

PORAVNAVA NIZOV IN DELANNOYJEVA ŠTEVILA

MARKO RAZPET

Pedagoška fakulteta v Ljubljani

Math. Subj. Class. (2010): 05A15, 40B05, 68R15, 92D20

Pokazali bomo, kako je poravnava nizov povezana z Delannoyjevimi števili $D(m, n)$, s katerimi zapišemo število mrežnih poti v množici $\mathbb{N} \times \mathbb{N}$ od točke $(0, 0)$ do dane točke (m, n) , pri čemer so dovoljeni koraki v smereh $(1, 0)$, $(0, 1)$ in $(1, 1)$. Vpeljali bomo Levenštejnovo razdaljo med nizoma in na kratko predstavili njeno uporabo v genetiki.

SEQUENCE ALIGNMENT AND DELANNOY NUMBERS

We will show how the sequence alignment is connected with the Delannoy numbers $D(m, n)$ which express the number of lattice paths in the set $\mathbb{N} \times \mathbb{N}$ with allowed steps $(1, 0)$, $(0, 1)$ and $(1, 1)$. The Levenshtein distance between two sequences is introduced and its application in genetics is briefly presented.

Uvod

Kdor piše z računalniškim urejevalnikom besedil, zagotovo pozna poravnavo vrstic: levo, desno, sredinsko in obojestransko. Po navadi imamo najraje obojestransko poravnano besedilo, pri čemer sta vnaprej določena levi in desni rob, ki nam kot vidna ali nevidna ravna črta ne dovoljujeta, da bi kakšna črka, številka, ločilo ali kakšen drug znak padel bolj v levo od levega roba oziroma bolj v desno od desnega roba. Pomembno vlogo pri tem igra tudi znak za presledek, kateremu je namenjena posebna tipka. S presledki na primer ločujemo besede in nakažemo nov odstavek, lahko pa si mislimo, da z njimi na desno poravnamo prekratko vrstico. Ker običajno prostori za grafične znake niso enake širine, del težav s poravnavo odpade, ker urejevalniki po potrebi sami rahlo zgoščajo in redčijo razmike med znaki.

Če pa izberemo znake, ki zahtevajo enake širine, potem imamo z obojestransko poravnavo take težave kot nekoč na klasičnih pisalnih strojih. In ravno taka poravnava nas bo v nadaljevanju najbolj zanimala. Spoznali bomo, da je poravnava besed oziroma nizov kar uporabna. Vpeljemo namreč lahko mero, koliko sta si niza blizu. Ker pa so znaki ali elementi niza lahko marsikaj, lahko preverjamo ne le podobnost besedil in s tem odkrivamo na primer plagiatorstvo, ampak tudi podobnost tonskih zapisov in nukleotidnih zaporedij v genetiki. Nekaj več o tem bomo povedali v zadnjem razdelku.