



TEHNOLOGIJA, INFORMATIKA I OBRAZOVANJE
ZA DRUŠTVO UČENJA I ZNANJA

6. Međunarodni Simpozijum, Tehnički fakultet Čačak, 3–5. jun 2011.

TECHNOLOGY, INFORMATICS AND EDUCATION
FOR LEARNING AND KNOWLEDGE SOCIETY

6th International Symposium, Technical Faculty Čačak, 3–5th June 2011.

UDK: 004.738.1

Stručni rad

RANGIRANJE PREČICA ADAPTIVNIH VEB SAJTOVA U FUNKCIJI EFIKASNOG PRISTUPA PODACIMA

Željko Eremić¹

Rezime: Adaptivni veb sajtovi mogu menjati svoju strukturu na osnovu informacija o dosadašnjim ponašanjima posetilaca. Shortcutting je pristup koji stvara prečice između stranica, koje nisu do tada postojale, a sve u cilju poboljšanja efikasnosti postojeće navigacije. Ovaj rad uključuje primenu vejpost (waypost) status stranica, dodeljene pojedinim stranicama, koje nisu ciljne, ali postoji pretpostavka da mogu sadržati važne informacije za navigaciju korisnika. Cilj ovog rada je da teoretski predstavi metodologiju čijim korišćenjem je moguće rangirati prečice ka vejpost ili krajnjim stranicama na osnovu stepena doprinosa efikasnosti navigacije adaptivnih veb sajtova. Rad uključuje deskriptivni metod, metod „poslednjeg dokumenta“ i Markovljevi model. Kao rezultat se očekuje teoretski objašnjen postupak za rangiranje prečica ka vejpost i ciljnim stranicama adaptivnog veb sajta.

Ključne reči: Vejpost, prečica, adaptivnost, navigacija, sajt

RANKING OF SHORTCUTS OF ADAPTIVE WEB SITES IN THE FUNCTION OF EFFICIENT INFORMATION ACCESS

Summary: Adaptive web sites may change their structure based on the information about the visitors' behavior. Shortcutting is an approach that creates a shortcut between the pages that were not previously existed, with the aim of improving the efficiency of the existing navigation. This paperwork includes the application of the waypost status pages, assigned only to the individual pages, and not the target ones, with the assumption that they may contain important information for users' navigation. The aim of this paperwork is to theoretically present methodology which can be used for ranking the shortcuts to waypost pages or target pages based on the level of contribution to the efficiency of adaptive web sites' navigation. Paperwork uses descriptive method, the end document method and Markov chains model. Theoretically explained procedure for ranking shortcuts which lead to certain waypost and target pages of adaptive web site is expected as a result.

Key words: Waypost, shortcut, adaptation, navigation, site

¹ Mr Željko Eremić, dipl. ing. inf., Visoka tehnička škola strukovnih studija u Zrenjaninu, Zrenjanin,
E-mail: zeljko.eric@gmail.com

1. UVOD

Inspiracija za ovaj rad potiče iz ranijih radova, pre svega [1], u kome je promovisana ideja da korisnici u digitalnom svetu mogu ostavljati digitalne tragove, poput tragova stopa u stvarnom svetu, i u [2] gde se prvi put javila ideja o vejpost (waypost) dokumentima. Već izvesno vreme postoji interesovanje vezano za pogodnosti koje pružaju adaptivni veb sajtovi. U osnovi ideje se nalazi mogućnost da se evidentiraju ponašanja korisnika prilikom njihovih poseta veb sajtu, i zatim da se takvi podaci obrade, a rezultati obrade iskoriste za poboljšanje navigacije na veb sajtu. Pri tome je poželjno da postoji što veći stepen automatskog korišćenja ovih procesa. Veoma često se unapređenja navigacije adaptivnih veb sajtova postižu na taj način što se dodaju prečice između stranica koje do tog trenutka nisu bile povezane. Ovaj pristup se naziva shortcutting.

Ovaj rad dokumentom smatra bilo veb stranicu u okviru sajta, bilo neki dokument u formatu poput pdf ili doc, koji se može preuzeti sa neke stranice veb sajta. U ovom radu se pravi razlika između stranica do kojih korisnici imaju potrebu da dođu (ciljni dokumenti), kao i onih stranica koje su samo usputne stanice do ovih ciljnih dokumenata. Standardno se prečice stvaraju ka ciljnim dokumentima, sa idejom da se smanji broj potrebnih aktivnosti (klikova) za dolazak do željenog dokumenta. Ovaj rad uzima u obzir i mogućnost da neki od međudokumenata sadrže korisne informacije za odluku o pristupu ciljnim dokumentima. U radu će biti objašnjen teoretski pristup za rangiranje značaja novostvorenih veza ka međudokumentima i ciljnim dokumentima kod adaptivnih veb sajtova.

2. ADAPTIVNI VEB SAJTOVI

Adaptivni veb sajtovi na prvi pogled izgledaju veoma slično običnim veb sajtovima. Razlika se ogleda u tome što ovakvi sajtovi imaju mogućnost da izmene svoj prikaz na osnovu evidentiranih prethodnih ponašanja korisnika. Adaptivni veb sajtovi su prema [3] sajtovi koji automatski unapređuju svoju organizaciju i prikaz kroz učenje iz šablona korisničkih poseta.

Adaptivni veb sajtovi imaju dva osnovna modula. Prvi (posmatrački) modul se bavi posmatranjem evidentiranih korisničkih poseta, iz kojih se mogu izvući određeni zaključci. Drugi (transformacioni) modul omogućava sprovođenje zaključaka dobijenih posle izvršenja prvog modula. U ovom radu će biti kratko objašnjena oba modula.

“Posmatrački modul posmatra korisnikovu interakciju sa sajtom i prikuplja važne statističke podatke o posećenim stranama, pređenim linkovima, praćenim putanjama, i problemima koji su se javili. Transformacioni modul na osnovu ovih podataka pravi promene na strukturi sajta.” [4]

3. LOG FAJLOVI, OBRADA PODATAKA, POČETNI, KRAJNJI I VEJPOST DOKUMENTI

Jedan od izvora informacija o ponašanju korisnika se može naći u log fajlovima. Log fajlovi su fajlovi koji se nalaze na veb serveru, i u njima su evidentirani korisnički zahtevi serveru. Ilustracija log fajlova, kao i najčešće korišćenog formata za log fajlove je prikazana u [5]. Teoretsko objašnjenje obrade zapisa iz log fajlova je razmatrano u [6]. Korisnička poseta veb sajtu podrazumeva obično pristup početnoj stranici, jednoj ili više međustranica, a završava se posetom krajnjem dokumentu (bilo stranici bilo nekom fajlu).

Ciljni dokumenti se biraju iz spiska krajnjih dokumenata, i oni predstavljaju dokumente za koje se osnovano pretpostavlja da su dokumenti koji zadovoljavaju potrebe korisnika za traženim informacijama, razmatrani su u [6]. Vejpst dokumenti se mogu dobiti iz spiska poseta korisničkih putanja, i to onih često posećenih dokumenata koji nisu ni početni, ni krajnji, ni ciljni dokumenti. Ideja o vejpost dokumentima je izložena u [6].

4. RANGIRANJE VEZA

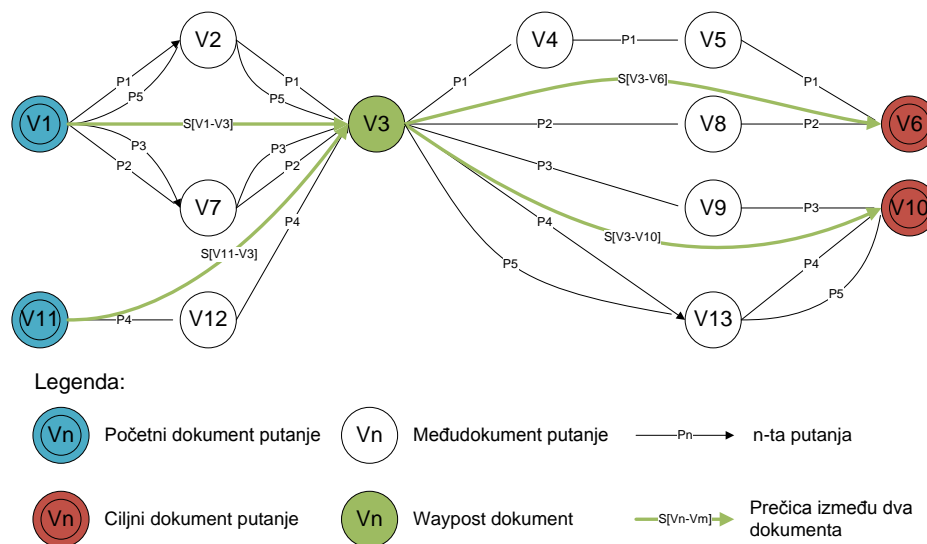
“Cilj shortcuttinga je smanjenje broj klikova koje posetilac mora učiniti u cilju pristupa njegovoj ciljnoj stranici“ [7]. Ovaj proces kao rezultat ima stvaranje prečica između parova dokumenta, koji do tada nisu bili povezani.

Kada su identifikovane korisničke putanje moguće je iz njihovog spiska izdvojiti početne, krajnje, ciljne i vejpost dokumente. Posle toga moguće je predlagati prečice između parova dokumenata. Neka se poseta n-tom dokumentu označi sa V_n , a sam dokument sa D_n , lista svih putanja korisnika sa LP , n-ta korisnička putanja sa P_n , lista vejpost dokumenata sa WL , a lista prečica sa SL . Neka je data lista korisničkih putanja $LP = (P_1, P_2, P_3, P_4, P_5)$, pri čemu svaka putanja predstavljaju niz poseta dokumentima (*sl. 1*):

- $P_1 = (V_1, V_2, V_3, V_4, V_5, V_6)$
- $P_2 = (V_1, V_7, V_3, V_8, V_6)$
- $P_3 = (V_1, V_7, V_3, V_9, V_{10})$
- $P_4 = (V_{11}, V_{12}, V_3, V_{13}, V_{10})$
- $P_5 = (V_1, V_2, V_3, V_{13}, V_{10})$

Kod ovih pet putanja postoji poseta međudokumentu D_3 , koja se pojavljuje u svakoj od njih (100% putanja iz klastera), i to je poseta V_3 . Njen dokument možemo proglasiti za vejpost dokument. Na osnovu dosadašnjih proračuna, moguće je kreirati četiri nove prečice: $S[V_1-V_3]$ koja ima primenu na četiri putanje, $S[V_3-V_6]$ i $S[V_3-V_{10}]$ koje imaju primenu u po dve ili više putanja iz klastera i jos jednu prečicu $S[V_{11}-V_3]$ koja zahteva posebno objašnjenje, i koja se odnose samo na putanju P_4 . Naime putanja P_4 je jedina putanja čiji je početni dokument V_{11} . Ona koristi status vejpost dokumenta V_3 , koji je on dobio na osnovu ostale četiri putanje. Putanja P_4 zapravo samo koristi prednosti prethodnih proračuna, odnosno pošto je V_3 proglašen za vejpost dokument, formira se prečica $S[V_{11}-V_3]$ za potrebe putanje P_4 . Kao rezultat dobijena je lista vejpost dokumenata (WL), kao i lista predloženih prečica (SL):

- $WL = (V_3)$
- $SL = (S[V_1-V_3], S[V_3-V_6], S[V_3-V_{10}], S[V_{11}-V_3])$



Slika 1: Primer određivanja vejpost dokumenta i prečica

Ukoliko se prihvati spisak predloženih prečica moguće je dodavati prečice na dokumentima, kako bi se uspostavila veza ka određivim dokumentima. Ukoliko postoji više od jedne prečice koju je potrebno dodati na dokument (stranicu), onda je korisno izvršiti rangiranje spiska prečica za dodavanje, jer ona veza koja je istaknutija (nalazi se na prvom mestu), privlači više pažnje, a ona prečica koja privlači više pažnje treba da vodi ka relevantnijem dokumentu. Teoretsko objašnjenje koje koristi lance Markova za potrebe predviđanja budućih ponašanja posetilaca, i koje predstavlja osnovu u onom radu je navedeno u [8]. U ovom radu spisak određivih dokumenata za prečice, pored ciljnih dokumenata se proširuje i na spisak vejpost dokumenata. Ovaj prošireni pristup će biti ilustrovan kroz primer (*sl. 1*), a procena veza među dokumentima je data u tabeli 1.

Tabela 1: Predviđanje navigacije korisnika zasnovano na lancima Markova

Početni dokument	Određivni dokument	Frekvencij a veze (I)	Ukupna frekv. početnog dokum. (II)	Odnos I/II u procentima	Zadovoljava
D1	D6	2	4	50	Y
D1	D10	2	4	50	Y
D11	D10	1	1	100	Y
D1	D3	4	4	100	Y
D11	D3	1	1	100	Y
D3	D6	2	5	40	Y
D3	D10	3	5	60	Y

U realnoj situaciji pojedine prečice bi mogle biti eliminisane ukoliko imaju učestalost ili procenat predviđanje ispod određenih vrednosti. U ovom primeru zbog jednostavnosti neće biti eliminacije prečica. Transformacioni modul ima zadatak da potrebne prečice ubaci na odgovarajuće stranice, kao što je ilustrovano (*sl. 2*):

Dokument (veb stranica) D1	Dokument(veb stranica) D11	Dokument (veb stranica) D3
<u>Veza ka dokumentu D3</u>	<u>Veza ka dokumentu D10</u>	<u>Veza ka dokumentu D10</u>
<u>Veza ka dokumentu D6</u>	<u>Veza ka dokumentu D3</u>	<u>Veza ka dokumentu D6</u>
<u>Veza ka dokumentu D10</u>
...	Sadržaj dokumenta D11	Sadržaj dokumenta D3
Sadržaj dokumenta D1
...		

Slika 2: Dodavanje novoformiranih veza na odgovarajuće dokumente

Na dokumentu D1 se mogu dodati 3 prečice: ka dokumentu D3 jer u sve četiri putanje koje počinju posetom dokumentu D1, one uključuju i posetu dokumentu D3 (4/4, odnosno 100%), ka dokumentu D6, jer 2 od 4 putanje koje počinju posetom dokumentu D1, one završavaju u dokumentu D6 (2/4, odnosno 50%), i ka dokumentu D10, jer 2 od 4 putanje koje počinju posetom dokumentu D1, one završavaju u dokumentu D10 (2/4, odnosno 50%).

Na dokumentu D11 se mogu dodati 2 prečice: ka dokumentu D10, jer jedina putanja koja počinje posetom dokumentu D11, završava u dokumentu D10 (1/1, odnosno 100%), i ka dokumentu D3 jer jedina putanja koja počinje u dokumentu D11 prolazi kroz vejpost dokument D3 (1/1, odnosno 100%).

Na vejpost dokumentu D3 se mogu dodati 2 prečice: ka dokumentu D10, jer 3 od 5 putanja koje prolaze kroz vejpost dokument D3 završavaju u dokumentu D10 (3/5, odnosno 60%), i ka dokumentu D6, jer 2 od 5 putanja koje prolaze kroz vejpost dokument D3 završavaju u dokumentu D6 (2/5, odnosno 40%).

Prikazani primer (sl. 2) ilustruje jedan mogući način za rangiranje prečica, gde se kriterijum za relevantnost prečice nalazi u broju i procentu dosadašnjih ostvarenih veza između posmatranih parova dokumenata, u evidentiranim putanjama primera. Pri tome se ciljani i vejpost dokumenti smatraju ravnopravnima.

5. EFIKASNOST NAVIGACIJE

Efikasnost jedne putanje se može naći o odnosu aktivnosti (broj klikova) koji je potrebno izvršiti korišćenjem predloženih prečica, i aktivnosti (klikova) potrebnih za realizaciju putanje kakva je ona bez korišćenja dodatnih prečica. Ukupna efikasnost sistema se nalazi u srednjoj vrednosti izmerenih efikasnosti svih zabeleženih putanja.

„Prosečna efikasnost svih operativnih trasa se može smatrati merom efikasnosti celog veb sajta, na primer sve operativne trase od početne do ciljne stranice“. [9].

Na primeru razmatranom u ovom radu efikasnost putanje P1 je 40%, jer se do ciljnog dokumenta stiže sa 2 klika (V3 i V6), dok je bez upotrebe prečica potrebno 5 klikova (V2, V3, V4, V5, i V6) za dolazak do ciljnog dokumenata. Efikasnost se može slično izračunati i za ostale putanje, a ukupna efikasnost navigacije sajta je srednja vrednost efikasnosti za sve evidentirane putanje.

6. ZAKLJUČAK

U ovom radu su teoretski navedeni metodi za optimizaciju navigacione strukture kod adaptivnih veb sajtova. Ova optimizacija se ostvaruje stvaranjem veza između dokumenata (veb stranica) koji do tada nisu bili povezani. Pošto je često slučaj da sa jednog dokumenta postoji više veza koje je potrebno dodati, potrebno je rangirati predložene veze i one najperspektivniji postaviti na najuočljivija mesta. U ovom radu se u obzir uzimaju i vejpst dokumenti – međudokumenti koji mogu sadržati važne informacije potrebne za odlazak na ciljne dokumente. Spisak odredišnih dokumenata pored ciljnih dokumenata se proširuje i na spisak vejpst dokumenata. Efikasnost sistema se ogleda u odnosu napora za dolazak do ciljnih dokumenata kod optimizovanog i ne optimizovanog veb sajta.

7. LITERATURA

- [1] Wexelblat, A., Maes, P.: *Footprints History-Rich Tools for Information Foraging*, ACM Press, 1999., pp. 270-277.
- [2] Bathumalai, G.: *Self adapting websites: mining user access logs*, The Robert Gordon University, 2008.
- [3] Perkowski, M., Etzioni, O.: *Adaptive web sites: an AI challenge*, In Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, 1997., pp. 16-23.
- [4] Perkowski, M., Etzioni, O.: *Adaptive sites: Automatically learning from user access patterns*, Technical report, Department of Computer Science and Engineering, University of Washington, 1997.
- [5] Pamnani R., Chawan P.: *Web Usage Mining: A Research Area In Web Mining*, Department of computer technology, VJTI University, Mumbai, 2010.
- [6] Eremić Ž., Radosav D., Markoski B.: *Mining User Access Logs to Optimize Navigational Structure of Adaptive Web Sites*, in: Proceedings of the CINTI 2010 : 11th IEEE International Symposium on Computational Intelligence and Informatics, 2010., pp. 271-275.
- [7] Brickell J., Dhillon I., Modha D.: *Adaptive Website Design using Caching Algorithms*, WebKDD'06 Proceedings of the 8th Knowledge discovery on the web international conference on Advances in web mining and web usage analysis, 2006.
- [8] Eremić Ž.: *Lanci Markova u funkciji poboljšanja navigacije adaptivnih veb sajtova*, I Naučno – stručni skup: PREDUZETNIŠTVO, INŽENJERSTVO I MENADŽMENT, Visoka tehnička škola strukovnih studija u Zrenjaninu, Zrenjanin, 2010., str. 187-194.
- [9] Lee, J., Shiu, W.: *An adaptive website system to improve efficiency with web mining techniques*, Advanced Engineering Informatics, 2004., Vol. 18, No. 3., pp. 129-142.