Luštění Vigenerovy šifry, příklad první

Otevřený text je v češtině a je napsán bez mezer. Byla použita Vigenerova šifra. Najděte otevřený text.

```
ZFFLN QATOO AVFTS GQKZN MUXXB FJVVZ FBEPO FQKTN ADTCB OFLEB
UQKZQ ATNKP HBGTV EQXNI GOTXB FFFLU UDDPP XZFAJ MEXGF PFODB
WQKPT LLHFN MOBKE MTXGF ELNEF OOHDU UTHFU QABNJ BPSOF VJLEB HBCTP
BSTGE AWRXJ YBMPN MUBVZ
```

Řešení. Celý text má 175 znaků. Nejdříve najdeme všechny opakované trigramy.

Trigram	poprvé	podruhé	Rozdíl
FFL	2	77	75
LEB	48	148	100
NMU	20	170	150
QAT	6	55	49
QKZ	17	52	35
XBF	24	74	50
XGF	93	118	25

Vidíme, že s výjimkou trigramu QAT jsou všechny ostatní opakované trigramy ve vzdálenostech, které jsou násobkem 5. Nejpravděpodobnější délkou klíče je proto 5. Šifrový text si napíšeme do pěti sloupců.

1	2	3	4	5
Z	F	F	L	N
Q	A	T	O	O
A	V	F	T	S
G	Q	K	Z	N
M	U	X	X	B
F	J	V	V	Z
F	B	E	P	O
F	Q	K	T	N
A	D	T	C	B
O	F	L	E	B
U	Q	K	Z	N
A	T	A	K	P
H	B	G	T	V
E	Q	X	A	I
G	O	T	X	B
F	F	F	L	U
U	D	D	P	P
X	Z	F	A	J
M	E	X	G	F
P	F	O	D	B
W	Q	K	P	T
L	L	H	F	N
M	O	B	K	E
M	T	X	G	F
E	L	A	E	F
O	O	H	D	U

U	T	H	F	U
Q	A	B	A	J
B	P	S	O	F
V	J	L	E	B
H	B	C	T	P
B	S	T	G	E
A	W	R	X	J
Y	B	M	P	N
M	U	B	V	Z

Četnosti písmen ve sloupcích:

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

Jako poslední krok řešení musíme rozhodnout, jaká posunutí abecedy byla použita v jednotlivých sloupcích. Tentokrát použijeme následující tabulku frekvencí pěti nejméně používaných písmen v jednotlivých jazycích a pěti nejvíce používaných písmen v jednotlivých jazycích.

Angl.	Franc.	Něm.	Čeština	Slov.
K:0,41	Y:0,21	P:0,54	G:0,48	G:0,40
Q:0,17	J:0,19	J:0,16	F:0,33	F:0,31
X:0,17	Z:0,07	Q:0,01	W:0,06	W:0,06
J:0,16	K:0,00	X:0,00	X:0,04	X:0,03
Z:0,05	W:0,00	Y:0,00	Q:0,00	Q:0,00
∑:0,96	∑:0,47	∑:0,71	∑:0,91	∑:0,80

Angl.	Franc.	Něm.	Čeština	Slov.
E: 12,86	E: 17,76	E: 19,18	E: 10,13	A: 9,49
T: 9,72	S: 8,23	N: 10,20	A: 8,99	O: 9,34
A: 7,96	A: 7,68	I: 8,21	O: 8,39	E: 9,16
I: 7,77	N: 7,61	S: 7,07	I: 6,92	I: 6,81
N: 7,51	T: 7,30	R: 7,01	N: 6,64	N: 6,34
R: 6,83	I: 7,23	T: 5,86	S: 5,74	S: 5,94
Σ: 52,65	Σ: 55,81	Σ: 57,53	Σ: 46,81	Σ: 47,08

Pro každý sloupec je třeba zjistit, o jakou vzdálenost je text posunutý. Použijeme např. kvantizační odchylku: od součtu četností znaků AEION se odečte součet četností znaků FGWXQ a zapíše se do prvního sloupce pro znak A . Hodnota pro znak B se určí odečtením součtu GHXYR od součtu BFJPO, atd. Pro ruční řešení si připravíme pomůcku v podobě proužku papíru, na který si do sloupce napíšeme jednotlivá písmena abecedy a vyznačíme si červeně pět nejčastěji se vyskytujících písmen v českých textech a modře pět nejméně často se vyskytujících písmen. Tento proužek papíru budeme postupně posunovat podél tabulky výskytu písmen v jednotlivých sloupcích a spočteme četnosti jednotlivých písmen vedle červených míst a od nich odečteme četnosti písmen vedle modrých míst. Dostaneme tak následující tabulku. Naším cílem je najít takové posunutí, při kterém je tento rozdíl maximální. Můžeme dokonce odhadnout, kolik asi by mělo toto maximum být. Protože součet frekvencí nejčastějších pěti písmen v českých otevřených textech, tj. písmen E, A, O, I a N je 41,07%, tj. 0,4107 a součet frekvencí pěti nejméně častých písmen je 0,91%, tj. 0,0091, je tento rozdíl pro jedno písmeno v průměru 0,4107-0,0091 = 0,4016. Protože naše sloupce mají vždy 35 písmen, mělo by se maximum rozdílů v jednotlivých sloupcích rovnat přibližně 35* 0,4016 = 14,056.

	1	2	3	4	5
A	-2	-4	-5	0	5
B	-1	-5	1	-5	-3
C	1	-7	1	-3	-8
D	5	3	9	0	1
E	0	-2	-5	1	-3
F	-5	-3	-5	3	0
G	-1	-6	-1	-4	1
H	2	-2	**15**	4	-2
I	6	4	1	-3	-1
J	2	3	-3	2	2
K	-1	8	-3	-2	0
L	-12	-3	2	8	-4
M	-1	1	-5	-8	3
N	7	-1	-3	-3	7
O	**15**	7	-2	2	1
P	-2	-2	-1	**12**	-2
Q	-7	-4	6	-1	-1
R	-8	-6	1	-8	-3
S	0	-1	-6	-7	-10

T	0	3	-3	-2	0
U	0	3	5	6	6
V	0	3	3	1	7
W	-2	-8	-3	0	-8
X	0	1	3	-3	-6
Y	1	4	2	8	0
Z	3	**14**	-1	2	**18**

Tučně vyznačená maxima se skutečně kolem odhadnuté velikosti průměrného maximálního rozdílu 14 pohybují. Šifrovému A tak v prvním sloupci nejspíše odpovídá otevřené O, ve druhém sloupci otevřené Z, ve třetím otevřené H, ve čtvrtém otevřené P a v pátém otevřené Z. V prvním případě tak otevřený text dostaneme z šifrového posunutím o 14, v druhém sloupci posunutím o 25, ve třetím posunutím o 7, ve čtvrtém posunutím o 15 a v pátém posunutím o 25. Klíč pro dešifrování je proto 14,25,7,15,25. Pokud tedy každou pětici šifrového textu posuneme podle tohoto klíče, dostaneme otevřený text

```
NEMAME ZADNOU MIRU PRO MATEMATICKY TALENT PRIMOCARA CESTA I
PRO POSUZOVANI USPECHU NA MATEMATICKE OLYMPIADE VEDE VSAK PRES
ZKOUMANI ZDA SE VE SKUTECNOSTI SOUTEZICI POZDEJI STAVAJI
OPRAVDOVYMI MATEMATIKY
```

Klíč pro šifrování tedy byl 12,1,19,11,1. Pokud jej převedeme na písmena, pak je to MBTLB. Všimněte si rovněž, že od okamžiku nalezení délky klíče jsme už postupovali zcela mechanicky pouze s využitím tabulky četností písmen v příslušném jazyce.