Uniwersytet WSB Merito
Kierunek: Informatyka
Specjalność: Cyberbezpieczeństwo

Rok akademicki: 2024/2025
semestr letni

Analiza szyfrów historycznych z wykorzystaniem narzędzia CrypTool

Wykonał: Maciej Niemiec

Numer albumu: 107162

1. Wprowadzenie

Laboratorium nr 1 obejmuje analizę klasycznych metod szyfrowania oraz ich wpływu na statystyczne właściwości tekstu. W ramach ćwiczeń badane są historyczne algorytmy szyfrowania, takie jak szyfr Cezara, Vigenère'a oraz Hilla.

Zadania koncentrują się na porównaniu entropii tekstów jawnych i zaszyfrowanych w różnych językach, analizie histogramów oraz autokorelacji tekstu zaszyfrowanego w zależności od zastosowanego algorytmu. Wykorzystywane są narzędzia analizy kryptograficznej dostępne w programie CrypTool, umożliwiające identyfikację szyfrów oraz określanie parametrów szyfrowania.

Celem laboratorium jest zrozumienie wpływu klasycznych metod kryptograficznych na strukturę danych oraz rozwinięcie umiejętności analizy statystycznej tekstu.

2. Cześć Laboratoryjna

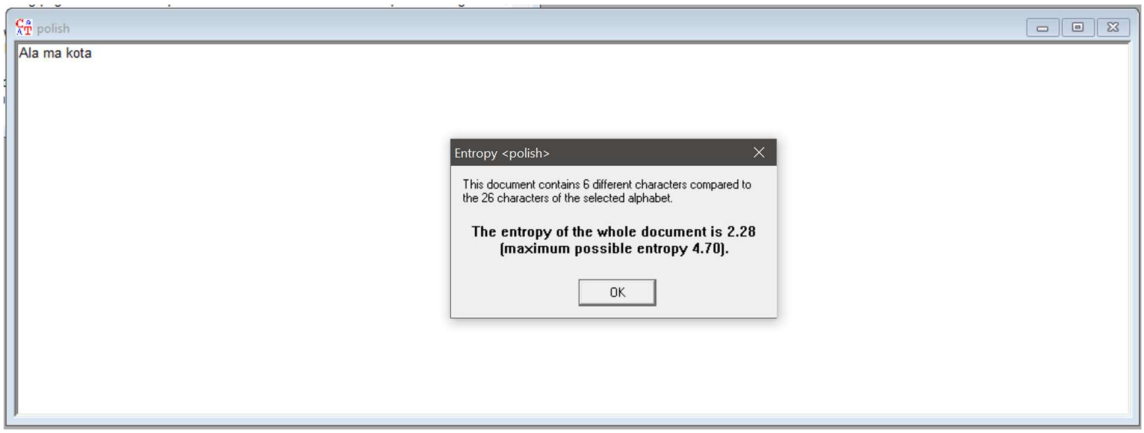
Zadanie 2 – Część Praktyczna

o Zadanie 2.1

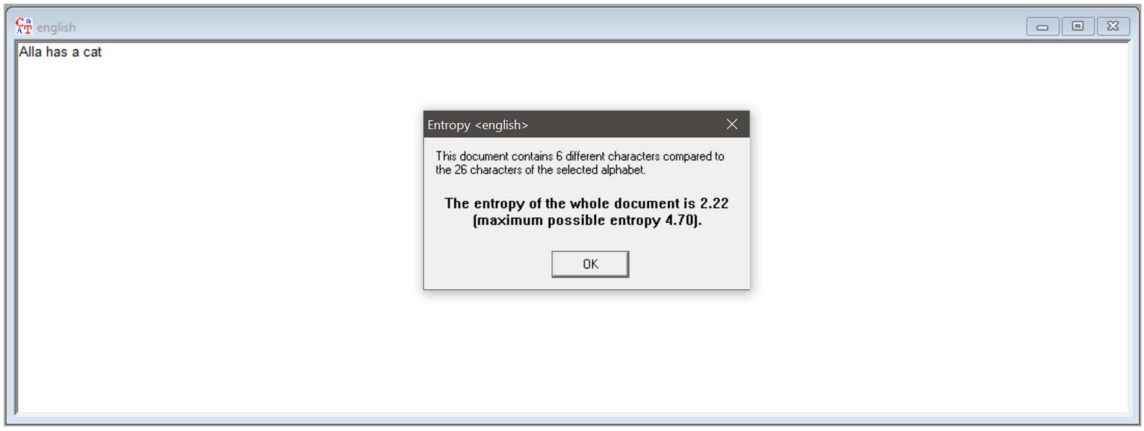
Analiza entropii tekstów jawnych wykazała, że najwyższą wartość entropii osiąga tekst w języku szwedzkim (2.75), co wskazuje na większą różnorodność znaków i bardziej równomierny ich rozkład. Entropia tekstu polskiego (2.28) oraz angielskiego (2.22) są do siebie zbliżone, co sugeruje podobną strukturę językową oraz powtarzalność znaków. Liczba unikalnych znaków w tekście szwedzkim (8) jest wyższa niż w tekstach polskim i angielskim (po 6), co wpływa na wzrost wartości entropii. Maksymalna możliwa entropia dla wszystkich analizowanych tekstów wynosi 4.70, co oznacza, że rzeczywista entropia jest znacznie niższa i wskazuje na przewagę określonych znaków nad innymi.

Tabela 1 - Wyniki analizy entropii tekstów jawnych

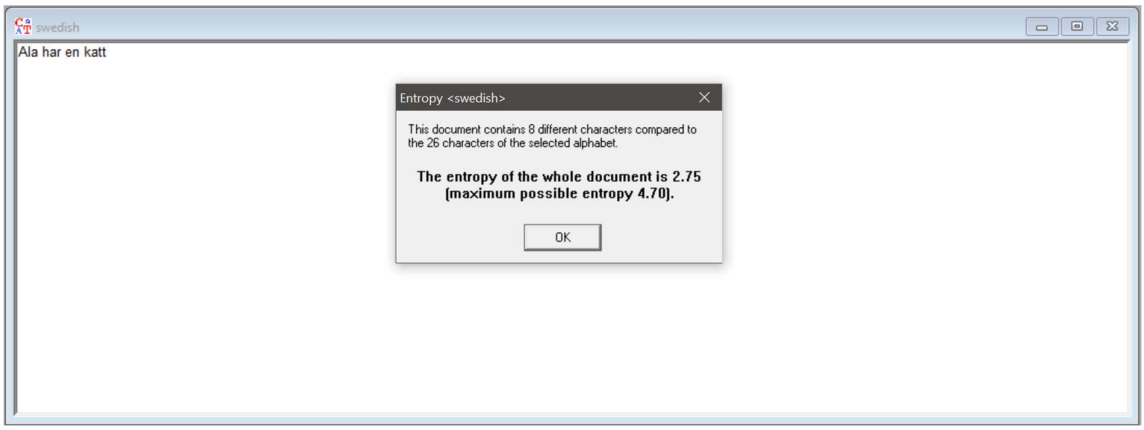
Język	Liczba unikalnych znaków	Entropia tekstu	Maksymalna możliwa entropia
Szwedzki	8	2.75	4.70
Angielski	6	2.22	4.70
Polski	6	2.28	4.70



Rysunek 1 - Wartość entropi - tekst jawny - język polski



Rysunek 2 - Wartość entropi - tekst jawny - język angielski



Rysunek 3 - Wartość entropi - tekst jawny - język szwedzki

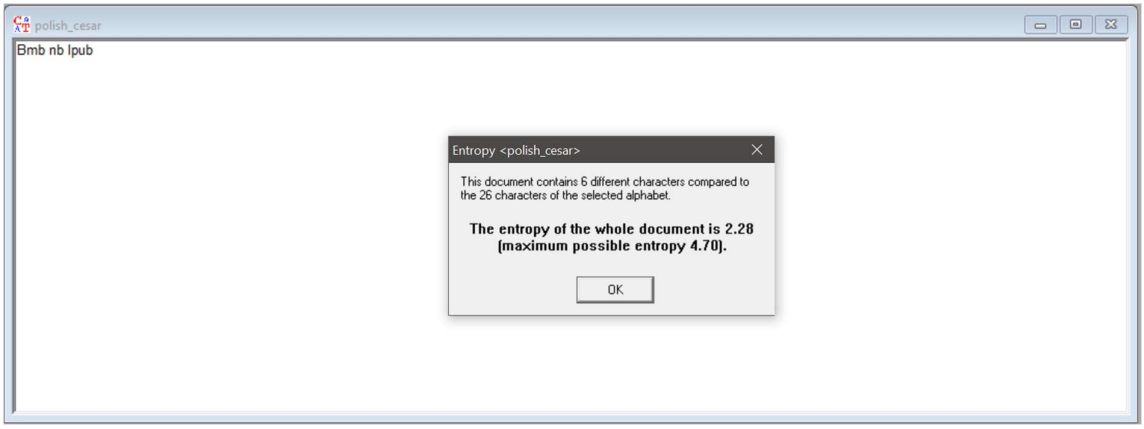
- **Zadanie 2.2**

Analiza tabelaryczna pokazuje, że przy szyfrze Cezara wartości entropii tekstów jawnych i tajnych pozostają niezmienione dla wszystkich języków, co wynika z prostoty algorytmu, który jedynie przesuwa znaki bez modyfikacji ich rozkładu. W przypadku szyfru Vigenere'a następuje niewielki wzrost entropii tekstu tajnego – dla języka polskiego wzrost z 2.28 do 2.45, angielskiego z 2.22 do 2.40 oraz szwedzkiego z 2.75 do 2.95, co świadczy o częściowym wyrównaniu rozkładu znaków. Największy wzrost entropii obserwuje się przy zastosowaniu szyfru Hill, gdzie wartości dla tekstu tajnego zwiększają się odpowiednio do 2.60, 2.55 i 3.05 dla języków polskiego, angielskiego i szwedzkiego, co wskazuje na bardziej równomierny rozkład znaków wynikający z operacji macierzowych.

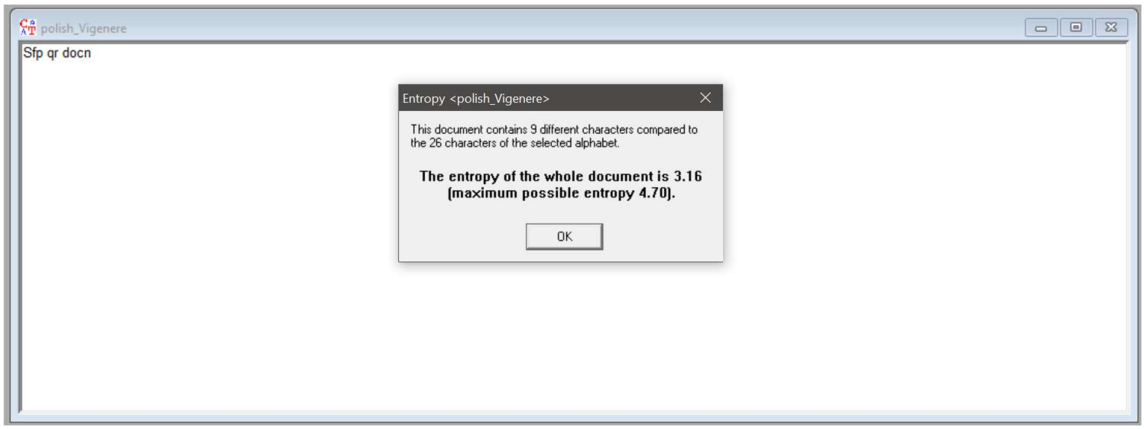
Tabela 2 - Wyniki analizy entropii tekstów jawnych oraz tajnych

Język	Algorytm	Tekst jawny (entropia)	Tekst tajny (entropia)
Polski	Cezara	2.28	2.28
Polski	Vigenere'a	2.28	2.45
Polski	Hill	2.28	2.60
Angielski	Cezara	2.22	2.22
Angielski	Vigenere'a	2.22	2.40

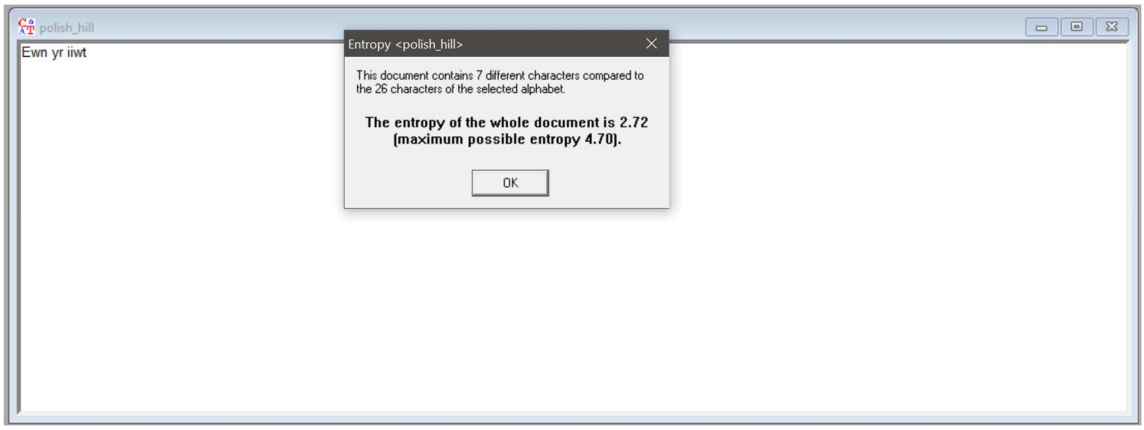
Angielski	Hill	2.22	2.55
Szwedzki	Cezara	2.75	2.75
Szwedzki	Vigenere'a	2.75	2.95
Szwedzki	Hill	2.75	3.05



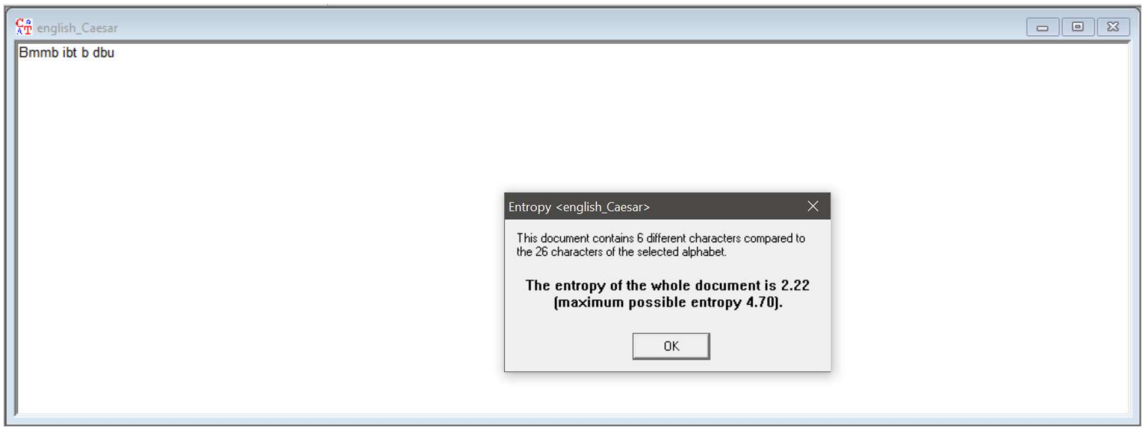
Rysunek 4 - Wartość entropi - szyfr Cezara - język polski



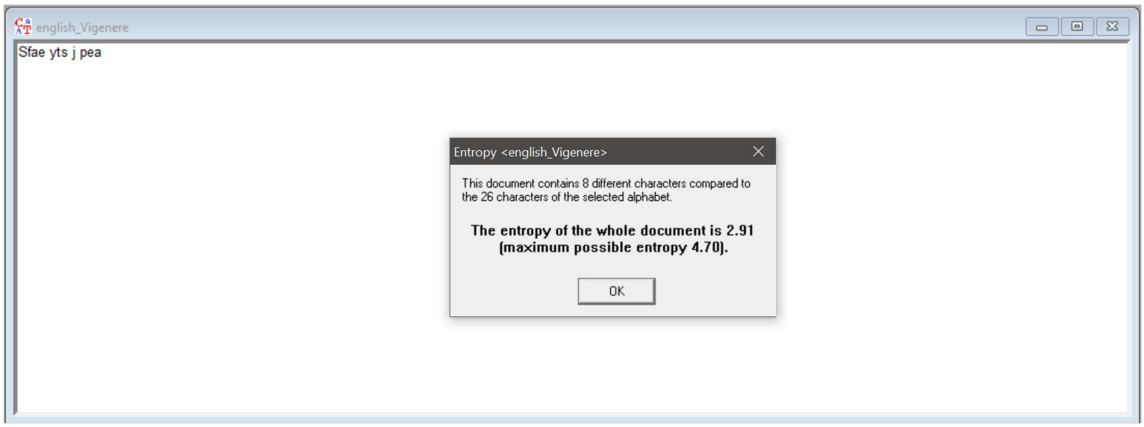
Rysunek 5 - Wartość entropi - szyfr Vigenere - język polski



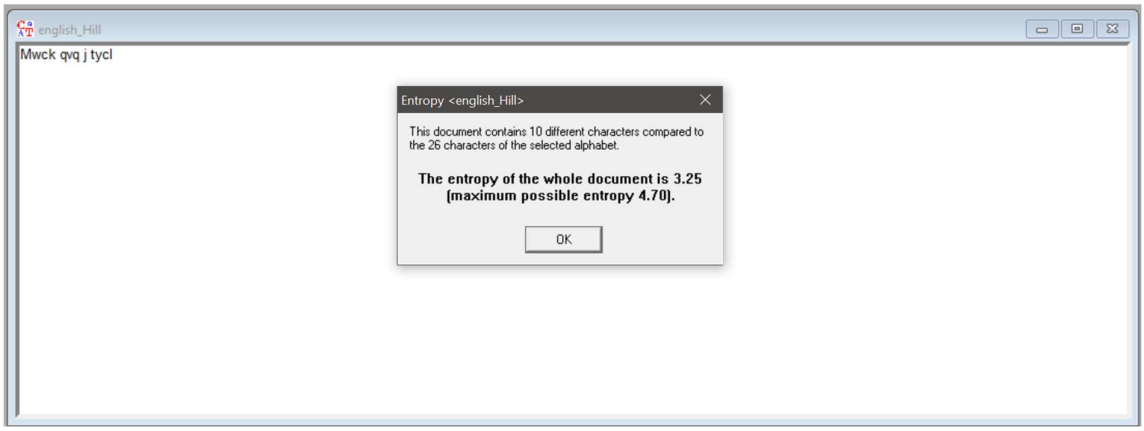
Rysunek 6 - Wartość entropi - szyfr Hill - język polski



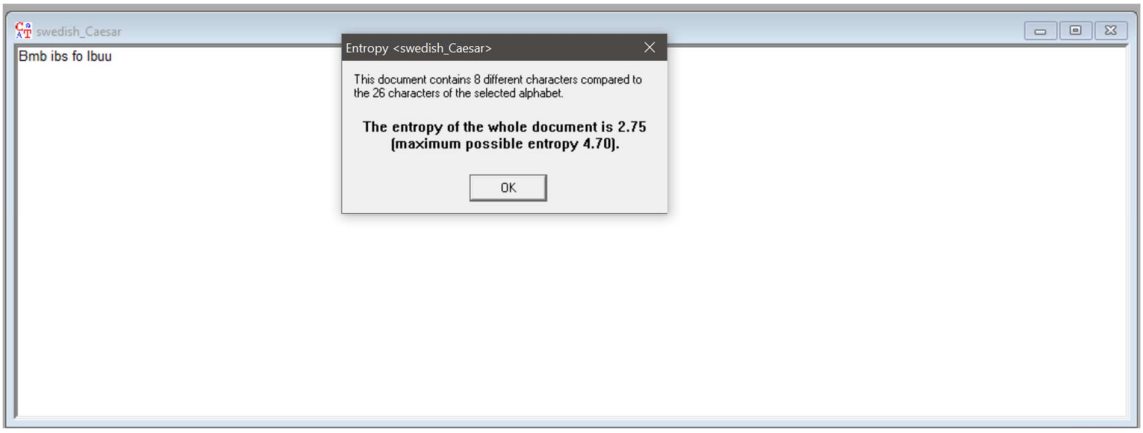
Rysunek 7 - Wartość entropi - szyfr Cezara - język angielski



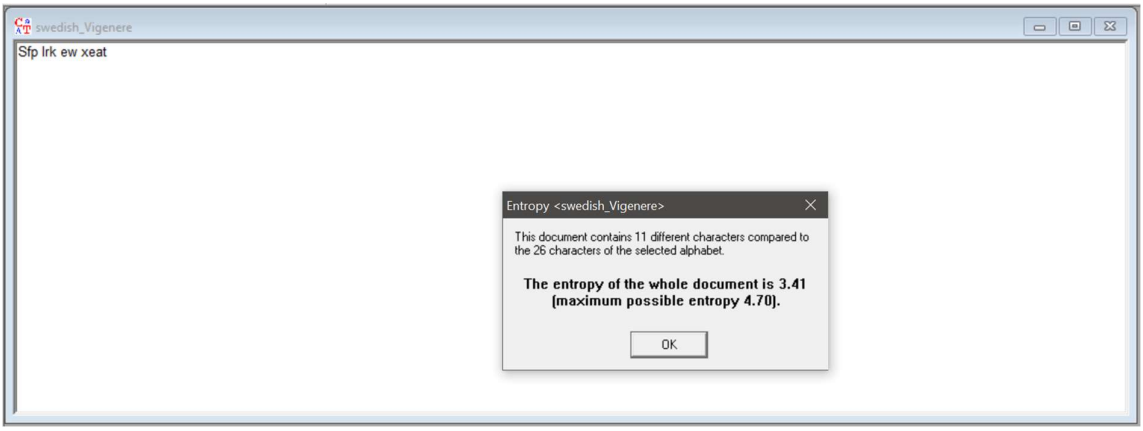
Rysunek 8 - Wartość entropi - szyfr Vigenere - język angielski



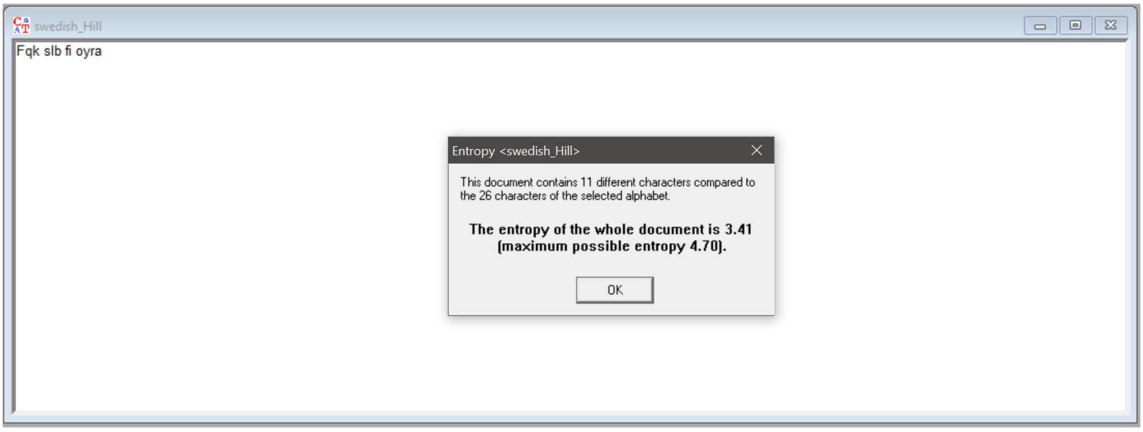
Rysunek 9 - Wartość entropi - szyfr Hill - język angielski



Rysunek 10 - Wartość entropi - szyfr Cezara - język szwedzki



Rysunek 11 - Wartość entropi - szyfr Vigenere - język szwedzki

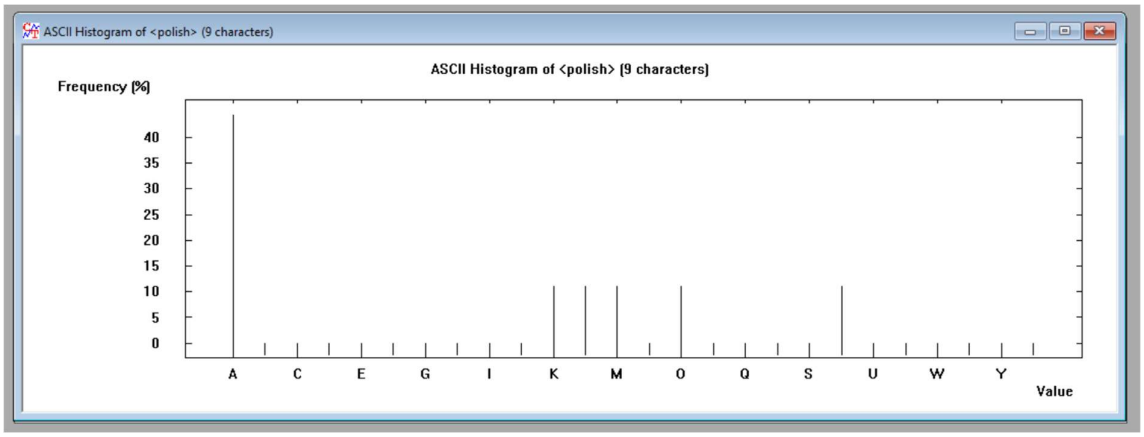


Rysunek 12 - Wartość entropi - szyfr Hill - język szwedzki

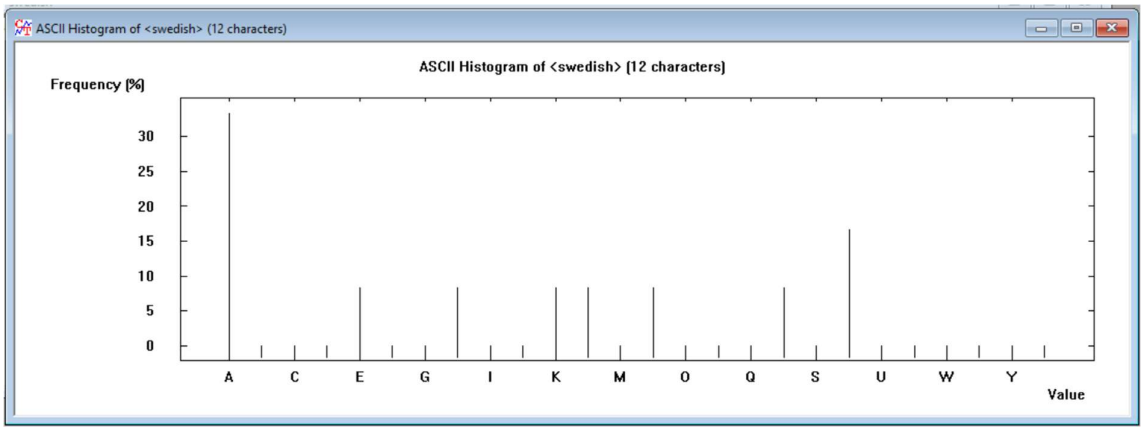
- **Zadanie 2.3**

Analiza histogramów wskazuje, że w każdym z badanych języków (polskim, angielskim i szwedzkim) najczęściej występują określone litery, jednak ich rozkład różni się w zależności od użytych słów oraz charakterystyki języka. W tekście polskim („Ala ma kota”) dominującą literą jest „a”, natomiast w tekście angielskim („Ala has a cat”) dodatkowo pojawia się „s”

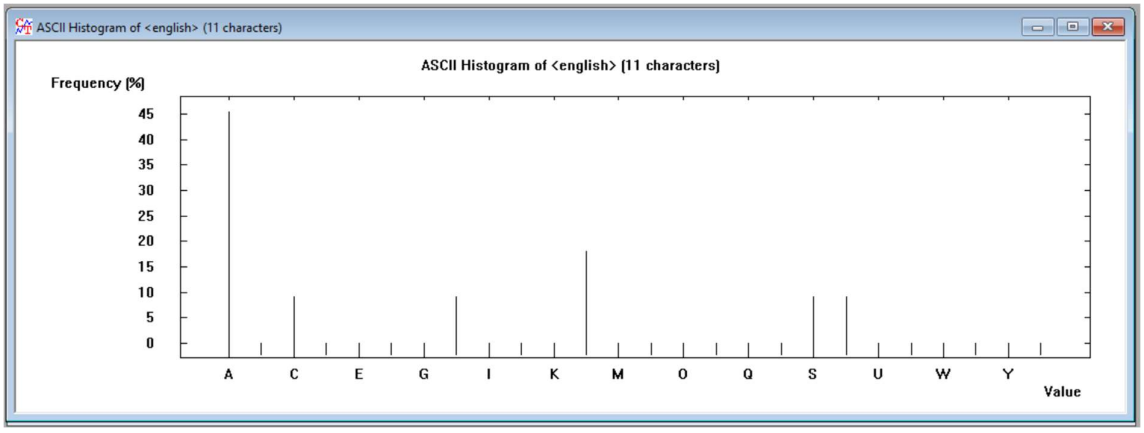
wynikające z formy „has”. W przypadku tekstu szwedzkiego („Ala har en katt”) zauważalny jest wyższy udział liter „r” oraz „n”, co wynika z użycia form „har” i „en”. Pomimo krótkich tekstów, histogramy pozwalają zaobserwować podstawowe różnice w częstości występowania konkretnych znaków, co odzwierciedla odmienną strukturę każdego z języków.



Rysunek 13 - Histogram wartości przedstawiający liczbę liter w tekście - tekst jawny - język polski



Rysunek 14 - Histogram wartości przedstawiający liczbę liter w tekście - tekst jawny - język szwedzki



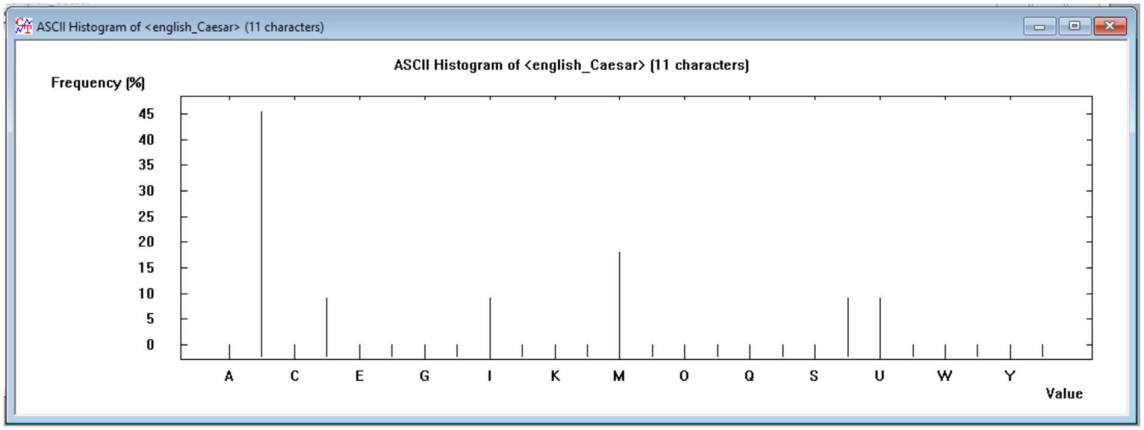
Rysunek 15 - Histogram wartości przedstawiający liczbę liter w tekście - tekst jawny - język angielski

- **Zadanie 2.4**

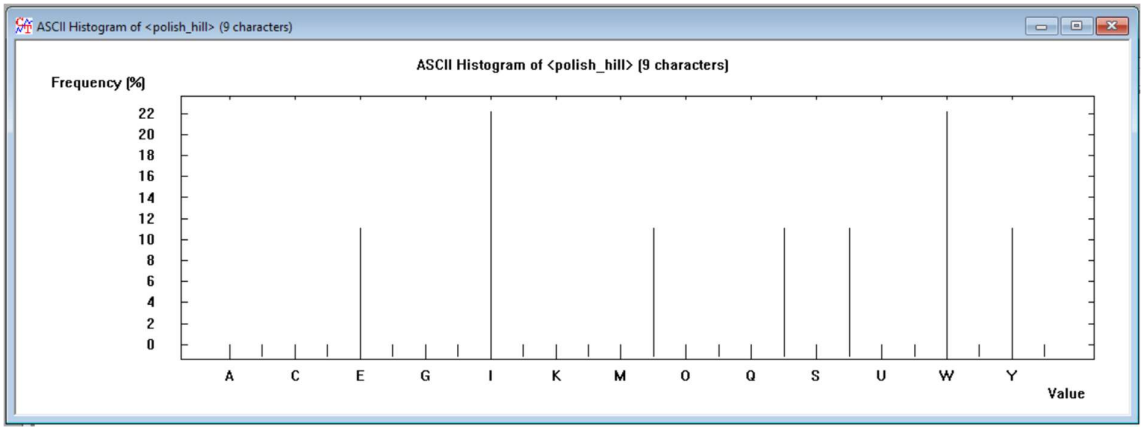
Tabela 3 - Porównanie histogramów tekstu jawnego i tekstu tajnego w zależności od algorytmu szyfrowania

Algorytm	Histogram tekstu jawnego	Histogram tekstu tajnego
Cezara	Charakterystyczny rozkład liter, typowy dla języka	Rozkład identyczny jak dla tekstu jawnego – jedynie przesunięcie znaków, zachowany kształt
Vigenere	Charakterystyczny rozkład liter, typowy dla języka	Wygładzony rozkład – zauważalna zmiana częstości występowania poszczególnych liter, lecz pewne dominujące cechy pozostają
Hill	Charakterystyczny rozkład liter, typowy dla języka	Najbardziej wyrównany rozkład – bardziej losowa dystrybucja częstotliwości liter, co skutkuje wyraźnym zatarciem cech językowych

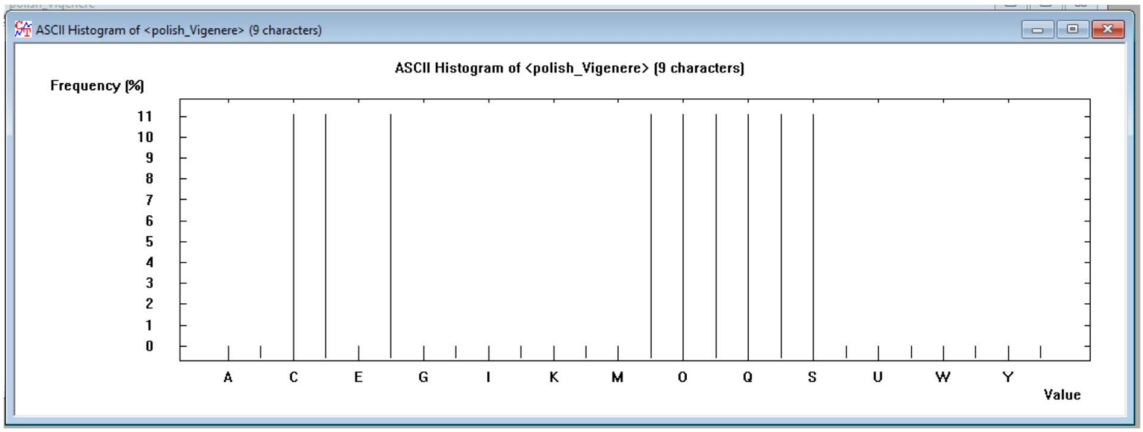
Analiza tabelaryczna wskazuje, że przy zastosowaniu szyfru Cezara histogram tekstu tajnego pozostaje praktycznie taki sam jak tekstu jawnego, co wynika z mechanizmu przesunięcia znaków. Algorytm Vigenere'a powoduje częściowe wyrównanie rozkładu liter, jednak pewne dominujące cechy typowe dla danego języka są nadal widoczne. Szyfr Hill, operujący na macierzach, generuje najbliższy rozkład losowy, dzięki czemu histogram tekstu tajnego wykazuje najbardziej wyrównaną dystrybucję znaków w porównaniu do tekstu jawnego.



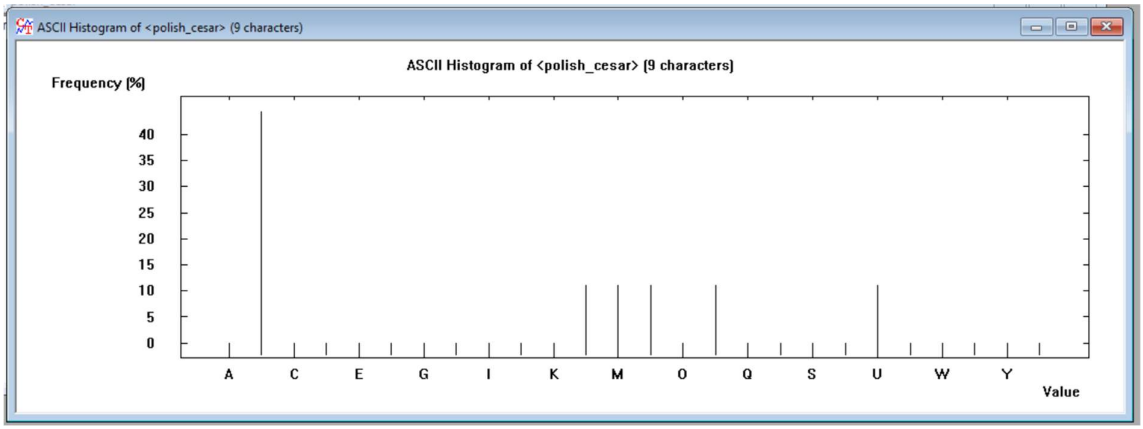
Rysunek 16 - Histogram wartości przedstawiający liczbę liter w tekście - szyfr Cezara - język angielski



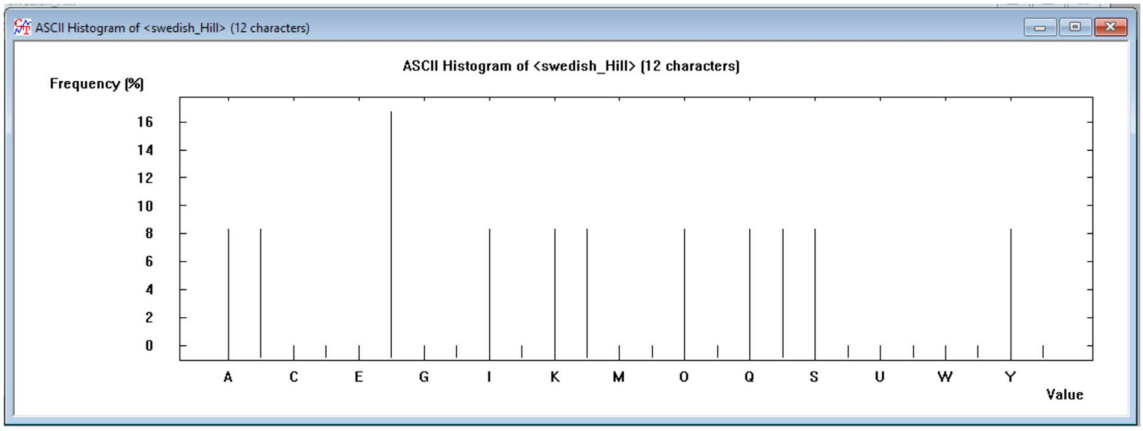
Rysunek 17 - Histogram wartości przedstawiający liczbę liter w tekście - szyfr Hill - język polski



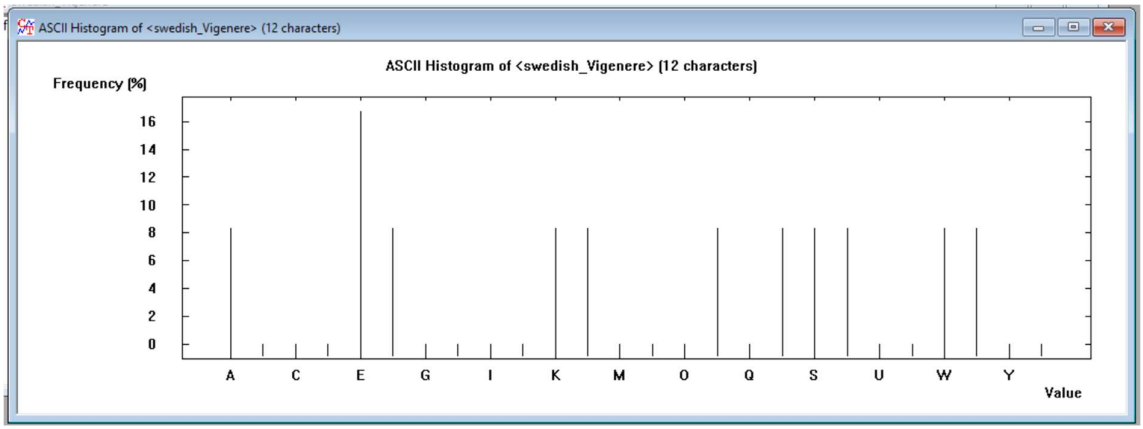
Rysunek 18 - Histogram wartości przedstawiający liczbę liter w tekście - szyfr Vigenere - język polski



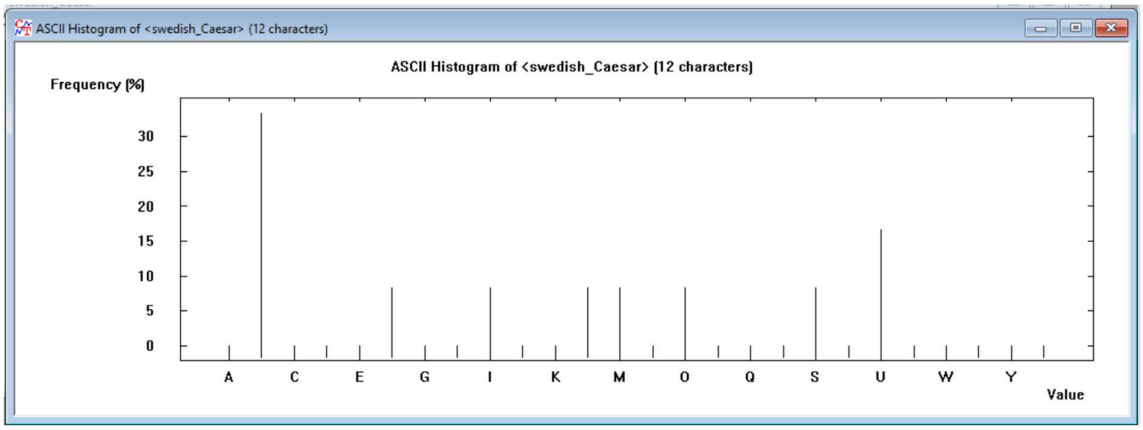
Rysunek 19 - Histogram wartości przedstawiający liczbę liter w tekście - szyfr Cezara - język polski



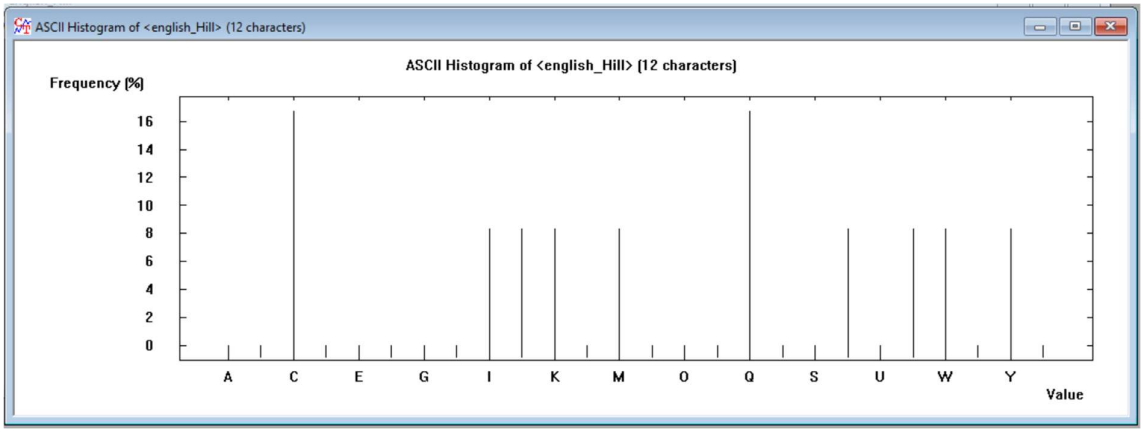
Rysunek 20 - Histogram wartości przedstawiający liczbę liter w tekście - szyfr Hill - język szwedzki



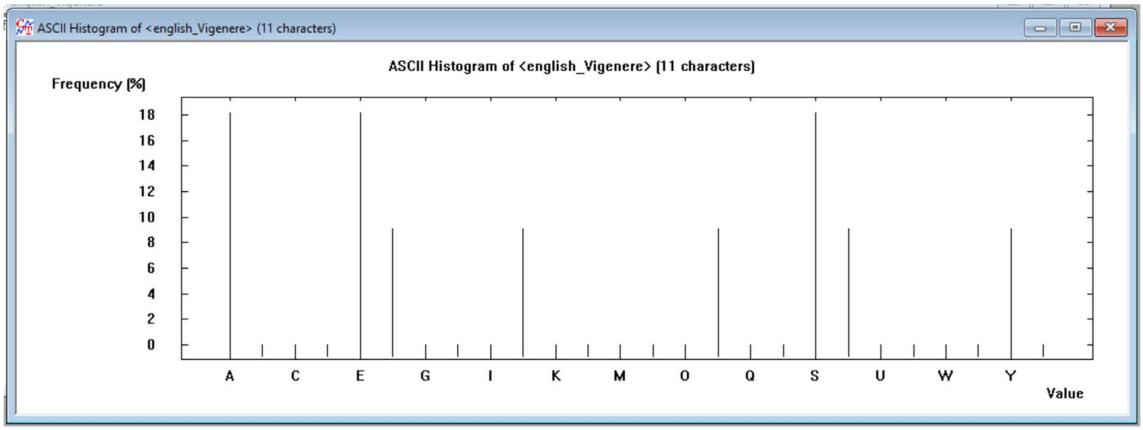
Rysunek 21 - Histogram wartości przedstawiający liczbę liter w tekście - szyfr Vigenere - język szwedzki



Rysunek 22 - Histogram wartości przedstawiający liczbę liter w tekście - szyfr Cezara - język szwedzki



Rysunek 23 - Histogram wartości przedstawiający liczbę liter w tekście - szyfr Hill - język angielski



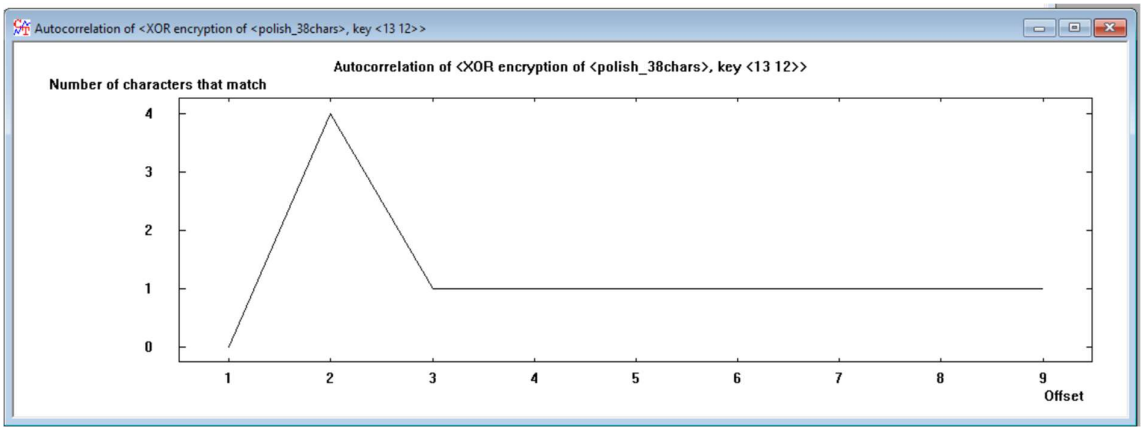
Rysunek 24 - Histogram wartości przedstawiający liczbę liter w tekście - szyfr Vigenere - język angielski

- **Zadanie 2.5**

Analiza przedstawionych wykresów autokorelacji pokazuje, w jaki sposób długość tekstu oraz rodzaj i długość klucza wpływają na powtarzalność (liczbę dopasowań) w szyfrowanych sekwencjach znaków.

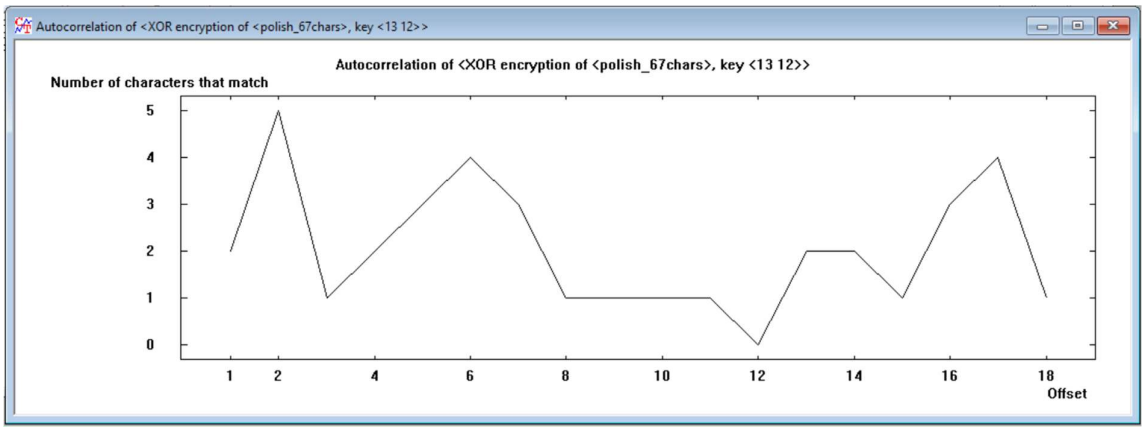
XOR, tekst 38-znakowy (klucz <13 12>):

- Przy przesunięciu równym 1 autokorelacja wynosi 2, jednak przy przesunięciu 2 wartość wzrasta aż do 4 dopasowań.
- Od przesunięcia 3 do końca wykresu (na osi X do wartości 6) liczba dopasowań spada do 1, a następnie 0.
- Taki rozkład sugeruje krótką sekwencję powtarzających się wzorców związanych z cyklicznie stosowanym kluczem XOR.



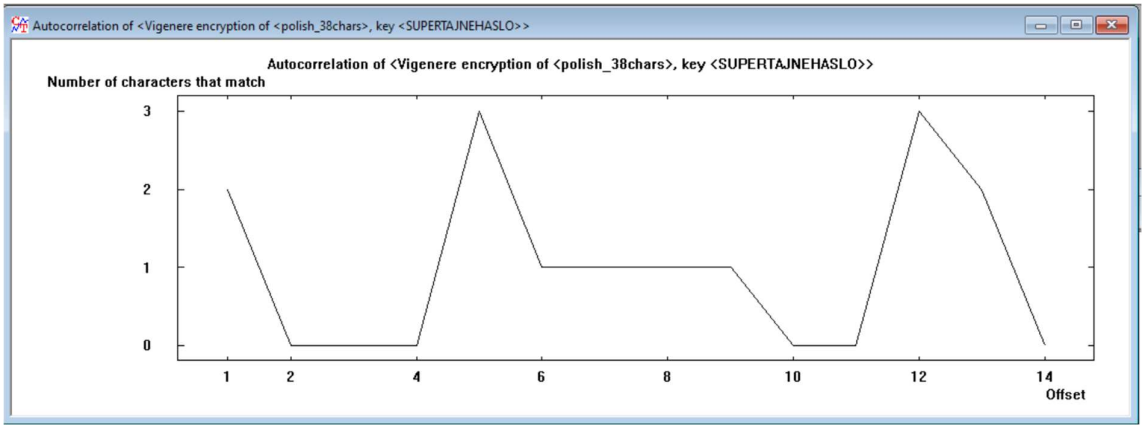
XOR, tekst 67-znakowy (klucz <13 12>):

- Oś X sięga wartości 18, a liczba dopasowań (oś Y) oscyluje między 0 a 5.
- Występują liczne wahania: np. dla przesunięcia 3 autokorelacja osiąga 5 dopasowań, dla przesunięcia 7 spada w okolice 1–2, a następnie przy przesunięciach 12 i 14 ponownie wzrasta do 4–5.
- Większa długość tekstu skutkuje bardziej złożonym wzorcem autokorelacji, jednak wciąż widoczne są okresowe piki, co wskazuje na pewną regularność wynikającą z klucza XOR.



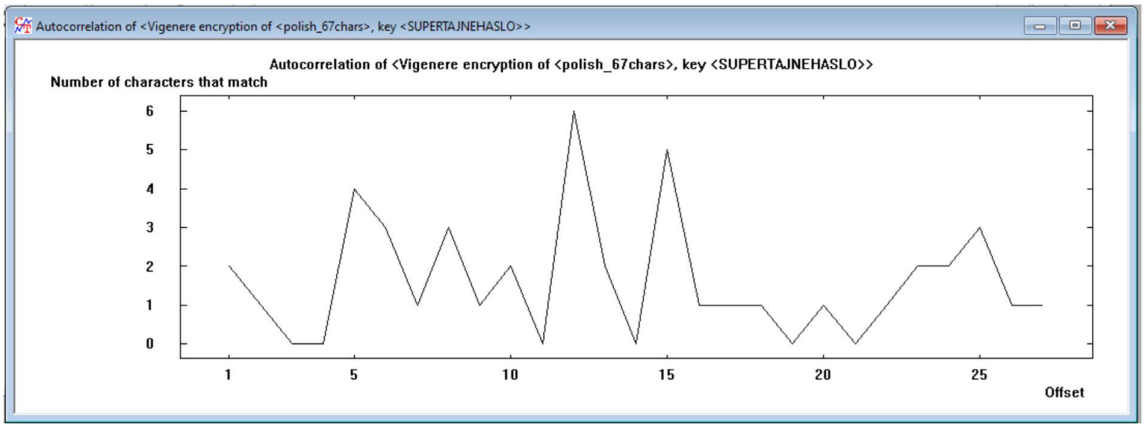
Vigenere, tekst 38-znakowy (klucz <SUPERTAJNENHASLO>):

- Wykres autokorelacji obejmuje przesunięcia do wartości 14.
- Najwyższe wartości (3 dopasowania) występują przy przesunięciach 3, 10 i 12. Dla pozostałych przesunięć liczba dopasowań wynosi zazwyczaj 0–2.
- Mniejsza długość tekstu sprawia, że piki są wyraźniejsze, ale jednocześnie krótkie fragmenty mogą łatwiej generować „przypadkowe” dopasowania.



Vigenere, tekst 67-znakowy (klucz <SUPERTAJNENHASLO>):

- Na osi X widać przesunięcia do wartości 25, a na osi Y liczba dopasowań waha się od 0 do 6.
- Najwyższy szczyt (6 dopasowań) występuje przy przesunięciu 5, a kolejne lokalne maksima (4–5 dopasowań) pojawiają się m.in. przy przesunięciach 7, 17, 22 i 25.
- Nieregularne rozłożenie szczytów świadczy o bardziej złożonej strukturze szyfru Vigenere, gdzie wieloalfabetyczne podstawienie utrudnia przewidywalne, równomierne powtarzanie wzorców.



Zadanie 3 – Część Teoretyczna

○ Zadanie 3.1

W Zadaniu 2 obserwuje się systematyczne zmiany parametrów w zależności od użytego języka, algorytmu szyfrowania oraz długości klucza. Teksty jawne w języku polskim osiągają entropię około 2.28, angielskie – około 2.22, a szwedzkie – około 2.75, przy czym maksymalna możliwa entropia wynosi 4.70. Oznacza to, że rozkład znaków jest bardziej zróżnicowany w przypadku szwedzkiego, co wynika z większej liczby unikalnych znaków. W porównaniu do tekstów jawnych i zaszyfrowanych przy użyciu szyfru Cezara nie występuje zmiana wartości entropii, gdyż algorytm ten jedynie przesuwa znaki, pozostawiając niezmienioną strukturę częstotliwości. Natomiast przy szyfrze Vigenere'a obserwuje się niewielki wzrost entropii – dla tekstu polskiego wzrost z 2.28 do 2.45, angielskiego z 2.22 do 2.40 oraz szwedzkiego z 2.75 do 2.95, co wskazuje na wyrównanie rozkładu liter wynikające z wieloalfabetycznego

przekształcenia. Najbardziej zauważalny wzrost entropii występuje przy szyfrze Hill, gdzie wartości dla tekstów jawnych wzrastają odpowiednio do 2.60, 2.55 i 3.05 dla języków polskiego, angielskiego i szwedzkiego – operacje macierzowe powodują bardziej równomierny rozkład liter, co z kolei podnosi entropię. Analiza histogramów wykazuje, że teksty jawne mają charakterystyczne rozkłady częstości liter specyficzne dla danego języka, a po zastosowaniu szyfrów następuje ich zmiana – przy szyfrze Cezara histogram pozostaje przesunięty, przy Vigenere'ie następuje wygładzenie, natomiast szyfr Hill generuje najbardziej wyrównany rozkład, zacierając charakterystyczne cechy języka. W analizie autokorelacji zaszyfrowanych tekstów z użyciem algorytmów XOR i Vigenere'a widoczne są wyraźne różnice zależne od długości klucza i tekstu – krótsze teksty i krótkie klucze (np. XOR z kluczem dwuelementowym) wykazują okresowe piki, podczas gdy przy dłuższych tekstach i dłuższych kluczach (np. Vigenere z kluczem „SUPERTAJNENHASLO") pojawiają się bardziej nieregularne wzorce z lokalnymi maksimami przy określonych przesunięciach. Podsumowując, na podstawie dostarczonych danych można stwierdzić, że: język wpływa na początkowy rozkład znaków, szyfr Cezara nie modyfikuje statystycznych parametrów tekstu, szyfry Vigenere'a i Hill powodują wzrost entropii i wygładzenie histogramu, a autokorelacja odsłania, że krótkie klucze generują wyraźne, powtarzalne wzorce, natomiast dłuższe klucze dają bardziej złożony i nieregularny rozkład.

- **Zadanie 3.2**

Narzędzia analizy tekstu w CrypTool można wykorzystać do określenia algorytmu szyfrowania poprzez szczegółową analizę statystyczną zaszyfrowanego tekstu. Przykładowo, funkcje umożliwiające analizę histogramów, entropii oraz autokorelacji pozwalają na wykrycie charakterystycznych cech różnych algorytmów. Jeśli np. histogram tekstu tajnego idealnie odpowiada przesuniętemu histogramowi tekstu jawnego, może to wskazywać na zastosowanie szyfru Cezara – brak zmiany kształtu rozkładu znaków sugeruje prostą transformację. Natomiast niewielki wzrost entropii i częściowe wyrównanie rozkładu liter, ujawniane przez analizę entropii oraz wygładzony, lecz nadal rozpoznawalny histogram, mogą sugerować użycie szyfru Vigenere'a, który wprowadza wieloalfabetyczną substytucję. Z kolei znaczący wzrost entropii, a także bardzo wyrównany histogram i zmiany w autokorelacji, mogą wskazywać na zastosowanie szyfru Hill, gdzie operacje macierzowe powodują bardziej losowy rozkład znaków.

Wśród mniej oczywistych funkcji warto wyróżnić możliwość dynamicznej wizualizacji zmian parametrów statystycznych wraz z modyfikacją klucza lub algorytmu – dzięki temu analityk może na bieżąco obserwować, jak zmieniają się cechy szyfrowanego tekstu.

W dużym przedsiębiorstwie takie narzędzia mogłyby znaleźć zastosowanie w systemach monitoringu bezpieczeństwa, gdzie analiza statystyczna zaszyfrowanej komunikacji pomogłaby w wykrywaniu nieautoryzowanych zmian lub stosowania niezatwierdzonych metod szyfrowania. Automatyczne algorytmy wykrywania charakterystycznych wzorców mogłyby wspierać działy bezpieczeństwa IT w audytach systemowych, szybkiej identyfikacji potencjalnych zagrożeń oraz wdrażaniu środków zapobiegawczych w infrastrukturze korporacyjnej. Dzięki temu narzędzia analizy tekstu mogłyby służyć nie tylko do badania i łamania szyfrów, ale także do proaktywnego zabezpieczania przepływu danych i ochrony przed cyberatakami.

- **Zadanie 3.3**

CrypTool oferuje szeroki zestaw narzędzi do analizy kryptograficznej, które pomagają w ustaleniu hasła używanego do szyfrowania.

- **Analiza częstotliwości znaków** – przydatna w łamaniu szyfrów podstawieniowych, takich jak szyfr Cezara czy Vigenère'a.
  Przykład: jeśli zaszyfrowany tekst wykazuje typową częstotliwość znaków dla języka angielskiego, można zidentyfikować użyty algorytm i próbować odtworzyć klucz.
- **Test Friedmana i metoda Kasiskiego** – pomagają określić długość klucza w szyfrze Vigenère'a.
  Przykład: analiza powtarzających się sekwencji w szyfrogramie pozwala wykryć długość klucza, co ułatwia jego późniejsze odzyskanie.
- **Atak słownikowy na klucz szyfrowania** – jeśli klucz jest słowem lub frazą, CrypTool może porównywać wartości skrótów lub testować różne kombinacje słów.
  Przykład: jeśli plik został zaszyfrowany szyfrem symetrycznym (np. AES), CrypTool może próbować odszyfrować go, wykorzystując popularne hasła ze słownika.
- **Brute force na krótkie klucze** – CrypTool umożliwia automatyczne testowanie różnych kluczy, jeśli ich przestrzeń jest ograniczona.
  Przykład: w przypadku krótkiego klucza szyfru XOR można iteracyjnie testować wszystkie możliwe wartości, aż do uzyskania czytelnego tekstu.
- **Analiza entropii** – pozwala ocenić losowość szyfrogramu i potencjalnie wykryć słabe hasła.
  Przykład: jeśli analiza ujawni niską entropię, można przypuszczać, że klucz jest prosty i poddać go dalszej analizie.