

CMI – Center za metodologijo in informatiko
FDV – Fakulteta za družbene vede, Univerza v Ljubljani
[http:// www.ris.org](http://www.ris.org)
email: info@ris.org

RIS2000

merjenje spletne obiskanosti

Ljubljana, 2001

POVZETEK

V javnosti pogosto naletimo na neustrezne navedbe o obiskanosti spletnih predstavitev. Še več, pokazati je mogoče, da je večina navedb v slovenskih medijih pravzaprav napačnih ali pa vsaj zavajajočih.

Pričujoče poročilo je zato posvečeno primerjavi merjenj spletne obiskanosti, v manjši meri tudi aktualnim meritvam. Sistematično je obravnavana terminologija in analizirana je pristranskost posameznega načina merjenja spletne obiskanosti. V okviru projekta RIS namreč spremljamo obiskanost spletnih strani na tri načine:

- a) Analiza zahtevkov na spletnih strežnikih - »*server-centric*« analiza logov.
- b)
- c) Anketiranje (»*user-centric*«) v reprezentativnih telefonskih anketah.

Poleg zgoraj navedenega poteka tudi primerjalna analiza »*ad-centric*« analize logov (d), to je merjenj, ki jih izvajajo oglaševalski posredniki. V teku je tudi evaluacija števcov, to je merjenj, ki jih izvajajo tretje strani (e). V pripravi pa je tudi pilotno zajemanje na osnovi PC-metrije (f).

Pričujoče poročilo se osredotoča predvsem na primerjave analize logov (a) in telefonskih anket (c). Analiziramo torej meritve, ki temeljijo na **log analizi spletnih strežnikov**, kjer so objekti merjenja zahtevki v log datotekah, ter meritve, ki temeljijo na **telefonskih anketah uporabnikov Interneta**.

Osnovna ugotovitev je, da sta oba pristopa pri pravilni obravnavi - kljub številnim specifičnostim - primerljiva. Multipla regresija namreč s koeficientom $R=0,86$ preslika mesečno število različnih IP števil ter število mesečnih obiskov (sessions) v število različnih mesečnih oseb/obiskovalcev.

Posebej podrobno so bile v log analize obravnavana razmerja naslednjih statistik:

- število dostopov do prve strani vs. število obiskov,
- število obiskov vs. število različnih obiskovalcev,
- število različnih obiskovalcev vs. število IP števil.

Analize kažejo, da zaradi tehnoloških specifičnosti in narave obiskanosti določenih spletnih predstavitev **ponekod nastajajo izjemno velike razlike v razmerjih** zgoraj navedenih statistik. Razlike so še večje, če upoštevamo bolj podrobne kategorije (npr. zahtevki), ki so zato **neprimerne za kakršnekoli primerjave in javno navajanje**.

Metodološke težave nastajajo tudi pri merjenju s pomočjo telefonskega priklica, vendar je za primerjave obiskanosti ta način **bistveno bolj ustrezen**. Za podrobnejšo analizo znotraj samih spletnih predstavitev pa je primernejša analiza logov.

KAZALO

I. OSNOVE ANALIZE LOGOV	5
1. Terminologija.....	5
1.1. Domene.....	5
1.2. Gostiteljski računalnik (host).....	6
1.3. Spletni objekti.....	7
1.4. Elementi spletne obiskanosti.....	9
2. Metodologija log analize.....	12
2.1. Kaj je log datoteka?.....	12
2.2. Kukiji.....	13
2.3. Logi, kukiji, računalniki in obiskovalci.....	14
2.4. Metodološki problemi pri analizi logov.....	16
3. Potvarjanje log datotek	19
II. RIS ANALIZA LOGOV	21
1. Metodologija analize logov projekta RIS.....	21
2. Rezultati analize logov projekta RIS.....	22
3. Razmerja log statistik.....	24
3.1. IP in različni obiskovalci.....	24
3.2. Dostopi na prvo strani in različni obiskovalci.....	26
3.3. Dostopi na prvo strani in število obiskov.....	27
3.4. Število obiskov in število različnih obiskovalcev.....	29
III. ANALIZA OBISKANOSTI - TELEFONSKE ANKETE	33
1. Metodologija.....	33
1.1 Anketa kot merski instrument.....	33
1.2. Merske napake	36
1.3. Telefonske ankete vs. log-analiza.....	37
2.1. Mesečne serije za izbrane predstavitve.....	39
2.2. Pregled obiskanosti po demografskih strukturah	39
IV. PRIMERJALNA ANALIZA MERITEV	40
1. Obiski, IP številke in različni uporabniki.....	40
2. Razvoj linearnega regresijskega modela	42

KAZALO TABEL

<i>Tabela 1: Rezultati log analize</i>	22
<i>Tabela 2: Razmerje - dostopi na HP / število različnih obiskovalcev</i>	26
<i>Tabela 3: Frekvence obiskov na prvi strani v okviru obiska spletne predstavitev</i>	28
<i>Tabela 5: Frekvenca dostopov različnih obiskovalcev na predstavitev</i>	30
<i>Tabela 7: Širina intervalov zaupanja</i>	34
<i>Tabela 8: Oznake koeficientov variacije</i>	35
<i>Tabela 9: Regresijske ocene</i>	43

KAZALO GRAFOV

<i>Graf 1: Shematska predstavitev delovanja proxy strežnika</i>	17
<i>Graf 2: Primerjava št. Različnih obiskovalcev (Unique Users) in št. različnih strežnikov (distinct hostes)</i>	24
<i>Graf 3: Različni obiskovalci (Unique Users) in različni strežniki (Distinct Hosts)</i>	25
<i>Graf 4: Primerjava št. dostopov na prvi strani (Home Page Hits) s številom obiskov na celi predstavitvi (Users sessions)</i>	27
<i>Graf 5: Število obiskov (users sessions) in število različnih obiskovalcev (unique users)</i>	29
<i>Graf 6: Primerjava obiskov (Users Sessions) in različnih obiskovalcev (Unique Users)</i>	31
<i>Graf 7: Primerjava obiskov (user sessions) in različnih obiskovalcev (unique users)</i>	32
<i>Graf 8: Primerjava log analize in telefonskih anket (a)</i>	40
<i>Graf 9: Primerjava log analize in telefonske ankete (b)</i>	41
<i>Graf 10: Primerjava meritev obiskanosti spletnih strani</i>	45
<i>Graf 11: Ocenjeno in izmerjeno število obiskovalcev</i>	46

I. OSNOVE ANALIZE LOGOV

1. Terminologija

Uvodoma si oglejmo nekatere najpogostejše pojme na tem razmeroma zapletenem področju.

1. 1. Domene

Domena je ime (npr. ris.org), ki enolično določa subjekte (npr. podjetja, organizacije in ostale pravne osebe, pa tudi projekte, fizične osebe in blagovne znamke) v omrežju Internet. Sestavljena je iz več komponent, zadnji del, .org, .com, .net - ali npr. .si v naslovu <http://www.ijs.si> - imenujemo **vrhnja domena**, predzadnji del, npr. »ijs« v prejšnjem naslovu, pa **vrhnja poddomena**, podobno tudi npr. oznaka »ris« v domeni ris.org.

Obstaja več načinov dodeljevanja vrhnjih poddomen znotraj posameznih držav. Ponekod, tipičen primer so ZDA, lahko kdorkoli registrira katerokoli poddomeno. Pravilo velja znotraj vseh vrhnjih domen .com, .org, .net, kjer se dnevno registrira tisoče novih domen, mnoge pa se tudi preprodajajo. V primeru spora, ko nekdo registrira domeno, ki jo npr. drugi subjekt uporablja kot blagovno znamko, se problem rešuje po sodni poti. V Evropi pogosteje naletimo na način dodeljevanja vrhnjih poddomen na osnovi registriranega pravnega subjekta in tako je tudi v Sloveniji - vsaj do leta 2001, ko se napovedujejo določene spremembe. Tako bi npr. lahko vrhnjo poddomeno *ris.si* registriralo le podjetje RIS, d.o.o ali podobna organizacija. Tak način preprečuje mnoge sodne spore, po drugi strani pa seveda **sili slovenske subjekte v registracijo v tujini**, kar med drugim bistveno **zmanjšuje (statistično) preglednost slovenskega spletnega prostora**. Povzroča tudi **določen odliv sredstev v tujino in manjšo nacionalno promocijo**. V vsakem primeru neusklajeni administrativni postopki dodeljevanja domen **otežujejo vsebinske interpretacije** njihovega števila ter mednarodne statistične primerjave. Tipičen primer je npr. selitev domen v Turkmenijo, ki ima liberalno regulativo, dodeljuje pa naslove z nadvse privlačno vrhnjo poddomeno »tm«, ki jo zahodni svet pozna predvsem kot pomemben simbol »™« (»trademark«) industrijske lastnine. Porast števila domen v Turkmeniji zato ne bo odražal širitve Interneta v tej državi.

V Sloveniji dodeljuje vrhnje poddomene ARNES (www.arnes.si) in po njegovih administrativnih podatkih obstaja 10.000 (začetek leta 2001) vrhnjih poddomen, ki pripadajo vrhnji domeni .si. Sproten mesečni pregled števila domen se nahaja na <http://www.ris.org/indikatorji/domene.html>, kjer se število domen spremlja tako administrativno kot tudi programsko. Seveda pa med dejanskim stanjem (aktivnih) domen in administrativnimi podatki obstajajo določene razlike.

S podobno hitrostjo raste tudi skupno število domen v svetu. Rast domen se je torej od leta 1993, ko je obstajalo 21.000 domen, odvijala eksponentno, vendar se je že nekoliko upočasnila. Podrobnosti lahko najdemo na straneh [NetworkWizard](http://www.nw.com) (<http://www.nw.com>), ki na osnovi raziskave Internet Domain Survey objavlja podrobno gibanje števila domen.

1. 2. Gostiteljski računalnik (*host*)

Še bolj pomembno vlogo za razumevanje ima pojem »host«-ov, t. j. posebnih gostiteljskih računalnikov, ki so vključeni v Internet in imajo tudi posebno »Internet protocol« IP številko. Pri tem veliko težav povzroča dejstvo, da »host« ne pomeni nujno računalnika z IP številko, saj jo ima lahko tudi klicni modem. Tako npr. več deset tisoč Siol-ovih dial-up uporabnikov vstopa v Internet od doma preko samo nekaj tisoč IP modemov. Posamezni računalnik s klicnim dostopom v takem primeru torej nima svoje IP številke, ampak privzame trenutni IP modema, ki se mu je ob vstopu alocirala slučajno izmed vseh trenutno prostih modemov. Uporabnik torej lahko v desetih uporabah dostopa preko desetih različnih IP številok oz. modemov.

Podobno ima lahko IP številko tudi omrežni strežnik, to je posebni računalnik, preko katerega dostopa do Interneta celo omrežje računalnikov v podjetju. V takem primeru lahko npr. deset računalnikov (uporabnikov) dostopa preko iste IP številke.

Mogoče je tudi obratno: da en sam fizičen računalnik služi kot osnova za več virtualnih »hostov«. Načeloma je torej mogoče ne enem računalniku gostiti deset različnih IP številok, vsak uporabnik ima v tem primeru svojega virtualnega »hosta« in v Internet vstopa s svojo IP številko.

Očitno je razmerje med računalniki, IP številkami in uporabniki nadvse zapleteno. Dodatne zaplete lahko pričakujemo, ko bodo IP številke začele dobivati tudi druge naprave, s katerimi bomo vstopali v Internet (interaktivna TV, mobilni telefoni), ali pa bodo kako drugače povezane z Internetom (gospodinjski aparati).

Število IP številok v omrežju se ugotavlja programske, s pošiljanjem signalov po Internetu. Na tem mestu velja omeniti European Hostcount ([RIPE](#)) in Internet Domain Survey ([Network Wizzard](#)), ki meri globalno število »host«-ov in tudi število aktivnih domen. Zaradi razlik v metodologiji se rezultati obeh meritev razlikujejo. Tovrstna merjenja se namreč srečujejo s številnimi tehničnimi težavami: zaščitenost računalnikov, podvojena imena, računalniki, ki so vključeni v Internet na nižjem nivoju poddomen ipd.

Omeniti velja, da skupno z vrhnjimi domenami *.mil*, *.org*, *.gov* in *.us* pripada komercialnim domenam, registriranim v ZDA (*.com*, *.net*) več kot polovica vseh računalnikov, vključenih v Internet. Seveda niso vsi ti računalniki tudi locirani v ZDA. Določen (manjši) del se nahaja tudi drugje, posebej velja to za vrhnje domene *.net*, *.com* in *.org*. Vrhnji domeni *.si* pripada 0.05% vseh računalnikov na svetu. Dodati velja, da dajejo meritve Network Wizzard za vse vrhnje domene bistveno nižje rezultate kot [RIPE](#), posebej izrazita pa je razlika pri Sloveniji.

Seveda pa se z novimi vrhnjimi domenami razmere na tem področju hitro spreminjajo.

Število »host«-ov se je od leta 1993, ko jih je bilo okoli milijon, povečalo v letu 1997 na več kot 20 milijonov, v letu 1998 na 30 milijonov in je v novem tisočletju preseglo 100 milijonov.

Število »host«-ov na 100.000 prebivalcev je kljub nekaterim pomanjkljivostim eden najbolj indikativnih pokazateljev rasti in razvitosti Interneta v določeni državi. Kot rečeno, pa nastajajo pri tovrstnih statistikah številne težave, od tehničnih problemov zaradi zaščite, ki lahko prepreči preverjanje obstoja določenega »hosta«, pa do težav zaradi različnih administrativnih postopkov registracije domen, kar je povezano z dodeljevanjem IP številok znotraj vrhnjih poddomen. Tako je npr. v večini držav OECD znaten (10-50%) del hostov registriran pod nenacionalnimi domenami. Podrobnosti o tem so na <http://www.ris.org/si/ris99/news/n311099.html>.

Tako so tudi za Slovenijo upoštevani le računalniki, ki so vključeni v Internet znotraj vrhne domene *.si*. Dodatnih nekaj tisoč računalnikov je vključenih v okviru drugih vrhnjih domen, predvsem *.com*, *.org* in *.net*. Slovenija je v tem pogledu posebej problematična, saj statistika o številu hostov kaže več kot zaskrbljujoče stanje, saj gostota hostov v Sloveniji že tri leta praktično stagnira. Podrobnosti o tem so na <http://ris.org/si/ris2000/novice/20010410.html>, kjer se redno spremlja gostota hostov za evropske države. Razlogi za to stagnacijo so kompleksni, vsekakor pa so k temu prispevale *visoke cene najetih vodov, restriktivna politika dodeljevanja domen*, deloma pa gre to pripisati tudi *nekaterim tehnoloških posebnostim v Sloveniji*.

1. 3. Spletni objekti

Čeprav obseg svetovnega spleta nezadržno narašča, iskalci (»search engines«) še vedno zagotavljajo osnovni pregled nad glavnino dokumentov, vključenih v WWW. Po nekaterih ocenah je bilo namreč v letu 2000 – brez zaščitene in ostalih nedosegljivih strani – na Internetu že več kot milijarda dokumentov (HTML strani). Zaradi naraščajočega števila in kompleksnosti dokumentov je v bodoče pričakovati precejšnje težave pri zagotavljanju učinkovitega pregleda. Po nekaterih ocenah namreč že danes večina iskalcev po WWW ne doseže niti 10% vseh strani, ki so vključene v Internet; le najboljši iskalci dosežejo tretjino. V tem okviru je karakteristično, da najdejo največ manj znani iskalci, kar vse pa se izredno hitro spreminja.

Za razumevanje obiskanosti svetovnega spleta (WWW) je nadvse pomembno ločevanje naslednjih kategorij oziroma pojmov:

- **Predstavitvena stran** oz. **domača stran** (»home page«) je osnovni zaokrožen prikaz določene vsebine na svetovnem spletu. Predstavitvena stran zavzema razmeroma širok spekter, tako lahko obravnavamo kot predstavitveno stran projekta RIS99 (www.ris.org/ris2000), predstavitveno stran projekta RIS2000 (www.ris.org/ris2000), ipd., lahko pa s predstavitveno stranjo razumemo tudi celoten projekt RIS (www.ris.org).

Včasih razumemo z izrazom domača stran tudi osnovno stran (*index.html*) celotne predstavitvene strani. Predstavitvene strani najdemo urejene v imenikih (npr. Yahoo, Matkurja.com in Slowwwenia.com), vendar številne vanje niso vključene, saj sistem vključevanja temelji na samoregistraciji. Po drugi strani lahko spletni iskalci (npr. Slovista, Najdi.si) najdejo bistveno več predstavitvenih strani, vendar pa te niso urejene v področja, kataloge in spiske.

Tako Altavista (<http://ris.org/indikatorji/altavista.html>) vključuje okoli 500.000 strani v domeni .si, iskalnik Najdi.si pa okoli milijon (<http://www.ris.org/si/ris2000/novice01/20010422.htm>). Seveda pa so mnoge od teh strani neuporabne, zastarele, nekatere strani pa kljub vsemu niso vključene.

– **Spletna stran (»page«)** je posamični dokument v posebnem standardiziranem formatu (html), ki ima na spletu tudi svoj enoličen, tako imenovani URL naslov. S tem je to tudi najosnovnejši element svetovnega spleta, ki ga je mogoče pregledovati, iskati in razvrščati. Predstavlja torej osnovni delec, iz katerega je sestavljen celoten svetovni splet (WWW). Posamezna stran je največkrat seveda le ena od strani v okviru določene predstavitevne strani. Poudariti velja, da je skoraj vsaka predstavitevna stran sestavljena iz nekaj (tudi več sto) posamičnih strani, čeprav obstajajo tudi poenostavljene domače strani (predstavitevne strani) z eno samo stranjo.

Običajno imajo spletne strani končnico "htm" ali "html". Strani, ki se generirajo dinamično (npr. iskalniki, novice) se od običajnih strani bistveno razlikujejo, saj ne nastopajo samostojno na spletu in niso vključene v iskalnike in imenike. Do njih je mogoče dostopati le preko interaktivnih aplikacij ponudnika teh informacij (tipičen primer so on-line dnevnik).

Za orientacijo velja torej ponoviti, da imajo iskalniki (npr. Najdi.si) registriranih že več kot milijon slovenskih strani (»page«). Po drugi strani pa lahko ocenimo, da je samih predstavitevni strani („*home page*“) v Sloveniji samo nekaj deset tisoč. Ob tem velja dodati, da je število v Sloveniji registriranih domen – torej spletnih naslovov tipa *www.xxx.si* - že preseгло 10.000.

Takoj velja dodati, da je na določeni domeni lahko večje število predstavitevni strani. V primeru osebnih predstavitevni strani na Arnesu na naslovu <http://www2.arnes.si/home/osebne.html> najdemo stotine osebnih predstavitevni strani in podobno velja tudi za domeno *siol.net*, v okviru katere ima predstavitevne strani veliko slovenskih podjetij. Kljub temu je še vedno najbolj običajno, da se predstavitevna stran povsem ujema z domeno.

Seveda pa je – kot smo že omenili – v okviru vrhnjih domen domen *com*, *net*, *org* dodatno registriranih veliko slovenskih predstavitevni strani, od *siol.net*, *amis.net*, pa do *podjetnik.com*, *ris.org*, *novogradnje.net*, *poljub.net* ipd. Število slovenskih spletnih predstavitev tudi zato bistveno presega število domen v okviru *.si*.

- **Predstavitev oz. spletna predstavitev (»site« oz. »website«)** označuje večji in kompleksnejši prikaz organizacije, podjetja, projekta ipd. na WWW. Tipičen primer je spletna predstavitev projekta RIS (*ris.org*), kjer lahko govorimo o samostojni domači strani imenika Sispleset (*sispleset.org*), o samostojni angleški domači strani ter o domačih straneh posameznih letnih projektov. Spletna predstavitev (*web-site*) se največkrat sicer omejuje le na eno samo predstavitevno/domačo stran (*home-page*), lahko pa vsebuje tudi večje število predstavitevni strani. Neredko se predstavitev ujema tudi z določeno domeno; predstavitev namreč pogosto obsega ravno vse strani znotraj določene vrhnje poddomene. Ker pa s predstavitvijo razumemo vsebinsko zaokrožene dokumente, se nekatere predstavitve lahko raztezajo na več domenah (npr. projekt RIS je delno na domeni *Arnes.si*, delno na *Ris.org*). Velja tudi obratno: v

okviru določene domene je lahko več neodvisnih spletnih predstavitev (home page), ki niso – taka kot v primeru projekta RIS – tesno povezane. Tipičen primer so domače strani podjetij, ki se nahajajo na domeni ponudnika dostopa do Interneta (npr. amis.net, siol, net), pa tudi domače strani javnih organizacij, ki domujejo v okviru vrhnje poddomene gov.si in so med sabo neodvisne in nepovezane

- **Spletna blagovna znamka (»web site property«)** je osnovna kategorija spremljanja obiskanosti spletnih strani in označuje vse predstavitve, ki sodijo pod določeno blagovno znamko (npr. Yahoo, Microsoft). Pri tem lahko povezujejo več kompleksnih spletnih predstavitev (»site«) in se pogosto, ne pa vedno, ujemajo tudi z domeno.

Očitno je, da so zgornje opredelitve razmeroma ohlapne, zato se njihove vloge pogosto prepletajo in zamenjujejo. Tako je lahko individualna predstavitev določene osebe le posamičen dokument (»page«), ena sama stran, v okviru predstavitvene strani (»home page«) določene organizacije, po drugi strani pa lahko isto stran upravičeno razumemo tudi kot predstavitveno stran (»home page«) te osebe.

Zaradi nelahke razmejitev med domeno, predstavitvijo, predstavitveno stranjo ter posamezno stranjo, je statistično spremljanje njihovega števila nadvse težavno. Kljub temu velja dodati, da so vsaj pojmi domene, strani ter predstavitvene strani (domača stran) med seboj razmeroma jasno razmejeni in med njimi ne bi smelo prihajati do prekrivanja. Naj ponovimo okvirno število osnovnih kategorij za slovenski Internet prostor. Pri tem se v grobem omejujemo na strani v slovenščini in na strani, na katerih se slovenska podjetja, organizacije in posamezniki predstavljajo v tujih jezikih.

- Število posameznih strani na spletu –	cca. milijon
- Število domen .si	10.000
- Število vseh domen	cca. 15 tisoč
- Število predstavitvenih strani	cca. 20 tisoč

Ostali pojmi so slabše opredeljeni. Spletnih predstavitev (web-site) je vsekakor manj kot predstavitvenih strani (home-page), čeprav se pojma v večini primerov pokrivata. Še manj je spletnih blagovnih znamk (web-site properties), čeprav se tudi slednje pogosto pokrivajo z spletno predstavitvijo ali kar z domeno.

1. 4. Elementi spletne obiskanosti

Zahtevek (request) je najbolj elementarna kategorija spletne obiskanosti. Z njim odjemalčev računalnik (klient) sproži nalaganje določenega objekta (npr. strani na spletu) iz gostiteljskega računalnika (hosta). Z vidika uporabnika pa to običajno pomeni, da si uporabnik pač ogleduje določeno spletno stran. Na eni spletni (html) strani je navadno več objektov (npr. grafik, logotipov, bannerjev).

Za vsak objekt je potreben poseben zahtevek, zato je dostop do strani lahko sestavljen iz več realiziranih zahtevkov (»requests«). Vsak tak zahtevek generira vrstico zapisa v posebni »log« datoteki, ki vsebuje osnovne podatke o računalniku, s katerega prihaja obiskovalec in druge karakteristik. Število zahtevkov je torej odvisno od števila

objektov na spletnih straneh, ki se prenašajo s strežnika na odjemalec, zato je število zahtevkov specifično za posamezno spletno stran. Statistika obiskanosti na osnovi zahtevkov torej ni medsebojno primerljiva in je zato kot indikator obiskanosti razmeroma neuporabna. Stran z npr. 10 slikami pri vsakem nalaganju generira dodatnih 10 zahtevkov plus enega za samo spletno stran (html), stran brez slik pa samo en zahtevek.

Dostopi (page impressions) do strani so zadetki na posamezni spletni strani (page). Nekateri jih imenujejo tudi **ogledi** ali **prikazi**. Gre za realizirane zahtevke po ogledu določene strani (s svojim URL naslovom), ki smo jih s klikom izvedli iz uporabniškega računalnika, realiziral pa jih je analizirani strežnik in ne nek »proxy« strežnik ali »cash« predpomnilnik, o čemer bomo še govorili. S posebnimi nastavitvami je mogoče slediti tudi dostopom, ki so jih posredovali različni posredniki (npr. proxy). Dostope je mogoče beležiti tudi z vsiljeno inštalacijo posebne datoteke na analizirano stran, ki uporabnika prisili, da jo naloži iz strežnika, kjer se opravi štetje. Število dostopov do določene strani je zato v veliki meri odvisno od organizacije in tehnologije spletne predstavitve, povečujejo pa jo tudi številni roboti, spiderji in ostali programi, s katerimi je mogoče umetno generirati poljubno število dostopov do določene strani.

Dostopi do prve strani (hits for home page) so zadetki – to so realizirani zahtevki - na prvo stran spletne predstavitve (navadno je to index.html).

Obiski (visit, user sessions) predstavljajo dostopanje do spletne predstavitve (»web site«) oziroma domače strani (»home page«) v okviru ene uporabe. Imenujemo jih tudi **seanse** ali **uporabniške seje**. Gre za dostopanje, ki ni nastalo kot ponovljeno vračanje v okviru ene »seanse« na spletu. Kadar se večkrat vrnemo na isto stran, se torej povečuje število dostopov, ne pa tudi število obiskov. Običajno se po 30-minutni prekinitvi ponovna vrnitev istega uporabnika na določeno predstavitev obravnava kot nov obisk (uporabniška seja) na tej predstavitvi. Obiski se pogosto navajajo kot osnovno merilo obiskanosti. S tem pa se lahko uporabnike zlahka zavede, da z npr. 5000 obiski razumejo 5000 obiskovalcev, kar je načeloma sicer povsem mogoče. Mogoče pa je tudi, da je vseh 5000 obiskov generiral en sam obiskovalec, kar je velika razlika.

Mnoge spletne predstavitve imajo v ta namen inštalirane tudi enostavne števec, ki kažejo kumulativno število obiskovalcev. Pri tem nekateri števec naraščajo celo z enostavnim osveževanjem strani (ctrl R), drugi pa v okviru ene seanse ne povečujejo števca obiskov. Nekateri zahtevnejši števec (npr. 100si) štejejo kot obisk le dostopanje iz iste IP številke v presledku več kot 30 minut, s čimer izločijo dostopanje iz iste IP številke znotraj 30-minutnega intervala, tudi če je pri tem prišlo do prekinitve. Z nekoliko poenostavitve lahko rečemo, da v takem primeru iz enega računalnika teoretično ni mogoče generirati dnevno več kot 48 obiskov. Spet drugi števec štejejo samo obisk iz enega kukija (če je bil uspešno inštaliran, sicer uporabijo IP številko) v enem dnevu samo enkrat (npr. obisk.net), kar nadalje zmanjšuje število obiskov za 10-20%. Očitno je torej, da obstaja precejšnja variabilnost pri merjenju števila obiskov. To je še posebej očitno, ko primerjamo rezultate različnih programov, ki se običajno razlikujejo do 10% ali celo za 20%, kar je odvisno tudi od nastavitve parametrov znotraj vsakega takega programa.

Različni uporabniki (unique users) – gre za število različnih uporabnikov, kot jih lahko med seboj ločijo programi za analizo obiskanosti (npr. Webtrends) na osnovi razpoložljivih informacij v log datoteki (uporabniško ime, IP številka računalnika, kukiji), o čemer bomo podrobneje govorili v nadaljevanju. V določenih primerih različnih uporabnikov seveda ne moremo ločiti, kot tudi ne moremo vedno preprečiti, da se isti uporabnik ne bi pojavljal v dveh različicah. Izkaže se, da v primeru, ko se kukiji ne uporabljajo, praktično ni razlik med številom IP številčk in različnimi uporabniki, saj razen IP številčk ni druge osnove za razločevanje. V primeru uporabe kukijev pa se število različnih obiskovalcev poveča za faktor 2-4, lahko pa tudi drugače, če gre za specifično okolje in nastavitve.

Različne IP številke (unique IPs) – kot smo že omenili gre za število gostiteljskih računalnikov (»host«), ki obiščejo spletno predstavitev (web site). Dostopanje z istim računalnikom ob različnih »seansah« tako povečuje dostope in obiske, ne pa tudi število različnih IP. Seveda oseba lahko dostopa prek različnih računalnikov in tudi različne osebe lahko dostopajo prek istih računalnikov. Pogosto namesto gostiteljskega računalnika kot IP številka nastopa modem ali pa poseben strežnik, ki služi več uporabnikom. Pričakovati je, da bo imelo v bodoče svojo IP številko vse več naprav (»devices«), ki bodo omogočale dostop do Interneta (npr. mobilni telefoni, TV, in drugi aparati). Zaradi takšnih primerov se interpretacije števila različnih IP številčk dodatno zapletejo. Kot rečeno pa je brez uporabe kukijev število različnih uporabnikov in število različnih IP številčk precej podobno, saj razen IP številke ni veliko drugih podatkov, na osnovi katerih bi ločevali uporabnike.

2. Metodologija log analize

V nadaljevanju obravnavamo sledi/zapise, ki nastanejo/ostanejo na strežniku (»serverju«) – to je računalnik, kjer se nahaja določen spletna predstavitev – kadar si oseba/uporabnik spletno predstavitev ogleduje in pri tem nalaga spletne strani s strežnika na svoj računalnik. Ker analiziramo sledi obiska na strežniku s spletno predstavitvijo, govorimo o »*server-centric*« pristopu.

2.1. Kaj je log datoteka?

Log datoteka je tekstovni zapis »log«-ov, to je dokumentacije o obisku zapisanem v vrstici posebne log datoteke. Gre za navaden tekstovni (ascii) zapis o posameznem zahtevku na strežniku. V posamezno log vrstico se torej zapisujejo razpoložljivi podatki o uporabniku, ki obiskuje določeno spletno stran.

Bolj natančno: v resnici seveda govorimo o odjemalcu (spletnemu brskalniku) v času komunikacije med brskalnikom, ki ga uporablja uporabnik, ter spletnim strežnikom, kjer se nahaja obiskana spletna predstavitev. Govorimo torej o dokumentiranem zapisu odnosa med dvema računalnikoma, čeprav gre pri tem le za formalen odsev odnosa, ki se v vsebinskem smislu odvija med osebo/uporabnikom in vsebino na svetovnem spletu (WWW), ki si jo oseba ogleduje. Log datoteka je torej računalniški odsev, ki priča o tem, da je določena oseba obiskala določeno spletno stran.

Obisk, ki ga je izvedel obiskovalec na določeni strani, se torej preslika v zapis komunikacije dveh računalnikov (log). Žal ta preslikava ni enolična, zato je sklepanje o obnašanju oseb na osnovi logov nadvse približno in težavno. Značilen zapis o obisku je praviloma sestavljen iz naslednjih podatkov:

1. IP številka gostiteljskega računalnika in/ali ime računalnika (hostname)
2. datum in čas obiska
3. podatki o zahtevkih po posameznih vsebinah (datotekah)
4. koda (pove, ali je bil vsak zahtevek uspešno izveden)
5. velikost datoteke, za katero je bil sprožen zahtevek
6. spletna stran, s katere je prišel obiskovalec
7. brskalnik in operacijski sistem obiskovalca

Spodaj predstavljamo tipičen zahtevek za prikaz spletne strani, ki ga je sprožil uporabnik Interneta prek spletnega brskalnika (primer spletne strani RIS).

```
209.167.195.11 - - [26/Jan/1999:04:47:38 +0100] "GET /ris98/surveys98.html
HTTP/1.1" 200 12022 "http://www.ris.org/" "Mozilla/4.0 (compatible; MSIE
4.01; MSN 2.6; Windows 95)"
```

V prvem polju je vedno zapisano ime gostujočega računalnika (hostname) ali IP številka računalnika (**209.167.195.11**). IP številka v teoriji enolično določa računalnik, ki je povezan v Internet in dostopa do določenega spletnega strežnika. Obširneje o problematiki števila hostov (IP številke) v Sloveniji razpravljamo na <http://www.ris.org/si/ris99/news/n311099.html>

Informacija o datumu in času obiska je nadvse pomembna in omogoča opazovanje datumskih in časovnih gostitev obiskov, tako v okviru enega dneva kot tudi meseca (in poljubnih časovnih intervalov). S pomočjo takšne informacije tudi razberemo, ali je npr. začetek oglaševalske akcije povzročil porast obiska na spletni predstavitvi oglaševalca. (**26/Jan/1999:04:47:38 +0100**).

S podatki o zahtevkih posameznih elementov vidimo, kolikokrat je bila sprožena zahteva po prikazu določene datoteke (**GET /ris98/surveys98.html**). Običajno vsaka stran vsebuje več elementov (datotek) in ko obiskovalec obiše stran, se sprožijo zahtevki zanje¹. Kadar je med zahtevanimi datotekami tudi oglas, se zabeleži vpogled. HTTP-koda pove, ali je bil prenos uspešen (**200**), razpolagamo pa tudi z velikostjo (**12022**) zahtevane datoteke.

Podatek o strani, s katere je prišel obiskovalec (**http://www.ris.org/**), je bistven z več vidikov. Če oglašujemo, imamo s tem nadzor nad številom preskokov. Vsak prišlek s strani, kjer oglašujemo, je kliknil na oglas in prek hiperpovezave prišel na naše (ciljne) strani, kar je razvidno iz log datoteke. Na koncu zapisa sledijo še podatki o brskalniku in operacijskem sistemu obiskovalčevega računalnika (**Mozilla/4.0compatible; MSIE 4.01; MSN 2.6; Windows 95**)

Posebej velja poudariti, da nekateri strežniki dodeljujejo obiskovalcem (njihovim računalnikom) tudi t.i. *kukije* (cookies), o čemer bomo govorili v nadaljevanju.

2.2. Kukiji

Kukiji (ang. cookies) - nekateri jih dobesedno prevajajo tudi v »piškote« ali »kolačke« - so majhne datoteke, ki jih spletni strežnik zapišejo na odjemalčev računalnik. V vsakem osebem računalniku je namreč poseben direktorij (npr. Windows/Cookies v sistemu Windows 98), kjer se shranjujejo omenjene datoteke.

Datoteke so običajno velike samo 1 kB in so npr. oblike pcpeter@ris.org, kar pomeni, da je uporabniški računalnik z imenom Pcpeter obiskal stran ris.org. Pri tem se generira tudi datum obiska. Ob naslednjem obisku bo torej strežnik, kjer je spletna predstavitev www.ris.org, uporabnika, ki prihaja iz navedenega računalnika, identificiral kot ponavljajočega obiskovalca, ki je stran obiskal že na določen datum.

S tem enostavnim postopkom lastniki spletne predstavitve – posebej če razpolagajo z več spletnimi predstavitvami – zbirajo informacije o načinu in pogostosti obiskovanja. Na osnovi strani, ki jih je uporabnik doslej obiskal, pa lahko ocenijo tudi profil obiskovalca in na tej osnovi npr. generirajo individualizirane oglase.

¹ Če zahtevamo spletno stran, na kateri je pet slik in se te tudi v celoti naložijo, se bo v log datoteko zapisalo šest zahtevkov (eden za html dokument in pet za slike).

Kljub vsemu velja poudariti, da strežnik sledi le lastne kukije, to je kukije, ki jih je sam inštaliral, ne pa tudi drugih kukijev. S tem je domet kukijev bistveno zožen. Večji domet imajo le kukiji oglaševalskih posrednikov, ki se inštalirajo iz večih spletnih predstavitev.

Seveda na osnovi kukijev ni mogoče ločiti različnih uporabnikov, ki dostopajo iz istega osebnega računalnika (razen če ne uporabljajo npr. svojih NT profilov). Ravno tako bo uporabnik, ki dostopa iz večih osebnih računalnikov, prejel različne kukije in videti bo, da je predstavitev obiskalo več različnih oseb.

Uporabnik lahko z nastavitvami v brskalniku sprejme ali zavrne namestitev kukijev na trdem disku, kadarkoli pa jih lahko tudi izbriše. Večina uporabnikov ima opcijo kukijev odprto, torej takšno, kot je siceršnja privzeta nastavev operacijskega Windows². Nekateri ocene govorijo o 80% uporabnikov, ki ne preprečujejo inštalacije kukijev, čeprav število uporabnikov, ki to preprečujejo, narašča. Narašča tudi število uporabnikov, ki kukije občasno izbrišejo. V anketi podjetij RIS2000 se je npr. izkazalo, da večina slovenskih podjetij ovira (bodisi preprečuje bodisi briše) uporabo kukijev. Vse to seveda vnaša v uporabnost kukijev velike težave.

Oglejmo si še zapis kukijev na strani strežnika. Kukije dodeljujemo s pomočjo posebnih nastavitvev na strežniku, kar lahko uredimo tudi v okviru standardiziranih programskih paketov za analizo logov (npr. WebTrends). V takem primeru predstavlja kuki dodaten zapis v log datoteki, ki ima lahko obliko **EGSOFT_ID=194.249.10.216-1248502736.29315859** - zapis, ki enolično določa uporabnikov računalnik. S pomočjo dodeljevanja kukijev torej lahko programski paket bistveno bolj natančno oceni število različnih in ponovnih obiskovalcev določene predstavitve, kot če kukiji niso nameščeni.

2.3. Logi, kukiji, računalniki in obiskovalci

V vsakem strežniku, kjer je spletna predstavitev, je mogoče z enostavnimi nastavitvami omogočiti, da se beleži in hrani datoteke logov. Seveda pa so lahko take datoteke nadvse obsežne in tudi nepregledne. V pomoč so nam programi za analizo logov kot npr. Webalizer, Webtrends, Analog, ipd., ki proizvajajo standardizirane rezultate. Nekateri so brezplačni, zahtevnejši programi pa so plačljivi. Različni programski paketi uporabljajo različne kategorije, različna poimenovanja in tudi različne parametre (npr. čas prekinitve med dvema obiskoma, ki pomembno določa število obiskov), kar vse močno otežuje primerjavo rezultatov. Vse to pomembno vpliva na statistično obiskanost.

Posebej kritično pa je vprašanje sklepanja o dejanskih obiskovalcev. Dokler gre samo za analizo zahtevkov – kjer je vsak zahtevk shranjen v svoji vrstici log datoteke – je analiza razmeroma enostavna. Vendar pa zahtevki kot tudi ostale log statistike nimajo jasne interpretacije v smislu obiskovalcev.

² Kukije onemogočimo, če v pregledovalniku, npr. *Internet Explorer*, izberemo *Internet Options*, nadaljujemo s *Security* tab, kliknemo na *Custom Level* - kar vključuje siceršnjo opcijo omejitev nivoju celega Interneta, ki jo izberemo brez posebnih klikov kot cone omejitve (»zone«), saj je to običajna default izbira - nakar izberemo pri obeh opcijah *Cookies*, kategorijo *Disable*.

Že iz dosedanje razprave je mogoče razbrati, da iz log analize ne moremo enostavno sklepati na osebe, to je na število uporabnikov in na njihovo obnašanje. Bistvena izboljšava pri tovrstnem sklepanju je vsekakor uporaba kukijev. Žal prinašajo kukiji – kot smo videli – tudi težave, ki jih podrobneje obravnavamo v nadaljevanju.

Na tem mestu velja poudariti tudi etično spornost kukijev zaradi ogrožanja zasebnosti. Nekateri indici celo govorijo, da bo v bodoče uporaba kukijev v EU celo prepovedana, zato se oglaševalski posredniki že pripravljajo na spletno oglaševanje brez kukijev.

V okviru same log analize je IP številka daleč najbolj stabilna in primerljiva kategorija obiskanosti, saj je neobčutljiva na tehnološke lastnosti strežnika in na nastavitve programov za merjenje obiskanosti, čeprav je res, da ima po drugi strani razmeroma nejasno interpretacijo. Enostavna predpostavka, da vsak IP naslov predstavlja posameznika, je seveda napačna, saj je v Sloveniji 10-krat več uporabnikov kot IP številke.

Ob uporabi kukijev («cookies») je mogoče uporabnike identificirati bistveno bolj natančno kot samo na osnovi IP številke. Še boljše je seveda, če nam uspe doseči, da se obiskovalec ob prihodu na spletne strani prijavi (izpolni obrazec, kjer se identificira z uporabniškim imenom in geslom). Žal je to izredno redko.

Seveda tudi kukiji ne omogočajo enačenja z osebami. Problem kukijev je v splošnem v tem, da precenjujejo število obiskovalcev. Za to sta v grobem dva razloga: po eni strani kukije brišejo uporabniki, po drugi strani pa jih sam spletni strežnik po določenem času (npr. v mesecu dni) inštalira na novo. Seveda prihaja tudi do določenega podcenjevanja, saj lahko isti računalnik/profil uporablja več uporabnikov (npr. v družini). V primerih, ko kukija ni mogoče inštalirati, pa uporabnikov iz iste IP številke sploh ni mogoče ločevati. Končni rezultat je torej razmeroma nejasen, čeprav primerjave kažejo, da kukiji običajno precenjujejo število obiskovalcev.

Predpostavka, da lahko identiteto računalnika izenačimo z identiteto uporabnika, je slaba tudi zaradi uporabe posebnih strežnikov – o čemer smo že govorili - kjer lahko veliko število uporabniških računalnikov dostopa do spletnih strežnikov z enako IP-številko. Problematični so tudi ponudniki klicnega dostopa do Interneta, ki svojim uporabnikom dinamično dodeljujejo IP-številke.

Po drugi strani pa tudi posamezniki obiskujejo spletne strani z več računalnikov, torej z različnih IP-naslovov, kar pomeni, da jih program za analizo logov vsakič evidentira kot različne uporabnike. Zveza med rezultati log-analize in številom oseb, ki so obiskale določeno stran, je torej nadvse zapletena.

Sklepanje o obnašanju oseb je vsekakor osnovni nerešen problem »server-centric« pristopa k merjenju spletne obiskanosti. Podatkov o ogledih (dostopih, prikazih) ali o klikih, ki so osnova oglaševalskega tarifiranja, namreč ni mogoče prevesti v siceršnje in z drugimi mediji primerljive kategorije medijskega planiranja (npr. doseg, rating, GRP), ki so zgrajene na statistiki oseb, to je »user centric« pristopih. Prava rešitev je le, če tudi spletno obiskanost merimo z »user centric« pristopi, ki pa so bistveno dražji.

2.4. Metodološki problemi pri analizi logov

Rezultati log analiz torej zaradi vseh navedenih metodoloških problemov močno odstopajo od dejanske obiskanosti. Navedimo oziroma povzemimo še enkrat nekaj najpomembnejših vzrokov:

Obiskovalci iz tujine

Obisk oseb, ki ne prebivajo v Sloveniji, je precej raznolik. V določeni meri to izraža visoko število različnih IP števil. V Sloveniji je po podatkih RIPE le okoli 25.000 različnih IP števil v okviru domene .si, po nekaterih ocenah je približno dodatna tretjina hostov registriranih še pod domenami .com, .net, .org. Strani z veliko tujega obiska, kot so na primer RTVSLO, Rumeni Internet, Matkurja ali RIS, so v telefonskih anketah podcenjene, saj tu ni upoštevan tudi obisk iz tujine. V log analizi pa so navedene strani seveda precenjene, kadar govorimo o obiskovalcih iz Slovenije. Ločevanje tujega obiska namreč ni enostavno.

Intenzivnejše obiskovanje določene strani

Delež obiskovalcev, ki se pogosto vračajo, je na nekaterih straneh večji, drugod večino predstavljajo samo enkratni obiskovalci. Tako npr. 10.000 obiskovalcev, ki obišejo stran 10-krat v mesecu, generira 10-krat več dostopov in obiskov kot 10.000 obiskovalcev, ki obišejo stran enkrat v mesecu, čeprav je v smislu različnih obiskovalcev obiskanost enaka. Pri analizi logov je – kljub uporabi kukijev – namreč težko ločevati različne obiskovalce. Posebej kritične so strani registriranih klepetalnic (npr. Interblod), kjer nekaj tisoč uporabnikov generira visoko število ponavljajočih obiskov.

Obiskovanje robotov, plazilcev, agentov

»Inteligentni agenti« ali »roboti« obiskujejo spletne predstavitve vsako uro in tako povečujejo število obiskov. Variacije »inteligentnih agentov« so t. i. pajki (spiders) in plazilci (crawlers), programi, ki obišejo skorajda vsako stran v spletu zato, da bi ustvarili indekse spletnih strani za potrebe iskalnikov (Altavista, Infoseek...). Ti programi so osredotočeni na iskanje in analizo tekstovnih vsebin, ki so pomembne za klasifikacijo strani in ustrezen vpis v indekse iskalnikov. Standardizirana programska orodja (npr. WebTrends) za analizo logov prepoznajo večino agentov in izločijo zahteve, ki so jih sprožili tovrstni programi.

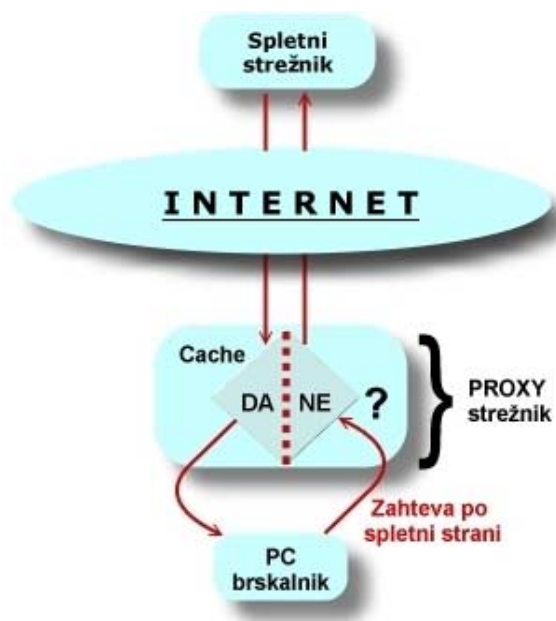
Posebna težava so »prirejeni roboti« (rogue bots) - programi za namerno povečevanje števila obiskov, programiranih za varanje kompleksnejših števcov, ki npr. evidentirajo vpogled oglasa le takrat, ko odjemalčev osebni računalnik ob priklicu strani strežniku sporoči svoj predhodni URL (s katere spletne strani je prišel) in IP številko računalnika - odjemalca. Prirejeni robot v tem primeru pošlje strežniku namišljeni URL in lažno IP številko in se na ta način predstavi kot veljavni obiskovalec. Obseg tovrstnega delovanja v Sloveniji ni poznan.

“Proxy servers” in “browser cash”

Kadar npr. oseba iz Slovenije dostopa do določene strani (npr. v ZDA) prek proxy strežnika, se ta stran shrani tudi v poseben »proxy« računalnik njegove organizacije ali ponudnika dostopa (v Sloveniji). Ob naslednjem obisku katerkoli osebe, ki uporablja isti proxy, se bo ta strani, če se medtem ni spremenila, posredovala kar iz omenjenega proxy strežnika (v Sloveniji) in ne več iz originalnega strežnika (v ZDA). S tem se prihranijo stroški mednarodnega prometa in razbremenijo komunikacije. Istočasno pa osnovni strežnik z vsebino (v ZDA) seveda ne more vedeti za ponovni obisk iz Slovenije. Dodati velja, da novejšie programske verzije proxy strežnikov omogočajo odpravo te pomanjkljivosti, vendar pa gre vseeno za osamljene pojave (večina je starejših proxyev, novejši pa te opcije nimajo nastavljene kot privzeto).

"Proxy" strežnik nadzira vse zahteve lokalnih odjemalcev do vseh zunanjih spletnih strežnikov in preverja, ali lahko določeni zahtevi zadosti že sam. Šele ko ugotovi, da zahtevanih vsebin nima v svojem pomnilniku (»cache«), zahtevo posreduje na izvorni spletni strežnik.

Graf 1: Shematska predstavitev delovanja proxy strežnika



Tehnično zahtevnejše strani (cgi, pl, asp) praviloma ne dovoljujejo shranjevanja v začasne pomnilnike ali pa je tovrstno shranjevanje omejeno. Enostavnejše predstavitve (htm, html) zato beležijo bistveno manj dostopov neposredno na samem strežniku.

Na splošno velja, da je aktivnost, ki je merjena na podlagi analize logov, zaradi vpliva »proxy« strežnikov podcenjena. Analize v primeru strani RIS kažejo, da je pri enostavnih HTML predstavitvah na vsakih 100 zahtevkov opravljenih še dodatnih 60 zahtevkov, ki jih logi na strežniku ne zabeležijo, ker se posredujejo prek različnih posrednikov. Ocena se ujema z ocenami vodilnih raziskovalnih organizacij v svetu: MatchLogic na primer ocenjujejo, da znaša promet, ki ga prikrijejo »proxy« strežniki, kar 78%.

Večje število uporabnikov na računalnik

V nekaterih primerih dostopa iz posameznega računalnika (IP številka) veliko različnih uporabnikov (npr. Univerza), kar vpliva tudi na obiskanost posameznih strani, še posebej, če so te obravnavane kot začetna oziroma privzeta spletna stran (»start page«) v vseh računalnikih. Če se npr. pri vsakem zagonu spletnega pregledovalnika (browserja) najprej in avtomatično obiše stran Univerze, čeprav želi uporabnik obiskati povsem neko drugo stran, potem bo imela stran Univerze umetno povečan obisk. Opisani problem je lahko aktualen pri vseh organizacijah, kjer se programska oprema instalira centralno. V določeno pomoč pri tem problemu so lahko kukiji.

Vpliv dinamičnih IP števil (modemski dostop)

Večje število uporabnikov nastopa tudi v primeru klicnega dostopa (dial-up), kjer določen modem v mesecu dni uporabijo številni uporabniki. Tako lahko en uporabnik v mesecu dni dostopa iz 30 klicnih modemov in torej generira 30 različnih IP števil. Pri večjih ponudnikih modemskega dostopa pa obstaja dodatna nevarnost, da uporabniki v telefonskih in spletnih anketah navajajo obiskovanje strani, kot so npr. Arnes ali Siol, čeprav do spletnih strani niso dostopali, ampak so se prek navedenega ponudnika le logirali na Internet.

Strežniki večjih sistemov

Večje organizacije (npr. sigov, n-lb) imajo posebne sisteme zaščite in nastavitve. Za enim samim strežnikom se tako lahko skriva več sto računalnikov in več tisoč uporabnikov. Pri tem običajno predpostavljamo, da so tretje predstavitev strani zaradi tega v približno enakem položaju: npr. strani RIS.org so v pogledu podcenjenega števila IP števil in različnih uporabnikov v enakem položaju kot strani Uni-lj.si, saj predpostavljamo, da uporabniki, ki dostopajo iz večjih strežnikov, enakomerno obiskujejo strani in torej strežniki prikrivajo obiskanost (v smislu IP števil) pri vseh straneh v enakem obsegu. Ob tej predpostavki torej lahko sklepamo, da so relativne primerjave med spletnimi predstavitvami dopustne, medtem ko so absolutne statistike (na nivoju IP števil) gotovo podcenjene.

Druge aktivnosti

Na strežnikih lahko potekajo specifične aktivnosti in posebnosti pri nastavitvah za beleženje logov z nepredvidljivim učinkom na statistiko logov:

- Znotraj same strani lahko potekajo avtomatična preusmerjanja, ki povečujejo obisk določenih strani znotraj predstavitve.
- Avtomatska obveščanja, ki potekajo tako, da se ob vsakem ažuriranju strani avtomatično obnovi tudi zadnji update pri uporabnikih, kar pomeni, da se registrira dostop do strani vseh naročnikov, ne da bi oni sami to vedeli ali sprožili. Program namreč sproži tak obisk avtomatično.
- Določena »pinganja« ali aktivnosti robotov, ki nekatere strani obiskujejo pogosteje kot druge ipd.

3. Potvarjanje log datotek

Načeloma je seveda mogoče, da si predstavitvene strani umetno večajo obiskanost. Na voljo je več načinov:

- a) Najbolj enostavna je izdelava programa, ki v določenih intervalih generira obisk, kar zmore izdelati praktično vsak računalniško orientiran uporabnik. Seveda je pri tem razvidno, da prihaja obisk iz enega in istega računalnika, na nivoju samega štetja obiskov pa to ustvarja vtis velike obiskanosti.
- b) Razmeroma enostavno je tudi umetno večanje obiska na osnovi sistema naročnikov. Tako npr. projekt RIS generira velik obisk vsakič, ko sporoči svojo novico v sistem dot.si. Vendar ne glede na to, ali se sporoči novica ali ne, pa se vsakih nekaj minut nekaj tisoč uporabnikov (ki so takrat priključeni v internet in imajo vključen program za spremljanje dot.si) sproži zahtevek za posebno datoteko, ki je na RISovem spletnem strežniku. Ta datoteka vsebuje podatke o novostih in novicah. Slednje močno vpliva na povečanje tako števila zahtevkov kot tudi ostalih statistik.
- c) Zgoraj omenjena načina je mogoče z natančnejšo analizo logov seveda izločiti. Težje pa je to v primeru, ko uporabnik izdela program za avtomatski dial-up dostop npr. iz Arnesa ali Siola. Pri tem vsakič dobi drugačno IP številko in če ob tem vsakič izbriše še kukije ter v intervale dostopanja in poti obiskovanja vnese komponento slučajnosti, potem so njegovi obiski praktično neugotovljivi. V slovenskih razmerah, kjer npr. 200 obiskovalcev dnevno že pomeni pomembno razliko v obiskanosti, za večino strani zadošča že en sam računalnik, ki bi neprekinjeno generiral take obiske.
- d) Najtežja oblika poneverjanja je celovito generiranje obiskov, ki poneverja prav vse informacije, od IP številke do kukija. Slednje je nekoliko bolj zahtevno in ga lahko izvede le bolje podkovan uporabnik. Če je izvedeno kvalitetno, je tako poneverjanje izredno težko ugotovljivo.
- e) Poleg zgoraj naštetih večanj pa ostajajo še drugi načini, ki so na meji lažnega/umetnega povečevanja obiskanosti:
 - Intenzivno nagovarjanje obiskovalcev ali vsaj ožjih sodelavcev, da stran inštalirajo kot začetno stran, s čimer se sproži obisk pri vsakem nalaganju pregledovalnika.
 - Vnašanje ključnih besed v predstavitev (npr. meta elementi, opisi slik), ki v resnici niso povezane z njo, npr. sex, s čimer prihajajo obiskovalci, ki na osnovi te besede izvajajo iskanje prek portalov in drugih iskalnih mehanizmov.
- f) Umetno obiskanost povzroča tudi poudarjeno in pretirano informiranje o določeni strani na portalih, s čimer je mogoče v Sloveniji generirati več sto obiskov dnevno. V veliki meri pa so lahko takšni obiskovalci zavedeni z naslovom in stran po kliku takoj zapustijo. Tipičen primer je navajanja »novic« na določeni strani na portalih, kot je npr. Matkurja, Slowwwenia, Eon, kjer morda sploh ne gre za novico, ampak le za pojavljanje starih vsebin

pod novim, umetnim naslovom strani. Na tej osnovi je mogoče povečati obisk za več sto seans (sessions) dnevno, kar je seveda daleč najcenejša promocija. V grobem namreč velja upoštevati, da je klik na spletno stran vreden okoli 100 SIT., kolikor bi morali plačati oglaševalskim posrednikom.

g) Poleg navedenih načinov je mogoče o obiskanosti zavajati tudi povsem neposredno z zavajajočim poročanjem:

- z enostavnim navajanjem napačnih rezultatov,
- s terminološkim manipuliranjem, t. j. z navajanjem statistik, ki so po obsegu velike (npr. število zahtevkov), nimajo pa nobene interpretacije v smislu števila obiskovalcev;
- posebej tipično je navajanje števila obiskov in ga razglašati kot število različnih obiskovalcev;
- z zamolčanjem določene tehnološke specifičnosti ali določene posebnosti pri nastavitvah, npr. zamolčanje podatka, s kakšnim intervalom se inštalirajo in obnovijo kukiji ali parametri (npr. dolžina prekinitve med dvema obiskoma), ki so bili uporabljeni, ko se je izvedla analiza logov z določenim programom.
- s poneverjanjem same log datoteke, kadar jo analizira nek neodvisni vir.

II. RIS ANALIZA LOGOV

1. Metodologija analize logov projekta RIS

V sodelovanju z Laboratorijem za Telekomunikacije (Fakulteta za Elektrotehniko) smo v mesecu januarju 2000 zbrali in s standardiziranimi programi (WebTrends, Analog) analizirali log datoteke večine spletnih predstavitev, ki mesečno dosežajo vsaj 5% obiskanost aktivnih uporabnikov Interneta (torej tistih, ki uporabljajo Internet vsaj mesečno), kar v absolutnem smislu pomeni vsaj 12.500 različnih obiskovalcev od skupno 250.000 - 300.000, kolikor obsega po ocenah RIS (na podlagi telefonskih anket) v tem času domači Internet avditorij.

V okviru projekta RIS smo s standardiziranimi programskimi orodji (WebTrends, Analog) analizirali log datoteke posameznih spletnih strani za določeno časovno obdobje (pri večini predstavitev za mesec januar 2000) in razbrali akumulirane statistike obiskanosti.

Pri analizah smo se omejili na naslednje kategorije obiskanosti:

- *število dostopov na prvo stran,*
- *število obiskov na spletni predstavitvi,*
- *število različnih obiskovalcev na predstavitvi,*
- *število različnih hostov (IP-številka), ki so »dostopale« do merjenih strežnikov.*

2. Rezultati analize logov projekta RIS

Spletna predstavitev	WebTrends			Analog
	Št. dostopov za prvo stran (page hits)	Št. obiskov za celo predstavitev (user sessions)	Št. različnih obiskovalcev (unique users)	Distinct hosts serverd (različni hosti)
APP	35864	208582	8285	8303
Arnes	148359	167623	27482	27344
Avto.net *	37044	47183	10696	10629
Borza	21817	30382	9156	9156
Boter *	3073	2528	1054	1047
COBISS	40774	46569	13509	14813
CVI	15570	56024	16321	28794
Delo **	8707	11279	5653	5638
Dnevnik	66609	68512	15848	15569
DZS ***	12581	27174	9204	9126
Eon *	185445	151218	14810	12560
HMZ *	77024	92672	17507	17309
Infomediji	5515	7201	3439	3408
Jerovšek	35833	27188	7874	7712
Kompas	31026	15029	5524	8598
Lj Kino	10376	11648	4555	4630
mail-kiss	1810	1424	773	774
MKZ	/	4449	2229	2278
MobiSux	45764	64108	10457	10382
Mobitel	304691	197170	20181	20506
NKMB	185468	144448	12777	12560
Pasadena	/	4424	2638	2623
RCUL	/	/	8905	4502
RCUM	34265	68231	21829	21098
RIS	4005	19975	6068	5966
RTVSlo - netra	22823	35276	10836	10567
RTVSlo - realsrv	59	23951	10939	10852
Rumeni Internet	14150	32090	14819	14422
Salomon	29828	40613	9882	9882
Siol ****	1095496	398285	13257	13243
SloKnjiga *****	713	2892	1737	1737
SloWWWenia (Cookies)	389032	278545	70049	15610
SloWWWenia	389032	241341	19219	15610
ŠOU	1517	4941	2211	2212
Telekom *****	/	/	10678	10678
TIS	/	135226	25277	21532
www-kissF	8309	39440	22385	22415
ZRSZ	10370	13157	5081	5031

Tabela 1: Rezultati log analize – OPOMBE NA NASLEDNJI STRANI

OPOMBE

- * Podatki za februar 2000, Avto.net, Boter, Eon, HMZ
- ** Podatki za časovno obdobje (18.4. - 30.04.) Delo
- *** Podatki za časovno obdobje (11.3. - 31.3.) DZS
- **** Podatki za časovno obdobje (21.3. - 20.4.) Siol
- ***** Podatki za časovno obdobje (3.4. - 11.4.) Slo Knjiga
- ***** Podatki za marec 2000; Telekom

V tabeli so zbrane analize vseh merjenih strani, kjer smo s programom Webtrends prešteli število dostopov na prvo stran, število obiskov na celotni predstavitvi in število različnih obiskovalcev na celotni predstavitvi. S programom Analog smo določili število različnih odjemalcev (IP števil), ki so dostopali do analiziranega strežnika.

Treba je poudariti, da primerjave med spletnimi stranmi na podlagi analize logov niso enostavne. Nasprotno, pogosto so obremenjene z napakami, ki izhajajo iz številnih tehničnih nastavitvev na strežnikih, kar v splošnem preprečuje neposredno primerjanje med predstavitvami. Za primerjavo med predstavitvami zato priporočamo predvsem rezultate telefonskih anket.

Pri straneh, za katere v Tabeli 1 ni podatka za število dostopov na prvi strani, je problem v tem, da ni bilo mogoče definirati vhodne (prve) strani spletne predstavitve, npr. vhodna stran Telekoma Slovenije (www.telekom.si je asp (active server page) stran <http://www.telekom.si/Enter.slo.3.asp?stran=uvod>, ki je netipična in jo program za analizo ne prepozna). Pri straneh, kjer ni prešteti obiskov (user sessions), program za analizo (WebTrends) zaradi neznanih vzrokov ni izračunal ustrezne statistike.

Poudariti velja, da je število IP števil (Analog - distinct hosts served) zelo podobno številu različnih obiskovalcev na predstavitvi (WebTrends - unique users), kadar so slednji prešteti brez kukijev. Predstavljeni rezultati ne upoštevajo kukijev, zato se število različnih uporabnikov približa številu različnih IP števil in ga presega kvečjemu za 20%.

V primeru specifične rabe kukijev pa je nadvse ilustrativna spletna predstavitev Slowwwenia.com, kjer število različnih obiskovalcev naraste od 19.219 brez kukijev na 70.049 s kukiji. Ob tem velja upoštevati, da je v tem času mesečni doseg strani Slowwwenia v telefonskih anketah znašal okoli pod 50.000.

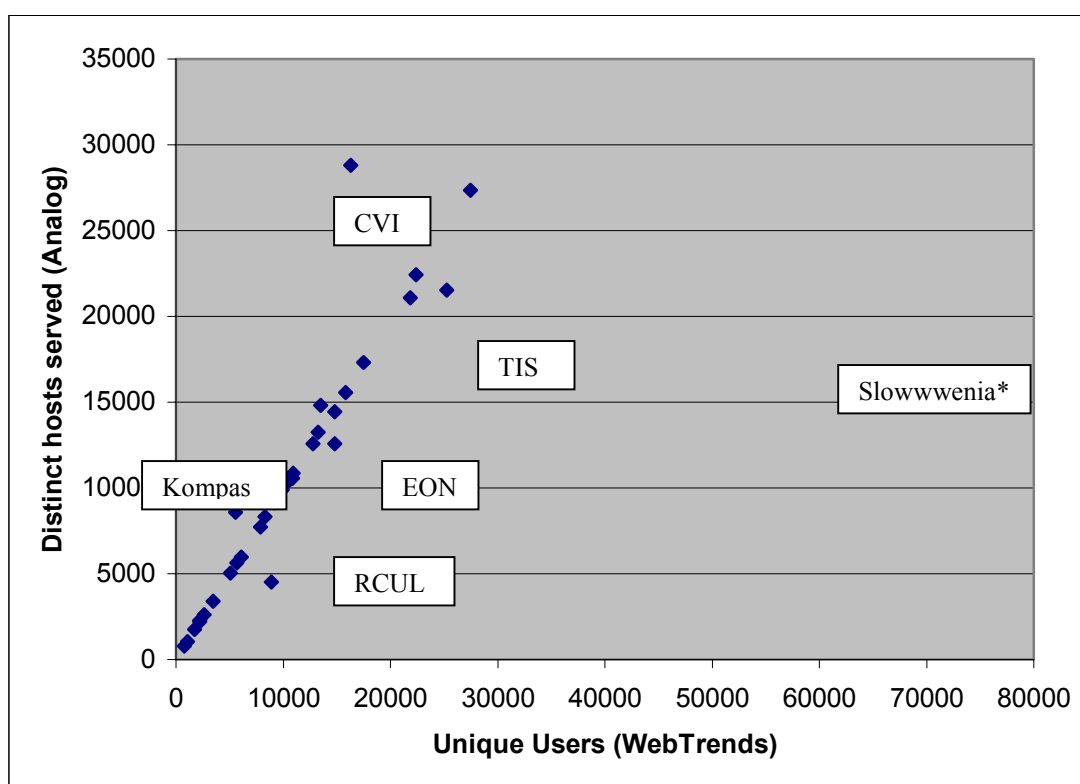
3. Razmerja log statistik

V nadaljevanju predstavljamo nekatera razmerja med indikatorji obiskanosti.

3. 1. IP in različni obiskovalci

V spodnjem pregledu vidimo, da skoraj ni razlik med številom različnih obiskovalcev in številom različnih IP števil (gostiteljskih strežnikov), kadar pri izračunu različnih obiskovalcev ne upoštevamo dodeljevanja kukijev.

Graf 2: Primerjava št. Različnih obiskovalcev (Unique Users) in št. različnih strežnikov (distinct hostes)



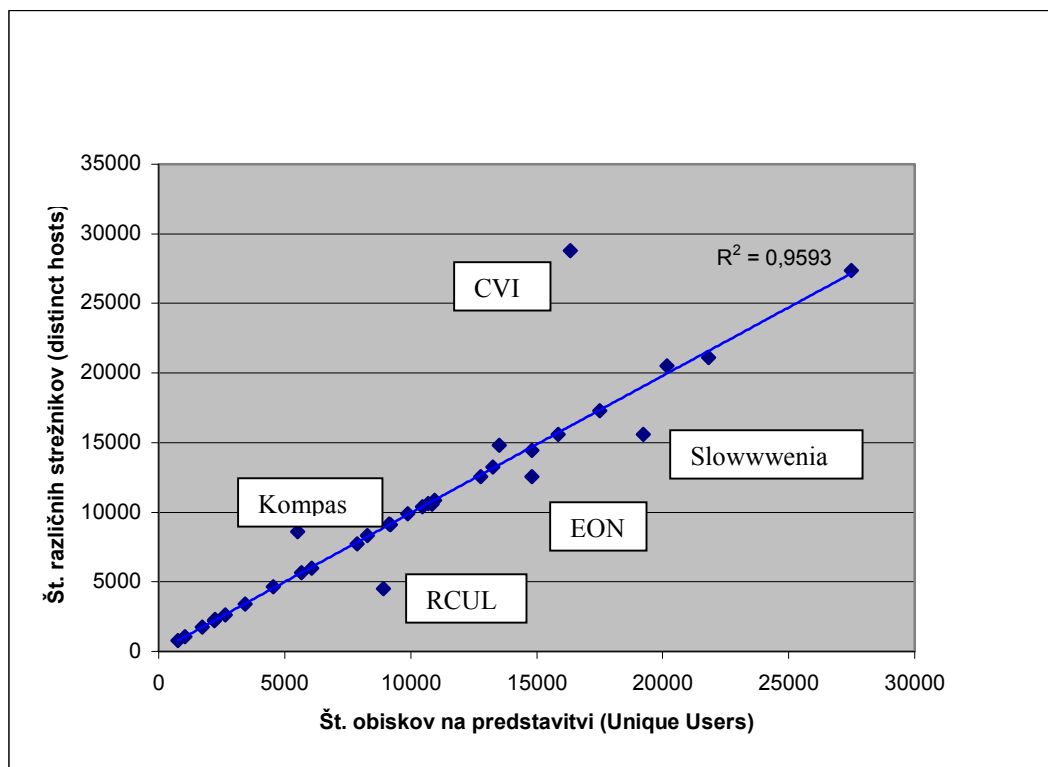
Gornja primerjava pokaže, da razen Slowwwenie (kjer so bili uporabljeni kukiji) spletne predstavitve skoraj ne odstopajo od linearne povezanosti.

Kot rečeno, poseben problem kukijev je v tem, da jih spletni strežnik dodeljuje spletnim brskalnikom, torej lahko vsak posameznik dobi toliko kukijev, kolikor različnih računalnikov in kolikor različnih brskalnikov uporablja pri dostopanju do posamezne spletne predstavitve. Po drugi strani lahko z enakim kukijem dostopa prek istega brskalnika več različnih obiskovalcev.

Ocenjevanje števila različnih uporabnikov na podlagi dodeljenih kukijev se še dodatno zaplete zaradi dejstva, da lahko uporabnik kukije izbriše ali onemogoči dodeljevanje teh na svojem računalniku.

V spodnjem prikazu smo odstranili odstopajočo Slowwwenio, kjer je število različnih obiskovalcev izračunano s kukiji in jo nadomestili z meritvijo brez kukijev. Stopnja linearne povezanosti je v tem primeru 96%.

Graf 3: Različni obiskovalci (Unique Users) in različni strežniki (Distinct Hosts)



Najizraziteje odstopa predstavitev Centra vlade za informatiko (CVI), kar lahko pojasnimo s pomanjkljivostjo podatkov, saj smo uspeli pridobiti samo četrtno – to je samo vsak četrto vrstico log datoteke - zapisov (log datoteke) za mesec januar.

V drugo stran nekoliko odstopajo predstavitve RCUL-a, Slowwwenie (brez kukijev) in TIS-a, za katere je značilno znatnejše število pogostih uporabnikov. V primeru RCUL-a fakultetne računalnike uporablja veliko študentov, kar pomeni, da ista IP številka pogosto dostopa do strani RCUL (Računalniški Center Univerze v Ljubljani), ki je pogosto osnovna stran na spletnem brskalniku pri teh računalnikih. Ker program WebTrends upošteva interval 30 minut (zaradi dinamičnega dodeljevanja IP številke s strani ponudnikov dostopa do Interneta pri modemskega (dial-up) dostopu), se obisk iste IP številke izven omenjenega intervala zapiše kot nov obisk (user session), kar vpliva tudi na število različnih obiskovalcev. Nekatere IP številke (uporabniki) pogosteje uporabljajo telefonski imenik, zato so učinki podobni.

3. 2. Dostopi na prvo strani in različni obiskovalci

V nadaljevanju bo oznaka * pri Slowwwenii pomenila analizo s kukiji, odsotnost zvezdice pa, da smo analizo izvedli brez kukijev.

Tabela 2: Razmerje - dostopi na HP / število različnih obiskovalcev

Spletna predstavitev	Število dostopov / število različnih uporabnikov
APP	4,3
Arnes	5,4
Avto.net	3,5
Borza	2,4
Boter	2,9
COBISS	3,0
CVI	1,0
Delo	1,5
Dnevnik	4,2
DZS	1,4
Eon	12,5
HMZ	4,4
Infomediji	1,6
Jerovšek	4,6
Kabi	3,2
Kompas	5,6
Lj Kino	2,3
mail-kiss	2,3
MobiSux	4,4
Mobitel	15,1
NKMB	14,5
Rumeni Internet	1,0
Salomon	3,0
Siol	82,6
SloKnjiga	0,4
SloWWWenia*	5,6
SloWWWenia	20,2

Različni obiskovalci predstavitve APP so npr. v mesecu januarju v povprečju obiskali prvo stran predstavitve Agencije za plačilni promet 4,3-krat.

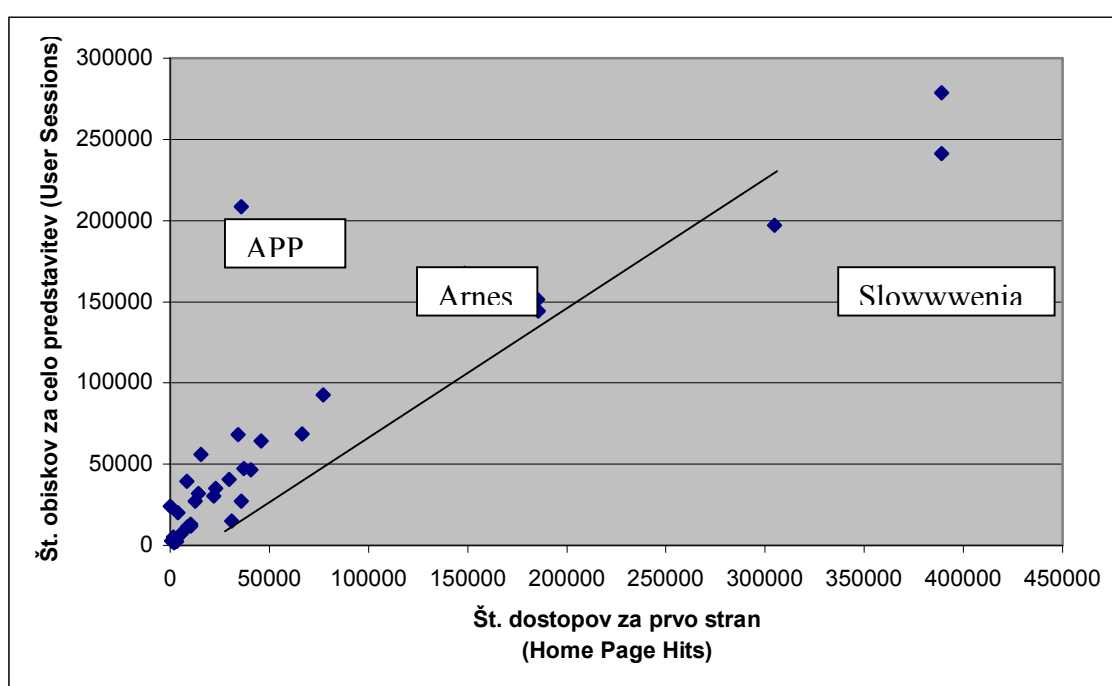
Gornja preglednica dobro odraža tudi vsebinsko strukturiranost spletnih predstavitev, saj so frekvence obiskov različnih obiskovalcev na prvih straneh zelo različne. Če je vstopna stran obenem tudi navigacija do večine podstrani, kot je to v navadi pri spletnih portalih (Slowwwenia, Eon), potem so frekvence obiskov na prvi strani bistveno višje, kot npr. pri predstavitev, na katerih je na prvi strani le uvodni logotip (intro) in povezava na naslednji nivo (npr. APP).

Med pogosteje obiskanimi sta tudi vhodni strani predstavitev Mobitela in Nove kreditne banke Maribor. Posebnost je Siol, ki kot ponudnik dostopa »prisiljuje« uporabnike, ki prek njega dostopajo do Interneta, da obišejo njihovo vhodno stran (odtod zelo visoka frekvenca 82,6).

3. 3. Dostopi na prvo strani in število obiskov

Podobno informativna je v tem pogledu tudi frekvenca obiskov prve strani v okviru ene uporabniške seje (obiska spletne predstavitev), ki pove, kolikokrat se v povprečju uporabniki vrnejo na prvo stran v okviru enega obiska (uporabniške seje).

Graf 4: Primerjava št. dostopov na prvi strani (Home Page Hits) s številom obiskov na celi predstavitvi (Users sessions)



Razvidno je, da se večina spletnih predstavitev nahaja sorazmerno blizu premice, ki predstavlja razmerje (1:1) - en obisk prve strani v okviru enega obiska spletne predstavitve. Tipičen primer je Arnes. Primera odstopanj sta Agencija za plačilni promet, kjer je frekvenca obiskov bistveno nižja od 1 in Slowwwenia, kjer uporabniki v okviru obiska spletne predstavitve pogosteje obiskujejo tudi prvo (vstopno) stran, kar pa je lahko tudi rezultat preusmerjanja na sami strani.

Zgornji prikaz predstavljamo tudi s tabelo:

Tabela 3: Frekvence obiskov na prvi strani v okviru obiska spletne predstavitev

Spletna predstavitev	Frekvenca obiskov na prvi strani v okviru obiska spletne predstavitev
APP	0,2
Arnes	0,9
Avto.net	0,8
Borza	0,7
Boter	1,2
COBISS	0,9
CVI	0,3
Delo	0,8
Dnevnik	1,0
DZS	0,5
Eon	1,2
HMZ	0,8
Infomediji	0,8
Jerovšek	1,3
Kompas	2,1
Lj Kino	0,9
mail-kiss	1,3
MobiSux	0,7
Mobitel	1,5
NKMB	1,3
RCUM	0,5
RIS	0,2
RTVSlo - netra	0,6
RTVSlo - realsrv	0,0
Rumeni Internet	0,4
Salomon	0,7
SloKnjiga	0,2
SloWWWenia *	1,4
SloWWWenia	1,6
ŠOU	0,3
www-kiss	0,2
ZRSZ	0,8

V povprečju očitno samo vsak peti obisk na strani APP vključuje tudi vhodno stran, torej večina obiskovalcev dostopa neposredno do storitev, ki so na podstraneh APP.

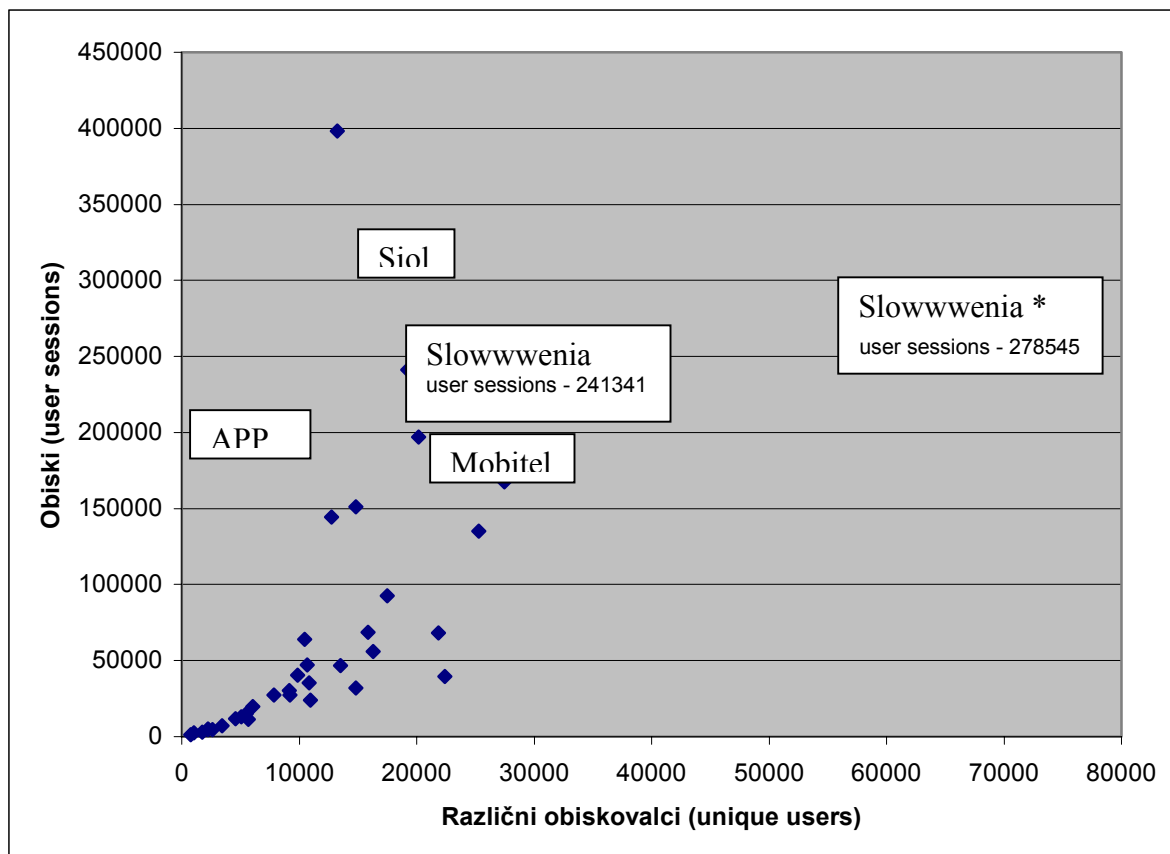
Nasprotno je pri predstavitvi Slowwwenie: vsak obisk v povprečju vključuje okoli 1,5 obiska prve strani.

Najvišje razmerje je razvidno pri Kompas, kjer povprečen obisk spletne predstavitev zajema dva obiska prve strani. Natančnejša analiza Kompasove spletne predstavitev pokaže, da večina obiskov oziroma sej (81%) vključuje tudi prvo stran.

3. 4. Število obiskov in število različnih obiskovalcev

Nad vse pomemben indikator pogostosti obiskovanja spletnih predstavitev s strani različnih obiskovalcev daje tudi spodnji prikaz. Razvidno je, da dodeljevanje kukijev pri Slowwwenia.com povzroča tudi razliko na nivoju števila obiskov (in ne le različnih obiskovalcev, kot smo že ugotovili).

Graf 5: Število obiskov (users sessions) in število različnih obiskovalcev (unique users)



Zgornji graf, ki kaže razpršenost in razporeditev v razsevnem grafikonu, je bolj informativen v tabelarnem prikazu, kjer je podano razmerje med obiski in obiskovalci, razvidna pa je tudi frekvenca dostopov na predstavitve s strani različnih obiskovalcev.

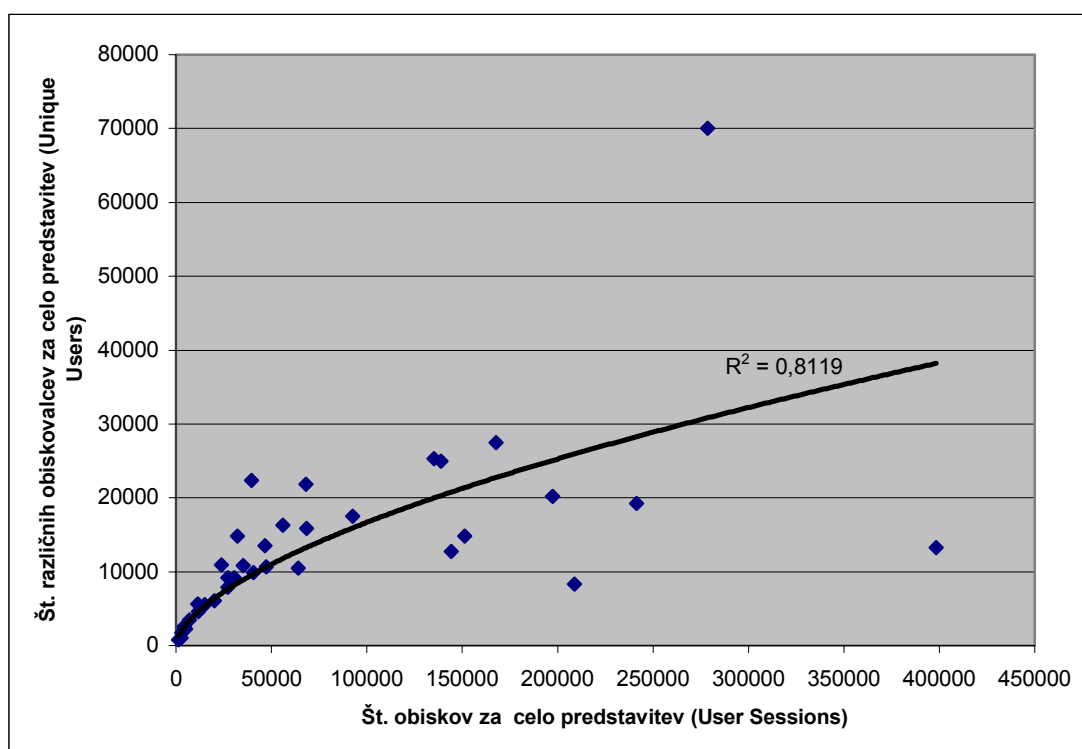
Tabela 5: Frekvenca dostopov različnih obiskovalcev na predstavitev

Spletna predstavitev	Obiski (User Sessions) / Različni obiskovalci (Unique Users)
APP	25,2
Arnes	6,1
Avto.net	4,4
Borza	3,3
Boter	2,4
COBISS	3,4
CVI	3,4
Delo	2,0
Dnevnik	4,3
DZS	3,0
Eon	10,2
HMZ	5,3
Infomediji	2,1
Jerovšek	3,5
Kabi	5,6
Kompas	2,7
Lj Kino	2,6
mail-kiss	1,8
MobiSux	6,1
Mobitel	9,8
NKMB	11,3
RCUM	3,1
RIS	3,3
RTVSlo - netra	3,3
RTVSlo - realsrv	2,2
Rumeni Internet	2,2
Salomon	4,1
Siol	30,0
SloKnjiga	1,7
SloWWWenia *	4,0
SloWWWenia	12,6
ŠOU	2,2
www-kiss	1,8
ZRSZ	2,6

Povprečni obiskovalec (unique user) predstavitve APP je v mesecu januarju obiskal predstavitev Agencije za plačilni promet (katerokoli stran v okviru predstavitve APP) 25,2-krat.

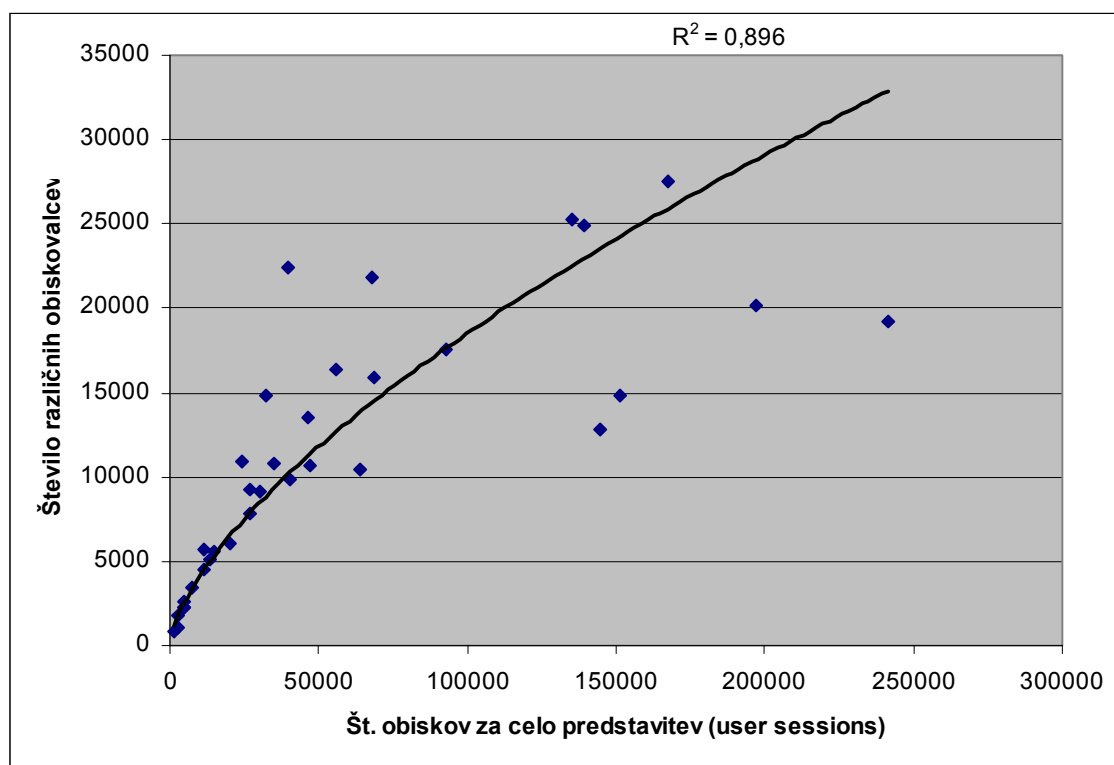
Ponovno lahko vidimo, da imajo visoke vrednosti spletni portali Slowwwenia, Eon in Mobitel, medtem ko Siol dosega visoko frekvenco zaradi uporabnikov, ki preko modemskega dostopa najprej naložijo Siolovo stran in s tem sprožijo obisk (uporabniško sejo), čeprav strani v resnici ne pregledujejo. Gornja razpredelnica je tudi dober prikaz »lojalnosti« občinstva, kjer po drugi strani lahko vidimo, da npr. povprečni obiskovalec Kiss-a mesečno obiše manj kot dve izmed strani v okviru Kiss-ovega strežnika, kjer so številne predstavitvene strani študentov, na katerih zagotovo ni rednega in zvestega občinstva.

Graf 6: Primerjava obiskov (Users Sessions) in različnih obiskovalcev (Unique Users)



S potenčno regresijsko krivuljo tudi pokažemo, da obstaja med pojavoma znatna povezanost, saj z regresijsko enačbo pojasnimo 81% delež variance. Če odvezemo še tri najbolj izstopajoče enote (Siol, Slowwwenia s kukiji in APP), dobimo spodnji prikaz:

Graf 7: Primerjava obiskov (user sessions) in različnih obiskovalcev (unique users)



Poudariti velja, da gre pri različnih obiskovalcih praktično za različne IP številke, saj so navedbe brez uporabe kukijev. Razberemo lahko tudi različne skupine strani, saj se na levi strani nakazuje skupina z razmerjem 1: 3, na desni pa skupina z razmerjem, ki je bližje 1: 2 oz. znaša celo manj.

III. ANALIZA OBISKANOSTI - TELEFONSKE ANKETE

Statistika logov vsekakor nudi koristne informacije o obiskanosti, vendar je mogoče zaenkrat zaradi vseh navedenih težav in specifičnosti, ki spremljajo merjenje spletne obiskanosti, primerjati obiskanost le na osnovi anket.

Pomembno izboljšavo pri analizi logov lahko predstavlja on-line beleženje obiskanosti (prestrežanje logov) in zahtevnejše metode beleženja obiska (auditing). Tovrstne metode merjenja obiskanosti pa se lahko razvijejo šele, ko je tržišče (spletni založniki, oglaševalci) pripravljeno tudi finančno podpreti razvoj ali nakup ustreznih merskih inštrumentov, metodologije in izvajalcev.

Glede na obseg spletnega oglaševanja, ki npr. zaostaja za oglaševanjem na TV za v grobem 100-krat, v Sloveniji – ob težavah financiranja telemetrije - ni pričakovati PC-metrije, saj je izredno draga. Pri PC-metriji ali Netmetriji se namreč vzorcu uporabnikov inštalira v njihov osebni računalnik posebnem program, pri čemer so potrebni – zaradi fragmentiranosti spletne obiskanosti (veliko malo obiskanih strani) – veliki vzorci.

V odsotnosti PC-metrije, kjer se reprezentativnemu vzorcu uporabnikov Interneta namesti v računalnik program za spremljanje obiskanosti, so za tovrstne primerjave pri nas na voljo predvsem telefonske ankete na osnovi priklica.

1. Metodologija

Merjenje obiskanosti na osnovi priklica je podobno merjenju gledanosti TV na osnovi telefonskega anketiranja (Day-after recall, DAR), z vsemi prednostmi in slabostmi.

1.1 Anketa kot merski instrument

Ocene izhajajo iz telefonskih anket RIS2000. Podatki iz drugih anket, predvsem WWW ankete (priklic in ocenjevanje), so v celoti objavljeni na Internetu <http://www.ris.org/www97/obisk98.html>, enako kot podrobnosti glede načina vključevanja strani v postopek ocenjevanja <http://www.ris.org/si-splet/index.html>.

a) Merjenje obiskanosti

V posamezno meritev je običajno vključenih 350 mesečnih uporabnikov Interneta. V primeru večkratnih meritev za določeno stran je vzorec - in natančnost - seveda večji. V merjenju je bilo skupno vključenih že prek 100 strani, ki so bile evidentirane kot najbolj obiskane v prejšnjih anketah ter nekatere strani, ki so bile vključene na osnovi kriterijev Si-spleta (<http://www.ris.org/si-splet/>).

Respondenti so bili v splošni in standardizirani telefonski anketi najprej vprašani o uporabi Interneta. Aktivni (mesečni) uporabniki so nato odgovarjali na vprašanje, kako pogosto obiskujejo našete WWW strani.

Možni odgovori so bili: *ne poznam, slišal sem za stran, obiskal sem samo enkrat, obiskujem občasno, obiskujem mesečno, obiskujem tedensko, obiskujem dnevno.*

Respondenti običajno ocenjujejo blok okoli deset do dvajset strani. Vsaka stran je navedena z njenim polnim nazivom, v nekaterih primerih pa je navedeno še dodatno pojasnilo.

Na tem mestu velja ponoviti, da se navedene ocene seveda razlikujejo - pravzaprav se dopolnjujejo - od ocen na osnovi analize logov, s katerimi razpolagajo same spletne strani.

Posebej pomembna kategorija so osebe, ki so izjavile, da določeno stran obiskujejo mesečno in pogosteje. Obstaja seveda majhna razlika med mesečnimi obiskovalci določene strani in osebami, ki so to stran obiskale v prejšnjem mesecu. Določene osebe, ki stran obiskujejo redkeje, lahko obisk namreč izvedejo, po drugi strani pa oseba, ki sicer dostopa mesečno, v določenem mesecu strani ni obiskala. Ker pa so navedene razlike zanemarljive, lahko mesečne obiskovalce praktično enačimo z osebami, ki v mesecu dni obišejo določeno stran.

Kadar mesečne obiskovalce izrazimo v odstotku aktivnih uporabnikov Interneta – to so tisti, ki izjavljajo, da Internet uporabljajo mesečno – govorimo o mesečnem dosegu ali ratingu določene strani. Mesečni doseg je v množici indikatorjev povsod po svetu uveljavljena osnovna mera za primerjavo med stranmi. Govorimo torej o odstotku aktivnih uporabnikov Interneta, ki so v določenem mesecu obiskali obravnavano spletno predstavitev, ali drugače, o mesečnem dosegu strani med aktivnimi uporabniki.

Seveda so strani, ki imajo veliko dnevni in tedenskih uporabnikov s tem nekoliko prizadete, saj imajo ob istem številu različnih mesečnih obiskovalcev bistveno večji obisk v smislu log statistike in tudi v smislu kumulativnega ratinga GRP (gross rating points).

b) Intervali zaupanja

Ker gre pogosto za majhne deleže, velja upoštevati dejstvo, da imajo 95% intervali zaupanja naslednjo širino:

Tabela 7: Širina intervalov zaupanja

%	N = 300		N = 600		N = 900	
	spodnja meja	zgornja meja	spodnja meja	zgornja meja	spodnja meja	zgornja meja
3	2.0	5.0	1.8	3.5	1.2	2.9
5	2,5	7,4	4.6	5.2	1.5	4.3
10	6,6	13,3	7.6	9.3	3.8	7.7
15	10,9	19,0	10.8	13.3	6.3	11.0
20	15,4	24,5	14.1	17.2	8.9	14.2
25	20,1	29,9	17.4	20.9	11.7	17.3
30	24,8	35,1	20.7	24.6	14.4	20.4
35	29,6	40,3	0.0	28.2	17.2	23.4

Interval zaupanja 95% pomeni, da se v 19 od 20 primerov ne motimo, kadar trdimo, da navedeni interval vsebuje populacijsko vrednost. Ocena 5% torej pomeni, da imamo pri vzorcu velikosti $n=300$ interval zaupanja 2,5 – 7,4%. Seveda pa je bolj verjetno, da je ocena bliže 5% kot 7%. Tako npr. v navedenem primeru s približno 60% gotovostjo trdimo, da je dejanska vrednost znotraj intervala 4-6%, zelo malo verjetno pa je, da je izven 2,5 – 7,5% intervala. Pri $n=300$ ocene za deleže pod 3% zato niso zanesljive, ker je vzorec premajhen. Seveda pa ocene, manjše od 3%, za katere eksplicitno ne podajamo intervala zaupanja, ne presegajo intervala zaupanja, ki ga lahko ima delež 3%.

Na osnovi zgornjih navedb je za ugotavljanje statistično značilne razlike med dvema vrednostma treba odstotke pomnožiti s faktorjem 1,4. Tako lahko – pri dveh neodvisnih vzorcih velikosti $n = 300$ – statistično značilno razliko odkrijemo šele pri npr. porastu vrednosti od 5% na 9%. Seveda se v primeru večjih vzorcev ocene zožijo: v primeru dveh vzorcev velikosti $n=300$ za faktor 0,7, v primeru treh vzorcev velikosti $n=300$ za faktor 0,58, v primeru štirih vzorcev velikosti $n=300$ pa za faktor 0,5.

c) Koeficient variacije

Podoben indikator kvalitete vzorčne ocene je tudi koeficient variacije (CV). Upoštevati velja, da je običajni kriteriji $CV=0,33$, da se ocena sploh objavi, $CV=0,20$ je meja za slabšo oceno, pod $CV=0,10$ pa se pričenjajo natančne ocene. Pri objavljanju zato ocene s $CV>0,33$ nadomestimo s piko ».«, kar pomeni: večje od nič, vendar nezanesljivo, ocene $0,20<CV<0,33$ z dvema oklepajema (()), ocene med 0,10-0,20 pa z enim oklepajem (), kar označuje previdnost pri njihovi interpretaciji. Pri eni meritvi ($n=300$) imamo naslednje koeficiente variacije in odgovarjajoče oznake:

Tabela 8: Oznake koeficientov variacije

odstotek	CV	oznaka
<3	>0,33	.
3	0,33	((3))
5	0,25	((5))
10	0,17	(10)
15	0,13	(15)
20	0,11	(20)
25	0,10	25
30	0,8	30
35	0,7	35

d) Koliko je uporabnikov Interneta?

Podatki iz ankete kažejo, da je Internet v Sloveniji v začetku leta 2000 uporabljalo 16% aktivnih (mesečnih ali pogostejših) prebivalcev, kar pomeni, da je aktivnih/mesečnih uporabnikov okoli 250.000. V začetku leta 2001 pa je bilo uporabnikov okoli 330.000. Vsi odstotki, ki izražajo *mesečni dosegi*, se torej nanašajo na celoto aktivnih uporabnikov Interneta v Sloveniji. Navedeno število uporabnikov se ujema tudi z ocenami anket Slovensko javno mnenje in Mediana.

e) Opombe k standardiziranim prikazom

- vsi podatki so prikazani kot odstotki;
- deleži manjši od 3% (pri eni ponovitvi je to 10 enot), s CV večjim od 0,33, se v splošnem izpuščajo in označujejo s piko ».«, kar pomeni večje od nič, vendar nezanesljivo (torej $0 < p < 3\%$), vendar jih v tokratnem prikazu kljub temu navajamo;
- nenatančni deleži z intervalom zaupanja širokim okoli tretjino vrednosti ocene (gre za CV v intervalu 0,1-0,2, oziroma za deleže 8-25%) so načeloma označeni v enojnih oklepajih, npr. (20%), česar pa tokrat nismo posebej označevali;
- zelo nenatančni deleži z intervalom zaupanja, širokim okoli polovice vrednosti ocene, so načeloma označeni v dveh oklepajih, gre za CV v intervalu 0,2-0,33 oziroma za deleže 3-8%, npr. ((5%)), česar tokrat nismo posebej označevali;
- pri vsaki strani je bilo odgovorov »ne vem« kvečjemu odstotek ali dva;
- oznaka »-« ali »0« pomeni, da v kategoriji ni bilo podatka, torej vrednost nič, kar seveda ne pomeni nujno, da je takšna tudi vrednost v populaciji.

1.2. Merske napake

Poleg napak, ki izhajajo iz samega vzorca, obstajajo tudi nekatere posebne merske napake, od vrstnega reda postavljenih vprašanj, vpliva specifičnega dneva anketiranja, pa do sistematičnega precenjevanja obiskanosti. V nadaljevanju navajamo nekatere dodatne posebnosti, ki lahko vplivajo na rezultate telefonskega merjenja na osnovi priklica.

- **Problem grafičnega vidika**

V telefonskih anketah je zanemarjen grafični vidik. Študije priklica na osnovi grafičnih logov namreč kažejo določene razlike v primerjavi s tekstualno navedbo. Težave te vrste se kažejo predvsem pri priklicu strani Matkurja in Slowwwenia. Zato je bilo v telefonski anketi posebej poudarjeno, da gre pri slednji za rdečkaste strani podjetja Inetis in ne za strani Matkurje.

- **Natančnost opredelitve obiska strani**

Upoštevati velja, da je pri nekaterih straneh priklic izrazitejši. Pri bolj splošnih straneh, kot so npr. TV sporedi, sporedi za kino, vremenska napoved, je natančnost opredelitve manjša, saj respondenti včasih ne vedo, na kateri strani so obiskali navedeno vsebino. V določeni meri velja to tudi za Ljubljansko borzo in Hidrometeorološki zavod Slovenije.

- **Pogostost obiska strani**

Število mesečnih obiskovalcev ne določa avtomatično dnevnega obiska, ki se lahko zaradi pogostih obiskovalcev pri enaki mesečni obiskanosti bistveno razlikuje. Nekatere strani imajo namreč veliko dnevnih obiskovalcev, ki se vračajo na stran večkrat dnevno, pri drugih straneh prevladujejo obiskovalci, ki stran obiščejo le občasno.

- **Pregledovanje ali dostop?**

Prikazani rezultati ne vključujejo korekcije za respondente, ki določene predstavitve v resnici ne pregledujejo, ampak prek teh strani samo dostopajo do ponudnika. Primer takšnih strani sta SiOL in Arnes. Zato je v telefonski anketi poleg pogostosti obiskovanja strani zastavljeno še vprašanje, ali uporabniki dejansko dostopajo do spletnih strani ali pa se morda le priključujejo do omrežja, ne da bi strani tudi pregledovali.

1.3. Telefonske ankete vs. log-analiza

Na osnovi doslej navedenega so neposredne primerjave telefonskih ocen s statistiko logov seveda nemogoče, saj nastopa vrsta tehničnih težav, ki smo jih navedli, od množice neprimerljivih programov za analizo logov, kjer ima vsak program še množico nastavitvenih opcij, do »proxy« strežnikov, ki pomembno prikrivajo obisk. Povzemimo se enkrat najpomembnejše probleme, ki genereirajo razhajanja.

Podrobna analiza npr. Arnesovega strežnika je namreč pokazala, da kar 40% vsega Arnesovega prometa poteka preko »proxy-ja«. S tem so posebej prizadete strani v statičnem html-ju, ki se redkeje ažurirajo. Relativno precenjeni so iskalniki in dinamično generirane (komercialne) strani.

Težave nastajajo tudi zaradi dinamičnih IP, lokalnih mrež z enim samim IP, zaradi obiska iz tujine, pa tudi zaradi več uporabnikov na istem računalniku (in obratno) ... Problematična je tudi intenziteta obiskov - tako imajo npr. strani, ki so namenjene članom (npr. Intereblod), veliko dnevnih (ponavljajočih se) obiskov svojih članov, vendar relativno majhen mesečni doseg.

Vse bolj kritičen za log-analizo je tudi pojav avtomatskega obiskovanja. Pri tem ne gre samo za ponarejanje obiska v smislu posebnih programov ali avtomatiziranega dostopanja preko dial-up (kjer se vsakič dodeli drugačen IP). Slednje namreč lahko – v primeru, da je narejeno dovolj spretno – »nedokazljivo« ponaredi statistiko logov (ne vpliva pa na telefonske ankete). Še bolj problematično je osveževanje naročniških servisov. Tako ima npr. stran RIS.org, ki je vključena v servis za avtomatsko obveščanje Dot.si, od skupno samo 100 naročnikov, ki prejemajo RIS novice prek Dot.si, od tam nesorazmerno veliko obiskov (sessions). Uporabniški računalnik (klient) namreč dostopa do strežnika RIS.org (in osvežuje vsebino) brez vednosti osebe (uporabnika). Nevarnost predstavlja tudi agresivno ponujanje vpisa strani kot »default« stran, ki sproži obisk pri vsakem zagonu. Kljub temu primerjave – vsaj za večje strani - kažejo visoko korelacijo ($R=0.85$) med mesečno obiskanostjo na osnovi

telefonskih anket ter log-analizo (IP številke). Podrobnosti o tem so na www.ris.org/doksis.htm, www.ris.org/si/oglas/obisk.htm in www.ris.org/topwww/.

Še enkrat velja poudariti, da so se doslej vsa razhajanja – in teh v nerazvitih pogojih začetniških napak, nesporazumov, nerazumevanja problematike, pa tudi potvarjanja ni bilo malo – skoraj v celoti pojasnila v korist telefonskih anket. Najpogosteje je bila težava v razumevanju tehnološke specifičnosti strežnika ali programa za log-analize. V poplavi različnih log statistik so se zato telefonske ankete izkazale kot zanesljiv indikator mesečnega dosega. Posebej kvalitetne so ocene za strani, za katere je bilo opravljenih več merenj. Navedene velikosti vzorcev so namreč za pomembnejše strani dovolj velike in povsem primerljive, npr. z vzorci volilnih anket ter z vzorci v raziskavah drugih medijev, vključno s telemetrijo.

V pogledu statistike logov velja dodati, da imajo šele strani s 500-1.500 dnevnimi obiski – kot jih izračuna Webtrends (podobno tudi 100.si) – možnost, da dosežejo 5% mesečni doseg. Pri tem je odločilno število različnih obiskovalcev, ki se posredno izraža tudi prek IP števil. Bliže 10.000 je mesečno število IP, večja je verjetnost, da ima stran 5% doseg³.

V pogledu kukijev se mnoge log analize za število različnih uporabnikov (distinct users) ne ujemajo z rezultati telefonskih anket. Razkorak še ni pojasnjen. Načeloma je mogoče, da je sistematična napaka v telefonskih anketah, vendar je to na osnovi dosedanjih izkušenj malo verjetno. Tovrstna statistika je namreč (za razliko od »obiskov«, števila IP-jev) še edina preostala log statistika, ki se ponekod ne ujema s telefonskimi anketami. Možni razlogi za razhajanja so lahko: specifične nastavitve kukijev, brisanje kukijev pri uporabnikih, prepoved inštalacije kukijev pri klientu ipd.

2. Analize na osnovi telefonskih ocen

Sprotne ocene obiskanosti so podane v posebnih kvartalnih poročilih in jih na tem mestu ne navajamo. Poudariti pa velja, da so razmerja razmeroma stabilna. Pojavljajo se sicer nekatere nove strani, v letu 2000 npr. Najdi, Mladina, Finance, ki prebijejo 5% ali celo 10% doseg, vendar je novih strani, ki bi izstopale z obiskanostjo, vse manj.

Ker se RIS anketno spremljanje obiskanosti izvaja že od leta 1998 je mogoče spremljati tudi trende v mesečnem dosegu najbolj obiskanih strani, kar je predstavljeno v posebnem poročilu, enako kot podrobna socio-demografska struktura spletnih strani.

Skupno je bilo v analizo obiskanosti vključenih že preko 150 spletnih strani. V nadaljevanju si oglejmo osnovne zaključke iz analize trendov in analize socio-demografskih profilov predstavitev strani.

³ V posebnem poročilu o spletni obiskanosti je predstavljen tudi poseben model za izračunavanje dosega na osnovi log statistike.

2.1. Mesečne serije za izbrane predstavitve

V splošnem se vsem stranem obiskanost veča, izjema so nekatere strani, kot je Arnes, si21. Seveda pa tu govorimo le o relativnem dosegu; absolutno število obiskovalcev se zaradi naraščanja števila uporabnikov Interneta veča tudi v primeru omenjenih strani.

Na vseh profilih obiskanosti so razvidne razlike med aktivnimi obiskovalci (torej tistimi, ki so določene strani obiskovali mesečno ali pogosteje), razvidna so tudi večja nihanja v različnih časovnih obdobjih.

Pri vseh straneh je seveda treba upoštevati naslednje:

- kljub konstantnemu deležu obiskanost strani seveda raste, saj je število uporabnikov večje vsako leto za 30-40%.
- Pri osciliranju vrednosti je treba upoštevati naslednje razloge:
 - Gibanje vzorčne napake, ki smo jo opisali zgoraj.
 - Mersko napako, saj uporabniki poročajo o obiskanosti določene strani razmeroma nezanesljivo, posebej kadar stran ni najbolj znana, pa tudi kadar ima stran nejasno identiteto, npr. Slowwwenia.
 - Akcije določenih strani v medijih oziroma v oglaševalskih medijih, kar lahko povzroči občasni porast v poznavanju in obiskanosti.

2. 2. Pregled obiskanosti po demografskih strukturah

Med uporabniki večine strani prevladujejo moški, le pri straneh Univerze v Ljubljani je delež moških in ženskih uporabnikov izenačen. Med najpogostejšimi uporabniki strani so respondenti stari med 21 in 29 let ter med 30 in 50 let. Na nekaterih straneh, kot so npr. Mobitel ali Arnes, so najpogostejši obiskovalci mlajši od 20 let, medtem ko je starejših obiskovalcev (nad 50 let) na vseh straneh manj.

Respondenti, ki imajo končano srednjo šolo ali tisti z dokončano višjo šolo ali več, so med pogostejšimi uporabniki. Tistih z dokončano poklicno izobrazbo je v večini primerov malo, v nekaterih primerih celo nihče od teh respondentov ni navedel obiska določenih strani.

Menedžerji, strokovnjaki in podjetniki so pogostejši obiskovalci na večini navedenih strani, na nekaterih straneh se pogosteje zadržujejo šolajoči (Mobitel, Slowwwenia, Matkurja, Arnes in Telekom). Na vseh straneh najdemo najmanj neaktivnih obiskovalcev.

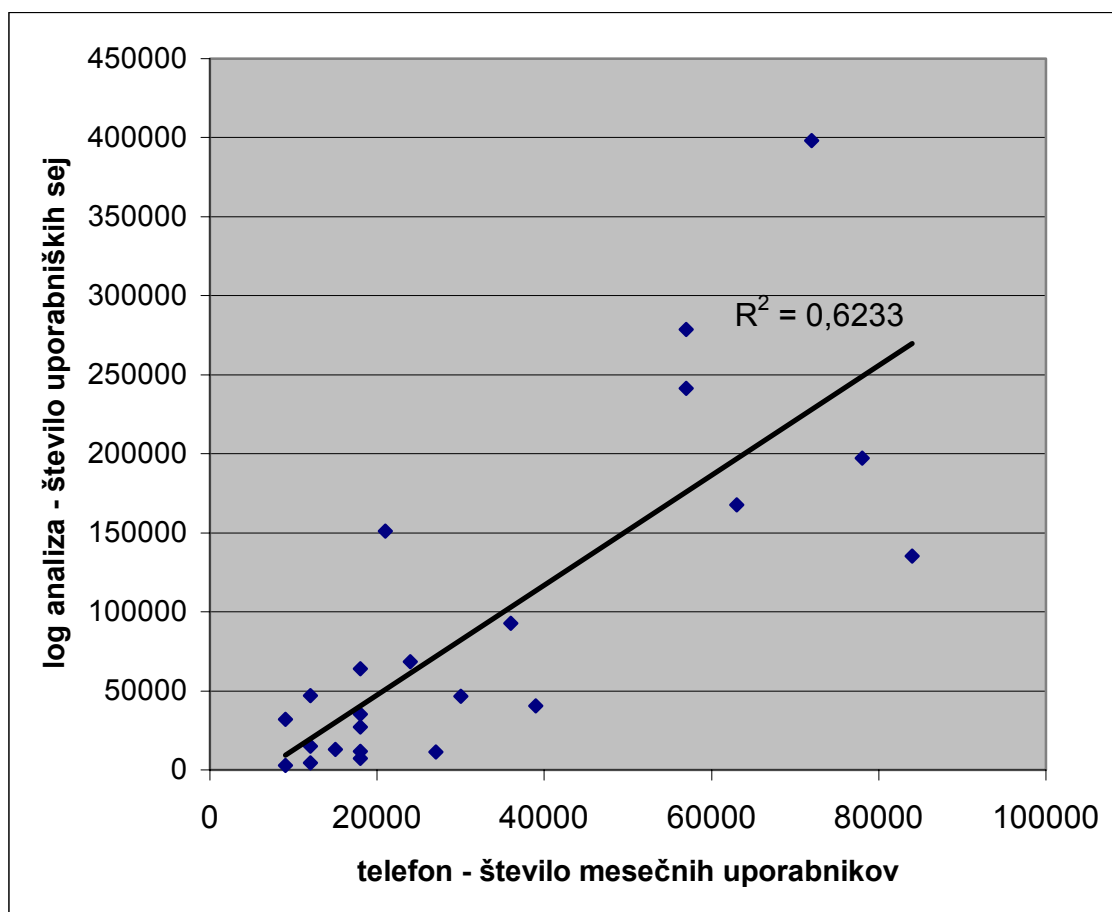
IV. PRIMERJALNA ANALIZA MERITEV

1. Obiski, IP številke in različni uporabniki

Zaradi številnih ovir, ki onemogočajo, da bi rezultate analize logov brez zadržkov priznali za objektivno in primerljivo meritev števila različnih obiskovalcev, smo natančneje in primerjalno raziskali analizo logov in telefonske ankete. Izhajamo iz predpostavke, da so telefonske ankete, kljub nekaterim metodološkim težavam, ki so povezane s priklicem spletnih strani, bistveno natančnejši indikator števila različnih obiskovalcev.

V spodnjih grafičnih prikazih je nekaj osnovnih primerjav.

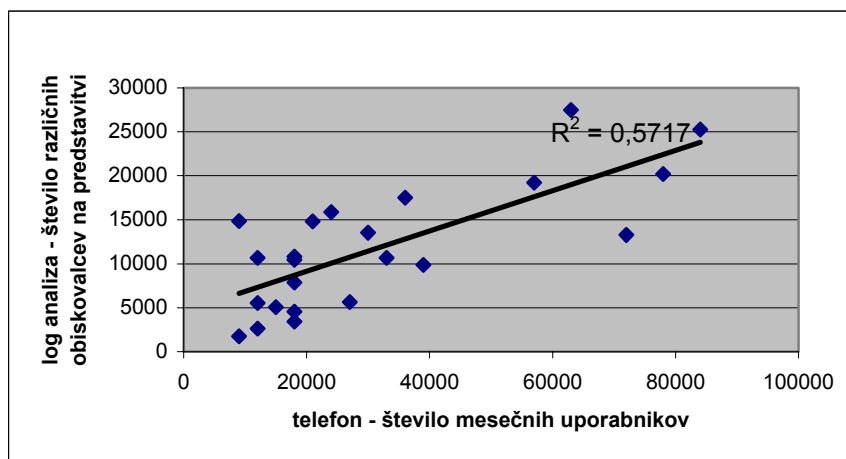
Graf 8: Primerjava log analize in telefonskih anket (a)



Korelacijski koeficient med številom uporabniških sej (log analiza) in številom različnih obiskovalcev (telefonska anketa) znaša 0,78, kar pomeni da med meritvama obstaja visoka linearna povezanost. Kadar se število uporabniških sej poveča za 1, se torej število mesečnih uporabnikov poveča za 0,78.

Podobno sliko opazimo pri razsevnem grafikonu, ko na ordinato postavimo število različnih obiskovalcev (unique users). Korelacijski koeficient med meritvama znaša 75%, kar ponovno kaže na visoko linearno povezanost med meritvama.

Graf 9: Primerjava log analize in telefonske ankete (b)



Na osnovi zgornjih dveh ocen bi lahko zapisali regresijski oceni, ki temeljita na samo eni statistiki, to je na številu obiskov:

$$\text{TEL} = -215000 + 4,3 * \text{US}$$

oziroma na številu različnih uporabnikov – v našem primeru kar različnih IP številčk:

$$\text{TEL} = 1200 + 0,3 * \text{UU}.$$

Na osnovi prve enačbe bi stran z 10000 različnimi IP številčkami imela 21000 obiskovalcev. Na osnovi druge enačbe pa izhaja, da bi stran z 30000 obiski imela 10000 različnih obiskovalcev.

Seveda pa navedene ocene veljajo predvsem za strani z več kot 7000 IP številčk (UU) oziroma z več kot 20000 obiskov (US). Za manjše strani je očitno potrebno izvesti poseben model.

Glede na to, da oba indikatorja analize logov (število različnih obiskovalcev in število uporabniških sej) visoko in pozitivno korelirata s številom obiskovalcev v telefonskih anketah, smo v nadaljevanju razvili linearni regresijski model, ki hkrati obravnava linearna vpliva obeh indikatorjev analize logov na odvisno spremenljivko, t.j. na število obiskovalcev po telefonskih anketah. Postavitev takšnega linearnega modela je tudi teoretsko upravičena, saj imamo po eni strani opravka s spletnimi stranmi, kjer stalno občinstvo generira veliko obiska (uporabniških sej), npr. APP, Arnes. Po drugi strani pa imamo tudi spletne strani, kjer je doseg večji (večje število različnih uporabnikov), vendar do strani dostopajo samo nekajkrat mesečno (npr. TIS).

2. Razvoj linearnega regresijskega modela

Odvisna spremenljivka v regresijskem modelu je bilo število mesečnih obiskovalcev v telefonski anketi (TEL), neodvisna spremenljivka pa število različnih uporabnikov (UU- unique users), ki se v primeru brez kukijev izenačuje s številom različnih IP števil in mesečno število obiskov oziroma uporabniških sej (US - users sessions).

Imamo naslednje rezultate:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,858 ^a	,737	,709	12789,33

a. Predictors: (Constant), US, UU

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3793,574	5388,656		,704	,490
	UU	1,475	,492	,446	2,998	,007
	US	,123	,036	,510	3,434	,003

a. Dependent Variable: TEL

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8,7E+09	2	4,3E+09	26,580	,000 ^a
	Residual	3,1E+09	19	1,6E+08		
	Total	1,2E+10	21			

a. Predictors: (Constant), US, UU

b. Dependent Variable: TEL

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3793,574	5388,656		,704	,490
	UU	1,475	,492	,446	2,998	,007
	US	,123	,036	,510	3,434	,003

a. Dependent Variable: TEL

Rezultati torej kažejo regresijsko enačbo:

$$TEL = 0,12 * US + 1,48 * UU.$$

Vsaka dodatna IP številka torej povečuje število različnih mesečnih uporabnikov v telefonskih anketah za 1,5 osebe, vsak obisk (uporabniška seja) pa za 0.12 osebe. Seveda moramo pri tem upoštevati vse omejitve obravnavane regresije.

Oglejmo si še regresijske ocene (četrti stolpec) v tabeli z ostalimi indikatorji obiskanosti ter številom mesečnih uporabnikov po telefonskih anketah (zadnji stolpec).

Tabela 9: Regresijske ocene

Spletna stran	Št. Obiskov za celo predstavitev (user sessions)	Št. Različnih obiskovalcev za celo predstavitev (unique users)	Ocenjeno število obiskovalcev na podlagi regresije $TEL = 1,48 * UU + 0,12 * US$	Dejansko število mesečnih uporabnikov (jesen 2000)
APP	208582	8285	37292	21000
Arnes	167623	27482	60788	63000
Avto.net	47183	10696	21492	12000
COBISS	46569	13509	25582	30000
Dnevnik	68512	15848	31676	24000
Eon	151218	14810	40065	21000
HMZ	92672	17507	37031	36000
Infomediji	7201	3439	5954	18000
Jerovšek	27188	7874	14916	18000
Kabi	138956	24966	53624	9000
Kompas	15029	5524	9979	12000
Lj Kino	11648	4555	8139	18000
MobiSux	64108	10457	23169	18000
Mobitel	197170	20181	53528	72000
Pasadena	4424	2638	4435	12000
Rumeni Internet	32090	14819	25783	9000
Salomon	40613	9882	19499	39000
Siol	398285	13257	67415	72000
SloKnjiga	2892	1737	2918	9000
SloWWenia	241341	19219	57405	57000
TIS	135226	25277	53637	84000
ZRSZ	13157	5081	9099	15000

Razlike med spletnimi stranmi variirajo; ilustrativen je grafičen prikaz, v grafih 10 in 11 na zadnjih dveh straneh.

Razvidno je, da le pri nekaterih straneh obstajajo večje razlike med ocenjenim (regresija) in izmerjenim (telefonske ankete) številom mesečnih obiskovalcev posameznih spletnih strani. Posebej veliko je to odstopanje na straneh, kjer stalno občinstvo generira veliko število obiskov (uporabniških sej).

Posebej izstopa Kabi (Si21), kjer je problem v ne-uporabniških dostopih programov (aktivnih kanalih), ki brez vednosti uporabnika dostopajo do spletnih strani.

Pri nekaterih najbolj obiskanih straneh (TIS, Mobitel) pa je izmerjeno število obiskovalcev celo višje od ocenjenega, kar gre po vsej verjetnosti na račun proxy strežnikov, ki se na gostitelju predstavljajo z enakimi IP številkami. Telefonski imenik Slovenije pogosto uporabljajo podjetja in druge organizacije, katerih omrežja so za proxy strežniki. To pomeni, da Telefonski imenik dejansko uporablja veliko uporabnikov, ki pa se strežniku predstavijo z enako IP številko, kar pomeni, da so v log datoteki zavedeni kot isti uporabnik. V veliki meri pa se obisk sploh ne pojavi, ker se zahtevek dostavi iz proxy strežnika.

Nekatere strani, ki so v telefonski anketi videti precenjene, pa so v časovnem zamiku (januar-september) bistveno povečale svojo obiskanost, kar v veliki meri pojasni razhajanje.

Oglejmo si še spodnji praktični zgled:

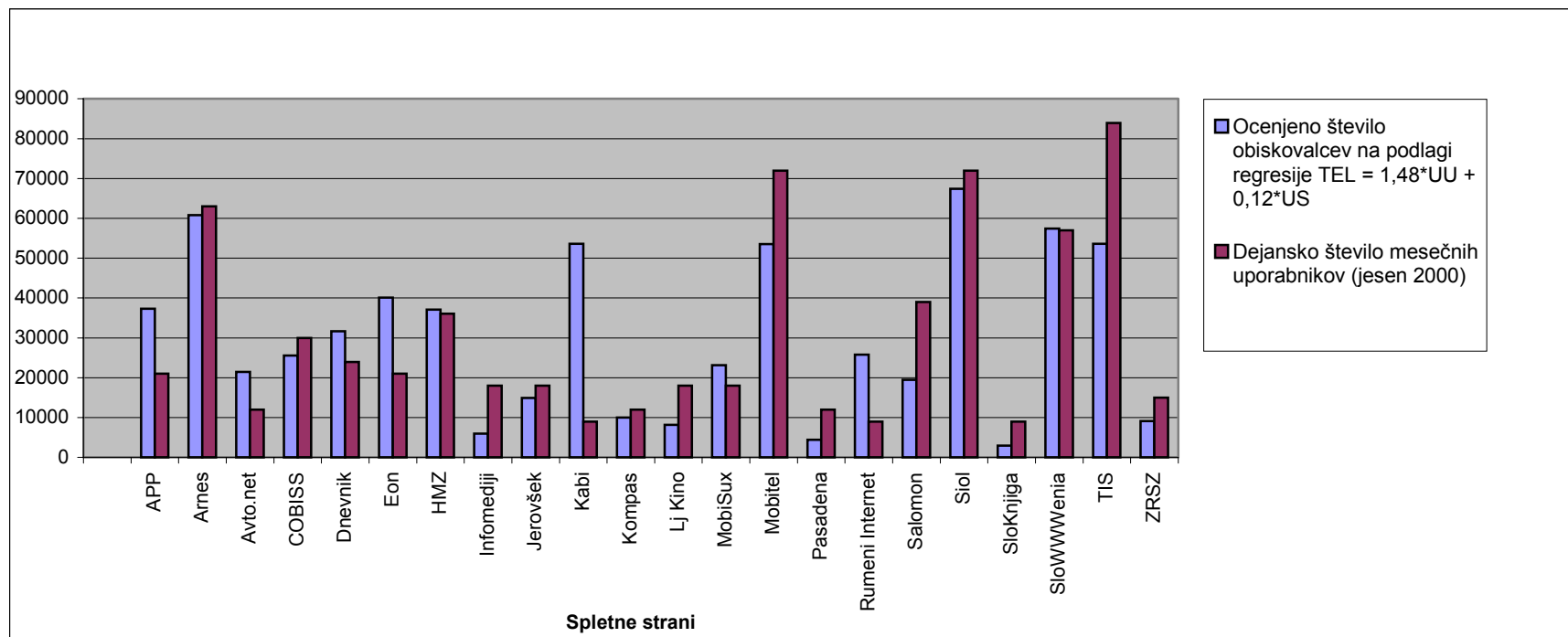
Če na spletni strani izmerimo (s programom WebTrends - www.webtrends.com) npr. 100.000 uporabniških sej (user sessions) in 60.000 različnih uporabnikov (unique users), tedaj je število različnih uporabnikov v skladu z regresijskim modelom na strani:

$$TEL = 0,12 * 100.000 + 1,48 * 60.000 = \mathbf{100.800}$$

Ponoviti velja, da gre pri UU za kategorijo brez kukijev, ki je praktično enaka številu **različnih IP številk mesečno**. Glede na 300.000 aktivnih uporabnikov Interneta v Sloveniji, bi to pomenilo, da je mesečni doseg strani 34%.

Seveda pa je zgornji izračun zgolj informativen, podvržen majhnemu vzorcu, napakam regresijskega modela, ki znaša 12.000 obiskovalcev (dve standardni napaki: torej 25.000 oseb) in tudi specifičnostim posamezne strani.

Kljub temu nudi zgornji izračun osnovno orientacijo za prehod od log statistike v statistiko uporabnikov. Oceno je mogoče bistveno izboljšati z vključitvijo večjega števila strani, z uporabo kukijev, z odstranitvijo najbolj problematičnih strani, z analizo na ločenih segmentih strani (bolj obiskane, manj obiskane) ter z uporabo podatkov, ki se nanašajo na isto obdobje ipd.



Graf 10: Primerjava meritev obiskanosti spletnih strani

Graf 11: Ocenjeno in izmerjeno število obiskovalcev

