

KAKO IŠČE GOOGLE?

MARJETA KRAMAR FIJAVŽ

Fakulteta za gradbeništvo in geodezijo
Univerza v Ljubljani

Math. Subj. Class. (2010): 15B48, 15A18, 47H10

Srce spletnega iskalnika Google je algoritem PageRank. V sestavku predstavimo osnovno idejo algoritma ter si ogledamo njegovo teoretično ozadje. Konvergenco algoritma bomo utemeljili na dva načina, s Perron-Frobeniusovo teorijo za pozitivne matrike ter s pomočjo Banachovega izreka o negibni točki.

HOW GOOGLE WORKS?

We present the main idea of the PageRank algorithm which is the core of the Google search engine. We concentrate on the theoretical background of the algorithm and prove its convergence in two different ways, by the Perron-Frobenius theory for positive matrices and using Banach fixed point theorem.

Uvod

Že od začetka leta 1989 (za ustanovitelja velja Tim Berners-Lee) se je svetovni splet zelo hitro širil in kmalu dobil glavno vlogo v prenosu informacij. Svetovni splet je ogromen¹ in neprestano raste. Poleg tega se stalno spreminja: 40 % strani spremeni vsebino tedensko, nastajajo nove in izginjajo stare strani. Splet je samoorganiziran s pomočjo raznovrstnih medsebojnih povezav. Gre torej za ogromno knjižnico podatkov, ki nima ne kataloga ne knjižničarjev. Kako se tu znajti?

Vzporedno z nastajanjem spleta so se razvili spletni iskalniki, ki uporabniku z vnosom ključnih besed pomagajo najti ustrezno stran. Med različnimi iskalniki je zadnja leta najbolj znan Google. Spletni iskalnik Google² sta leta 1998 zagnala Sergey Brin in Larry Page, takrat doktorska študenta na Stanfordski univerzi v Kaliforniji.

Vsak spletni iskalnik ima svojo bazo spletnih strani, ki se seveda neprestano spreminja. Gradi jo s pomočjo avtomatskega programa, ki po spletu stalno pošilja virtualne robote, imenovane *pajki*. Pajki potujejo po spletnih povezavah ter vsako obiskano stran oštevilčijo in indeksirajo njeno vsebino (naslov, ključne besede, imena povezav, sidra ipd.). Tako nastane baza spletnih strani s stvarnim kazalom. Ko uporabnik v iskalnik vtipka poizvedbo,

¹19. 1. 2014 obstaja vsaj 1.75 milijarde spletnih strani, <http://www.worldwidewebsite.com/>.

²Ime Google naj bi izviralo iz angleške besede »googol« (sl. gúgol), ki pomeni število 10^{100} .