ÚLOHA C

Vyluštěte text a zjistěte kdo je jeho autorem.

Použita je jedna ze základních klasických šifer.

```
LFJIV LYXBF NGYTE ZTVBF NXAWX UKJMV TUMVT YBYSL VMVGB FNAIZ
WYTBU SZWLY KYGHY CXUJI ULNXJ FYSTZ TNGSZ BFNCL YGTNX ZBFNS
TZLKY AHVFN IULNX VULBU TYLXY CNWUL HFNIU LNXMJ LJGHZ GHXZA
NEKNA YAZGB USFUA NHNEM JLJGH ZGHXY CGNKY AHVFN IULNX VJWMJ
LJFUW YAIKY TZZXY STNLX ZUMCY KNEIU LXUHJ GHYGH NZTYS LVWMJ
LYGBF NAIZW YHKFJ WXUJL UMJXY MJLJX NSLVK YLYHS LVGYE ZEHYG
NHXZH KJCBF NAIUL CYHFY MZWZA IUKZK ZHFZL AUCYH UFZLW YBHZT
GYEZT VBFNX ANHUC YXYAU IULXY WZBUE YXJHY IUULB UKYLY TZTNG
SZHUA UULTN GJCYC YLYXL YXULL FJIYI UCYLX JIULN XJULL FJIYE
UCNTU KANXZ BFNST ZLWZA IUKZK ZCNHZ SYFZL SZWLV AHKFH YSAIU
LNHZX ANHGL YKAZH VWKYG XNAYS ZWLVA HKFHY SCYBF YSFZG XVLYX
EUIJC NHXZB FUAIZ WSJZW LUKNX NAYSL VMVTU KANHZ XANTN SLVSU
TNKKG YAIXV LXVMV GYBUL UMZTV CYLYX LFJIY EJZCZ MVAIX YEYTZ
KJMYA BFZWL XNXVH ZSGNE ZTVBF NXAUA IUANT TNGSJ ZTYBF NMTNW
NTZGY IULNX ZULAI ULJZG NMJLJ BTZSZ HGHVG STZGN TNGSZ HUCYH
KZKNX ZFYST EZTVB FNXAX YAIHY TCGYE HNJMT NWNHZ TYHVG AIHYT
ZZMVA IGNHY UAIUA NTUKG YEULB UKYLY TZTNG SZHNE XNAXY WNGSZ
GWNGS ZEKWB UEYXG NXZMZ FKJUM NTNZL ULZTZ CLNGY BULNK ZHCYG
HYCYL XUJXZ FJWYB UAIUB NGWYH ZHKZC YCYLN XZXZG KYHYB FNCLY
GENLZ HGMUI YEZCZ HNLZE LZFYS HZCYE GHKN
```

Otázka (6 bodů)

- Kdo je autorem textu (zadejte velkými písmeny pouze příjmení a bez diakritiky):
  správná odpověď: **********

OT / abc def ghi jkl mno pqr stu vwx yz
ŠT / ZMA LYP RIN CST EXU BDF GHJ KOQ VW

Řešení:

Jde o klasickou šifru:

- jednoduchá záměna
- transpoziční šifra
- periodická šifra
- …

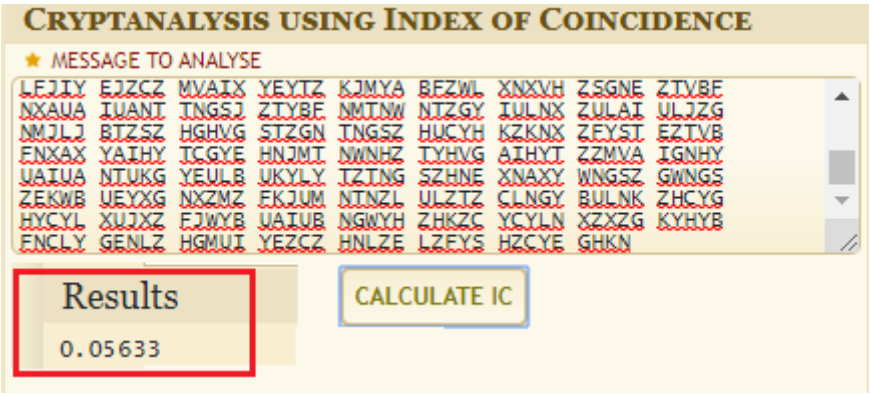
Typ / jazyk

How to calculate a coincidence index?

Index of coincidence uses the formula:

$$IC = \sum_{i=A}^{i=Z} \frac{n_i(n_i - 1)}{N(N-1)}$$

with n_i the number of occurrences of the letter i in the text and N the total number of letters.



jazyk	hodnota IC (κ_p)
Němčina	0,0824
Francouzština	0,0801
Španělština	0,0769
Italština	0,0754
Angličtina	0,0676
Slovenština	0,0581
Čeština	0,0577
Ruština	0,0470
Náhodný text (κ_r)	0,0385

Index koincidence pro vybrané druhy jazyků

Y	Z	N	L	U	H	X	G	T	A	F	J	K
95	87	81	71	67	55	54	48	45	43	38	38	35
9.7%	8.8%	8.2%	7.2%	6.8%	5.6%	5.5%	4.9%	4.6%	4.4%	3.9%	3.9%	3.6%
I	**S**	**V**	**B**	**C**	**W**	**E**	**M**	**D**	**O**	**P**	**Q**	**R**
33	32	32	30	28	26	23	23	0	0	0	0	0
3.4%	3.3%	3.3%	3.0%	2.8%	2.6%	2.3%	2.3%	0.0%	0.0%	0.0%	0.0%	0.0%

Celkem	precteno:		984	
četnost	v procentech			znak
43	4,37	0,043699190		**A**
30	3,05	0,030487800		**B**
28	2,85	0,028455280		**C**
0	0,00	0,000000000		**D**
23	2,34	0,023373980		**E**
38	3,86	0,038617890		**F**
48	4,88	0,048780490		**G**
55	5,59	0,055894310		**H**
33	3,35	0,033536590		**I**
38	3,86	0,038617890		**J**
35	3,56	0,035569110		**K**
71	7,22	0,072154470		**L**
23	2,34	0,023373980		**M**
81	8,23	0,082317070		**N**
0	0,00	0,000000000		**O**
0	0,00	0,000000000		**P**
0	0,00	0,000000000		**Q**
0	0,00	0,000000000		**R**
32	3,25	0,032520330		**S**
45	4,57	0,045731710		**T**
67	6,81	0,068089430		**U**
32	3,25	0,032520330		**V**
26	2,64	0,026422760		**W**
54	5,49	0,054878050		**X**
95	9,65	0,096544720		**Y**
87	8,84	0,088414630		**Z**
984	**100,00**			

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

d) Bigramy
ST, PR, SK, CH, DN, TR

e) Zvláštnosti frekventních souhláskových bigramů v češtině

ST :
- S a T má přibližně stejnou frekvenci
- existuje i bigram TS
- je součástí velkého počtu souhláskových trigramů STR, STN, STL, STV ...
- vyskytuje se uprostred i na konci slova

PR :
- P má přibližně poloviční frekvenci než R
- obrácený bigram RP se téměř nevyskytuje (chrpa)
- zpravidla nelze rozšířit "dozadu" na souhláskový trigram (PRV)
- lze rozšířit dopředu na samohláskový trigram (SPR, ZPR, ...)
- zpravidla stojí na počátku slov

CH:
- H má jen o něco menší frekvenci než C (při krátkých textech nemusí platit)
- bývá zpravidla na konci slov spolu se samohláskami Y,I,A,E (YCH, ICH, ACH, ECH)
- většinou platí : předchází-li CH souhláska, následuje po něm samohláska a naopak (OBCHOD, NECHTĚL)

f) Trigramy
PRO, UNI, OST, STA, ANI, OVA, YCH, STI, PRI, PRE, OJE, REN, IST, **STR**(nejběžnější souhláskový trigram !) , EHO , TER, RED, ICH, ...

Celkem	Znaků ŠT	984	ŠT		frekvence	pořadí		OT
95	9,65	0,096544720	Y		E	1		E
87	8,84	0,088414630	Z		A	2		A
81	8,23	0,082317070	N		I	3		I
71	7,22	0,072154470	L		O	4		D
67	6,81	0,068089430	U		N	5		O
55	5,59	0,055894310	H		S	6		T
54	5,49	0,054878050	X		T	7		N
48	4,88	0,048780490	G		R	8		S
45	4,57	0,045731710	T		V	9		L
43	4,37	0,043699190	A		U	10		C
38	3,86	0,038617890	F		L	11		R
38	3,86	0,038617890	J		Z	12		U
35	3,56	0,035569110	K		D	13		V
33	3,35	0,033536590	I		K	14		H
32	3,25	0,032520330	S		P	15		K
32	3,25	0,032520330	V		M	16		Y
30	3,05	0,030487800	B		C	17		P
28	2,85	0,028455280	C		Y	18		J
26	2,64	0,026422760	W		H	19		Z
23	2,34	0,023373980	E		J	20		M
23	2,34	0,023373980	M		B	21		B
0	0,00	0,000000000	D		G	22		Q
0	0,00	0,000000000	O		F	23		W
0	0,00	0,000000000	P		X	24		F
0	0,00	0,000000000	Q		W	25		X
0	0,00	0,000000000	R		Q	26		G
984	100,00 Kč							

OT / abc def ghi jkl mno pqr stu vwx yz

ŠT / ZMA LYP RIN CST EXU BDF GHJ KOQ VW

GENLZ HGMUI YEZCZ HNLZE LZFYS HZCYE GHKN

i*a *** e*a*a *i*a* *a*e* *a*e* ***i

PR, CH, ST

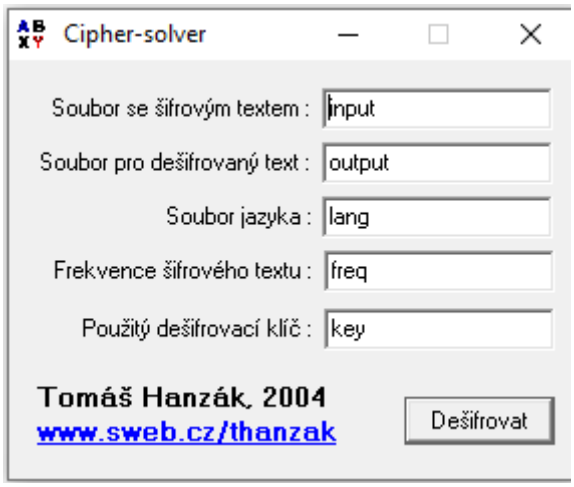
Frekvence bigramů [%]

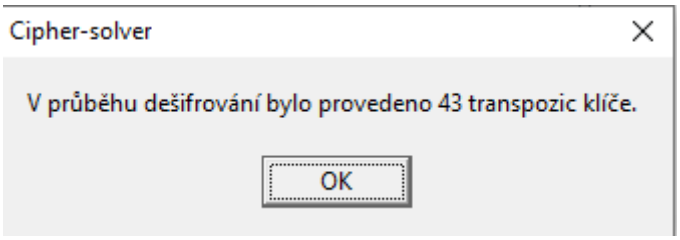
*	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	0	1	0	0	0	0	0	4	17	0	0	0	0	9	0	0	0	0	0	0	4	0	1	2	3	2
B	0	0	0	0	0	16	0	1	0	0	0	0	0	1	0	0	0	0	0	1	10	0	0	0	1	0
C	0	1	0	0	0	0	2	0	0	0	0	3	0	4	0	0	0	0	0	0	0	0	0	1	15	2
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	1	2	1	1	2	1	1	1	0	0	0	0	0	0	3	0	0	1	3	6
F	0	0	0	0	0	0	0	2	0	6	1	0	0	16	0	0	0	0	0	0	3	0	0	0	5	5
G	1	3	0	0	1	0	0	10	0	1	1	1	1	7	0	0	0	0	8	1	0	0	2	2	9	0
H	0	0	1	0	0	3	3	0	0	1	7	0	0	5	0	0	0	0	1	0	4	5	0	4	11	10
I	0	0	0	0	0	0	1	2	0	1	1	0	0	0	0	0	0	0	0	0	18	1	0	2	4	3
J	0	1	3	0	0	2	3	1	6	0	0	7	3	0	0	0	0	0	0	0	2	0	3	3	0	4
K	3	0	0	0	0	3	2	0	0	4	1	0	0	5	0	0	0	0	0	0	0	0	1	0	9	7
L	2	3	1	0	0	4	0	1	0	6	1	2	0	10	0	0	0	0	1	1	4	8	2	8	13	4
M	0	0	1	0	0	0	0	0	0	7	0	0	0	1	0	0	0	0	0	2	1	7	0	0	1	3
N	7	0	2	0	5	0	11	10	3	1	3	3	2	0	0	0	0	0	4	7	0	0	3	17	0	2
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	1	0	1	0	0	2	1	1	0	2	0	6	0	0	0	0	0	0	0	6	1	0	0	0	1	10
T	0	1	1	0	2	0	1	0	0	0	0	0	0	13	0	0	0	0	0	1	4	5	0	0	5	12
U	8	1	5	0	2	1	0	1	3	3	9	22	5	0	0	0	0	0	2	2	2	0	1	0	0	0
V	4	4	1	0	0	2	5	1	0	1	1	3	3	0	0	0	0	0	1	3	1	0	2	0	0	0
W	0	1	0	0	0	0	0	0	0	0	1	5	2	4	0	0	0	0	1	0	1	0	0	2	6	3
X	6	1	0	0	1	0	1	0	0	4	0	2	1	5	0	0	0	0	0	0	6	6	0	0	9	12
Y	8	8	6	0	9	1	8	8	3	0	3	9	2	0	0	0	0	0	11	8	1	0	2	8	0	0
Z	3	5	6	0	3	4	9	11	0	0	4	6	3	0	0	0	0	0	3	13	2	0	9	4	0	2

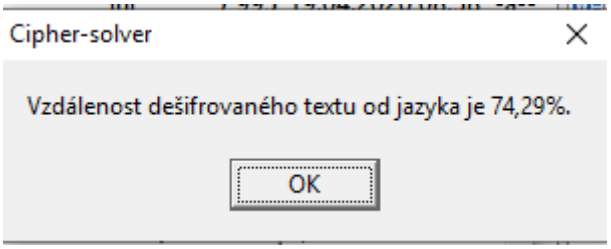
Lze využít i strojové luštění jednoduché záměny. Pro český jazyk však tyto nástroje na webu není lehké vyhledat.

Jedním z takových SW, které využívá mimo frekvence znaků a i bigramové frekvence je program Tomáše Hanzáka, který jej vytvořil již v roce 2004 při studiu na MFF UK - Cipher-solver http://thanzak.sweb.cz/

Tento program po zadání šifrového textu nalezne zcela správný otevřený text.








```
druhy denpr iselm alypr inczn ovuby lobyl epekd ybysp richa zelpo
kazde veste jnouh odinu rekla liska prijd eslin aprik ladve ctyri
hodin yodpo ledne jizod triho dinbu dusta stnac imvic ecasp okroc
itimb udust astne jsive ctyri hodin yuzbu duroz echve laane klidn
aobje vimho dnotu stest ialek dyzbu despr ichaz etvru znoud obune
budun ikdyv edetk dysem amtes itnat vujpr ichod jetre bazac hovav
atrad cojet oradz eptal semal yprin citoj eneco hodne zapom enute
hoodp ovede lalis katoc oodli sujej edend enodd ruheh ojedn uhodi
nuodd ruhem ojilo vcina prikl adzac hovav ajita kerad kazdy ctvrt
ekcho ditan citsd evcat yzves nicek azdyc tvrte kjepr ekras nyden
mohuj itnap rocha zkuaz dovin icekd ybylo vcita ncili kdyko livvs
echny dnyby sepod obaly jeden druhe muaja bychn emela vubec prazd
ninyt aksim alypr incoc hocil lisku alepr ibliz ilase hodin aodch
```

```
oduas ibudu plaka tstys klasi liska tojet vavin arekl malyp rincn
echte ljsem tiubl izita letys chtel aabyc hsite ochoc ilovs emodp
ovede lalis katim nicne ziska szisk amvzp omens inaba rvuob iliad
odala jdise podiv atjes tejed nouna ruzep ochop iszet atvaj ejedi
nanas vetep rijde smida tsboh emaja tidam darek tajem stvi
```

Otevřený text

```
DRUHY DENPR ISELM ALYPR INCZN OVUBY LOBYL EPEKD YBYSP RICHA
ZELPO KAZDE VESTE JNOUH ODINU REKLA LISKA PRIJD ESLIN APRIK
LADVE CTYRI HODIN YODPO LEDNE JIZOD TRIHO DINBU DUSTA STNAC
IMVIC ECASP OKROC ITIMB UDUST ASTNE JSIVE CTYRI HODIN YUZBU
DUROZ ECHVE LAANE KLIDN AOBJE VIMHO DNOTU STEST IALEK DYZBU
DESPR ICHAZ ETVRU ZNOUD OBUNE BUDUN IKDYV EDETK DYSEM AMTES
ITNAT VUJPR ICHOD JETRE BAZAC HOVAV ATRAD COJET ORADZ EPTAL
SEMAL YPRIN CITOJ ENECO HODNE ZAPOM ENUTE HOODP OVEDE LALIS
KATOC OODLI SUJEJ EDEND ENODD RUHEH OJEDN UHODI NUODD RUHEM
OJILO VCINA PRIKL ADZAC HOVAV AJITA KERAD KAZDY CTVRT EKCHO
DITAN CITSD EVCAT YZVES NICEK AZDYC TVRTE KJEPR EKRAS NYDEN
MOHUJ ITNAP ROCHA ZKUAZ DOVIN ICEKD YBYLO VCITA NCILI KDYKO
LIVVS ECHNY DNYBY SEPOD OBALY JEDEN DRUHE MUAJA BYCHN EMELA
VUBEC PRAZD NINYT AKSIM ALYPR INCOC HOCIL LISKU ALEPR IBLIZ
ILASE HODIN AODCH ODUAS IBUDU PLAKA TSTYS KLASI LISKA TOJET
VAVIN AREKL MALYP RINCN ECHTE LJSEM TIUBL IZITA LETYS CHTEL
AABYC HSITE OCHOC ILOVS EMODP OVEDE LALIS KATIM NICNE ZISKA
SZISK AMVZP OMENS INABA RVUOB ILIAD ODALA JDISE PODIV ATJES
TEJED NOUNA RUZEP OCHOP ISZET ATVAJ EJEDI NANAS VETEP RIJDE
SMIDA TSBOH EMAJA TIDAM DAREK TAJEM STVI
```

Druhý den přišel malý princ znovu.

„Bylo by lépe, kdybys přicházel pokaždé ve stejnou hodinu," řekla liška. „Přijdeš-li například ve čtyři hodiny odpoledne, již od tří hodin budu šťastná. Čím více čas pokročí, tím budu šťastnější. Ve čtyři hodiny už budu rozechvělá a neklidná; objevím hodnotu štěstí! Ale když budeš přicházet v různou dobu, nebudu nikdy vědět, kdy se mám těšit na tvůj příchod... Je třeba zachovávat řád."
„Co je to řád?" zeptal se malý princ.

„I to je něco hodně zapomenutého," odpověděla liška, „to, co odlišuje jeden den od druhého, jednu hodinu od druhé. Moji lovci například zachovávají také řád. Každý čtvrtek chodí tančit s děvčaty z vesnice. Každý čtvrtek je překrásný den! Mohu jít na procházku až do vinice. Kdyby lovci tančili kdykoliv, všechny dny by se podobaly jeden druhému a já bych neměla vůbec prázdniny."

Tak si malý princ ochočil lišku. Ale přiblížila se hodina odchodu.

„Asi budu plakat...," stýskla si liška.

„To je tvá vina," řekl malý princ. „Nechtěl jsem ti ublížit, ale tys chtěla, abych si tě ochočil..."

„Ovšem," odpověděla liška.

„Tím nic nezískáš!"

„Získám, vzpomeň si na barvu obilí." A dodala: „Jdi se podívat ještě jednou na růže. Pochopíš, že ta tvá je jediná na světě. Přijdeš mi dát sbohem a já ti dám dárek – tajemství."

Text je ukázkou z knihy **Malý princ**. Text je natolik známý, že by nemělo činit studentům zjistit, že jde o ukázku z této knihy a podle toho určit příjmení autora.

V textu se slovo „Malý princ“ nachází. Určení knihy + autora by tedy nemělo dělat potíže.

Otázka (6 bodů)

- Kdo je autorem textu (zadejte velkými písmeny pouze příjmení a bez diakritiky):
  správná odpověď: EXUPERY