dziesiętnych i z ewentualnymi zerami końcowymi
- Znak procentów (%)

A: 5.37%  
B: 2.01%  
C: 2.68%  
D: 2.68%  
E: 15.44%  
F: 1.34%  
G: 4.03%  
H: 4.03%  
I: 4.03%  
J: 0.67%  
K: 0.67%  
L: 3.36%  
M: 1.34%  
N: 10.74%  
O: 8.05%  
P: 1.34%  
Q: 0.67%  
R: 6.71%  
S: 3.36%  
T: 10.74%  
U: 4.03%  
V: 2.68%  
W: 0.67%  
X: 0.67%  
Y: 2.01%  
Z: 0.67%