Osječki matematički list 17 (2017), 51-62

Rangiranje web stranica

Damir Horvat[*] Dušan Munđar[†]

Sažetak

U ovom članku opisana je matematička pozadina PageRank algoritma kojeg Google koristi kod rangiranja web stranica. Objašnjene su dvije metode u radu algoritma. Prva metoda je metoda potencija koja je iterativna metoda, a druga metoda se svodi na rješavanje sustava linearnih jednadžbi. Obje metode povezane su s problemom pronalaska svojstvenog vektora pridruženog dominantnoj svojstvenoj vrijednosti odgovarajuće matrice. Funkcioniranje algoritma je pokazano na jednom malom primjeru s četiri web stranice.

Ključne riječi: *Google PageRank, metoda potencija, stohastička matrica, Perron-Frobeniusov teorem*

Ranking websites

Abstract

In this paper we describe the mathematical foundations of the Google's PageRank algorithm. We explain two methods used by the algorithm. The first one, the method of powers, is an iterative method. The second method is founded on solving a system of linear equations. Both methods are related to the problem of finding an eigenvector of the dominant eigenvalue of the corresponding matrix. Functioning of the algorithm is illustrated on a small example of four web pages.

Keywords: *Google PageRank, power method, stochastic matrix, Perron-Frobenius theorem*

[*]Fakultet organizacije i informatike, Varaždin, email: damir.horvat1@foi.hr
[†]Fakultet organizacije i informatike, Varaždin, email: dusan.mundjar@foi.hr

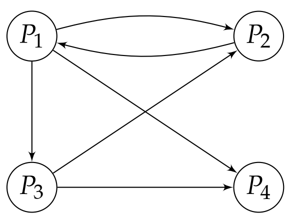
1 Uvod

U današnje vrijeme Internet je dio naše svakodnevice. U svakom trenutku možemo vrlo brzo doći do željene informacije tako da u tražilici upišemo odgovarajući pojam. Na temelju unešenog pojma tražilica daje popis relevantnih web stranica. Prve tražilice ranih devedesetih su upravo funkcionirale na taj način: prolaskom kroz indeksirani popis web stranica brojale su koliko se puta unešeni pojam pojavljuje na svakoj od stranica. Stranice koje su sadržavale najveći broj pojavljivanja unešenog pojma ponuđene su korisniku. Međutim, s ovakvim pristupom je bilo dosta problema jer web stranica može sadržavati veliki broj pojavljivanja nekog unešenog pojma, a da pritom ne sadrži nikakve korisne informacije o tom pojmu.

Današnje moderne tražilice koriste drugačiji pristup. Osim unešenog pojma prilikom pretraživanja uzimaju u obzir relevantnost, važnost ili popularnost pojedine stranice. Relevantnije stranice tražilica stavlja na vrh popisa, a manje relevantne na dno popisa i time omogućuje korisniku da što brže dođe do pravih i pouzdanih informacija. Jedan od takvih algoritama pretraživanja je PageRank algoritam kojeg koristi Google. Isti algoritam može se koristiti i u druge svrhe, primjerice kod rangiranja ekipa u nogometu [2] ili rukometu [8]. U nastavku ćemo ukratko opisati matematičku pozadinu tog algoritma.

2 PageRank algoritam

PageRank algoritam bazira se na pretpostavci da je relevantna web stranica svaka ona stranica na koju se referira veliki broj drugih više ili manje relevantnih web stranica. Ovo je sasvim prirodna pretpostavka jer niti jedna ozbiljna web stranica neće postaviti link na neku neprovjerenu i nekvalitetnu web stranicu. Stoga PageRank algoritam relevantnost web stranice određuje na temelju relevantnosti web stranica koje imaju linkove na promatranu stranicu.



Slika 1: digraf D predstavlja poveznice web stranica

Kako bismo lakše objasnili ideju algoritma, pogledajmo pojednostavljeni model Interneta s četiri web stranice prikazanog na slici 1. Model je prikazan pomoću usmjerenog grafa (digrafa) D u kojemu vrhovi predstavljaju web stranice P_1, P_2, P_3 i P_4, a od vrha P_i postoji usmjereni brid prema vrhu P_j jedino ako stranica P_i ima poveznicu na stranicu P_j. Na primjer, u našem slučaju stranica P_3 ima poveznice na stranice P_2 i P_4. *Izlazni stupanj vrha v* u digrafu je ukupni broj bridova koji izlaze iz tog vrha, a označavamo ga s $d^+(v)$. U našem slučaju izlazni stupnjevi vrhova su $d^+(P_1) = 3$, $d^+(P_2) = 1$, $d^+(P_3) = 2$, $d^+(P_4) = 0$. Nakon što smo konstruirali usmjereni graf koji reprezentira strukturu međusobno povezanih web stranica na Internetu, konstruiramo kvadratnu matricu $\mathbf{H} = [h_{i,j}]$ reda n pri čemu je $h_{i,j}$ jednak vjerojatnosti da korisnik klikom na neku od poveznica sa stranice P_j ode na stranicu P_i, a n je ukupni broj web stranica. Ukoliko stranica P_j ne sadrži poveznicu na stranicu P_i, tada je $h_{i,j} = 0$. U protivnom je $h_{i,j} = \frac{1}{d^+(P_j)}$. Za matricu $\mathbf{H}$ kažemo da *pripada* digrafu D. U našem slučaju matrica $\mathbf{H}$ je

$$\mathbf{H} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} & 0 \end{bmatrix}.$$

Na primjer, $h_{4,1} = \frac{1}{3}$ jer postoji poveznica sa stranice P_1 na stranicu P_4 i pritom je $d^+(P_1) = 3$. Isto tako, $h_{1,3} = 0$ jer ne postoji poveznica sa stranice P_3 na stranicu P_1.

Rang web stranice P je realni broj

$$r(P) = \sum_{Q \in B_P} \frac{r(Q)}{d^+(Q)}$$

pri čemu je B_P skup svih web stranica koje imaju poveznicu na web stranicu P. Ako je $B_P = \emptyset$, tada je $r(P) = 0$. Rang web stranice je njezina mjera relevantnosti u PageRank algoritmu. Rang stranice P ovisi o rangovima svih stranica Q koje imaju poveznice na stranicu P i o ukupnom broju poveznica na pojedinoj stranici Q. Što je veći broj poveznica na stranici Q, to njezin rang daje manji doprinos rangu stranice P. Nadalje, možemo uočiti da je definicija ranga rekurzivna. Ukoliko web stranice P_i i P_j imaju poveznice jedna na drugu, tada rang stranice P_i ovisi o rangu stranice P_j i obrnuto, rang stranice P_j ovisi o rangu stranice P_i. Stoga se definira iterativni postupak za računanje ranga web stranice P kao

$$r_{k+1}(P) = \sum_{Q \in B_P} \frac{r_k(Q)}{d^+(Q)} \tag{1}$$

pri čemu je $r_k(P)$ rang web stranice P nakon k iteracija. Pritom je početni rang web stranice P jednak $r_0(P) = \frac{1}{n}$ gdje je n ukupni broj web stranica. Drugim riječima, u inicijalnom koraku svim stranicama dajemo istu relevantnost.

Ilustrirajmo opisani postupak računanja ranga na našem primjeru prikazanom na slici 1. Na početku je

$$r_0(P_1) = \frac{1}{4}, \quad r_0(P_2) = \frac{1}{4}, \quad r_0(P_3) = \frac{1}{4}, \quad r_0(P_4) = \frac{1}{4}.$$

U prvoj iteraciji dobivamo

$$\begin{aligned}
r_1(P_1) &= \frac{r_0(P_2)}{d^+(P_2)} = \frac{\frac{1}{4}}{1} = \frac{1}{4}\\
r_1(P_2) &= \frac{r_0(P_1)}{d^+(P_1)} + \frac{r_0(P_3)}{d^+(P_3)} = \frac{\frac{1}{4}}{3} + \frac{\frac{1}{4}}{2} = \frac{1}{12} + \frac{1}{8} = \frac{5}{24}\\
r_1(P_3) &= \frac{r_0(P_1)}{d^+(P_1)} = \frac{\frac{1}{4}}{3} = \frac{1}{12}\\
r_1(P_4) &= \frac{r_0(P_1)}{d^+(P_1)} + \frac{r_0(P_3)}{d^+(P_3)} = \frac{\frac{1}{4}}{3} + \frac{\frac{1}{4}}{2} = \frac{1}{12} + \frac{1}{8} = \frac{5}{24}.
\end{aligned}$$

U drugoj iteraciji dobivamo

$$\begin{aligned}
r_2(P_1) &= \frac{r_1(P_2)}{d^+(P_2)} = \frac{\frac{5}{24}}{1} = \frac{5}{24}\\
r_2(P_2) &= \frac{r_1(P_1)}{d^+(P_1)} + \frac{r_1(P_3)}{d^+(P_3)} = \frac{\frac{1}{4}}{3} + \frac{\frac{1}{12}}{2} = \frac{1}{12} + \frac{1}{24} = \frac{1}{8}\\
r_2(P_3) &= \frac{r_1(P_1)}{d^+(P_1)} = \frac{\frac{1}{4}}{3} = \frac{1}{12}\\
r_2(P_4) &= \frac{r_1(P_1)}{d^+(P_1)} + \frac{r_1(P_3)}{d^+(P_3)} = \frac{\frac{1}{4}}{3} + \frac{\frac{1}{12}}{2} = \frac{1}{12} + \frac{1}{24} = \frac{1}{8}.
\end{aligned}$$

Postupak ponavljamo tako dugo dok komponente vektora

$$\big(r_k(P_1), r_k(P_2), r_k(P_3), r_k(P_4)\big)$$

ne „počnu konvergirati“, tj. dok se svaka komponenta ne približi dovoljno blizu nekoj specifičnoj vrijednosti. Međutim, konvergencija se nužno ne mora dogoditi pa treba napraviti daljnje zahvate na matrici **H** kako bismo osigurali konvergenciju.

Ako stavimo

$$Z_k = \begin{bmatrix} r_k(P_1) \\ r_k(P_2) \\ \vdots \\ r_k(P_n) \end{bmatrix},$$

tada iterativni postupak (1) možemo zapisati u matričnom obliku

$$Z_{k+1} = \mathbf{H} Z_k, \quad k = 0, 1, 2, \ldots \tag{2}$$

Na promatranom primjeru za prve dvije iteracije imamo

$$Z_1 = \mathbf{H} Z_0 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ \frac{5}{24} \\ \frac{1}{12} \\ \frac{5}{24} \end{bmatrix},$$

$$Z_2 = \mathbf{H} Z_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{4} \\ \frac{5}{24} \\ \frac{1}{12} \\ \frac{5}{24} \end{bmatrix} = \begin{bmatrix} \frac{5}{24} \\ \frac{1}{8} \\ \frac{1}{12} \\ \frac{1}{8} \end{bmatrix}.$$

Relacija (2) može se zapisati preko početnog vektora Z_0 i potencija matrice **H** u obliku

$$Z_k = \mathbf{H}^k Z_0, \quad k = 1, 2, 3, \ldots \tag{3}$$

Ako niz (Z_k) konvergira i ako je

$$\lim_{k \to \infty} \mathbf{H}^k Z_0 = \tilde{Z},$$

tada su komponente vektora $\tilde{Z}$ upravo rangovi web stranica $P_1, P_2, \ldots, P_n$. Drugim riječima, nakon određenog broja koraka komponente vektora Z_k približit će se dovoljno blizu komponentama vektora $\tilde{Z}$ pa će vektor Z_k predstavljati dobru aproksimaciju vektora $\tilde{Z}$. No, kao što smo već ranije rekli, matrica **H** ne osigurava konvergenciju.

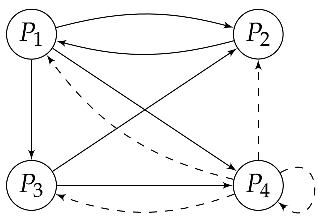
Iterativni postupak (3) poznat je pod nazivom *metoda potencija*. Kako bi metoda potencija konvergirala, matrica **H** mora imati određena svojstva.

Stoga ćemo u dva koraka matricu **H** modificirati u matricu **G** koja će osigurati konvergenciju.

U prvom koraku matricu **H** modificiramo u matricu **S** koja je stohastička po stupcima. Za kvadratnu matricu kažemo da je *stohastička po stupcima* ako su svi njezini elementi nenegativni i suma elemenata u svakom stupcu je jednaka 1. Ako pogledamo našu matricu **H**, vidimo da je problematičan jedino zadnji stupac koji sadrži samo nule pa suma elemenata u tom stupcu nije jednaka 1. To se dogodilo zbog toga što web stranica P_4 ne sadrži poveznice niti na jednu drugu stranicu pa korisnik s te stranice ne može preko poveznice otići na neku drugu stranicu. Međutim, korisnik u web pregledniku može takvu stranicu napustiti na druge načine, npr. pisanjem web adrese. Stoga sve nulstupce u matrici **H** zamijenimo vektorom **v** koji ima nenegativne komponente čija suma je jednaka 1. Komponente tog vektora predstavljaju vjerojatnosti da korisnik s promatrane stranice bez poveznica ode na neku drugu stranicu ili ostane na toj istoj stranici. Klasični izbor je vektor $\mathbf{v} = \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$. Vektor **v** je u engleskoj literaturi poznat pod nazivom *personalization vector*. Za našu matricu **H** stavimo da je $\mathbf{v} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ pa dobivamo matricu

$$\mathbf{S} = \begin{bmatrix} 0 & 1 & 0 & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{3} & 0 & 0 & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix}.$$

Modifikacijom matrice **H** u matricu **S**, usmjereni graf sa slike 1 modificira se u usmjereni graf prikazan na slici 2.



Slika 2: digraf kojem pripada matrica **S**

U drugom koraku moramo još osigurati ireducibilnost matrice **S**. Kvadratna matrica koja pripada nekom usmjerenom grafu je ireducibilna ako i samo ako je pripadni usmjereni graf jako povezan. Za usmjereni graf kažemo da je *jako povezan* ako između svaka dva njegova vrha postoji usmjereni put. Usmjereni graf na slici 1 nije jako povezan jer na primjer ne postoji usmjereni put od vrha P_4 do vrha P_2. Drugim riječima, šetajući se po usmjerenim bridovima i poštivajući njihovu orijentaciju, ne možemo iz vrha P_4 doći do vrha P_2. Stoga naša matrica $\mathbf{H}$ nije ireducibilna. Usmjereni graf na slici 2 jest jako povezan pa je naša matrica $\mathbf{S}$ ireducibilna. Međutim, matrica $\mathbf{S}$ ne mora općenito biti ireducibilna. Kako bismo osigurali ireducibilnost, na svakoj web stranici dodamo poveznice prema svim ostalim web stranicama. Drugim riječima, na usmjerenom grafu koji je određen matricom $\mathbf{S}$ iz svakog vrha dodamo usmjerene bridove prema svim ostalim vrhovima bez obzira postoje li već takvi bridovi ili ne. Novodobiveni usmjereni graf reprezentiramo matricom $\mathbf{S}+\mathbf{E}$ pri čemu su stupci matrice $\mathbf{E}$ jednaki vektoru $\mathbf{v}$. U našem primjeru je

$$\mathbf{S}+\mathbf{E} = \begin{bmatrix} 0 & 1 & 0 & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{3} & 0 & 0 & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} + \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{5}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{7}{12} & \frac{1}{4} & \frac{3}{4} & \frac{1}{2} \\ \frac{7}{12} & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{7}{12} & \frac{1}{4} & \frac{3}{4} & \frac{1}{2} \end{bmatrix}.$$

Matrica $\mathbf{S}+\mathbf{E}$ je ireducibilna jer je pripadni usmjereni graf jako povezan, ali nije stohastička po stupcima. Stoga odaberemo neki $\alpha \in \langle 0,1 \rangle$ i definiramo matricu $\mathbf{G} = \alpha\mathbf{S} + (1-\alpha)\mathbf{E}$ koja je stohastička po stupcima i ireducibilna. Sve ovako konstruirane matrice $\mathbf{G}$ osiguravaju konvergenciju iterativnog postupka

$$Z_k = \mathbf{G}^k Z_0, \quad k = 1,2,3,\ldots \tag{4}$$

Zainteresirani čitatelj više informacija o navedenoj tvrdnji može pronaći u [3] i [5]. Parametar α zapravo govori u kojem postotku matrica $\mathbf{S}$, u kojoj je spremljena struktura Interneta, utječe na rangiranje web stranica.
Klasični izbor vrijednosti parametra α je 0.85. U našem primjeru je

$$\mathbf{G} = 0.85 \cdot \mathbf{S} + 0.15 \cdot \mathbf{E} = \begin{bmatrix} \frac{3}{80} & \frac{71}{80} & \frac{3}{80} & \frac{1}{4} \\ \frac{77}{240} & \frac{3}{80} & \frac{37}{80} & \frac{1}{4} \\ \frac{77}{240} & \frac{3}{80} & \frac{3}{80} & \frac{1}{4} \\ \frac{77}{240} & \frac{3}{80} & \frac{37}{80} & \frac{1}{4} \end{bmatrix}.$$

Izbor vrijednosti vektora $\mathbf{v}$ i parametra α mogu utjecati na konačne rangove web stranica, ali ovdje nećemo ulaziti u detaljnije analize.

Primijenimo li iterativni postupak (4) na našem primjeru, u prvih devet iteracija dobivamo vektore Z_k prikazane u tablici 1.

k	Z_k	k	Z_k
0	$\begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}$	5	$\begin{bmatrix} 0.3085513078901503 \\ 0.2556888666930023 \\ 0.1800709587238453 \\ 0.2556888666930023 \end{bmatrix}$
1	$\begin{bmatrix} 0.303125 \\ 0.2677083333333333 \\ 0.1614583333333333 \\ 0.2677083333333333 \end{bmatrix}$	6	$\begin{bmatrix} 0.3091694208613148 \\ 0.2557869121987731 \\ 0.1792567547411389 \\ 0.2557869121987731 \end{bmatrix}$
2	$\begin{bmatrix} 0.3219401041666666 \\ 0.2488932291666666 \\ 0.1802734374999999 \\ 0.2488932291666666 \end{bmatrix}$	7	$\begin{bmatrix} 0.3092735942111965 \\ 0.2556368421845959 \\ 0.1794527214196118 \\ 0.2556368421845959 \end{bmatrix}$
3	$\begin{bmatrix} 0.3019490559895833 \\ 0.2582223849826389 \\ 0.1816061740451389 \\ 0.2582223849826389 \end{bmatrix}$	8	$\begin{bmatrix} 0.3091141448211331 \\ 0.2557177539274007 \\ 0.1794503473240656 \\ 0.2557177539274007 \end{bmatrix}$
4	$\begin{bmatrix} 0.3118612840440538 \\ 0.2551071133083767 \\ 0.1779244893391927 \\ 0.2551071133083767 \end{bmatrix}$	9	$\begin{bmatrix} 0.3092001135478632 \\ 0.2556887613549549 \\ 0.179422363742227 \\ 0.2556887613549549 \end{bmatrix}$

Tablica 1:

Vidimo da se u posljednje četiri promatrane iteracije prve tri decimale ne mijenjaju. Ako se ovdje zadovoljimo aproksimacijom na tri decimale, za traženi vektor $\tilde{Z}$ možemo uzeti

$$\tilde{Z} = (0.309, 0.256, 0.179, 0.256). \tag{5}$$

Uočimo da je suma njegovih komponenata jednaka 1. Naravno, zbog aproksimacije u općenitom slučaju ta suma će biti približno jednaka 1. Promatrane web stranice P_1, P_2, P_3 i P_4 imaju rangove prikazane u tablici 2. Rangovi tih stranica jednaki su komponentama vektora $\tilde{Z}$. Dakle, P_1 je najbolja web stranica, a P_3 je najlošija web stranica.

web stranica	rang
P_1	0.309
P_2	0.256
P_3	0.179
P_4	0.256

Tablica 2:

2.1 Rangiranje preko sustava linearnih jednadžbi

Osim iterativnim postupkom (4), do vektora $\tilde{Z}$ može se doći rješavanjem sustava linearnih jednadžbi. Ne ulazeći preduboko u problem svojstvenih vrijednosti, spomenimo samo da za vektor $\tilde{Z}$ i matricu **G** vrijedi

$$\mathbf{G}\tilde{Z} = \tilde{Z}. \tag{6}$$

Drugim riječima, umnožak matrice **G** i vektora $\tilde{Z}$ jednak je vektoru $\tilde{Z}$. U tom slučaju se govori da je 1 svojstvena vrijednost matrice **G** i da je $\tilde{Z}$ jedan svojstveni vektor za tu svojstvenu vrijednost. Zainteresirani čitatelj više informacija o problemu svojstvenih vrijednosti može pronaći u [1], [4] i [6], a jedan dokaz relacije (6) može pronaći u [9]. Množenjem matrica i sređivanjem (6) se svodi na homogeni sustav linearnih jednadžbi. Pokažimo kako to izgleda s našom matricom **G**. Ako stavimo $\tilde{Z} = (z_1, z_2, z_3, z_4)$ iz (6) slijedi

$$\begin{bmatrix} \frac{3}{80} & \frac{71}{80} & \frac{3}{80} & \frac{1}{4} \\ \frac{77}{240} & \frac{3}{80} & \frac{37}{80} & \frac{1}{4} \\ \frac{77}{240} & \frac{3}{80} & \frac{3}{80} & \frac{1}{4} \\ \frac{77}{240} & \frac{3}{80} & \frac{37}{80} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix},$$

odnosno nakon množenja matrica

$$\begin{bmatrix} \frac{3}{80}z_1 + \frac{71}{80}z_2 + \frac{3}{80}z_3 + \frac{1}{4}z_4 \\ \frac{77}{240}z_1 + \frac{3}{80}z_2 + \frac{37}{80}z_3 + \frac{1}{4}z_4 \\ \frac{77}{240}z_1 + \frac{3}{80}z_2 + \frac{3}{80}z_3 + \frac{1}{4}z_4 \\ \frac{77}{240}z_1 + \frac{3}{80}z_2 + \frac{37}{80}z_3 + \frac{1}{4}z_4 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix}. \tag{7}$$

Iz (7) dobivamo homogeni sustav linearnih jednadžbi

$$
\begin{aligned}
-\frac{77}{80}z_1 + \frac{71}{80}z_2 + \frac{3}{80}z_3 + \frac{1}{4}z_4 &= 0\\
\frac{77}{240}z_1 - \frac{77}{80}z_2 + \frac{37}{80}z_3 + \frac{1}{4}z_4 &= 0\\
\frac{77}{240}z_1 + \frac{3}{80}z_2 - \frac{77}{80}z_3 + \frac{1}{4}z_4 &= 0\\
\frac{77}{240}z_1 + \frac{3}{80}z_2 + \frac{37}{80}z_3 - \frac{3}{4}z_4 &= 0
\end{aligned}
$$

kojeg riješimo Gaussovim postupkom. Ako varijablu z_4 uzmemo za parametar, dobivamo opće rješenje

$$z_1 = \frac{1769}{1463}p, \quad z_2 = p, \quad z_3 = \frac{40}{57}p, \quad z_4 = p$$

gdje je $p \in \mathbb{R}$ proizvoljni realni broj. Dakle, nismo dobili jedinstveno rješenje za traženi vektor $\tilde{Z}$. Međutim, dobili smo jedinstveno rješenje do na skalar $p \in \mathbb{R}$, tj.

$$\tilde{Z} = p \cdot \left(\frac{1769}{1463}, 1, \frac{40}{57}, 1\right).$$

Biramo $p \in \mathbb{R}$ tako da suma komponenata vektora $\tilde{Z}$ bude jednaka 1. Kako je

$$\frac{1769}{1463} + 1 + \frac{40}{57} + 1 = \frac{17165}{4389},$$

uzmemo $p = \frac{4389}{17165}$. Stoga je traženi vektor $\tilde{Z}$ jednak

$$\tilde{Z} = \frac{4389}{17165} \cdot \left(\frac{1769}{1463}, 1, \frac{40}{57}, 1\right) = \left(\frac{5307}{17165}, \frac{4389}{17165}, \frac{616}{3433}, \frac{4389}{17165}\right). \tag{8}$$

Usporedimo li (5) i (8), vidimo da je (5) dobra aproksimacija od (8).

U realnoj situaciji Internet se sastoji od ogromnog broja web stranica koji se mjeri u milijardama. Metoda potencija u svega tridesetak iteracija dolazi do vektora $\tilde{Z}$ s točnošću na dvije decimale u slučaju kada je $\alpha = 0.85$. Općenito, brzina konvergencije metode potencija ovisi o samoj matrici, ali ovdje nećemo dublje ulaziti u detalje. Spomenimo samo da konvergencija i brzina konvergencije metode potencija ovisi o svojstvenim vrijednostima pripadne kvadratne matrice. U praksi je teško pronaći sve svojstvene vrijednosti zadane kvadratne marice. Međutim, Perron-Frobeniusov teorem daje sigurnu konvergenciju metode potencija za široku klasu kvadratnih matrica koje su važne u primjenama, a da pritom ne moramo unaprijed ništa znati o svojstvenim vrijednostima takvih matrica. Jedna takva klasa matrica su upravo nenegativne ireducibilne matrice koje na glavnoj dijagonali imaju barem jedan element različit od nule. Matrica **G** je jedan specijalni primjer takvih matrica koja je pritom još i stohastička. Zainteresirani čitatelj više informacija o predivnoj Perron–Frobeniusovoj teoriji može pronaći u [6].

Literatura

[1] N. Elezović, *Linearna algebra*, Element, Zagreb, 2006.

[2] A. Y. Govan, C. D. Meyer, *Ranking national football league teams using google's pagerank*, `https://www.ncsu.edu/crsc/reports/ftp/pdf/crsc-tr06-19.pdf`, datum pristupa: 10.01.2016.

[3] A. Howard, C. Rorres, *Elementary Linear Algebra: Applications Version 8th Edition*, John Wiley & Sons, New Jersey, 2000.

[4] K. Horvatić, *Linearna algebra (II. dio)*, Matematički odjel PMF-a, Sveučilište u Zagrebu, Zagreb, 1995.

[5] J. G. Kemeny, J. L. Snell, *Finite Markov Chains*, Springer-Verlag, New York, 1976.

[6] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, `http://saba.kntu.ac.ir/eecd/sedghizadeh/Ebooks/Matrix_Analysis.pdf`, datum pristupa: 10.01.2016.

[7] A. N. Langville, C. D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, `http://geza.kzoo.edu/~erdi/patent/langvillebook.pdf`, datum pristupa: 10.01.2016.

[8] D. Munđar, D. Horvat, *Rangiranje ekipa i prognoziranje ishoda u rukometu korištenjem PageRank algoritma*, Poučak: časopis za metodiku i nastavu matematike, 67 (2016), 35 – 42.

[9] Nepoznati autor, `https://info.maths.ed.ac.uk/assets/files/Projects/6.pdf`, datum pristupa: 10.01.2016.

[10] Nepoznati autor, `http://people.inf.ethz.ch/arbenz/ewp/Lnotes/chapter7.pdf`, datum pristupa: 10.01.2016.

[11] Y. Saad, *Numerical methods for large eigenvalue problems*, http://www-users.cs.umn.edu/~saad/eig_book_2ndEd.pdf, datum pristupa: 10.01.2016.

[12] R. Tanase, R. Radu, *The Mathematics of Web Search*, http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/index.html, datum pristupa: 10.01.2016.

[13] R. S. Wills, *Google's pagerank: the math behind the search engine*, http://www.cems.uvm.edu/~tlakoba/AppliedUGMath/other_Google/Wills.pdf, datum pristupa: 10.01.2016.