

Analiza częstotliwości

Przykład

Jacek Rogowski

Instytut Matematyki
Politechniki Łódzkiej

Łamanie monoalfabetycznego szyfru podstawieniowego

Jest to element kryptoanalizy polegający na odczytaniu zaszyfrowanej wiadomości bez znajomości klucza szyfru.

Testowanie wszystkich możliwych kluczy jest nieefektywne. Dla alfabetu złożonego z 26 liter i prostego szyfru podstawieniowego istnieje ponad

400 kwadrylionów

możliwych kluczy.

Analiza częstotliwości

Metoda porównująca częstotliwość występowania znaków tekstu jawnego z częstotliwością występowania liter w tekstach danego języka – w kolejnych próbach pod najczęściej występujące znaki kryptogramu podstawia się najczęściej występujące litery.

Tabela częstotliwości dla alfabetu polskiego (w %)

a/ą	9,90	g	1,42	m	2,80	t	3,98
b	1,47	h	1,08	n/ń	5,72	u	2,50
c/ć	4,36	i	8,21	o/ó	8,60	w	4,65
d	3,25	j	2,28	p	3,13	x	0,02
e/ę	8,77	k	3,51	r	4,69	y	3,76
f	0,30	l/ł	3,92	s/ś	4,98	z/ż/ź	6,53

Najczęściej występujące samogłoski (w %)

a/ą	9,90	g	1,42	m	2,8	t	3,98
b	1,47	h	1,08	n/ń	5,72	u	2,50
c/ć	4,36	i	8,21	o/ó	8,60	w	4,65
d	3,25	j	2,28	p	3,13	x	0,02
e/ę	8,77	k	3,51	r	4,69	y	3,76
f	0,30	l/ł	3,92	s/ś	4,98	z/ż/ź	6,53

Najczęściej występujące spółgłoski (w %)

a/ą	9,90	g	1,42	m	2,8	t	3,98
b	1,47	h	1,08	n/ń	5,72	u	2,50
c/ć	4,36	i	8,21	o/ó	8,60	w	4,65
d	3,25	j	2,28	p	3,13	x	0,02
e/ę	8,77	k	3,51	r	4,69	y	3,76
f	0,30	l/ł	3,92	s/ś	4,98	z/ż/ź	6,53

Zadanie

Korzystając z prostego szyfru podstawieniowego otrzymano tekst:

XFJDTE SODOBT PYJMKO ZTXUOZ
DCIJES NAOISO TDYPOZ JDLTED
HWEJMK OJOKST SNXSOL SCONXM
ESCUOI SOJIIU NA

Należy odczytać wiadomość jawną.

Etap I: Zliczanie

A	2	G	0	M	3	T	6
B	1	H	1	N	4	U	3
C	3	I	5	O	13	W	1
D	6	J	7	P	2	Y	2
E	5	K	3	R	0	Z	3
F	1	L	2	S	9	X	4

Etap I: Częstotliwości (w %)

A	2,3	G	0	M	3,5	T	7,0
B	1,2	H	1,2	N	4,7	U	3,5
C	3,5	I	5,8	O	15,1	W	1,2
D	7,0	J	8,1	P	2,3	Y	2,3
E	5,8	K	3,5	R	0	Z	3,5
F	1,2	L	2,3	S	10,5	X	4,7

Etap II: Wstępna analiza

- Tekst nie jest długi

Etap II: Wstępna analiza

- Tekst nie jest długi.
- Użycie symbolu X sugeruje zaznaczanie spacji.

Etap II: Wstępna analiza

- Tekst nie jest długi.
- Użycie symbolu X sugeruje zaznaczanie spacji.
- Hipoteza: Spacja jest szyfrowana najczęściej występującym symbolem O (15,1% wszystkich liter).

Etap II: Wstępna analiza

- Tekst nie jest długi.
- Użycie symbolu X sugeruje zaznaczanie spacji.
- Hipoteza: Spacja jest szyfrowana najczęściej występującym symbolem O (15,1% wszystkich liter).
- Usuwamy symbol O.

Etap III: Pierwsza litera

- Hipoteza: Następny, co do ilości wystąpień, symbol **S** (10,5% wszystkich liter) oznacza prawdopodobnie literę **a**.

Etap III: Pierwsza litera

- Hipoteza: Następny, co do ilości wystąpień, symbol **S** (10,5% wszystkich liter) oznacza prawdopodobnie literę **a**.
- Podstawiamy **a** w miejsce **S**.

Etap IV: Kolejne litery

- Symbol I występuje 5 razy (5,8% wszystkich liter), a w ostatnim wyrazie pojawia się zdublowany.

Etap IV: Kolejne litery

- Symbol **l** występuje 5 razy (5,8% wszystkich liter), a w ostatnim wyrazie pojawia się zdublowany.
- Hipoteza: **l = n**

Etap IV: Kolejne litery

- Symbol **l** występuje 5 razy (5,8% wszystkich liter), a w ostatnim wyrazie pojawia się zdublowany.
- Hipoteza: **l = n**.
- Hipoteza ta jest potwierdzana przez dwukrotne pojawienie się słowa **na**.

Etap IV: Kolejne litery

X	F	J	D	T	E	S		D		B	T	P	Y	J	M	K	
						a											
Z	T	X	U		Z	D	C	I	J	E	S	N	A		I	S	
								n			a				n	a	
T	D	Y	P		Z	J	D	L	T	E	D	H	W	E	J	M	K
	J		K	S	T	S	N	X	S		L	S	C		N	X	M
				a		a			a			a					
E	S	C	U		I	S		J	I	I	U	N	A				
	a				n	a			n	n							

Etap IV: Kolejne litery

X	F	J	D	T	E	S		D		B	T	P	Y	J	M	K	
						a											
Z	T	X	U		Z	D	C	I	J	E	S	N	A		I	S	
								n			a				n	a	
T	D	Y	P		Z	J	D	L	T	E	D	H	W	E	J	M	K
	J		K	S	T	S	N	X	S		L	S	C		N	X	M
				a		a			a			a					
E	S	C	U		I	S		J	I	I	U	N	A				
	a				n	a			n	n							

Etap IV: Kolejne litery

- Symbole **D** i **J** występują jako spójniki.

Etap IV: Kolejne litery

- Symbole **D** i **J** występują jako spójniki.
- Częstotliwości ich występowania (7,0% i 8,1%) odpowiadają częstotliwościom samogłosek **o**, **i** lub spółgłoski **z**.

Etap IV: Kolejne litery

- Symbole **D** i **J** występują jako spójniki.
- Częstotliwości ich występowania (7,0% i 8,1%) odpowiadają częstotliwościom samogłosek **o**, **i** lub spółgłoski **z**.
- Symbol **J** rozpoczyna ostatni wyraz, a po nim znajdują się dwie litery **n**.

Etap IV: Kolejne litery

- Symbole **D** i **J** występują jako spójniki.
- Częstotliwości ich występowania (7,0% i 8,1%) odpowiadają częstotliwościom samogłosek **o**, **i** lub spółgłoski **z**.
- Symbol **J** rozpoczyna ostatni wyraz, a po nim znajdują się dwie litery **n**.
- Hipoteza: **J = i**.

Etap IV: Kolejne litery

X	F	J	D	T	E	S		D		B	T	P	Y	J	M	K	
		i				a								i			
Z	T	X	U		Z	D	C	I	J	E	S	N	A		I	S	
								n	i		a				n	a	
T	D	Y	P		Z	J	D	L	T	E	D	H	W	E	J	M	K
						i									i		
	J		K	S	T	S	N	X	S		L	S	C		N	X	M
	i			a		a			a			a					
E	S	C	U		I	S		J	I	I	U	N	A				
	a				n	a		i	n	n							

Etap IV: Kolejne litery

X	F	J	D	T	E	S		D		B	T	P	Y	J	M	K	
		i				a								i			
Z	T	X	U		Z	D	C	I	J	E	S	N	A		I	S	
								n	i		a				n	a	
T	D	Y	P		Z	J	D	L	T	E	D	H	W	E	J	M	K
						i									i		
	J		K	S	T	S	N	X	S		L	S	C		N	X	M
	i			a		a			a			a					
E	S	C	U		I	S		J	I	I	U	N	A				
	a				n	a		i	n	n							

Etap IV: Kolejne litery

- Symbol **D** może oznaczać **o** lub **z**.

Etap IV: Kolejne litery

- Symbol **D** może oznaczać **o** lub **z**.
- Częstotliwość występowania litery **o** (8,60%) jest większa od częstotliwości występowania litery **z** (6,53%),

Etap IV: Kolejne litery

- Symbol **D** może oznaczać **o** lub **z**.
- Częstotliwość występowania litery **o** (8,60%) jest większa od częstotliwości występowania litery **z** (6,53%), a więc częstsze pojawienie się litery **o** jest bardziej prawdopodobne.

Etap IV: Kolejne litery

- Symbol **D** może oznaczać **o** lub **z**.
- Częstotliwość występowania litery **o** (8,60%) jest większa od częstotliwości występowania litery **z** (6,53%), a więc częstsze pojawienie się litery **o** jest bardziej prawdopodobne.
- Hipoteza: **D = o**.

Etap IV: Kolejne litery

X	F	J	D	T	E	S		D		B	T	P	Y	J	M	K	
		i	o			a		o						i			
Z	T	X	U		Z	D	C	I	J	E	S	N	A		I	S	
						o		n	i		a				n	a	
T	D	Y	P		Z	J	D	L	T	E	D	H	W	E	J	M	K
	o					i	o				o				i		
	J		K	S	T	S	N	X	S		L	S	C		N	X	M
	i			a		a			a			a					
E	S	C	U		I	S		J	I	I	U	N	A				
	a				n	a		i	n	n							

Etap IV: Kolejne litery

X	F	J	D	T	E	S		D		B	T	P	Y	J	M	K	
		i	o			a		o						i			
Z	T	X	U		Z	D	C	I	J	E	S	N	A		I	S	
						o		n	i		a				n	a	
T	D	Y	P		Z	J	D	L	T	E	D	H	W	E	J	M	K
	o					i	o				o				i		
	J		K	S	T	S	N	X	S		L	S	C		N	X	M
	i			a		a			a			a					
E	S	C	U		I	S		J	I	I	U	N	A				
	a				n	a		i	n	n							

Etap IV: Kolejne litery

- Ostatnim słowem może być **innemu**, **innych** lub **innego**, ale **o** zostało już rozszyfrowane.

Etap IV: Kolejne litery

- Ostatnim słowem może być **innemu**, **innych** lub **innego**, ale **o** zostało już rozszyfrowane.
- Częstotliwości występowania liter w trigramie **UNA** (3,5 – 4,7 – 2,3) lepiej pasują do częstotliwości liter w trigramie **ych** (3,76 – 4,36 – 1,08) niż w trigramie **emu** (8,77 – 2,80 – 2,50).

Etap IV: Kolejne litery

- Ostatnim słowem może być **innemu**, **innych** lub **innego**, ale **o** zostało już rozszyfrowane.
- Częstotliwości występowania liter w trigramie **UNA** (3,5 – 4,7 – 2,3) lepiej pasują do częstotliwości liter w trigramie **ych** (3,76 – 4,36 – 1,08) niż w trigramie **emu** (8,77 – 2,80 – 2,50).
- Hipoteza: **U = y**, **N = c**, **A = h**.

Etap IV: Kolejne litery

X	F	J	D	T	E	S		D		B	T	P	Y	J	M	K	
		i	o			a		o						i			
Z	T	X	U		Z	D	C	I	J	E	S	N	A		I	S	
			y			o		n	i		a	c	h		n	a	
T	D	Y	P		Z	J	D	L	T	E	D	H	W	E	J	M	K
	o					i	o				o				i		
	J		K	S	T	S	N	X	S		L	S	C		N	X	M
	i			a		a	c		a			a			c		
E	S	C	U		I	S		J	I	I	U	N	A				
	a		y		n	a		i	n	n	y	c	h				

Etap IV: Kolejne litery

X	F	J	D	T	E	S		D		B	T	P	Y	J	M	K	
		i	o			a		o						i			
Z	T	X	U		Z	D	C	I	J	E	S	N	A		I	S	
			y			o		n	i		a	c	h		n	a	
T	D	Y	P		Z	J	D	L	T	E	D	H	W	E	J	M	K
	o					i	o				o				i		
	J		K	S	T	S	N	X	S		L	S	C		N	X	M
	i			a		a	c		a			a			c		
E	S	C	U		I	S		J	I	I	U	N	A				
	a		y		n	a		i	n	n	y	c	h				

Etap IV: Kolejne litery

- Symbol **X** występuje cztery razy, w tym dwukrotnie w zestawieniu **cX**, a raz w zestawieniu **Xy**.

Etap IV: Kolejne litery

- Symbol **X** występuje cztery razy, w tym dwukrotnie w zestawieniu **cX**, a raz w zestawieniu **Xy**
- Hipoteza: **X = z**.

Etap IV: Kolejne litery

X	F	J	D	T	E	S		D		B	T	P	Y	J	M	K	
z		i	o			a		o						i			
Z	T	X	U		Z	D	C	I	J	E	S	N	A		I	S	
		z	y			o		n	i		a	c	h		n	a	
T	D	Y	P		Z	J	D	L	T	E	D	H	W	E	J	M	K
	o					i	o				o				i		
	J		K	S	T	S	N	X	S		L	S	C		N	X	M
	i			a		a	c	z	a			a			c	z	
E	S	C	U		I	S		J	I	I	U	N	A				
	a		y		n	a		i	n	n	y	c	h				

Etap V: Uzupelnianie luk

- Zestawmy wyniki przeprowadzonej analizy.

Etap V: Uzupełnianie luk

- Zestawmy wyniki przeprowadzonej analizy.
- Dla ułatwienia zaznaczmy odgadnięte litery na niebiesko..

Etap V: Uzupełnianie luk

- Zestawmy wyniki przeprowadzonej analizy.
- Dla ułatwienia zaznaczmy odgadnięte litery na niebiesko.
- Następnie, wśród pozostałych znaków kryptogramu wskażmy te o największych częstotliwościach.

Etap V: Uzupełnianie luk

A=h	2,3	G	0	M	3,5	T	7,0
B	1,2	H	1,2	N=c	4,7	U=y	3,5
C	3,5	I=n	5,8	O=_	15,1	W	1,2
D=0	7,0	J=i	8,1	P	2,3	Y	2,3
E	5,8	K	3,5	R	0	Z	3,5
F	1,2	L	2,3	S=a	10,5	X=z	4,7

Etap V: Uzupełnianie luk

A=h	2,3	G	0	M	3,5	T	7,0
B	1,2	H	1,2	N=c	4,7	U=y	3,5
C	3,5	I=n	5,8	O=_	15,1	W	1,2
D=0	7,0	J=i	8,1	P	2,3	Y	2,3
E	5,8	K	3,5	R	0	Z	3,5
F	1,2	L	2,3	S=a	10,5	X=z	4,7

Etap V: Uzupełnianie luk

- Analogiczne działania przeprowadźmy dla liter alfabetu polskiego

Etap V: Uzupełnianie luk

a/ą=S	9,90	g	1,42	m	2,80	t	3,98
b	1,47	h=A	1,08	n/ń=I	5,72	u	2,50
c/ć=N	4,36	i=J	8,21	o/ó=D	8,60	w	4,65
d	3,25	j	2,28	p	3,13	x	0,02
e/ę	8,77	k	3,51	r	4,69	y=U	3,76
f	0,30	l/ł	3,92	s/ś	4,98	z/ż/ź=X	6,53

Etap V: Uzupełnianie luk

- Przynajmniej niektóre znaki ze zbioru
C, E, K, M, T, Z
odpowiadają pewnym literom spośród
e/ę, r, s/ś, t, w
- W szczególności jeden z powyższych
znaków powinien reprezentować literę e/ę

Etap V: Uzupełnianie luk

X	F	J	D	T	E	S		D		B	T	P	Y	J	M	K	
z		i	o			a		o						i			
Z	T	X	U		Z	D	C	I	J	E	S	N	A		I	S	
		z	y			o		n	i		a	c	h		n	a	
T	D	Y	P		Z	J	D	L	T	E	D	H	W	E	J	M	K
	o					i	o				o				i		
	J		K	S	T	S	N	X	S		L	S	C		N	X	M
	i			a		a	c	z	a			a			c	z	
E	S	C	U		I	S		J	I	I	U	N	A				
	a		y		n	a		i	n	n	y	c	h				

Etap V: Uzupelnianie luk

X	F	J	D	T	E	S		D		B	T	P	Y	J	M	K	
z		i	o			a		o						i			
Z	T	X	U		Z	D	C	I	J	E	S	N	A		I	S	
		z	y			o		n	i		a	c	h		n	a	
T	D	Y	P		Z	J	D	L	T	E	D	H	W	E	J	M	K
	o					i	o				o				i		
	J		K	S	T	S	N	X	S		L	S	C		N	X	M
	i			a		a	c	z	a			a			c	z	
E	S	C	U		I	S		J	I	I	U	N	A				
	a		y		n	a		i	n	n	y	c	h				

Etap V: Uzupełnianie luk

X	F	J	D	T	E	S		D		B	T	P	Y	J	M	K	
z		i	o			a		o						i	e		
Z	T	X	U		Z	D	C	I	J	E	S	N	A		I	S	
		z	y			o		n	i		a	c	h		n	a	
T	D	Y	P		Z	J	D	L	T	E	D	H	W	E	J	M	K
	o					i	o				o				i	e	
	J		K	S	T	S	N	X	S		L	S	C		N	X	M
	i			a		a	c	z	a			a			c	z	e
E	S	C	U		I	S		J	I	I	U	N	A				
	a		y		n	a		i	n	n	y	c	h				

Etap V: Uzupełnianie luk

- Przyjmijmy więc, że **$M=e$** .

Etap V: Uzupełnianie luk

- Przyjmijmy więc, że **M=e**.
- Słowo **czeEaCy** sugeruje, że
C = m, E = k.

Etap V: Uzupelnianie luk

X	F	J	D	T	E	S		D		B	T	P	Y	J	M	K	
z		i	o		k	a		o						i	e		
Z	T	X	U		Z	D	C	I	J	E	S	N	A		I	S	
		z	y			o	m	n	i	k	a	c	h		n	a	
T	D	Y	P		Z	J	D	L	T	E	D	H	W	E	J	M	K
	o					i	o			k	o			k	i	e	
	J		K	S	T	S	N	X	S		L	S	C		N	X	M
	i			a		a	c	z	a			a	m		c	z	e
E	S	C	U		I	S		J	I	I	U	N	A				
k	a	m	y		n	a		i	n	n	y	c	h				

Etap V: Uzupełnianie luk

X	F	J	D	T	E	S		D		B	T	P	Y	J	M	K	
z		i	o		k	a		o						i	e		
Z	T	X	U		Z	D	C	I	J	E	S	N	A		I	S	
		z	y			o	m	n	i	k	a	c	h		n	a	
T	D	Y	P		Z	J	D	L	T	E	D	H	W	E	J	M	K
	o					i	o			k	o			k	i	e	
	J		K	S	T	S	N	X	S		L	S	C		N	X	M
	i			a		a	c	z	a			a	m		c	z	e
E	S	C	U		I	S		J	I	I	U	N	A				
k	a	m	y		n	a		i	n	n	y	c	h				

Etap V: Uzupełnianie luk

- Stosując metodę prób i błędów dla najbardziej prawdopodobnych liter dostajemy **Z = p, T = r.**

Etap V: Uzupełnianie luk

X	F	J	D	T	E	S		D		B	T	P	Y	J	M	K	
z		i	o	r	k	a		o			r			i	e		
Z	T	X	U		Z	D	C	I	J	E	S	N	A		I	S	
p	r	z	y		p	o	m	n	i	k	a	c	h		n	a	
T	D	Y	P		Z	J	D	L	T	E	D	H	W	E	J	M	K
r	o				p	i	o		r	k	o			k	i	e	
	J		K	S	T	S	N	X	S		L	S	C		N	X	M
	i			a	r	a	c	z	a			a	m		c	z	e
E	S	C	U		I	S		J	I	I	U	N	A				
k	a	m	y		n	a		i	n	n	y	c	h				

Etap V: Uzupełnianie luk

- Teraz można już bez trudu uzupełnić resztę tekstu.

Etap V: Uzupelnianie luk

X	F	J	D	T	E	S		D		B	T	P	Y	J	M	K	
z	b	i	o	r	k	a		o		d	r	u	g	i	e	j	
Z	T	X	U		Z	D	C	I	J	E	S	N	A		I	S	
p	r	z	y		p	o	m	n	i	k	a	c	h		n	a	
T	D	Y	P		Z	J	D	L	T	E	D	H	W	E	J	M	K
r	o	g	u		p	i	o	t	r	k	o	w	s	k	i	e	j
	J		K	S	T	S	N	X	S		L	S	C		N	X	M
	i		j	a	r	a	c	z	a		t	a	m		c	z	e
E	S	C	U		I	S		J	I	I	U	N	A				
k	a	m	y		n	a		i	n	n	y	c	h				