Sprawozdanie Podwójny Vigenere

kryreneus

17.04.2024

—

Kryptografia

—

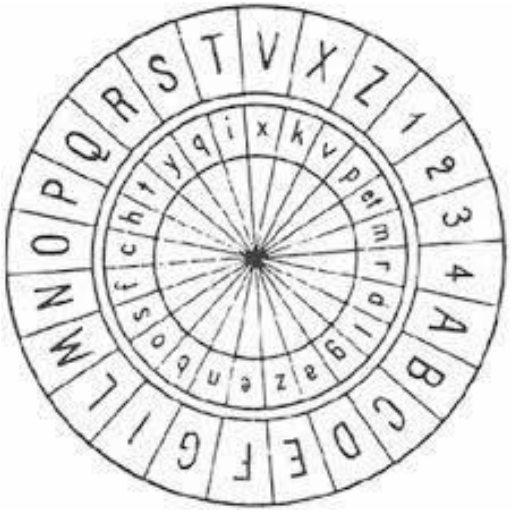
ODPOWIEDZI

Poniżej przedstawiono klucz oraz odszyfrowany tekst. W dalszej części sprawozdania omówiona została metodologia postępowania i analiza szyfru.

Wstęp teoretyczny

Szyfry Polialfabetyczne

Szyfr Vigenere'a należy do rodziny szyfrów polialfabetycznych. Oznacza to, wykorzystane w nim jest wiele alfabetów tajnych, z których każdy koduje pojedynczy znak tekstu jawnego (oczywiście uzywane cyklicznie, więc po wyczerpaniu wszystkich powraca się do pierwszego i cała sytuacja się powtarza. Za pierwszego twórcę szyfru polialfabetycznego uważa się włoskiego architekta z XV wieku, Leone Battiste Albertiego. Stosowana przez niego tarcza szyfrująca została ukazana na obrazie po prawej.

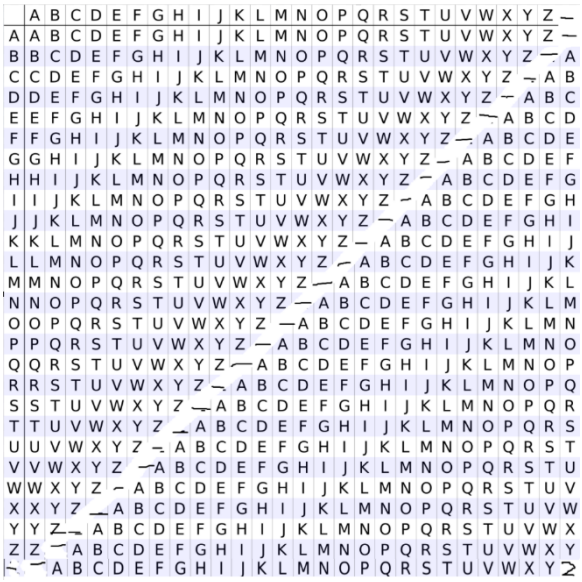


Kolejny etap historii – tabula recta

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1.	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
2.	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
3.	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
4.	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
5.	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
6.	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
7.	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
8.	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
9.	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
10.	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
11.	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
12.	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
13.	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
14.	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
15.	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
16.	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
17.	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
18.	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
19.	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
20.	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
21.	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
22.	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
23.	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
24.	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
25.	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
26.	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

Pół wieku po Albertim inny uczony, tym razem pochodzący z Niemiec Johannes Trithemius, zaproponował zupełnie inny sposób tworzenia szyfru polialfabetycznego. Bazował on na tabeli nazwanej przez Trithemiusa tabula reacta. Zawierała ona pewną liczbę kolumn, z których każda była przesunięciem w prawo względem poprzedniej o jeden znak alfabetu. Kiedy używano tego szyfru, litera tekstu jawnego zamieniana była na jej odpowiednik z pierwszego wiersza, kolejna na odpowiednik z drugiego wiersza itd. Dzięki temu każda litera szyfrowana była przy użyciu innego alfabetu, co znacznie zwiększało bezpieczeństwo. Taki system znacząco chroni poufną wiadomość przed atakirm polegającym na analizie częstotliwości wystepowania liter (jednak, jak się okazuje, niewystarczająco).

Vigenere – klucz do szyfrowania – le chiffre indechiffrable



Blaise de Vigenere, urodzony w 1523 roku, francuski dyplomata i duchowny, podczas misji dyplomatycznej w Rzymie w 1549 roku po raz pierwszy spotkał się z kryptografią. Temat ten na tyle go zaciekawił, że postanowił opuścić dwór i zacząć bardziej pogłębiać swoją wiedzę, by w końcu zająć się pisaniem książek. W taki oto sposób 1586 rok zaowocował opublikowaniem *Traktatu o szyfrach*, w którym opisał stworzony przez siebie szyfr polialfabetyczny. Polegał on na szyfrowaniu kolejnych liter tajnego tekstu za pomocą różnych wierszy tablicy Trithemiusa. Dzięki zastosowaniu tej innowacji, znajomość samego systemu przestała wystarczać do odszyfrowania tekstu. Aby ułatwić zapamiętanie których wierszy użyć, można zastosować prosty klucz, który np. literze A przypisze wartość 0, B przypisze wartość 1 itd. Dzięki temu osoba, której zadaniem jest zdeszyfrowanie tekstu w prosty sposób może określić, którego wiersza tabeli ma użyć.

Vigenere stworzył jeszcze dwa inne systemy szyfrownia oparte na koncepcji autoklucza, gdzie do odszyfrowania wiadomości konieczna była znajomość jedynie pojedynczej litery stanowiącej tzw. klucz pierwotny.

Na obrazie przedstawiona została zmodyfikowana wersja tablicy Trithemiusa (będącej podstawą szyfru Vigenere'a), zgodnie z założeniami konkursu.

One-Time Pad

Kolejnym krokiem w dziedzinie kryptologii był stworzony przez Vernama i Mauborgne'a szyfr, który polega na szyfrowaniu tekstu jawnego z wykorzystaniem losowo wygenerowanego klucza o długości odpowiadającej długości tekstu jawnego.

Aby szyfr był realnie nie do złamania, muszą być spełnione następujące założenia:

1. Losowość klucza
2. Jednorazowość klucza
3. Odpowiednia długość klucza
4. Bezpieczna dystrybucja klucza

Analiza szyfru i metodologia badań

Słowo wstępu – odrobina historii

Pierwszym, który z wymiernym sukcesem podjął się złamania szyfru nie do złamania był Charles Babbage urodzony w 1792 roku. Od małego interesował się szyframi. Przejawiał w tej dziedzinie spore zdolności. W 1854 roku podjął się zadania kryptoanalizy szyfru Vigenere'a. Zauważyć należy, że w tamtych czasach szyfr ten uznawany był jako nie do złamania. Babbage zauważył, że gdy uda się poznać długość klucza, to odszyfrowanie tekstu staje się znaczenie łatwiejsze, gdyż zostaje nam do rozważenia n (długość klucza) ciągów, gdzie w obrębie jednego ciągu wszystkie litery zaszyfrowane są tym samym alfabetem, czyli standardowym szyfrem podstawieniowym.

Problemem zostaje jedynie odnalezienie długości klucza, jednakże i na to są pewne sposoby. Można m.in., bazując na założeniu, że w dłuższych tekstach często zdarzają się powtórzenia wyrazów (lub ich fragmentów), policzyć odległości pomiędzy powtarzającymi się fragmentami tekstu. W takim przypadku długość klucza wspólnym dzielnikiem obliczonych wartości. Przyładowo dla odstępów 12, 16, 8, 20 można domniemywać, że długość klucza będzie równa 4.

Wyżej opisana technika została rozwinięta i usystematyzowana przez Kasiskiego, który opisał ją w swojej książce *Die Geheimschriften und die Dechiffrir-kunst,* gdzie dokładnie opisał metodologię łamania szyfrów polialfabetycznych, zaczynając od określania długości klucza, a kończąc na analizie uzyskanych szyfrów monoalfabetycznych.

Metoda Friedmana – bazująca na indeksie koincydencji, polegająca na porównywaniu szyfrogramu z nim samym, po przesunięciu go względem orygiału o pewną liczbę liter.

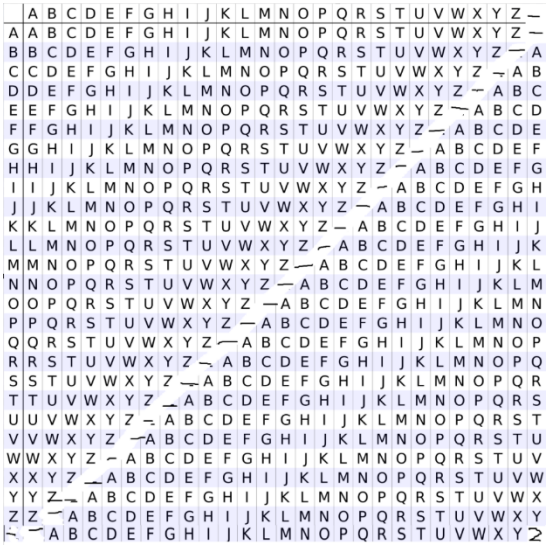
Podwójny Vigenere, czyli jak to właściwie działa?

Na wstępie zaznaczyć należy, że podwójnego Vigenere'a rozpatrywać można tak samo jak pojedynczego. Wystarczy jedynie zauważyć, że długość jego klucza będzie równa najmniejszej wspólnej wielokrotności długości kluczy składowych. Dzieje się tak, ponieważ Vigenere bazuje na przesunięciach, gdzie każda litera klucza odpowiada jakiemuś przesunięciowi względem tekstu oryginalnego. W taki oto sposób jeżeli np. mamy pierwszą literę tekstu jawnego „a" i zostanie ona poddana działaniu najpierw klucza „b", a potem klucza „c" to jest to to samo, jakby została poddana działaniu jedynie jednego klucza „b"+"c"="d". To samo zachodzi dla kolejnych liter klucza. Najmniejsza wspólna wielokrotność bierze się natomiast z faktu, że w najgorszym wypadku (gdy długości obu słów są np. liczbami pierwszymi) każda litera pierwszego klucza musi być połączona z każdą literą drugiego klucza (na zasadzie podwójnej pętli). Prościej mówiąc, gdyby rozrysować połączenia pomiędzy literami obu kluczy, stworzyłby się pełny graf dwudzielny.

Znając powyższe informacje wystarczy jedynie wyliczyć jakie składowe dają NWW równe długości klucza, a następnie spróbować słownikowo dopasować klucze składowe (zakładając, że są one słowami pochodzącymi np. z języka polskiego).

Programowanie do zwycięstwa, czyli komputer w służbie człowieka – metodologia postępowania i analiza szyfru

W celu rozwiązania zadania konkursowego napisany został program, który ma za zadanie znaleźć jak najlepszy klucz i odszyfrować nim zaszyfrowany tekst.



Na wstępie należy zaznaczyć iż do zaszyfrowania tekstu użyta została zmodyfikowana tabula recta (przedstawiona na obrazie po prawej). Sprawiło to, że odszyfrowanie stało się nieco trudniejszym wyzwaniem ze względu na trudność określenia, gdzie w oryginalnej wiadomości użyte były spacje, a przez to brak automatycznego rozróżnienia słów składających się na tekst oryginalny. Pominięto także polskie znaki diakrytyczne.

Do analizy statystycznej użyte zostały następujące tabele:

Prawa pochodzi z opracowania Ziółko, B., Gałka, J., Manandhar, S., Wilson, R.C., Ziółko, M. (2009). Triphone Statistics for Polish Language. In: Vetulani, Z., Uszkoreit, H. (eds) Human Language Technology.

Lewa pochodzi z analizy statystycznej tekstu w próbce Narodowego Korpusu Jezyka Polskiego z 2012 liczącej ponad 250 milionów słów

a	ą	b	c	ć	d	e
8,965%	1,021%	1,482%	3,988%	0,448%	3,293%	7,921%
ę	f	g	h	i	j	k
1,131%	0,312%	1,377%	1,072%	8,286%	2,343%	3,411%
l	ł	m	n	ń	o	ó
2,136%	1,746%	2,911%	5,600%	0,185%	7,590%	0,823%
p	q	r	s	ś	t	u
3,101%	0,003%	4,571%	4,263%	0,683%	3,966%	2,347%
v	w	x	y	z	ź	ż
0,034%	4,549%	0,019%	3,857%	5,620%	0,061%	0,885%

Table 1: *Phoneme transcription in Polish - SAMPA (Demenko et al., 2003)*

SAMPA	example	transcr.	occurr.	%
#		#	23,810,956	16.086,7
a	pat	pat	13,311,163	8.993
e	test	test	11,871,405	8.020,3
o	pot	pot	10,566,010	7.138,4
s	syk	sIk	5,716,058	3.861,8
t	test	test	5,703,429	3.853,2
r	ryk	rIk	5,171,698	3.494
p	pik	pik	5,150,964	3.48
v	wilk	vilk	5,025,050	3.394,9
j	jak	jak	4,996,475	3.375,6
i	PIT	pit	4,994,743	3.374,4
I	typ	tIp	4,974,567	3.360,8
n	nasz	naS	4,602,314	3.109,3
l	luk	luk	4,399,366	2.972,2
u	puk	puk	4,355,825	2.942,8
k	kit	kitk	4,020,161	2.716
z	zbir	zbir	3,602,857	2.434,1
m	mysz	mIS	3,525,813	2.382
d	dym	dIm	3,267,009	2.207,2
n'	koń	kon'	3,182,940	2.150,4
f	fan	fan	2,030,717	1.372
ts	cyk	tsIk	1,984,311	1.340,6
g	gen	gen	1,949,890	1.317,3
S	szyk	SIk	1,739,146	1.175
b	bit	bit	1,668,103	1.127
x	hymn	xImn	1,339,311	0.904,84
tS	czyn	tSIn	1,285,310	0.868,36
dz	dzwoń	dzvon'	692,334	0.467,74
ts'	ćma	ts'ma	690,294	0.466,36
dz'	dźwig	dz'vik	589,266	0.398,11
Z	żyto	ZIto	536,786	0.362,65
s'	świt	s'vit	531,402	0.359,02
o~	wąs	vo~s	306,665	0.207,18
N	pęk	peNk	184,884	0.124,91
w	łyk	wIk	144,166	0.097,399
z'	źle	z'le	66,518	0.044,94
dZ	dżem	dZem	27,621	0.018,661
e~	gęś	ge~s'	1,011	0.000,683
w~	ciąża	ts'ow~Za	sampa extension	
j~	więź	vjej~s'	sampa extension	
c	kiedy	cjedy	sampa extension	
J	giełda	Jjewda	sampa extension	

Schemat działania programu do szukania połączonego klucza (standardowy Vigenere)

1. Pobranie zaszyfrowanego tekstu od użytkownika.
2. Iteracja przez wszystkie możliwe długości klucza (przykładowo 1-20)
    a. Wycięcie z tekstu jedynie wartości odpowiadających literom lub spacji
    b. Iteracja przez wszystkie indeksy klucza
        i. Wycięcie z tekstu co n-tej litery, gdzie początkową jest dany indeks (n – długość klucza)
        ii. Obliczenie wyników różnych przesunięć dla danego indeksu klucza:
            1. Deszyfracja tekstu (dodanie do każdej litery wartości przesunięcia 0-26)
            2. Podliczenie częstotliwości występowania kolejnych liter (i spacji)
            3. Obliczenie sumy wartości bezwzględnych różnic pomiędzy częstotliwościami występowania liter dla danego klucza a domniemanymi wg wcześniej pokazanych tabel i zapamiętanie jej
        iii. Dla każdego indeksu klucza wybranie najlepszego przesunięcia (z wcześniej wyliczonych)
    c. Zapamiętanie najlepszego klucza dla badanej długości klucza
3. Posortowanie najlepszych kluczy (po jednym dla każdej długości) wg ich wyniku (obliczenie sumy wartości bezwzględnych różnic pomiędzy częstotliwościami występowania liter dla danego klucza a domniemanymi wg wcześniej pokazanych tabel)
4. Zostawienie 3 najbardziej prosperujących kluczy
5. Odszyfrowanie i wypisanie tekstu bazując na najlepszych kluczach $(Z_i - K_i + 27) \bmod 27$
6. Wybór prawidłowego pozostawiony użytkownikowi

Schemat działania programu rozbijającego znaleziony klucz na składowe

1. Pobranie klucza od użytkownika
2. Obliczenie długości klucza
3. Wyszukanie par liczb, których NWW jest równe długości klucza
4. Dla każdej pary:
    a. Wycięcie ze słownika wszystkich słów odpowiadających długością pierwszej oraz drugiej liczbie z pary
    b. Iteracja przez przez słownik i przejście przez wszystkie pary wyrazów:
        i. Wydłużenie długości słów z pary do długości klucza będącego kluczem pobranym od użytkownika
        ii. Zsumowanie kolejnych liter zgodnie z tabelą Vigenere'a
        iii. Jeżeli wynik jest taki sam jak klucz pobrany od użytkownika to działanie programu się zakańcza i wypisywane są klucze składowe.

Podsumowanie i wnioski

Mimo że Blaise de Vigenère wierzył w niezłomność swojego szyfru, historia udowodniła, że żaden system szyfrowania nie jest całkowicie odporny na ataki. Analiza statystyczna okazała się kluczowym narzędziem w pokonaniu tego wyzwania. Gdy udaje się poznać długość klucza, odszyfrowanie tekstu staje się znacznie łatwiejsze. Wtedy to mamy do czynienia z n ciągami, gdzie każdy ciąg składa się z liter zaszyfrowanych tym samym alfabetem, tworząc swoisty standardowy szyfr podstawieniowy. Dodatkowo, analizując częstotliwość występowania poszczególnych liter, można bazować na n-gramach charakterystycznych dla danego języka, co znacznie ułatwia proces odszyfrowywania.
Jeśli zaś chodzi o podwójny Vigenere, to rozpatrywać go można tak samo jak pojedynczego. Wystarczy jedynie zauważyć, że długość jego klucza będzie równa najmniejszej wspólnej wielokrotności długości kluczy składowych.

Bibliografia:

1. Marcin Karbowski „Podstawy kryptografii” wydanie III
2. Ziółko, B., Gałka, J., Manandhar, S., Wilson, R.C., Ziółko, M. (2009). Triphone Statistics for Polish Language. In: Vetulani, Z., Uszkoreit, H. (eds) Human Language Technology.
3. Kahn David „Łamacze kodów – historia kryptologii”
4. https://www.guballa.de/bits-and-bytes/implementierung-des-vigenere-solvers
5. http://practicalcryptography.com/cryptanalysis/stochastic-searching/cryptanalysis-vigenere-cipher/
6. https://pl.wikipedia.org/wiki/Alfabet_polski
7. http://www.crypto-it.net/pl/proste/szyfr-vigenere.html