

# PRESEK

List za mlade matematike, fizike, astronome in računalnikarje

ISSN 0351-6652

Letnik 5 (1977/1978)

Številka 1

Strani 40-42

Tomaž Pisanski:

## SKRIVNOSTNO SPOROČILO

Ključne besede: matematika, verjetnostni račun, tajna pisava, relativna frekvenca.

Elektronska verzija: <http://www.presek.si/5/5-1-Pisanski-Skri.pdf>

© 1977 Društvo matematikov, fizikov in astronomov Slovenije

© 2010 DMFA - založništvo

Vse pravice pridržane. Razmnoževanje ali reproduciranje celote ali posameznih delov brez poprejšnjega dovoljenja založnika ni dovoljeno.

SKRIVNOSTNO SPOROČILO

Zagotovo ste že razmišljali, kako lepo bi bilo, če bi imeli na voljo tajno pisavo. Lahko bi pisali pisma po mili volji. Le tisti, ki bi poznal skrivnost tajne pisave, bi jih znal prebrati. Ali res? Pred nami je sporočilo, napisano v taki tajni pisavi.



$$\{0, e, i, o\} = \{-, \check{C}, J, H\}$$

(e is also written as e) (i is also written as i)

Č obliko - H, J, Ć nato ni namygnol



J+ = iz (11)

(D), ob, od

up, os, uk

11

11

11

11

11

Na prvi pogled je nemogoče prodreti v skrivnost pisave. Kot pravi detektivi začnimo z najpreprostejšo možnostjo. Domnevajmo, da je sporočilo napisano v slovensščini. Domnevajmo, da je vsaka črka zamenjana z določenim znakom, npr. namesto A piše +, namesto B stoji -, namesto C je S in podobno. Seveda ne vemo, s katerim znakom je v resnici zamenjana črka. Če bi to vedeli, bi preprosto prebrali, kaj piše. Če bi hoteli preskusiti vse možnosti, bi obupali. Znak + lahko pomeni katerokoli od 25 črk ali presledek, torej 26 možnosti. Znak - lahko nadomešča kateragakoli od preostalih 25 znakov. Naslednji znak lahko pomeni enega od 24 znakov, naslednji enega od 23 znakov in tako dalje. Preizkusiti bi torej morali

$26 \times 25 \times 24 \times \dots \times 3 \times 2 \times 1 = 40329146112665635584000000$  možnosti. Na videz neresljivi problem. Vendar ni tako. Na pomoč bomo poklicali *statistiko*. Razporednica kaže, kako pogoste so posamezne črke v slovensščini.

presledek	E	A	I	O	N	R	S	L	J	T	V	D
173	89	84	74	73	57	44	43	39	37	37	33	30
K	M	P	U	Z	B	G	Č	H	Š	C	Ž	F
29	27	26	18	17	15	12	12	9	9	6	6	1

Pod vsako črko oziroma presledkom je napisana njena *relativna frekvenca*, izražena v promilih. Preden povemo, kako lahko tabelo uporabimo, povejmo, kako so jo sestavili.

Zbrali so nekaj slovenskih besedil. Tako so dobili *povprečno slovensko besedilo*. V tem besedilu so prešteli vse znake (črke in presledek). Potem pa so prešteli kolikokrat se posamezni znak pojavi v besedilu. Seveda je dolgočasno nalogo seštevanja opravil računalnik. Ulomek

#### število izbranih znakov v besedilu

število vseh znakov besedila

imenujemo *relativna frekvenca znaka*. Števec ulomka pa je *absolutna frekvenca znaka*. Če relativno frekvenco pomnožimo s 1000, dobimo izražavo v promilih. Zdaj pa pomislimo, kaj nam pove številka 84 pod črko A v tabeli. Če pogledamo katerikoli slovenski tekst, bo med 1000 znaki približno 84 znakov A. Po drugi strani pa vidimo, da je A poleg presledka in črke E najpogostejši znak v slovenščini. Pri reševanju naše naloge ne bomo potrebovali relativnih frekvenc znakov. Dovolj je že, da iz tabele razberemo, kako si sledijo po pogostosti presledek, črke E, A, I, O, N in tako naprej do črke F, ki je najmanj uporabljana črka v slovenščini.

Zdaj moramo sestaviti tabelo absolutnih frekvenc znakov skrivnostnega sporočila. K vsakemu znaku zapišemo število, ki pove, kolikokrat se je znak ponovil v sporočilu. Tabela uredimo tako, da bo na prvem mestu najpogostejši znak, za njim drugi najpogostejši znak in tako naprej. Zdaj najpogostejši znak v sporočilu zamenjamo s presledkom, drugega najpogostejšega s črko E in tako naprej. Seveda se lahko zgodi, da se vrstni red pogostosti znakov našega besedila ne ujema točno z vrstnim redom v povprečnem slovenskem besedilu. Takrat moramo uporabiti naše male sive celice.

Pri reševanju naloge vam bodo v pomoč tudi tile podatki:

- na začetku besed so najpogostejše črke  
N, S, K, T, J, L in tako naprej;
- najpogostejše končnice pa so  
E, A, I, O, U, R in N

