

# iz naših knjižnica

Uređuje: Danko Škare

## Analiza pristupa mrežnom-poslužitelju knjižnice Instituta Ruđer Bošković

A. Vodopijevac

Institut Ruđer Bošković, Bijenička c. 54, 10000 Zagreb  
alen.vodopijevac@irb.hr

### Uvod

Internet je postao nužnost, a prisutnost na mreži nezaobilazni dio gotovo svake djelatnosti. Knjižnice, u skladu s time, ulažu mnogo truda u pripremi te prezentaciji vlastitih sadržaja i usluga na internetskim stranicama. Na taj način omogućen je pristup knjižničnim "online" katalozima (OPAC), raznim elektroničkim izvorima (elektroničke knjige, e-časopisi) te obrazovnim materijalima. Svrha je tih napora olakšati korisniku pristup informaciji koja je za njega relevantna.

Internetske stranice knjižnice Instituta Ruđer Bošković nastale su davne 1994. godine, a izradilo ih je osoblje knjižnice. Na stranicama se godinama skupljalo sve više informacija kako bi se zadovoljili svi korisnici. Isto tako, bilježimo i sve veći broj posjeta, tako da se i struktura posjetitelja promijenila. Danas, s više od 1000 posjeta dnevno moramo se zapitati tko nas to posjećuje i što im skladu s time možemo ponuditi. Budući da je u planu redizajn cjelokupnog internetskog sadržaja, napravili smo malu studiju pristupa internetskom poslužitelju knjižnice kako bismo nešto saznali o korisnicima i njihovim preferencijama.

### Metodologija

Poslužitelj bilježi komunikaciju s korisnikom u obliku "log" datoteka, gdje svaki redak označava jedan izdvojeni zahtjev (hit) korisničkog internetskog preglednika za učitavanje jednog objekta s mrežnog poslužitelja. Objekti o kojima govorimo uglavnom su html ili php datoteke, a mogu biti i dokumenti s nastavcima doc, ppt, xls, pdf i txt. Bilježe se i oni zahtjevi koji proizvode standardnu pogrešku (npr. u slučaju nepostojanja izvora koji je naveden u HTTP zaglavju).

194.186.163.241 - - [13/Feb/2003:15:00:46 +0100]

"GET /izvjestaj/1993/1993.txt HTTP/1.0" 206 9853

"http://www.google.com/search?num=50&hl=en&q=

R-matrix+synthesis+organic&btnG=Google+Search"

"Mozilla/4.0 (compatible; MSIE 5.00; Windows 98)"

Naveli smo primjer takozvanog kombiniranog "log" zapisa generiranog od strane Apache mrežnog poslužitelja verzije 1.3.29. Kao što vidimo, mnogo toga se može iščitati iz ovakvih zapisa, no isto tako moramo znati da su to velike količine podataka ako uzmemo u obzir prosječni broj dnevnih hitova, koji se kreće oko 25 tisuća. Zapis nam nudi sljedeće podatke:

- IP adresu s koje je stigao zahtjev
- Korisničko ime, ukoliko je za određeni resurs bila potrebna autorizacija

- Vrijeme
- Primijenjenu GET ili POST metodu
- Datoteku na koju se zahtjev odnosi
- Povratnu informaciju (response code) mrežnog poslužitelja (npr. 404 – NOT FOUND)
- Količinu podataka poslanu korisniku
- Referentni URL
- "User agent" – entitet koji je generirao zahtjev (mrežni preglednik, robot ..) itd.

Radi analize tih podataka, upotrijebili smo dva alata – Webalizer i AWStats. Webalizer je već uobičajena aplikacija za praćenje statistika pristupa mrežnom poslužitelju (može obrađivati i druge vrste logova), dok je AWStats nešto recentniji uradak i glavna mu je prednost bolje identificiranje takozvanih robota (npr. automatiziranih zahtjeva generiranih od strane mrežnih tražilica radi indeksiranja podataka o internetskim stranicama) od ljudskih korisnika.

U pripremi fokusirali smo se na konfiguraciju (ugađanje) ovih alata kako bismo izvukli što relevantnije podatke. Tako smo, na primjer, izdvojili korisnike s IRB-a od ostalih korisnika, a isto tako iz analize smo izuzeli pristupe s računala djelatnika knjižnice (riječ je o javnim IP adresama) koji, osim ostalog, i uređuju dotične internetske sadržaje.

### Rezultati

Analizirano je razdoblje od jedne godine (od siječnja do prosinca 2003.) s posebnim osvrtom na mjesec travanj, koji smo uzeli kao reprezentativan budući da se broj posjeta kreće oko godišnjeg prosjeka.

Tablica 1.

Ukupno za 2003.	
Broj hitova	8628766
Broj posjeta	397661
Količina podataka (u KB)	56298820

Odvojili smo hitove (pojedinačne zahtjeve) od posjeta. Jedan posjet je standardno određen bilo kojim brojem zahtjeva s iste IP adrese u razdoblju od pola sata.

Tablica 2.

Mjesečne statistike za travanj 2003.		
Hitovi	756863	
Posjete	32185	
	Avg	Max
Hitova po satu	1051	4914
Hitova po danu	25228	40055
Posjeta po danu	1072	1448
Broj hitova prema povratnoj informaciji		
Code 200 – OK	420045	
Code 304 – Not Modified	239308	
Code 404 – Not Found	82704	

## Tko nas posjećuje

Prilično je teško odrediti strukturu posjetitelja, a suočili smo se i s raznim problemima u definiranju određenih grupa korisnika. Prvi zadatak bio je odvojiti korisnike koji pristupaju lokalno – s Instituta. Nakon toga, ostale korisnike moramo grupirati što smo ostvarili na osnovi podataka o internetskim domenama s kojih nam pristupaju.

Zanimljiv je podatak da oko 24,5 % ukupnih zahtjeva dolazi od strane korisnika koji koriste modemsku vezu (HTNet, Iskon, Globalnet i dr.). Taj oblik veze koristi se najčešće za spajanje od kuće. Sličan postotak (23,9 %) odnosi se na korisnike s Instituta Ruđer Bošković. Dakle, te dvije grupe korisnika ostvaruju nam gotovo 50 % prometa.

S hrvatskih sveučilišta dolazi nam oko 10 % od ukupnih zahtjeva, dok ostali instituti i bolnice sudjeluju s 2,5 %. Djelatnici većih hrvatskih tvrtki (Pliva, Ina, Podravka) posjete nas oko 170 puta mjesečno.

30 % zahtjeva dolazi nam s inozemnih adresa (izvan hr. domene).

## Kako se dolazi do nas

U logovima nalazimo i podatak o izvoru trenutačnog zahtjeva, dakle, koji je to URL prethodio našem posjetu.

Direktni zahtjev znači da je osoba manualno upisala adresu u pregledniku ili je adresa postavljena kao početna stranica (Homepage). Kao što smo i pretpostavljali, Google je najzastupljenija tražilica, a velik broj korisnika pristupa nam preko linkova na stranicama Instituta Ruđer Bošković.

Izvor	Hitovi
Direktni zahtjev	122632
Google tražilica (www.google.com)	11249
IRB web (www.irb.hr)	3961
Yahoo tražilica (search.yahoo.com)	1486
CROSBİ (priroda.irb.hr)	1080
MSN tražilica (search.msn.com)	994

## Koji sadržaji su popularni

Najveći broj posjeta bilježe stranice koje pružaju informacije i pristup "online" bazama podataka (Ovid, Wos), a prate ih stranice namijenjene lokanim korisnicima (novosti, događaji, izvještaji o radu Instituta i dr.). Visoko na top-listi su i "Virtualna referentna zbirka", stranice časopisa "Kemija u industriji" te "Znanost na Internetu". Navedeni sadržaji nalaze se među prvih deset prema količini podataka prenesenih između poslužitelja i korisnika.

## Zaključak

"Log" datoteke nam pružaju dobar uvid u strukturu i potrebe naših posjetitelja. Možemo uočiti dvije glavne grupe temeljene na broju posjeta i preferencijama u odnosu na sadržaj.

Prva grupa odnosi se na osobe koje traže specifične informacije vezane uz Institut Ruđer Bošković i našu knjižnicu. Korisnici iz druge grupe naše stranice koriste isključivo kao "odskočnu dasku" u potrazi za znanstvenim informacijama i njihovim izvorima. Dakle, u dizajniranju nove strukture internetskih stranica potrebno je veću pažnju pokloniti razdvajanju tih dviju vrsti sadržaja te u skladu s time olakšati navigaciju kako bi se vrlo brzo stiglo do željene informacije.

## Bibliografija

1. D. Fichter, Server logs: making sense of the cyber tracks. *Online* 2003; Sept/Oct: 47–50.
2. M. Schayler, Meaningful Web statistics? How you're hitting on something? *InfoToday* 2003; Jan: 38–40.
3. S. Ketchpel, Build a better site by understanding customer behaviour 2003; Jan/Feb: 14–16.
4. C. Morris, How log files work. *Web Developer's Virtual Library* 1999; Aug. [http://wdvl.internet.com/Internet/Management/how\\_work.html](http://wdvl.internet.com/Internet/Management/how_work.html) [2003-09-15].
5. Glenn Fleishman. *Web Log Analysis W: Who's Doing What, When?* *Web Developer magazine* 1996; 2(2).