

**Akademia Techniczno-Humanistyczna**

**w Bielsku-Białej**

Wydział Budowy Maszyn i Informatyki

**Bezpieczeństwo technologii informatycznych**

**(ćwiczenia laboratoryjne)**

**Ćwiczenie numer:**

**1**

**Czas realizacji zajęć:** 90 min.

**Temat ćwiczenia:**

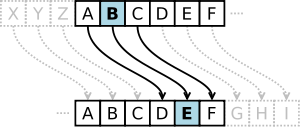
**Szyfr Cezara**

**1. Wstęp teoretyczny.**

**Szyfr Cezara**

**Szyfr Cezara** (zwany też szyfrem przesuwającym, kodem Cezara lub przesunięciem Cezariańskim) – jedna z najprostszych technik szyfrowania. Jest to rodzaj szyfru podstawieniowego, w którym każda litera tekstu jawnego (niezaszyfrowanego) zastępowana jest inną, oddaloną od niej o stałą liczbę pozycji w alfabecie, literą (szyfr monoalfabetyczny), przy czym kierunek zamiany musi być zachowany. Nie rozróżnia się przy tym liter dużych i małych. Nazwa szyfru pochodzi od Juliusza Cezara, który prawdopodobnie używał tej techniki do komunikacji ze swymi przyjaciółmi.

Algorytm szyfrowania zastosowany w kodzie Cezara bywa fragmentem bardziej złożonych systemów szyfrowania, takich jak szyfr Vigenère'a. Współcześnie szyfru Cezara używa się z przesunięciem 13 (ROT13), będącego prostym i szybkim sposobem na ukrycie treści. Obecnie szyfr Cezara, jak każda technika podmieniająca pojedyncze litery alfabetu na inne, nie oferuje żadnego bezpieczeństwa komunikacji.

[](https://pl.wikipedia.org/wiki/Plik:Caesar3.svg)

Szyfr Cezara zastępuje każdą literę tekstu jawnego inną, przesuniętą względem litery kodowanej o stałą liczbę pozycji w alfabecie. Na rysunku szyfr z przesunięciem równym 3, tak więc B w tekście jawnym jest podmieniane w szyfrogramie na E (rozpatrywany jest alfabet łaciński).

**Przykład**

Sposób szyfrowania może być przedstawiony za pomocą diagramu dwóch ciągów z odpowiadającymi sobie kolejnymi literami alfabetu. Te same litery drugiego ciągu są przesunięte względem ciągu pierwszego o określoną liczbę pozycji, zwaną parametrem przesunięcia (tutaj 2) i pełniącą funkcję [klucza](https://pl.wikipedia.org/wiki/Klucz_(kryptografia)) szyfru:

Alfabet: AĄBCĆDEĘFGHIJKLŁMNŃOÓPRSŚTUWYZŹŻ

Szyfr: CĆDEĘFGHIJKLŁMNŃOÓPRSŚTUWYZŹŻAĄB

Należy przy tym zauważyć, że ostatnim literom alfabetu w górnym ciągu odpowiadają początkowe litery w ciągu dolnym (alfabet został „zawinięty”). Chcąc zaszyfrować wiadomość, należy każdą jej literę zastąpić odpowiednikiem z szyfru (wiadomość w przykładzie jest zapisana [wersalikami](https://pl.wikipedia.org/wiki/Majusku%C5%82a), aczkolwiek szyfr jest niewrażliwy na wielkość liter):

Tekst jawny: MĘŻNY BĄDŹ, CHROŃ PUŁK TWÓJ I SZEŚĆ FLAG

Tekst zaszyfrowany: OHBÓŻ DĆFĄ, EKTRP ŚZŃM YŹSŁ L UAGWĘ INCJ

Deszyfrowanie polega na odwróceniu tej operacji.

**Ujęcie matematyczne**

Operację szyfrowania i deszyfrowania można wyrazić w języku [arytmetyki modularnej](https://pl.wikipedia.org/wiki/Arytmetyka_modularna). W tym celu wystarczy każdej literze alfabetu jednoznacznie przypisać jej numer według schematu A↔0, Ą↔1, B↔2, …, Ż↔31. Wygodnie jest też przyjąć, że klucz n jest pewną liczbą z zakresu 0...31 (jest to numer zaszyfrowanej litery A).

Szyfrowanie można wtedy zdefiniować za pomocą [kongruencji](https://pl.wikipedia.org/wiki/Kongruencja_(algebra)):

E_n(x) \equiv x + n \pmod {32},

gdzie x jest numerem litery tekstu jawnego w alfabecie, E_n(x) – numerem litery szyfrogramu w alfabecie.

Podobnie deszyfrowanie tekstu można zapisać jako:

D_n(y) \equiv y - n \pmod {32}

gdzie y jest numerem litery szyfrogramu w alfabecie, D_n(y) – numerem litery tekstu jawnego w alfabecie.

Na podstawie własności kongruencji i tego, że E_n(x), D_n(y) są z przedziału 0...31:

* jeśli przy wyznaczaniu E_n(x) wartość wyrażenia x + n przekroczy 32 – 1, to należy ją zmniejszyć o 32.
* jeśli przy wyznaczaniu D_n(x) wartość wyrażenia y - n będzie ujemna, to należy ją zwiększyć o 32.

Operacje E_n(x), D_n(y) są do siebie odwrotne, bowiem przesuwanie w prawo o k jest zarazem przesuwaniem w lewo o 32-k.

Liczba 32 powyżej jest liczbą liter w [alfabecie polskim](https://pl.wikipedia.org/wiki/Alfabet_polski). Dla [alfabetu łacińskiego](https://pl.wikipedia.org/wiki/Alfabet_%C5%82aci%C5%84ski) należy przyjąć liczbę 26

**Kryptoanaliza.**

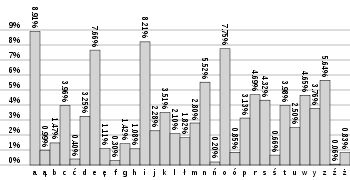
Nie istnieją żadne źródła mówiące o technice złamania prostych szyfrów podstawieniowych w starożytności. Pierwsze udokumentowane metody łamania takich szyfrów autorstwa arabskiego filozofa [Al-Kindi](https://pl.wikipedia.org/wiki/Al-Kindi) pochodzą z IX wieku, gdy zaczęto stosować [analizę częstościową](https://pl.wikipedia.org/wiki/Atak_statystyczny).

|  |  |
| --- | --- |
| **Przesunięcie deszyfrujące** | **Otrzymana treść** |
| 0 | DŹDŃŚADRMI |
| 1 | ĆZĆNSŻĆPŁH |
| 2 | CYCMRŹCÓLG |
| 3 | BWBŁPZBOKF |
| 4 | ĄUĄLÓYĄŃJĘ |
| 5 | ATAKOWANIE |
| 6 | ŻŚŻJŃUŻMHD |
| … | |
| 29 | FĄFPWCFTOL |
| 30 | ĘAĘÓUBĘŚŃK |
| 31 | EŻEOTĄESNJ |

Obecnie wypracowane są techniki łamania szyfru Cezara. Można go bardzo łatwo złamać nawet wtedy, gdy dostępny jest wyłącznie szyfrogram, o ile zachodzi jedna z poniższych możliwości:

1. wiadomo (lub przypuszcza się), że zastosowano jakiś prosty szyfr podstawieniowy, ale nie wiadomo, czy jest to szyfr Cezara;
2. wiadomo, że zastosowano szyfr Cezara, ale nieznane jest przesunięcie, jakiego użyto do zakodowania wiadomości.

W pierwszym przypadku szyfr może zostać łatwo złamany przez zastosowanie tych samych technik, których używa się do łamania innych szyfrów podstawieniowych, jak np. [atak statystyczny](https://pl.wikipedia.org/wiki/Atak_statystyczny). Jest bardzo prawdopodobne, że osoba chcąca rozszyfrować tekst szybko spostrzeże pewną prawidłowość w szyfrogramie i wywnioskuje, że do zakodowania użyto szyfru Cezara.

[](https://pl.wikipedia.org/wiki/Plik:Polish_letters_frequencies.svg)

Wykres częstości liter polskiego alfabetu w tekście napisanym po polsku ma charakterystyczny i przewidywalny kształt. Szyfr Cezara „przesuwa” ten rozkład, dzięki czemu przy dostatecznie długim tekście możliwe jest ustalenie przesunięcia szyfrującego poprzez przyrównanie rozkładu liter w szyfrogramie do tego wykresu.

W drugim przypadku szyfr Cezara jest jeszcze prostszy do złamania. Ponieważ istnieje skończona liczba możliwych przesunięć (32 w języku polskim), każda z kombinacji może być przetestowana [atakiem *brute force*](https://pl.wikipedia.org/wiki/Atak_brute_force). Jednym ze sposobów uczynienia tego jest wypisanie fragmentu zaszyfrowanego tekstu w tabeli razem z jego wszystkimi możliwymi przesunięciami – technika zwana czasem "układaniem elementu jawnego" (ang.*completing the plain component*). Niech przykładowym szyfrogramem będzieDŹDŃŚADRMI. Bez wnikliwej analizy można od razu zauważyć, że rozkodowanie tekstu następuje przy przesunięciu 5. Istnieje jeszcze jeden sposób wykorzystania takiej tabeli. Pod każdą literą szyfrogramu wypisywane są w odwrotnej kolejności poprzedzające ją litery w alfabecie. Można wykorzystać ten fakt w celu szybszego rozkodowania tekstu, używając zestawu pasków z pionowo wypisanymi literami alfabetu także w odwrotnej kolejności. Po dopasowaniu poszczególnych liter z pasków do szyfrogramu, rozkodowany tekst będzie widoczny w którymś z ułożonych rzędów.

Innym rodzajem ataku *brute force* jest dopasowanie rozkładu częstości liter. Tworząc wykres częstości liter w szyfrogramie oraz znając rozkład tych liter w języku, w którym został zapisany tekst jawny, można z łatwością rozpoznać wartość przesunięcia szyfrującego poprzez zaobserwowanie przemieszczenia układu najwyższych słupków wykresu odpowiadającym najpopularniejszym literom. Na przykład w języku polskim najczęściej używanymi w tekstach są litery A, I, O, E, a charakterystyczny utworzony przez nie układ jest widoczny w rozkładzie większości dostatecznie długich szyfrogramów. W bardziej zaawansowanej wersji można użyć [metod statystycznych](https://pl.wikipedia.org/wiki/Statystyka#Metody_statystyczne) dla wyliczenia w jakim stopniu cały rozkład częstości w [próbie](https://pl.wikipedia.org/wiki/Pr%C3%B3ba_statystyczna)odpowiada rozkładowi oczekiwanemu, na przykład za pomocą [testu zgodności chi-kwadrat](https://pl.wikipedia.org/wiki/Test_zgodno%C5%9Bci_chi-kwadrat). Takimi badaniami posługuje się [kryptoanaliza statystyczna](https://pl.wikipedia.org/wiki/Kryptoanaliza_statystyczna).

Dla tekstu w [języku naturalnym](https://pl.wikipedia.org/wiki/J%C4%99zyk_naturalny) najczęściej tylko jeden klucz daje zrozumiały wynik, aczkolwiek pewnych bardzo krótkich wiadomości nie można jednoznacznie złamać bez znajomości klucza. Na przykład zakodowany tekst "WĄĘ" może być rozkodowany na słowo "KOT" lub "RYB" (przyjąwszy, że językiem tekstu jawnego jest polski), podobnie szyfrogram "ŻYAH" na "TRUĆ" lub "ROSĄ", a tekst "DIFBA" na "MROKI" lub "[RYTON](https://pl.wikipedia.org/wiki/Ryton)".

Wielokrotne kodowanie tej samej treści szyfrem Cezara nie zwiększa bezpieczeństwa, ponieważ dwie operacje szyfrowania, na przykład z przesunięciem 3 i z przesunięciem 5 są równoważne kodowaniu z przesunięciem 3+5=8. Używając terminologii matematycznej można to wyrazić następująco: zbiór wszystkich szyfrów Cezara z różnymi kluczami tworzy[grupę cykliczną](https://pl.wikipedia.org/wiki/Grupa_cykliczna) ze względu na [złożenie](https://pl.wikipedia.org/wiki/Z%C5%82o%C5%BCenie_funkcji) operacji szyfrowania.

**2. Plan wykonania ćwiczenia.**

1. Napisz program, który szyfruje tekst zgodnie z algorytmem szyfru Cezara.
2. Napisz program, który odszyfrowuje tekst zgodnie z algorytmem szyfru Cezara.
3. Napisz sprawozdanie.