

## Matija Remec, Vasja Vehovar: RAZISKOVANJE INTERNETA

### 1. Uvod

Sistematične in primerljive raziskave o Internetu je mogoče zaslediti šele v letu 1994, ko so velika podjetja v ZDA opazila v novi tehnologiji priložnost za konkurenčno prednost. Kakšen je povprečen uporabnik Interneta, koliko je uporabnikov, kakšne so njihove demografske značilnosti, razmišljanja in navade, so bila najpogostejša vprašanja, s katerimi so se ukvarjale prve raziskave o Internetu. S hitro širitvijo Interneta se je skokovito razmahnilo tudi raziskovanje in postalo danes že skoraj nepregledno.

Raziskave izvajajo že uveljavljene raziskovalne agencije (npr. *Nielsen Media Research*) kot tudi nova specializirana podjetja (npr. *MIDS*, *NUA*), ki so svoje poslanstvo usmerila zgolj v raziskovanje novega medija. Neprofitnih in akademskih raziskav je v svetu razmeroma malo – posebej velja omeniti globalno WWW anketo *GVU*<sup>1</sup> ter *Project 2000* pod vodstvom Donne Hoffman<sup>2</sup>.

Ločujemo globalne in lokalne raziskave Interneta. Pri slednjih se omejujemo na državo, regijo ali drugače začrtano celoto. V pogledu vsebine pa se raziskave Interneta osredotočajo predvsem na kazalce demografskih struktur ter indikatorje širitve Interneta, v zadnjem času tudi na vidike ekonomske učinkovitosti. Posebno mesto zavzema raziskovanje WWW, predvsem obiskanost WWW strani. Manj pogosto je (zaenkrat) raziskovanje samega konteksta Interneta, vsebin in njihovih značilnosti.

Že uvodoma velja poudariti, da lahko Internet uporabimo tudi kot raziskovalno orodje – v tem primeru govorimo o raziskovanju *po* Internetu. Interaktivnost, ki jo omogoča Internet, namreč razširja domet že uveljavljenih raziskovalnih tehnik (npr. fokusne skupine, delfi študije, »brainstorming«), omogoča pa tudi povsem nove načine raziskovanja, kot npr. anketiranje po WWW in različne *computer-to-computer* tehnike.

V nadaljevanju si bomo najprej ogledali tipologijo raziskav, kjer Internet nastopa kot predmet proučevanja (2. poglavje), nato bomo prikazali Internet kot raziskovalno orodje (3. poglavje), se dotaknili ključnih metodoloških problemov (4. poglavje) ter strnili osnovne ugotovitve (5. poglavje).

---

<sup>1</sup> <http://www.gvu.org>

<sup>2</sup> <http://www2000.osm.vanderbild.edu>

## ***2. Internet kot predmet raziskovanja***

Internet je globalno omrežje računalnikov, ki so povezani s standardiziranim komunikacijskim protokolom. V širšem smislu pa Internet ne povezuje le računalnikov, ampak tudi uporabnike, informacije in vsebine, zato lahko kot objekt statističnega proučevanja nastopajo najrazličnejši vidiki Interneta. V nadaljevanju bomo sistematično predstavili najpomembnejše vsebine, ki se pojavljajo v okviru raziskovanja Interneta (uporabniki, infrastruktura, obiskanost WWW strani in pretok informacij).

### ***2.1 Uporabniki***

Najpogostejši predmet raziskovanja Interneta so končni uporabniki. Na eni strani so to osebe in gospodinjstva, na drugi strani pravne osebe oziroma organizacije (podjetja, institucije ipd.). Pri uporabnikih so predmet raziskovanja številne vsebine, ki jih lahko strnemo v štiri velika področja.

#### **a) Osnovni demografski kazalci**

Pri končnih uporabnikih nas največkrat zanima razširjenost uporabe (delež enot, ki uporablja Internet), obseg uporabe (pogostost ter število ur uporabe), vsebina in struktura uporabe (zabavne, izobraževalne, poslovne vsebine ipd), uporaba posameznih storitev (elektronska pošta, WWW, FTP...), način dostopa do Interneta (lokacija, način plačevanja) ter omejitve pri uporabi WWW oz. elektronske pošte (stroški, čas, hitrost ipd). V tem okviru so se izoblikovali nekateri najpogostejši kazalci uporabe Interneta, kot npr:

- delež prebivalstva, ki je v zadnjih treh mesecih uporabil Internet,
- delež prebivalstva, ki redno uporablja Internet,
- delež uporabnikov z lastnim elektronskim naslovom,
- delež uporabnikov, ki nakupujejo po Internetu,
- delež gospodinjstev z dostopom do Interneta,
- delež žensk med uporabniki Interneta;

kot tudi:

- delež podjetij z dostopom do Interneta,
- delež podjetij z lastno predstavitvijo na WWW,
- delež podjetij z interaktivno predstavitvijo na WWW,
- delež podjetij z on-line prodajo.

Pregled osnovnih kazalcev za Slovenijo je v Prilogi I, ažurirano stanje pa se nahaja na WWW<sup>3</sup>.

#### **b) Informacijsko/komunikacijska opremljenost uporabnikov**

---

<sup>3</sup> <http://www.ris.org/si/index97-1.html>

Poleg osnovnih demografskih kazalcev nas tako pri organizacijah kot pri gospodinjstvih zanima tudi tehnološka opremljenost, saj le-ta v veliki meri opredeljuje način uporabe Interneta. Pri podjetjih gre predvsem za indikatorje, ki osvetljujejo vpogled v informacijsko (računalniki, programska oprema) in komunikacijsko opremljenost (omrežja, ISDN, ATM ipd.) ter podrobnejši uvid v integriranost Interneta in elektronskega poslovanja v poslovni proces.

Podobno nas zanima pregled informacijsko/telekomunikacijske infrastrukture (računalniki, modemi, mobilni telefoni, kabelska in satelitska televizija ipd.) tudi v primeru gospodinjstev. Podrobni podatki o tem so v Prilogi I, ažurirano stanje pa se nahaja na zgoraj omenjenem naslovu na WWW.

### c) Komercialni vidik uporabe Interneta

Internet skriva izjemne poslovne priložnosti (čeprav obstajajo o dosednji profitnosti razmeroma protislovne informacije), zato je spremljanje finančne učinkovitosti Interneta nadvse aktualno. Pri posameznikih in gospodinjstvih nas v tem pogledu zanima predvsem podrobno spremljanje nakupnega obnašanja: pogostost, obseg in vsebina nakupov. Z vidika ponudnikov Internet dostopa pa je pomemben tudi čas, ki ga uporabniki porabijo na Internetu.

Iz podobnega izhodišča so zanimivi tudi poslovni učinki v podjetjih – posebej pri on-line trgovinah, kjer so aktualni vsi pokazatelji on-line prodaje in oglaševanja kot tudi pregled finančnih transakcij preko Interneta in drugih omrežij.

Sam pregled on-line ponudbe proizvodov/storitev v Sloveniji je v Prilogah II in III<sup>4</sup>, nekateri osnovni vsebinski kazalci za Slovenijo pa so navedeni v tretjem poglavju (Raba Interneta) pričujoče publikacije.

### d) Odnos do Interneta

Poleg vsebin, ki se nanašajo na dejstva oziroma obnašanja uporabnikov, je pogost predmet raziskovanja tudi odnos (stališče) uporabnikov do Interneta. Pri tem so najpogostejši naslednji sklopi: bodoča uporaba Interneta, pomembnost Interneta, pričakovanja, varnost, cenzura, jezik, oglaševanje, elektronsko poslovanje, kot tudi običajni psiho-socialni profil uporabnikov. Obširen pregled aktualnih mnenjskih vsebin se nahaja v WWW anketah GVVU<sup>5</sup>. Tako se je npr. izkazalo, da v letu 1998 najbolj kritično vprašanje Interneta ni več cenzura, ampak varnost. (V Sloveniji je nekoliko drugače, najbolj kritično vprašanje je hitrost.) Vsa pomembnejša področja so bila vključena tudi v WWW raziskave RIS96<sup>6</sup>, RIS97<sup>7</sup> in RIS98<sup>8</sup>.

<sup>4</sup> Pregled on-line trgovin in njihovih aktivnosti je na <http://www.ris.org/si/trgovina.html>, pregled oglaševalcev pa na <http://www.ris.org/si/oglas.html>.

<sup>5</sup> <http://www.gvu.gatech.edu/>.

<sup>6</sup> <http://www.ris.org/si/ris96.html>.

<sup>7</sup> <http://www.ris.org/si/ris97.html>.

<sup>8</sup> <http://www.ris.org/si/ris98.html>.

## 2.2 Infrastruktura

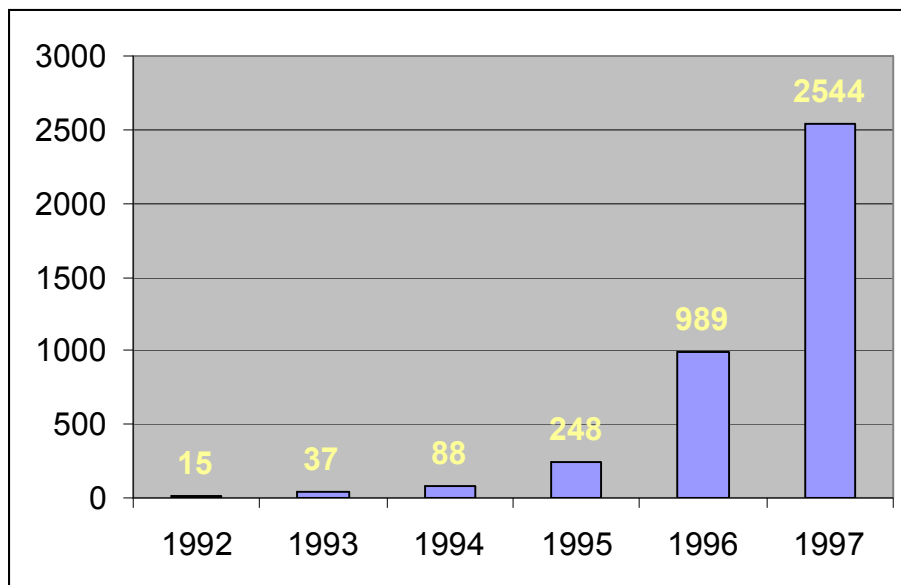
Spremljanje tehnološke infrastrukture Interneta najbolj natančno odslukuje njegovo hitro rast. V tem okviru lahko spremljamo vrsto najrazličnejših komponent oziroma elementov omrežja Internet.

### a) 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 večih komponent, zadnji del npr. *org*, *com*, *net* ali npr. *si* v naslovu (<http://www.ijs.si>) imenujemo **vrhnja domena**, predzadnji del pa **vrhnja poddomena**, kot npr. oznaka *ijs* v **domeni** *ijs.si*.

Obstaja več načinov dodeljevanja vrhnjih poddomen znotraj posameznih držav. Ponekod, tipičen primer so ZDA, lahko kdorkoli registrira katerokoli poddomeno, kar torej velja znotraj vseh vrhnjih domen *com*, *org*, *net*, kjer se dnevno registrira več tisoč novih domen. V primeru spora, npr. ko nekdo registrira ime, ki ga 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. Tako bi npr. lahko vrhnjo poddomeno *ris.si* registriralo le podjetje RIS, d.o.o. 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 WWW prostora. V vsakem primeru neusklajeni administrativni postopki dodeljevanja domen otežujejo vsebinske interpretacije njihovega števila ter mednarodne statistične primerjave. Tipičen primer je najnovejša 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 »<sup>TM</sup>« (»trademark«) industrijske lastnine. Porast števila domen v Turkmeniji zato ne bo odražal širitve Interneta v tej državi.

V Sloveniji dodeljuje vrhnje poddomene ARNES in po njegovih administrativnih podatkih obstaja (konec leta 1997) okoli dva tisoč vrhnjih poddomen, ki pripadajo vrhnji domeni *si*. V tabeli 1 je prikazano število vseh registriranih domen v Sloveniji. Število domen se spremlja v okviru projekta RIS tudi programsko (Priloga V), seveda pa med dejanskim stanjem (aktivnih) domen in administrativnimi podatki obstajajo določene razlike.



**Tabela 1:** Število registriranih vrhnjih poddomen (\*.si) v Sloveniji (Vir: ARNES)

S podobno hitrostjo raste tudi skupno število domen v svetu in je v začetku leta 1998 že preseglo 2 milijona. Rast domen se je torej od leta 1993, ko je obstajalo 21.000 domen, odvijala eksponentno, vendar se je v zadnjem letu nekoliko upočasnila. Podrobnosti lahko najdemo na straneh NetworkWizzard<sup>9</sup>, ki na osnovi raziskave Internet Domain Survey objavlja podrobno stanje o gibanju števila domen.

#### **b) Računalniki, vključeni v omrežje (»host«)**

Še bolj aktualni kot domene so podatki o številu »host«-ov, to je računalnikov, ki so vključeni v Internet in imajo tudi posebno IP številko. Njihovo število se ugotavlja programsko, s pošiljanjem signalov po Internetu. Na tem mestu velja omeniti European Hostcount – RIPE<sup>10</sup> 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.

Network Wizzard je zaradi metodoloških težav januarja 1998 v celoti zamenjal metodologijo merjenja, saj je zaradi naraščajoče zaščite računalnikov (»fire-wall«) dosegel le 76% vseh ustreznih računalnikov. Nov način štetja kaže za januar 1998 skupno 29.669.611 računalnikov, od tega jih kar 28% pripada vrhnji domeni *com*, nadaljnjih 18% vrhnji domeni *net* in 13% vrhnji domeni *edu*. Skupno z vrhnjimi domenami *mil*, *org*, *gov* in *us* pripada domenam, registriranim v ZDA, kar dve tretjini

<sup>9</sup> <http://www.nw.com>

<sup>10</sup> <http://www.ripe.net>

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 (Tabela 2), posebej izrazita pa je razlika pri Sloveniji, kjer Network Wizzard navaja za januar 1998 le 15.432 računalnikov vključenih v Internet.

Število »host«-ov raste nekoliko počasneje kot število domen in se je od leta 1993, ko jih je bilo okoli milijon, povečalo v letu 1997 na več kot 20 milijonov, v letu 1998 pa bo preseglo 30 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 številčk znotraj vrhnjih poddomen.

V Tabeli 2 so prikazani podatki RIPE za število ter 'gostoto' računalnikov, ki so vključeni v Internet v evropskih državah (domenah). Razberemo izstopanje skandinavskih držav ter – z nekaj izjemami – izrazito zaostajanje tranzicijskih držav, še posebej in v celoti velja to za države, ki na tem mestu sploh niso navedene; (izčrpen pregled rezultatov je v Prilogi VIII).

Razberemo lahko, da v skoraj vseh državah stopnja rasti v prvih letih hitro narašča in doseže višek, ki presega stoo odstotno rast, nakar se umiri in postopno zmanjšuje. V zadnjem letu zato zaznavamo rahlo ustavljanje rasti (še posebej izrazito je to v Sloveniji), čeprav je letno naraščanje seveda še vedno izjemno visoko. Kljub tej splošni tendenci pa zadnja kolona, ki za razliko od ostalih govori le o polletni rasti, pri mnogih državah nakazuje dodatni vzpon rasti.

KAZALEC ⇒ ↓ DRŽAVA	Računalniki vključeni v Internet	Računalniki na 100.000 prebivalcev	Rast (%) julij 94 julij 95	Rast (%) julij 95 – julij 96	Rast (%) julij 96 – julij 97	Rast (%) julij 97 – januar 98
Avstrija	110518	1378	111	72	33	8
Belgija	111883	1100	91	82	103	24
Bolgarija	7000	81	–	230	150	26
Češka	59801	579	101	125	32	35
Danska	172531	3287	190	120	71	25
Estonija	16881	1157	250	170	89	37
Finska	501221	9816	122	169	32	26
Francija	385443	661	53	73	60	23
Grčija	26339	250	123	90	84	20
Hrvaška	8164	163	170	158	106	17
Irska	40886	1146	95	125	55	23
Islandija	19240	7126	107	58	30	36
Italija	250713	436	135	136	108	1
Jugoslavija	5365	51	–	–	113	55
Litva	4171	114	419	400	101	53
Madžarska	71296	713	80	140	81	63
Nemčija	1140068	1365	180	40	55	22
Nizozemska	401296	2578	120	59	60	17
Norveška	295115	6732	73	85	73	41
Poljska	92823	240	127	163	80	24
Romunija	13943	64	111	276	65	13
Rusija	109985	74	746	430	143	63
Portugalska	42625	432	96	75	132	17
Slovaška	15246	284	80	168	93	35
Slovenija	19400	994	206	200	65	13
Španija	201685	515	93	106	201	17
Švedska	360643	4052	91	74	55	22
Švica	192183	2666	33	60	48	23
V. Britanija	1058207	1809	110	67	55	18

**Tabela 2:** Število računalnikov (»host«-ov) vključenih v Internet – januar 1998 ter izbrani kazalci rasti 1994-1998 (Vir: RIPE, ARNES)

Ponoviti velja, da so v zgornji tabeli za Slovenijo upoštevani le računalniki, ki so vključeni v Internet znotraj vrhnje domene *si*. Dodatnih nekaj tisoč računalnikov je vključenih v okviru drugih vrhnjih domen, predvsem *com*, *org* in *net*. Približno 30-40% računalnikov, ki so vključeni v okviru vrhnje domene *si*, pripada javnemu omrežju ARNES.

### c) Tehnološke značilnosti Internet strežnikov in povezav

Nekateri računalniki se vključujejo v Internet ne le pasivno (to je z dostopom do vsebin na Internetu), ampak nudijo vsebino tudi sami. Tehnične karakteristike takih računalnikov – **Internet strežnikov** – so za uporabnike nadvse pomembne. V prvi vrsti gre seveda za hitrost, pa tudi za dostopnost in stabilnost. Zanimive pa so lahko tudi blagovne znamke strojne ter programske opreme. V okviru Internet strežnikov seveda

ločujemo različne strežnike glede na storitev, ki jo opravljajo: WWW, e-mail, FTP, ... V nadaljevanju se bomo omejili na WWW strežnike, ki v celotni strukturi strežnikov tudi prevladujejo.

– **Hitrost** merimo na več nivojih. Hitrost (klicnega) dostopa zadeva podatkovno in komunikacijsko hitrost med ponudnikom dostopa do Interneta ter uporabnikom, ki preko njega vstopa v omrežje. V veliki meri je ta hitrost določena predvsem s tehničnimi lastnostmi modema, konfiguracijo računalnika in s specifičnostmi lokalnega telefonskega omrežja. Pomembna pa je tudi 'medmrežna' hitrost med posameznimi strežniki, ki meri zmogljivost vmesnih poti. Seveda sta v primeru, ko uporabnik dostopa z modemom in pregleduje strani na nekem strežniku, vključeni obe hitrosti.

– **Dostopnost in stabilnost** sta sorodna pojma, saj gre pri obeh za preverjanje določenih točk v Internetu. Pri *stabilnosti* gre za stabilnost strežnika (navadno gre za WWW strežnik, ni pa to nujno) na lokalni ravni. Odvisna je predvsem od delovanja oziroma ustavitve delovanja ('sesutja') računalnika. Z *dostopnostjo* pa označujemo stabilnost na globalni ravni, saj nam pove, kako dostopen je določen strežnik z ene ali več točk v Internetu. Na slabšo dostopnost vpliva tudi nestabilnost nekega drugega strežnika, preko katerega potujejo signali. V okviru projekta RIS se redno meri dostopnost strežnikov<sup>11</sup> tako za ponudnike dostopa do Interneta<sup>12</sup> kot tudi za nekatere druge pomembnejše strežnike v Sloveniji. Rezultati kažejo (Priloga VI), da večina strežnikov dosega več kot 98-odstotno dostopnost.

– **Pot povezave** (»traceroute«) nam pove, po kakšni poti (preko katerih vozlišč in s kakšno hitrostjo) potujejo podatkovni paketi med posameznimi točkami v Internetu. Za doseganje hitrih povezav je namreč izbira optimalne poti kar najbolj pomembna. V tem okviru obstajajo programi, ki razstavijo na komponente celotno pot in tudi hitrost pretoka informacij po omrežju<sup>13</sup>.

– **Hitrost in zmogljivost povezav** je pri povezovanju računalnikov oziroma omrežij nadvse pomembna, odvisna pa je predvsem od tehnoloških karakteristik komunikacijskih kanalov. Omeniti velja, da je npr. omrežje ARNES povezano s tujino z znano povezavo TEN34 zmogljivosti 10Mbytov na sekundo, z ZDA pa s povezavo 2 Mbytov/s. Manjšim ponudnikom dostopa do Interneta lahko zadošča že povezava 128 Kbytov/s ali celo 64 Kbytov/s. Razvoj tehnologije seveda išče nove in nove rešitve, ki zvišujejo hitrosti in zmogljivosti pretoka podatkov. V vsakem primeru so podatki o hitrosti in zmogljivosti povezav za uporabnike – in tudi za raziskovalce in analitike – izjemno pomembni, posebej z vidika tehnološke tekme med različnimi oblikami prenosa podatkov (satelitski, kabelski, ISDN, ...).

<sup>11</sup> Februarja 1998 je bilo na seznamu za testiranje 189 strežnikov, meritve pa so potekale iz dveh računalnikov (iz omrežij SiOLa in S-neta). Podrobnejša metodologija je predstavljena v poglavju 3.3.b.

<sup>12</sup> <http://www.ris.org/pro97/>.

<sup>13</sup> Programov za tovrstno spremljanje "paketov" v omrežju Interneta je precej. Eden boljših je NeoTrace <http://www.neoworx.com/neotrace/>, saj poleg grafičnega prikaza povezav podaja še osnove informacije o domenah, preko katerih potujejo paketi.



#### d) Elektronski naslovi

O razširjenosti in strukturi Interneta govori tudi število e-mail naslovov, posebej v zgodnjih fazah njegove širitve. V Sloveniji večina uporabnikov – govorimo o osebah, ki so že kdaj uporabile Internet – še vedno uporablja skupinski e-mail (predvsem v podjetjih, šolskih zavodih) oziroma e-mail druge osebe (prijatelji, sorodniki), kar ostaja specifičnost za Slovenijo tudi v letu 1997. Običajno razmerje (kot ga npr. navaja NUA<sup>14</sup>) med osebami z lastnim e-mail naslovom ter osebami, ki so uporabile Internet v zadnjih treh mesecih, je namreč 1:3. V Sloveniji je bilo navedeno razmerje v letih 1996 in 1997 skoraj 1:4, vendar se v zadnjem času hitro znižuje.

Danes je mogoče z razmeroma enostavnimi programskimi posegi sistematično zbirati e-mail naslove – na ta način je v elektronskem imeniku Telekoma Slovenije zbranih 18.722 (januar 1998) naslovov, slovenski iskalec TOBI<sup>15</sup> pa poroča o 20.000 zbranih elektronskih naslovih. Dodati velja, da je po ocenah v anketah RIS to še vedno manjšina vseh e-mail naslovov, ki jih imajo slovenski uporabniki Interneta, še posebej, ker ima tretjina vseh uporabnikov Interneta z e-mail naslovom po več e-mail naslovov. Seveda se pri vzpostavljanju e-mail imenikov pojavljajo številna etična vprašanja, predvsem problem zasipanja s komercialnimi sporočili (»spamming«).

#### e) WWW: predstavitev, predstavitvene strani, strani

Čeprav obseg WWW nezadržno narašča, iskalci (»*search engines*«) še vedno zagotavljajo osnovni pregled nad glavnino dokumentov, vključenih v WWW. Po nekaterih ocenah<sup>16</sup> je bilo namreč v začetku leta 1998 – brez zaščitene in ostalih nedosegljivih strani – na Internetu že več kot 320 milijonov 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 najde največ dokumentov malo znani iskalec HotBot, ki pa hkrati izstopa tudi z največjim deležem neuporabnih in zastarelih dokumentov.

V tem okviru je lahko ilustrativno že naraščanje števila novosti na What's New – Mat'Kurja<sup>17</sup>: v začetku (decembra 1994) je bilo mesečno po 8 novosti, decembra 1995 že 17, v decembru 1996 je bilo 141 novosti, decembra 1997 pa že 384. Seveda gre v zadnjem času – v začetku so omenjene novosti vključevale izključno predstavitvene strani organizacij – vse bolj pogosto za oglaševanje, za objavljanje novic ter za najrazličnejše ponavljanje/obnavljanje starih predstavitvenih strani. Po drugi strani velja dodati, da številne predstavitvene strani v Sloveniji na omenjene novosti niso vključene, saj temelji sistem vključevanja na samoregistraciji. Tako se je v postopku pregledovanja predstavitvenih strani za potrebe slovenskega WWW imenika WEB.SI

<sup>14</sup> <http://www.nua.ie>

<sup>15</sup> <http://www.i-rose.si>

<sup>16</sup> <http://www.nua.ie>

<sup>17</sup> <http://www.ijs.si/slovenia/news>

izkazalo, da novosti na straneh Mat'Kurje ne vsebujejo približno tretjine predstavitev strani v Sloveniji.

V vsakem primeru je izredno pomembno ločevati različne oblike dokumentov oziroma predstavitev na WWW. Predvsem je potrebno ločevati »home-page« – to je *predstavitevno stran* ali *domačo stran*, s katero razumemo zaokrožen prikaz določene vsebine na WWW – od posamezne *strani* (»page«), to je dokumenta, ki je lahko le eden izmed mnogih v okviru določene predstavitevne strani. Skoraj vsaka predstavitevna stran je namreč sestavljena iz nekaj (tudi več sto) posamičnih strani. Tako ima npr. slovenski iskalec TOBI – čeprav ne obsega vseh strani/dokumentov v slovenskem WWW prostoru – vključenih sredi leta 1998 že več kot 300.000 strani (»page«), kar se ujema tudi s številom strani, ki jih vključujejo večji iskalniki (npr. AltaVista)<sup>18</sup>. Po drugi strani pa lahko na osnovi projekta WEB.SI ocenimo, da je bilo konec leta 1997 skupno še vedno manj kot 10.000 predstavitev strani (»home page«), ki pripadajo slovenskim subjektom.

Nadalje je potrebno ločevati tudi izraz »site« oziroma »website« – prevajamo ga kot *predstavitev* oziroma *WWW predstavitev* –, ki označuje večji in kompleksnejši prikaz organizacije, podjetja, projekta ipd. na WWW. Pogosto se seveda predstavitev omejuje na eno samo predstavitevno stran, lahko pa vsebuje tudi večje število predstavitev strani. Neredko se predstavitev ujema tudi z določeno domeno; predstavitev namreč pogosto obsega ravno vse strani znotraj določene vrhnje poddomene.

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, ena sama *stran*, v okviru *predstavitevne strani* določene organizacije, po drugi strani pa lahko taisto *stran* upravičeno razumemo tudi kot samostojno *predstavitevno stran* te osebe.

Zaradi nelahke razmejitve med domeno, predstavitevjo, predstavitevno stranjo ter posamezno stranjo je statistično spremljanje njihovega števila nadvse težavno. Kljub temu velja dodati, da so pojmi domene, strani ter predstavitevne strani (domača stran) med seboj razmeroma jasno razmejeni in med njimi ne bi smelo prihajati do prekrivanja. Spremljanje njihovega števila je lahko zelo indikativno, natančno razmejevanje med navedenimi pojmi pa je nadvse pomembno pri merjenju obiskanosti WWW strani.

## f) Ponudniki Internet storitev

Pomemben kazalec razvoja Interneta so ponudniki Internet storitev, ki se raztezajo od ponudnikov dostopa in ponudnikov predstavitev pa do ponudnikov zahtevnejših storitev, kot je posredovanje informacij, oglaševalski zakup ipd. Posebej pomembno je ločevanje tistih ponudnikov dostopa do Interneta, ki nudijo lastne mednarodne povezave. Struktura in število ponudnikov dostopa, predvsem pa hitre sprememb,

<sup>18</sup> Mednarodne primerjave števila WWW strani so na <http://www.ris.org>.

nazorno ilustrirajo dinamičnost na tem področju. V Sloveniji se je število samostojnih ponudnikov dostopa od prvih začetkov hitro povečalo in se s prihodom SIOL-a, Telekom Slovenije (ter združitvijo z do tedaj največjim slovenskim omrežjem Sinet–Eunet) v letu 1997 ustalilo na 33 (Priloga IV). Podobno velja tudi za t.i. hrbtenice (»backbone«), ki omogočajo povezavo v tujino. Trenutno je takih povezav v Sloveniji sedem. Za primerjavo, v Veliki Britaniji je bilo konec leta 1997 prek 200 ponudnikov in 20 hrbtenic, v ZDA pa je več kot 3200 ponudnikov dostopa do Interneta.

### g) Splošna infrastruktura

Z vidika meddržavnih primerjav so pomembni tudi najširši indikatorji informacijsko/telekomunikacijske razvitosti. Raztezajo se od pokritosti s telefonskim omrežjem in stopnje njegove digitalizacije pa do informacijske opremljenosti šol in splošnega nivoja računalniške pismenosti. Nekatere tipične indikatorje najdemo v okviru projekta ISPO<sup>19</sup>, za Slovenijo pa so zbrani v Prilogi I oziroma so ažurirani na straneh projekta RIS<sup>20</sup>.

### 2.3 Obiskanosti WWW strani

Z obiskanostjo WWW strani mislimo na dogodke, ki so povezani s komunikacijo med računalniki; predvsem gre za odnos strežnik-klient. Opišemo jih lahko kot: zahteve, obiske, dostope in podobno.

V nadaljevanju bomo z *dostopom* (»impression«) do strani razumeli ogled oziroma zadetek, to je realiziran zahtevek po ogledu določene strani, ki smo ga s klikom izvedli iz uporabniškega računalnika. Dodati velja, da je lahko dostop do strani sestavljen iz več realiziranih *zahtevkov* (»request«) po dokumentih, saj se lahko na določeni strani nahaja npr. večje število slik – pri dostopanju do strani pa sprožimo za vsako sliko poseben zahtevek. Vsak tak zahtevek generira eno vrstico v »log« datoteki, ki vsebuje osnovne podatke o obiskovalcu<sup>21</sup>.

Z *obiskom* (»session«) določene strani razumemo dostop do strani v okviru ene uporabe. Gre torej za dostop, ki ni nastal kot ponovljeno dostopanje v okviru ene uporabe (seanse) na WWW. V primeru torej, ko se pri pregledovanju na WWW večkrat vrnemo na isto stran, se število dostopov povečuje, ne pa tudi število obiskov. Če pa uporabo WWW prekinemo – v formalnem smislu pomeni to izstop iz pregledovalnika – in ponovno dostopimo do iste strani, se bo povečal tudi števec obiskov.

Seveda se število obiskov hitro povečuje tudi v primeru, če ista oseba pogosto obiskuje določeno stran. Posebej je zato zanimivo opazovanje števila *različnih*

<sup>19</sup> <http://www.ispo.cec.be/>

<sup>20</sup> <http://www.ris.org/si/index97-1.html>

<sup>21</sup> Podrobneje o tem v naslednjem poglavju 2.3.a.

**obiskovalcev**, kar običajno – kljub mnogi slabostim – izražamo predvsem kot število različnih IP števil računalnikov (»unique IP numbers«), ki so obiskali določeno stran npr. v enem mesecu. Slednje se seveda ne ujema z dejanskim številom različnih obiskovalcev, predvsem zaradi težav s *proxy* strežniki, z zaščitami (*fire-wall*), z dinamično alociranimi modemskimi dostopi, z različnimi uporabniki, ki dostopajo iz istega računalnika, kot tudi zaradi uporabnikov, ki dostopajo do iste strani iz različnih računalnikov.

Posebej izrazit je problem zaradi uporabe »*proxy*« strežnikov. Le-ti v mnogočem poenostavljajo promet na Internetu, saj npr. shranjujejo določene strani ter jih ob zahtevku samostojno posredujejo uporabniku/klientu, s čimer prihranijo doseganje/pretakanje informacij iz osnovne (oddaljene) lokacije na WWW. Ravno tako je mogoče preko njih bistveno povečati hitrost dostopanja do oddaljenih WWW strani. Ko se torej uporabnikom naloži stran iz ali preko »*proxy*« strežnika, lahko števec na strežniku, kjer se nahaja WWW stran, beleži le IP naslov »*proxy*« strežnika, ne pa tudi resničnega IP naslova obiskovalca. Mogoče je tudi, da se na osnovnem strežniku ne zabeleži ničesar, ker informacijo v celoti posreduje »*proxy*« strežnik, včasih pa celo uporabnikov računalnik (klient). Dejstvo, ali bo uporabnik dostopil neposredno do osnovnega strežnika, je v mnogočem odvisno od nastavitve na samem uporabniškem računalniku (klient), predvsem pa od nastavitve na »*proxy*« strežniku, ki določajo, kdaj se zahtevek prenese na osnovni strežnik. Dodajmo, da v tem okviru seveda ne govorimo o »*proxy*« strežniku na strani ponudnika, kjer tak strežnik le razbremenjuje osnovni strežnik, vendar še vedno omogoča beleženje obiskanosti, ampak o »*proxy*« strežniku na strani uporabnika.

Pri analizah in primerjavah obiskanosti WWW strani zato velja tiha predpostavka, da je delež zahtevkov posredovanih prek »*proxy*«-jev približno enak pri vseh WWW straneh. Predpostavka je razmeroma realna, saj ni posebnih razlogov, da bi obiskovalci, ki dostopajo prek takih strežnikov, v svojih karakteristikah pregledovanja WWW vsebin kakorkoli izstopali. Podobna predpostavka velja tudi za ostale zgoraj navedene tehnične težave, ki motijo merjenje obiskanosti. Tovrstni problemi zato ne ovirajo samih primerjav obiskanosti, motijo (podcenjujejo) pa absolutno število obiskov. Pri tem ima seveda tovrstno podcenjevanje različen učinek pri IP številkah, pri obiskih in pri dostopih.

Nekatere radikalne ocene navajajo, da je zaradi delovanja »*proxy*« strežnikov na uporabnikovi strani dejansko število obiskov – v primerjavi z analizami na osnovi *log* datoteke – večje celo za 76%<sup>22</sup>. Slednje je v splošnem previsoka ocena, povsem verjetno pa je, da podcenjevanje obiskov obsega tretjino ali celo polovico obiskov, ki jih zaznajo običajne *log* datoteke. Navedene ocene seveda dramatično spreminjajo obiskanost strani, predvsem ocene klikov na oglasih (*click-through rate*) in stroške oglaševanja (CPM – *costs per thousand impressions*). V tem okviru so razvite številne tehnike, kako lahko strežnik, kjer se npr. nahaja oglas, »prevara« *proxy* strežnik (*cache busting*) in prisili uporabnika k neposrednemu dostopu, kar seveda dodatno obremenjuje omrežje. Opisane tehnike še niso v množični uporabi, vendar v bodoče

<sup>22</sup> [http://www.matchlogic.com/news/prel\\_01.htm](http://www.matchlogic.com/news/prel_01.htm)

lahko to pričakujemo, predvsem na straneh oglaševalcev. Ugodnejše so tehnike, pri katerih strežnik s posebnimi nastavitvami omogoči pravilno štetje tovrstnih dostopov (*cache counting*) in v tej smeri lahko pričakujemo tudi nadaljnji razvoj merskih standardov<sup>23</sup>.

Obravnavana problematika je nadvse kompleksna in dotaknili smo se le njenih osnovnih razmejitev; izčrpen in sistematičen pregled lahko najdemo v delu Donne Hoffman (1996a), v zadnjem času pa se pojavljajo tudi novejši smeri razvoja<sup>24</sup>

Pri merjenju elementarne obiskanosti strani je na voljo veliko število javno dostopnih programov – števec – za enostavno štetje obiskanosti posamične strani. Seveda ni vseeno, ali štejemo obiske ali kar vse dostope. Ravno tako je pomembno ločevati, ali štejemo dostope na eni sami strani (»page«), na celi predstavitveni strani (»homepage«), na celi predstavitvi (»site«) ali na celotni domeni. Nekateri števeci namreč štejejo kar vse zahteve na celotni predstavitvi. Ko torej poročamo o obiskanosti strani, je na voljo velika izbira nesporazumov, pa tudi zavajanj, čeprav je treba priznati, da večina števecv (ne pa tudi vsi) šteje predvsem obiske (»sessions«) določene (posamične) WWW strani. Po drugi strani je res tudi dejstvo, da je enostavnih števecv – predvsem pri zahtevnejših predstavitev – na WWW vse manj. Pogosto sicer naletimo na obsežne grafično podprte statistike obiskanosti, ki večinoma temeljijo na obdelavi zahtevkov (»requests«) – slednji pa o obiskanosti povedo razmeroma malo.

STATISTIKA ⇒ ↓ OBJEKT	(URL naslov)	Zahtevki	Dostopi	Obiski	Različni obiskovalci
<i>Angleški termin (WebTrends)</i>	<i>(URL adress)</i>	<i>Successful Requests</i>	<i>Page Impressions</i>	<i>User Sessions</i>	<i>Unique IP</i>
Predstavitev <i>Site</i>	<a href="http://www.ris.org">www.ris.org</a>	<b>65.203</b>	<b>17.176</b>	<b>4.404</b>	<b>2.455</b>
Predstavitvena stran <i>Home page</i>	<a href="http://www.ris.org/si">www.ris.org/si</a>	<b>16.400</b>	<b>6.474</b>	<b>2.066</b>	<b>1.328</b>
Stran <i>Page</i>	<a href="http://www.ris.org/trgovina.html">www.ris.org/trgovina.html</a>	<b>922</b>	<b>922</b>	<b>769</b>	<b>589</b>

**Tabela 3:** Statistike obiskanosti predstavitev projekta RIS na WWW, januar 98

V Tabeli 3 smo primerjali štiri osnovne kazalce obiskanosti v okviru predstavitve aktivnosti projekta RIS na WWW, gre za predstavitev (»site«) »Internet in Slovenia«. Zaradi večje jasnosti so v prikazu navedeni tudi angleški izrazi, kot jih izpiše eden od programov za analizo obiskanosti (WebTrends). Prikazane so torej štiri različne statistike obiskanosti za naslednje tri objekte na WWW:

- Predstavitev (»site«) »Internet in Slovenia«, ki se v tem primeru omejuje na strani v okviru domene *ris.org*. Strani (predvsem rezultati projekta RIS96<sup>25</sup>), ki se

<sup>23</sup> <http://www.iab.net/advertise/metricsource.html>

<sup>24</sup> [http://www.accipiter.com/press/releases/pr\\_milwardb.htm](http://www.accipiter.com/press/releases/pr_milwardb.htm)

<sup>25</sup> <http://www2.arnes.si/ris/>

nahajajo v okviru vrhnje poddomene [arnes.si](http://arnes.si), zato v zgornje števec niso vključene, zato je celotna predstavitev v resnici nekoliko obsežnejša, z nekoliko več obiska. V tem primeru torej obstaja tudi določena razlika med domeno in predstavitvijo. Dodati velja, da števci na predstavitvi lahko narastejo, ne da bi uporabnik dostopil do same začetne strani [www.ris.org](http://www.ris.org), saj lahko vstopi na katerikoli sklop predstavitve neposredno preko drugih (pod)strani.

- Predstavitvena stran (»home page«) *Internet v Sloveniji* v slovenščini. Znotraj celotne predstavitve (»site«), ki v tem primeru obsega vse strani znotraj domene *ris.org*, se nahaja vrsta vsebinsko zaokroženih predstavitvenih strani. Poleg predstavitvene strani v slovenščini lahko navedemo še npr. predstavitvene strani letnih projektov RIS98, RIS97 in RIS96, kot tudi predstavitvene strani posameznih članov projekta RIS.
- Stran (»page«), ki prikazuje stanje on-line trgovin v Sloveniji. Gre le za eno, nadvse obsežno WWW stran, ki vsebuje številne povezave na posamezne trgovine.

Iz zgornje ilustracije je očitno, da štetje npr. samih zahtevkov za analizo obiskanosti nikakor ne zadošča. Pri tem velja dodati, da razmerij med različnimi kategorijami v Tabeli 3 ne moremo posploševati na druge predstavitve, čeprav je res, da npr. med zahtevki in obiski določene predstavitve pogosto naletimo na približno razmerje 10:1.

Ravno tako ne moremo na osnovi tovrstne statistike obiskanosti natančneje sklepati o dejanskem številu obiskovalcev, saj zgoraj obravnavane statistike govorijo le o kontaktih med računalniki. Na prvi pogled bi sicer lahko ocenili, da različnih uporabnikov ne more biti manj kot različnih IP števil in ne več od skupnega števila obiskov, zaradi tehničnih problemov merjenja obiskanosti pa je treba tako oceno povečati še za npr. tretjino. V primeru predstavitve projekta RIS se tako dobljena ocena enomesečnega obiska v grobem ujema z rezultati iz telefonske ankete o skupnem številu oseb, ki so obiskale predstavitev projekta RIS. Ocena iz telefonske ankete je zaradi majhnosti vzorca razmeroma nenatančna, vendar nakazuje, da je stran doslej obiskalo okoli dvajset tisoč uporabnikov.

Poglobljene analize obiskanosti WWW strani običajno opravimo z analizo »log« datoteke, ki se dopolnjuje pri vsakem zahtevku. Za tovrstne analize je na voljo vrsta posebnih programov, npr. WebTrends<sup>26</sup>, NetIntellect, WebCounter, Netrics itd. Za kar najbolj objektivno merjenje obiskanosti pa potrebujemo (podobno kot pri merjenju TV ratingov) nepristranske opazovalce. V ta namen so se v tujini že oblikovali posebni agenti, ki s pomočjo prestrezanja »log«-ov opravijo tovrstno štetje (Hoffman, 1996b).

Na osnovi štetja obiskov nastajajo tudi različni sezname najbolj obiskanih WWW strani<sup>27</sup>. Poudariti velja, da najpomembnejše razvrstitve ne izhajajo iz zgoraj opisane

<sup>26</sup> Navedni program uporabljamo tudi pri projektu RIS <http://www.ardi.si/webtrends/>.

<sup>27</sup> <http://www.hitbox.com/wc/world.html>.

statistike zahtevkov (»log«-ov) posameznih strani, ampak iz anketiranja slučajno izbranih respondentov<sup>28</sup>. Na anketni način so izdelane tudi najbolj obiskane strani<sup>29</sup> (Priloga VII) v okviru projekta RIS. Zanimivo je, da se najbolj obiskane tuje strani v Sloveniji v veliki meri ujemajo z globalnim vrstnim redom najbolj obiskanih predstavitev, ki so navedene v Tabeli 4.

	Domena	št. oseb nad 12 let v 1000
1	yahoo.com	26.726
2	Netscape.com	20.723
3	Microsoft.com	15.674
4	excite.com	12.502
5	Infoseek.com	11.696
6	aol.com	11.243
7	Geocities.com	10.498
8	lycos.com	6.787
9	Altavista.digital.com	6.764
10	msn.com	6.315

**Tabela 4:** Deset najbolj obiskanih predstavitev (domene \*.com) za mesec december 1997 (Vir: RelevantKnowledge Inc)

Spremljanje obiskanosti se torej razvija v podobni smeri kot telemetrija, kjer se v TV sprejemnik vgradi posebna kartica, ki omogoča beleženje TV programov, ki jih v gospodinjstvih gledajo. Podobno se pri spremljanju obiskanosti WWW slučajno izbranim respondentom vgradi v računalnik program, ki spremlja gibanje po Internetu<sup>30</sup>. Problem nesodelovanja se rešuje podobno kot pri telemetriji, z nagradami ter z dodatno analizo nerespondentov.

Kljub temu ostajajo analize *log* datotek izredno dragocene za merjenje obiskanosti. Samo na tej osnovi je namreč mogoče opraviti vrsto pomembnih analiz, kot npr. analiza poti obiskovalca, analiza gostitve obiskov ter obremenjenosti strežnika, analiza domen, s katerih prihajajo obiskovalci, odstotek klikov na oglas (*click-through rate*) ipd.

Seveda velja ponoviti, da – kljub hitremu napredku v smeri poenotenja – zaenkrat še vedno ni splošno sprejetih standardov<sup>31</sup>, kako meriti obiskovanje WWW strani, zato lahko navedemo:

»Nesporno je in ni dvoma o tem, da so oglaševalske agencije spoznale Mrežo za nov medij. To pa še ne pomeni, da jim je ta medij všeč ali da vedo, kako ga obvladati. Za

<sup>28</sup> <http://www.mediametrix.com>

<sup>29</sup> <http://www.ris.org/www97/>

<sup>30</sup> <http://www.RelevantKnowledge.com/>

<sup>31</sup> Seveda pa potekajo intenzivna prizadavanja na tem področju, npr. <http://www.ipro.com>, <http://www.netcount.com>, <http://www.accessabc.com>, <http://www.nielsen.com>, <http://www.iab.net>.

kupce medijskega prostora je Internet ustvaril kaotično, nedefinirano bojno polje brez kakršnih koli industrijskih standardov in z oznako »*under construction*«<sup>32</sup>.« (Internet Marketing & Technology Report, 1996:4).

## **2.4 Pretok informacij na Internetu**

Podatke o prometu in pretoku informacij navadno merijo ponudniki Internet storitev in s tem omogočajo vpogled v trende in razmerja med tokovi informacij. V tem okviru je npr. nadvse ilustrativno, da je prek omrežja ARNES v decembru 1997 potekal mednarodni promet v obsegu 900 Gbytov, ki so ga povzročili uporabniki omrežja ARNES, njihov notranji promet (v Sloveniji) pa je bil približno 2-3 krat večji. Promet, ki pa so ga uporabniki iz tujine povzročili na straneh ARNESa, predstavlja približno tretjino navedenega obsega. Za predstavbo o celotnem prometu v Sloveniji velja ponoviti oceno, da predstavlja omrežje ARNES približno 30-40% celotnega prometa na Internetu v Sloveniji.

Še nekaj podatkov o prometu na virtualni predstavitvi Slovenije (Mat'Kurja), ki je najbolj obiskana predstavitev v slovenskem WWW prostoru. Skupni mesečni promet je 8–9 Gbytov, pri čemer dve tretjini prometa pripadata domačim uporabnikom. Realiziranih je (gre za podatke konec leta 1997) približno milijon zahtevkov po HTML dokumentih (straneh) mesečno, in tudi v tem pogledu prihajata dobri dve tretjini zahtevkov s strani slovenskih uporabnikov. Obrnjeno razmerje med domačimi in tujimi uporabniki je pri skoraj 30.000 različnih IP številkah. Le četrtnina IP številka namreč pripada računalnikom z vrhno domeno *si*, kar lahko pojasnimo predvsem z večjo intenzivnostjo domačih uporabnikov. Večina WWW strani je namreč v slovenščini in zato imajo tujci – ko predstavitev obišejejo – bistveno manj možnosti za nadaljnje surfanje. Pomemben razlog za tako razmerje je tudi dejstvo, da se domači uporabniki pogosteje vračajo; po podatkih RIS98 npr. desetina slovenskih uporabnikov obiskuje navedene strani tedensko, nekaj odstotkov celo dnevno. Načeloma so lahko vmešani tudi tehničnimi problemi tovrstnega merjenja (npr. *proxy* strežnik, *fire-wall*), vendar ni posebnega razloga, da bi se tovrstni problemi pri tujih uporabnikih pojavljali pogosteje. Povsem zanemarljivi pa niso niti računalniki slovenskih uporabnikov, ki so registrirani pod ne-slovensko vrhno domeno (npr. *com*, *org*, *net*).

## **3. Raziskovanje po Internetu**

Internet lahko proučujemo s klasičnimi raziskovalnimi tehnikami, ki vključujejo predvsem anketiranje ter običajne analize tehnoloških, administrativnih in finančnih virov ter s specifičnimi »*on-line*« raziskovalnimi tehnikami.

---

<sup>32</sup> *Under construction* je eden izmed pogostih žargonov na WWW straneh. Napis se nahaja na predstavitvenih straneh in opozarja obiskovalce, da so strani še v izgradnji, zato so možne napake.



Med »on-line« raziskovalne tehnike uvrščamo tri skupine: anketiranje po WWW, anketiranje po elektronski pošti ter »computer-to-computer« merjenje, kjer gre – za razliko od ostalih tehnik – za tipično štetje bitov<sup>33</sup>, to je stabilnosti in hitrosti prenosov informacij.

Za »on-line« tehnike merjenja je značilno, da so poceni, hitre, enostavne in natančne, procesiranje podatkov in rezultatov pa je ves čas računalniško vodeno. Vse te odlike so privedle do tega, da se uporablja »on-line« tehnike – z izjemo »computer-to-computer« tehnik – tudi za raziskovanje pojavov, ki niso povezani z Internetom.

Pričakujemo lahko, da bo nadaljnji tehnološki razvoj (predvsem v pogledu interaktivnih storitev) močno preobrazil številne klasične raziskovalne tehnike, kot npr. anketiranje po telefonu, poglobljene intervjuje, skupinske diskusije ipd.

### **3.1 Anketiranje po elektronski pošti**

Anketiranje po elektronski pošti je podobno klasičnemu, pisnemu anketiranju. Respondent dobi vprašalnik v svoj elektronski poštni predal, nanj odgovori (ali pa tudi ne) in ga odpošlje na zelen naslov. Tovrstno anketiranje se ni posebej razširilo in težko bi našli uspešen raziskovalni projekt, ki je uporabil tovrstno tehniko. Velik tehnični problem namreč predstavlja samo editiranje vprašalnika s strani respondenta, težko pa je dobiti tudi kvalitetne sezname elektronskih naslovov za ciljno populacijo.

Celo v primeru anketiranja uporabnikov v okviru enotne organizacije (s celovitim seznamom elektronskih naslovov, s podobno strojno ter s povsem enako programsko opremo) se izkaže, da so stroški ankete po »e-mailu« večji kot stroški navadne pisemske ankete (Couper, 1998). Osnovni razlog za to so predvsem nepredvideni zapleti na strani uporabnikov ter visoki stroški računalniškega programiranja.

Elektronska pošta se zato bolj pogosto uporablja le kot sredstvo za privabljanje respondentov k anketi na WWW. Izkušnje kažejo, da v primeru, ko izbiramo elektronske naslove iz javnih imenikov in izbiramo mlajšo generacijo (npr. študentske uporabnike KISS), odgovori približno tretjina povabljenih oseb (Brečko, 1997).

### **3.2 Anketiranje po WWW**

Anketiranje po WWW je že uveljavljen način raziskovanja Interneta. Tovrstno raziskovanje se je tako razmahnilo, da t.i. forme – eden izmed delov HTML<sup>34</sup>, ki omogočajo preprosto anketiranje, niso več zadostovale. Z uporabo CGI<sup>35</sup> in JAVA skript ter različnih programskih jezikov, postane anketiranje na WWW natančnejše,

<sup>33</sup> Negroponte N., ameriški strokovnjak s področja informacijskih tehnologij, avtor knjige *Being Digital*, loči med biti in atomi. To ponazori s primerom računalniške in navadne knjižnice: če sposodimo knjigo (atome), na knjižni polici ne ostane nič, če pa določeno datoteko prenesemo drugam, le-ta še vedno (lahko) ostane na računalniškemu strežniku.

<sup>34</sup> HTML - Hyper-Text Markup Language, preprost programski jezik za pisanje predstavitev strani.

<sup>35</sup> CGI - Common Gateway Interface.

prijaznejše za uporabnika in raziskovalca, predvsem pa bolj prilagodljivo. Anketiranje po WWW je naslednji korak v razvoju računalniško podprtega anketiranja in je konceptualno podobno DBM (»disk by mail«) anketiranju, kjer respondent prejme (po pošti) vprašalnik na disketi, ga izpolni in po pošti odpošlje. Tehnične možnosti WWW anketiranja danes ne le zagotavljajo kontrole kvalitete odgovarjanja (kontrola ranga, preskoki, konsistentnost), ki obstaja v ostalih računalniško podprtih sistemih (DBM, CATI, CASI), ampak jih v pogledu upravljanja s podatki bistveno presega. Podrobneje je metodologija anketiranja po WWW predstavljena v naslednjem tekstu pričujoče publikacije (Batagelj in Vehovar: Metodologija WWW anketiranja), zato je na tem mestu podrobneje ne obravnavamo.

### 3.3 »Computer-to-computer« merjenje

Značilnost »computer-to-computer« merjenj je popolna avtomatizacija procesa zbiranja podatkov ter dejstvo, da na vprašanja ne odgovarja človek, ampak računalnik. Zaradi slednjega pri tovrstnem merjenju pravzaprav težko govorimo o anketiranju, bolj o opazovanju ali celo o poskusu, čeprav gre za primarno zbiranje podatkov, ki (pogosto) temelji na vzorčenju.

a) Najbolj razširjena »computer-to-computer« tehnika, predvsem v okviru raziskovanja WWW, je **logiranje**<sup>36</sup>.

»Zaporedje dogodkov, ki jih izvede računalnik, se avtomatično zapiše (logira) v datoteko. Navadno so ti dogodki zahteve (po priklicu dokumenta/programa, zagonu skripte) ali pa ustrezni odgovori. Logi WWW strežnikov pa so le eden izmed primerov logiranja. Tehnično gledano se lahko logira brez vednosti uporabnikov (običajen pojav pri današnjih WWW strežnikih), čeprav bi lahko, legalno in etično gledano, take aktivnosti prepovedali (Abrams in Williams, 1996:101).«

Seveda dobimo iz logov razmeroma skopo informacijo o obiskovalcu. Beleži se predvsem čas zahteve oz. odgovora za določen dokument, ime *hosta* (DNS<sup>37</sup>), IP naslov računalnika, prek katerega se je uporabnik priključil v Internet. Navedene osnovne informacije lahko razširimo s posebnimi nastavitvami in beležimo tudi podatke o tem, kakšno programsko opremo uporablja obiskovalec, kako in odkod je obiskovalec prišel do določene strani ipd.

b) Med »online« tehnikami velja posebej omeniti »**pinging**«. Gre za program, ki uporablja TCP/IP protokol za pošiljanje sporočil na omrežne vmesnike, z odzivi (»**pongi**«) pa dobimo informacijo o tem, ali določen računalnik, vključen v omrežje (»**host**«), obstaja ali ne. Vendar to za kvalitetne podatke o tem, ali je nek WWW

<sup>36</sup> Logiranje - iz angleščine *logging*, pomeni zapisovanje, snemanje, beleženje priklicanih informacij v datoteko (log).

<sup>37</sup> DNS - *Domain Name System* je sistem, ki ga vzdržuje ponudnik Interneta, njegova funkcija pa je prevajanje naslovov v Internetu v numerični zapis (ime *hosta* ilusion.ris.org ima svoj IP naslov 195.250.206.98).

sistem dostopen ali ne, še ni dovolj. V okviru projekta RIS se redno izvaja testiranje (Priloga VI) slovenskih WWW strežnikov<sup>38</sup>. Program, ki preverja dostopnost, se v določenih časovnih intervalih poveže na izbrani sistem in zahteva stran, ki je praviloma HTML dokument. S takim pristopom dejansko preverimo delovanje WWW sistema in ne samo osnovno dostopnost na nivoju omrežja, kot to počne običajni »pinging«, ki uporablja TCP/IP protokol za pošiljanje sporočil med računalniki v omrežju. Lahko se namreč zgodi, da WWW sistem ne deluje, čeprav računalnik deluje. V takem primeru bi s »pinging«-om dobili odziv, z opisanim postopkom pa ne. Pri vsakem poskusu dostopa program poveča števec vseh poskusov in pri uspešnem prenosu tudi število uspešnih dostopov – tako dobimo dostopnost sistema, ki ga preverjamo.

#### **4. Metodološki problemi proučevanja Interneta**

Pri raziskovanju Interneta se poleg siceršnjih metodoloških problemov, ki spremljajo vsako kvantitativno raziskavo (operacionalizacija, merjenje, zbiranje podatkov,...), pojavljajo še dodatne težave, ki izvirajo iz narave proučevanega fenomena. V nadaljevanju bomo obravnavali nekatere najbolj pogoste metodološke probleme proučevanja Interneta.

##### **4.1 Nerazumevanje in nepoznavanje Interneta**

V Tabeli 5 je prikazana osnovna konceptualna struktura Interneta. Razvidno je, da vsaki storitvi pripada njej lastna sfera (informacijska, komunikacijska ali interakcijska), za dostop do storitev pa se uporablja različne aplikacije in orodja (programe).

<i>STORITVE</i>	<i>SFERA</i>	<i>APLIKACIJE</i>	<i>PROGRAMI</i>
World Wide Web	informacijska	Yahoo, search engines	Netscape, MSIE, Mosaic, Lynx...
Elektronska pošta, USENET	komunikacijska	mail in news strežniki	Pegasus Mail, Eudora...
IRC, MOOD, MOO, Video konference	interakcijska	IRC strežniki...	IRCII, mIRC, Iphone, CU SeeMe...

**Tabela 5:** Struktura konceptov Interneta

Zaradi izredne preprostosti uporabe Interneta se uporabnikom ni potrebno širše seznanjati z njegovim ozadjem. Ko pa hočemo raziskati posamezne značilnosti Interneta, se hitro zgodi, da se zamenjujejo sfere, storitve, programska orodja, aplikacije in celo infrastruktura, povezana z Internetom (npr. ponudnike dostopa do

<sup>38</sup> <http://www.ris.org/pro97/>

Interneta). Tako je v raziskavah v okviru projekta RIS prihajalo do številnih neuporabnih odgovorov kot npr.:

- Internet je samo World Wide Web,
- Elektronska pošta je storitev, ki ni del Interneta,
- Yahoo je ponudnik dostopa do Interneta.

V širši perspektivi sicer lahko rečemo, da so tovrstni problemi pogosti pri raziskovanju tehnoloških fenomenov. Tako npr. ne preseneča, da v telefonski anketi RIS96 skoraj tretjina uporabnikov Interneta ni vedela, preko katerega ponudnika se od doma priključujejo na Internet; v anketi RIS98 pa je navedeni delež padel na petino. Podobno visoke odstotke nepoznavanja srečamo v gospodinjstvih tudi pri vprašanjih o tehničnih karakteristikah računalnikov, modemov in mobilne telefonije.

#### **4.2 Družbena zaželenost**

Poleg zgoraj navedenih težav lahko opazimo tudi splošnejši pojav, ki ga velja posebej poudariti. Gre za pojav 'navideznih uporabnikov', ki se v prvih pregledovalnih vprašanjih opredeljujejo za uporabnike Interneta, kasneje, ko so vključeni v podrobnejšo analizo, pa se izkaže, da Interneta pravzaprav ne uporabljajo. Delež 'navideznih uporabnikov' je bil v prvih letih razmeroma visok (RIS96 – skoraj tretjina). Gre verjetno za tipičen primer družbene zaželenosti. V času pregledovanja (delno je tako tudi danes) je namreč Internet ena izmed 'modnih muh' in marsikdo – tudi če je le opazoval 'surfanje' svojih bližnjih, se zato hitro uvrsti med uporabnike Interneta.

Podoben pojav najdemo tudi pri oceni »on-line« nakupovanja. Tako npr. 12% oseb, ki so že uporabile Internet, pozitivno odgovarja na vprašanje o nakupovanju prek Interneta v Sloveniji. Vendar pa že vprašanje o vsebini takega nakupovanja razkrije, da jih najmanj tretjina na tak način v resnici še ni nakupovala.

Poleg same družbene zaželenosti lahko na precenjevanje uporabe Interneta vplivajo tudi drugi dejavniki (npr. učinek spomina, nerazumevanje tehnologije). V vsakem primeru je treba priznati, da ankete nekoliko precenjujejo obseg uporabe Interneta. Zanimivo je, da je tovrstno precenjevanje konsistentno; v okviru RIS96 je namreč povsem enak delež uporabnikov Interneta pokazalo poleg same ankete projekta RIS še pet različnih kontrolnih anket, v katere je bilo vključeno standardizirano vprašanje o uporabi Interneta, in sicer telefonske ankete: Delo STIK, GRAL d.o.o, FDV–Politbarometer, Statistični urad Republike Slovenije – Anketa o mnenju potrošnikov ter terenske ankete FDV–Slovensko javno mnenje in ITEO (Omnibus gospodinjstev). Natančni obseg precenjevanja je težko opredeliti. Za to bi potrebovali obsežnejši eksperiment, vendar glede na nekatere kontrolne spremenljivke (posebej število modemskih dostopov za gospodinjstva s strani večjih ponudnikov dostopa) precenjevanje ne presega več kot 10%.

### 4.3 Problem definicij

Internet je dinamičen koncept in zato ga je težko raziskovati, saj se znova in znova pojavljajo novi pojmi, povezani z Internetom, nekateri pa tudi hitro izginjajo. Zato je izredno pomembno pravilno opredeljevanje predmeta raziskovanja. Tipičen primer je že ugotavljanje števila uporabnikov Interneta, saj glede tega še vedno ni splošno sprejete definicije.

Zato je posebej pomembno, da se v vsaki raziskavi, ne glede na njeno velikost in namen, v poročilu in interpretacijah jasno poda kriterij opredelitve uporabnika Interneta. Najbolj pogoste so definicije uporabnika, ki temeljijo na naslednjem:

- število dostopov (npr. uporabnik Interneta, ki ga je uporabil vsaj trikrat),
- uporaba v časovnem obdobju (npr. uporaba v zadnjih treh mesecih),
- uporaba določenih storitev Interneta (npr. elektronske pošte ali WWW),
- posedovanje uporabniškega imena (elektronski naslov),
- samoopredelitev za uporabnika Interneta.

Ko primerjamo rezultate različnih raziskav, moramo biti torej pozorni na kriterije, na osnovi katerih so definirani posamezni predmeti raziskovanja. V Tabeli 6 je tipičen prikaz različnih definicij in tudi različnih rezultatov v raziskavah v ZDA.

<b>VIR</b>	<b>DATUM</b>	<b>DEFINICIJA UPORABNIKA</b>	<b>MIO</b>
Louis Harris & Assoc.	nov-96	Odrasli uporabniki Interneta v ZDA	35
International Data Corp. (IDC)	okt-96	Uporabniki WWW	31,4
FIND/SVP	okt-96	Uporabniki v ZDA (18+), ki so uporabili katero koli storitev Interneta razen elektronske pošte	27
Mediamark Research Inc. (MRI)	okt-96	Katera koli uporaba storitev v <i>Cyberspace</i> v ZDA	27
Intelliquest	jul-96	Uporabniki Interneta v ZDA	35
Louis Harris	maj-96	Uporabniki Interneta v ZDA	29
International Data Corp. (IDC)	maj-96	WWW surferji (maj 1996)	23
Computer Intelligence	maj-96	Uporabniki Interneta v ZDA konec leta 1995	15
Hoffman/Novak	apr-96	Uporabniki Interneta v ZDA	16,4
Wall Street Journal	mar-96	Uporabniki Interneta z dostopom doma ali v službi v S. Ameriki	17,6
Morgan Stanley	feb-96	Uporabniki WWW / Interneta v letu 1995	9
Matrix	feb-96	Uporabniki Interneta z vsega sveta v letu 1995	26,4

FIND/SVP	jan-96	Uporabniki iz ZDA, ki so uporabili katero koli Internet storitev razen e-pošte	9,5
Nielsen Media Research	okt-95	S. Američani (16+), ki so uporabili Internet v zadnjih 3 mesecih	23
O'Reilly & Assoc.	jul-95	Američani z direktnim dostopom na Internet	5,8

**Tabela 6:** Definicije uporabnika Interneta.

Vir: Cyberatlas (1997), <http://www.cyberatlas.com/>.

V primeru raziskave RIS97 lahko problem ilustriramo z dejstvom, da je do začetka leta 1998 uporabilo Internet okoli 250.000 oseb v Sloveniji. Od tega jih je v zadnjih treh mesecih uporabilo Internet tri četrtine. Uporabnikov, ki ga uporabljajo mesečno, je približno dve tretjini, tedensko ga uporablja slaba polovica, približno tretjina pa ima tudi svoj lastni elektronski naslov. Na osnovi navedenih števil je razvidno, da se lahko zaradi različnih opredelitev ocene močno razlikujejo.

Dodati velja, da se je v okviru globalnega spremljanja, ki ga izvaja NUA<sup>39</sup>, vzpostavil standard uporabnika Interneta, ki je »Internet uporabil v zadnjih treh mesecih«. Navedena opredelitev je seveda še vedno razmeroma ohlapna, vendar omogoča vsaj osnovne primerjave.

Seveda pa najdemo podobne nedorečenosti glede osnovnih opredelitev tudi na drugih področjih; posebej pereče je to, kar smo že obravnavali, pri vprašanju obiskanosti WWW strani.

#### **4.4 Terminološki problemi**

Problem pomanjkljivih definicij se nemalokrat zastruje tudi s terminološko nedorečenostjo. V tem pogledu je posebej tipična prav nejasnost na področju merjenja obiskov WWW strani. Tako celo uveljavljeni programi za analizo obiskanosti uporabljajo povsem različne izraze za najbolj elementaren pojem, kot je npr. obisk (»hit«, »visit«, »impression«).

Poleg splošnih terminoloških problemov, ki ostajajo nerešeni na globalni ravni, so pereča tudi vprašanja, ki so specifična za slovenski prostor. Verjetno smo eden redkih narodov, ki je za najosnovnejše izraze, povezane z Internetom, izredno hitro našel 'ustrezni' prevod. Najbrž so le redkokje tolikokrat prevedli enega najpogostejših izrazov, *World Wide Web* ali kratko WWW. Poleg originalnih poimenovanj (World Wide Web, WWW, Web) namreč uporabljamo zanj še izraze, kot so: mreža, splet, svetovni splet, spletne strani, pajčevina, ... Podobno smo za programe (»browser«), ki omogočajo pregledovanje WWW, dobili vrsto izrazov, kot so pregledovalnik, spletni brkljalnik, brkljalnik, brskalnik, bralnik.

<sup>39</sup> <http://www.nua.ie/>

Z metodološke plati je seveda neugodno vprašati respondenta o določenem pojavu ali stvari, če imamo zanj štiri ali več oznak, ter hkrati slediti načelom dobrega vprašalnika. Respondent bo lahko pojem pod enim imenom prepoznal, pod drugim pa ne, nekdo drug pa ravno obratno. Delna rešitev je več navedb v vprašanje, vendar s tem kršimo načela optimalnega sestavljanja vprašalnika.

#### **4.5 Problem neodgovorov v anketah**

Problem nesodelovanja je prisoten v vseh vrstah anket. Pri proučevanju Interneta pa smo posebej zaskrbljeni nad možnostjo, da se osebe, ki niso bile vključene v raziskavo, v pogledu Interneta razlikujejo. Tovrstni pomisleki so seveda utemeljeni, saj je npr. po podatkih RIS97 med osebami, ki v gospodinjstvu nimajo telefona, štirikrat manj uporabnikov Interneta. V primeru visoke pokritosti s telefoni – in Slovenija se že bliža 90% pokritosti – to pravzaprav niti ni več tako pereč problem, čeprav še vedno ni zanemarljiv. Še naprej pa obstaja nevarnost, da so ne-respondenti (nedosegljive osebe, osebe, ki zavračajo sodelovanje, ipd.) v pogledu Interneta nekoliko drugačni. Dodati velja, da grob pregled deleža uporabnikov Interneta med respondenti, ki v telefonski anketi odgovorijo po prvem, drugem itd. kontaktu, sicer kaže na določeno nevarnost tovrstne napake, vendar le-ta ne more pomembneje vplivati na rezultate.

Posebej aktualna so vprašanja neodgovorov v tistih WWW anketah, ki temeljijo na samoizboru. Ko se oseba odloči, da bo sodelovala pri anketi na WWW, se namreč izbere sama; raziskovalec nima nobene možnosti prepričevanja. Velik potencialni problem se pojavi, če je nesodelovanje značilno le za določen segment populacije, kar bi lahko privedlo celo do tega, da raziskava nekaterih drugih segmentov sploh ne bi vključevala. Zagovorniki WWW anket sicer trdijo, da gre 'le' za primer neodgovorov, ki smo mu priča tudi pri klasičnem načinu anketiranja.

»Vse raziskave se do neke mere soočajo s tem problemom. Če potencialni respondent odloži slušalko med telefonsko anketo, ali ne odpošlje pisemske ankete, se pojavi samoselekcija.« (Kehoe in Pitkow, 1996a:79).

Seveda pa je problem, ki se pojavlja pri WWW raziskavah, nadvse resen in ga ni mogoče odpraviti z zgornjo ugotovitvijo. Problematika je podrobneje obdelana v naslednjem tekstu pričujoče publikacije, ki je v celoti posvečen metodologiji WWW anket.

#### **4.6 Problemi »computer-to-computer« tehnik**

Omejili se bomo na nekaj najpogostejših problemov tovrstnega raziskovanja:

a) Tipičen problem, ki se pojavlja pri vseh »computer-to-computer« tehnikah, je **stabilnost povezav**. Če namreč strežnik, s katerega opravljamo določene meritve, zaradi kakršnih koli razlogov preneha delovati ali pride do izpada povezav, pride tudi

do prekinitve merjenja. Ker se aplikacije, ki merijo obiske na WWW, pogosto nahajajo na strežnikih v ZDA, nastopi tu dodatna možnost za večjo nestabilnost, saj se lahko povezava med Evropo in ZDA prekine. Če našo stran 'nadzoruje' aplikacija, ki se nahaja na nekem strežniku v ZDA, se dostopi v času 'komunikacijskega mrka' seveda ne bodo šteli.

b) Naslednji problem je povezan s **preobremenjenostjo** povezav. Pogosto namreč zaradi preobremenjenosti povezav v Internetu prihaja do izpadov. Lahko se zgodi, da se zaradi prevelikega prometa na strežniku, na katerem je aplikacija, ki beleži dostope oz. obiske na naši strani, kljub dobri povezavi obiski ne beležijo.

Zaradi zgoraj navedenih razlogov lahko izgubimo precej podatkov, pri čemer gre večinoma za nevidno izgubo, ker so podatki o prekinitvah, preobremenjenosti ipd. težko ali sploh nedostopni, saj prihaja do teh težav na različnih točkah.

c) Obstoječi načini merjenja obiskov so v mnogih pogledih sporni, saj je metodologija, ki omogoča beleženje obiskov in ostalih meritev s »*computer-to-computer*« *tehnikami, v veliki meri nestandardizirana. Obstaja sicer dokument z oznako RFC*<sup>40</sup> *1262, ki podaja osnovna priporočila pri merjenju na Internetu. Vendar to ni standard, ampak so samo napotki, ki naj bi jih raziskovalci pri merjenju upoštevali. Osnovno vodilo predvsem usmerja dejavnost zbiranja podatkov k strokovnosti, brez negativnih posledic za uporabnike in storitve omrežja.*

d) Upoštevati je potrebno tudi **etična načela** raziskovanja. Zbrani podatki ne bi smeli kršiti določil o zasebnosti, varnosti in uporabi. Če podatki predstavljajo potencialno nevarnost vdora v zasebnost, bi moral raziskovalec zaščititi zasebnost teh podatkov; na primer z omejevanjem objavljenih statistik na tak način, da individualni uporabniki ali institucije ne morejo biti identificirane. Če obstaja možnost, da se lahko določena aktivnost zbiranja podatkov tolmači kot kršitev varnosti podatkov, bi morali raziskovalci o tem obvestiti CERT<sup>41</sup> in zaprositi za nasvet. Zbiranje podatkov se ne sme izvajati niti v primeru, ko bi zaradi takega merjenja prihajalo do nepravilnosti in napak v omrežju ali na priključenih računalnikih. V takem primeru je treba izbrati možnost 'neagresivnega' merjenja. Raziskovalni cilji, metodologija in načrti morajo biti dostopni javnosti tako, da jih je mogoče hitro najti in priklicati. Če obstaja možnost, da bi merjenje povzročilo neprimerno obremenitev za računalnik ali omrežje, na katerem poteka aplikacija, pa se raziskovalna dejavnost ne sme izvajati brez eksplicitnega dovoljenja (Cerf, 1991).

---

<sup>40</sup> RFC - (*Request for Comment*) je dokument, ki ga navadno izda ena izmed delovnih skupin IETF (*Internet Engineering Task Force*, ki predstavlja eno izmed tehnično najbolj pomembnih delovnih skupin IAB - *Internet Activities Board*, ki je zadolžena za razvoj tehničnih rešitev za reševanje kakršnih koli problemov v zvezi z Internetom). Namen teh dokumentov je privabiti odgovore vseh, ki jih zanima tovrstna problematika. Najdemo lahko več katalogov z vsemi do sedaj izdanimi RFC, eden izmed takih je dosegljiv na (<ftp://src.doc.ic.ac.uk/rfc/>), dokumenti so tudi na slovenski domeni na naslovu (<ftp://ftp.arnes.si/standards/rfc>).

<sup>41</sup> CERT - (*Computer Emergency Response Team*) je organizacija, ki so jo zaradi zaskrbljenosti glede varnosti v Internetu ustanovili leta 1982; v tem okviru obstaja tudi slovenska skupina SI-CERT.



Etična načela niso aktualna le v primeru varovanja osebnih podatkov in tajnosti osebnega početa na Internetu, ampak tudi v smislu komuniciranja med računalniki. Postavi se lahko npr. vprašanje, ali lahko nenajavljeno opazujemo in merimo določen strežnik v namene zbiranja statistik v raziskovalne namene ali pa v namene npr. industrijske špijonaže.

e) Računalniško spremljanje statistik in s tem povezana vprašanja (posebej število obiskov določene WWW strani) kljub vsemu ostajajo **pomanjkljiva**. Manjkajo namreč vsi podatki o osebi, ki je izvedla obisk. Družboslovno, posebej pa marketinško raziskovanje ne more biti uspešno, če predmet proučevanja niso tudi ključni sociodemografski podatki o osebah. Pomembno je torej, da se zavedamo, da je merjenje obiskov in dostopov šele polovica celotne 'zgodbe', saj nam taka merjenja malo povedo o vplivu obiskov na stališča in nakupne odločitve (Young, 1996:2).

#### **4.7 Problem globalnih statistik**

Mednarodne primerjave števila uporabnikov Interneta, števila računalnikov v omrežju ter ostalih značilnosti Interneta so postale zanimiv predmet najrazličnejših analiz. Dostop do Interneta in njegova uporaba je v zadnjih letih postala pomemben indikator tehnološke, predvsem pa informacijske razvitosti posameznih držav. Kljub aktualnosti problematike je kvalitetnih primerjav razmeroma malo, saj je težko primerjati rezultate, ki so pridobljeni z različnimi raziskavami oziroma raziskovalnimi metodami.

Trenutno težko govorimo o globalnih raziskavah o Internetu. Temu se v določeni meri približujejo:

- a) največja WWW anketa na svetu, ki jo izvaja *GVU*; rezultati so seveda omejeni zgolj na uporabnike WWW;
- b) globalnega značaja so tudi raziskave na nivoju infrastrukture, ki s pomočjo »*computer-to-computer*«*»* tehnik merijo število »hostov«, domen in omrežij v Internetu (npr. Internet Domain Survey, RIPE, MIDS);
- c) omeniti velja tudi sekundarne raziskave, ki povezujejo rezultate različnih raziskav. Objekta primerjav sta najpogosteje dva: število (delež) uporabnikov Interneta ter število »hostov«*»* (na 100.000 prebivalcev) v posameznih državah (npr. NUA, CyberAtlas, NetworkWizzard, EITO, C-I-A);
- d) pomembna so tudi prizadevanja večjih (internacionalnih) raziskovalnih organizacij (npr. Harris, Nielsen, Gallup, pa tudi Eurobarometer), ki stremijo h globalnim standardiziranim anketam med končno populacijo. V tem okviru lahko pričakujemo ne le uveljavitev novih (predvsem WWW) anket, ampak tudi umeščanje standardiziranih vprašanj glede uporabe Interneta v mednarodne ankete splošne populacije;

- e) podobno vzpostavljanje mednarodno primerljivih podatkov poteka tudi na področju podjetij; izvajajo ga večje raziskovalne organizacije, ki so specializirane za informacijsko tehnologijo (npr. IDC, Dataquest, IRIS).

V Tabeli 7 je prikazana tipična mednarodna primerjava za nekatere evropske države. V tabeli so navedeni uporabniki Interneta (osebe, ki so uporabile Internet v zadnjih treh mesecih) ter delež uporabnikov v celotni populaciji. V zadnjem stolpcu je na osnovi podatkov iz tabele 2 izračunano še razmerje med skupnim številom uporabnikov Interneta (glede na datum, naveden v tabeli 7) ter številom računalnikov, vključenih v omrežje (Januar 1998).

	Datum	Skupno število uporabnikov Interneta v tisočih (000)	Uporabniki Interneta v celotni populaciji (%)	Št. uporabnikov Interneta na en računalnik vklj. v Internet
Avstrija	Nov 1997	500	6,3	4,5
Belgija	Sep 1997	200	1,9	1,8
Češka	Sep 1997	200	1,9	3,3
Danska	Sep 1997	600	11,5	3,5
Estonija	Jul 1997	108	7,2	6,4
Finska	Nov 1997	1040	20,4	2,1
Francija	Sep 1997	400	0,7	1,0
Grčija	Nov 1997	100	1,0	3,8
Irska	Sep 1997	100	2,9	2,4
Islandija	Feb 1998	122	45,0	6,3
Italija	Jan 1998	700	1,2	2,8
Madžarska	Mar 1997	100	1,0	1,4
Nemčija	Feb 1998	5800	7,0	5,1
Nizozemska	Sep 1997	1000	6,4	2,5
Norveška	Nov 1997	1400	32,5	4,7
Poljska	Nov 1997	700	1,8	7,5
Rusija	Okt 1997	600	0,4	5,5
Portugalska	Sep 1997	200	2,0	4,7
Slovaška	Nov 1997	190	3,5	12,5
Slovenija	Feb 1998	*180	*9,0	*9,2
Španija	Nov 1997	1340	3,4	6,6
Švedska	Jul 1997	1900	21,3	5,3
Velika Britanija	Aug 1997	6000	10,3	5,7

**Tabela 7:** Uporabniki Interneta v evropskih državah (Vir: NUA, ARNES, RIS);

\* za Slovenijo je dodatno metodološko pojasnilo v nadaljevanju teksta.

Zgornji rezultati kljub določenim slabostim ilustrirajo razširjenost Interneta in se v grobem ujemajo tudi s številom računalnikov (»host«) vključenih v Internet (Tabela 2). Razmerje uporabnikov in računalnikov je nadvse različno, delno zaradi časovnega razkoraka med podatkom (npr. Madžarska), delno zaradi specifičnosti uporabe

Interneta v posameznih državah, v določeni meri pa tudi zaradi neusklajenih metodologij.

Slovenija v tem pogledu – zaradi specifične širitve Interneta – izrazito izstopa: šolajoče generacije namreč še vedno predstavljajo večino uporabnikov Interneta v Sloveniji. Prav ti uporabniki višajo navedeno razmerje, saj vstopajo v Internet z računalniki, ki imajo veliko uporabnikov (npr. šola, cyber-cafe, knjižnice, prijatelji). Tako razlago potrjuje tudi relativno majhen delež uporabnikov Interneta, ki imajo lastni e-mail naslov.

V primeru Slovenije je treba poudariti, da so bili kot uporabniki – v skladu s priporočili, ki jih daje NUA – upoštevane osebe, ki so uporabile Internet v zadnjih treh mesecih. V primeru vseh oseb, ki so kdaj uporabile Internet, je razmerje bistveno višje, npr. skoraj 13 uporabnikov na en računalnik (»host«), kar seveda ni primerljivo z ostalimi državami. V resnici je ravno nasprotno: kljub priporočilu, ki ga daje NUA o opredelitvi uporabnika Interneta, več držav (npr. ZDA, Nemčija, Velika Britanija) poroča samo o aktivnih (rednih) uporabnikih, pri čemer je definicija aktivnosti nadvse različna. V primeru, da štejemo v Sloveniji kot aktivne samo uporabnike, ki dostopajo do Interneta vsaj mesečno, se delež znatno zniža, in sicer populacijski delež uporabnikov na 6.5%, razmerje uporabnik-host pa na 6.6. Ocenjujemo, da je z vidika primerljivosti podatkov v Tabeli 7 taka ocena za Slovenijo tudi najbolj ustrezna. Če pa bi kot aktivne razumeli samo uporabnike, ki dostopajo do Interneta tedensko, sta odgovarjajoči vrednosti 4.5% in 4.6.

Iz zgornjega prikaza je tudi razvidno – z izrazito izjemo skandinavskih držav – veliko zaostajanje Evrope za ZDA, kjer je bilo v februarju 1998 že 62 milijonov rednih uporabnikov Interneta, kar predstavlja 23% celotne populacije ZDA.

Zgornja tabela razkriva tudi tipične težave mednarodnih primerjav. Tako so očitni neusklajeni datumi raziskav, ki lahko pri pojavu s skoraj stodontno letno rastjo povzročajo velike razlike (npr. Madžarska). O nadaljnjih razlikah v metodologiji lahko le ugibamo; le delno si namreč lahko pomagamo z WWW predstavitvijo nacionalnega vira podatkov, ki je naveden poleg podatkov na Internetu<sup>42</sup>. Zgornjo tabelo je zato potrebno razumeti predvsem ilustrativno, saj ji neusklajenost metodologij in pomanjkljiva dokumentacij, jemljeta resnejšo veljavnost. V tem okviru so značilne tudi različne ocene, kadar sta za isto državo na voljo dva vira (tovrstne razlike najdemo tako v Rusiji kot v ZDA), kar nadalje potrjuje težavnosti tovrstnih primerjav.

Posebej izstopa npr. visok odstotek uporabnikov pri Slovaški, posebej v primerjavi s Češko, pa tudi glede na število računalnikov, vključenih v Internet. Velika verjetnost je, da gre za oceno, ki metodološko ni usklajena z ostalimi. Izstopajo tudi visoki deleži uporabnikov v Estoniji in Sloveniji ter nizki odstotki v Belgiji in Franciji, za kar pa obstajajo znani vsebinski razlogi.

---

<sup>42</sup> <http://www.nua.ie>

## 5. Sklep

Internet je dinamičen koncept, ki povezuje ljudi, organizacije, računalnike ter nepregledne množice najrazličnejših vsebin. Število uporabnikov, dejanski obseg rabe Interneta in tudi njegova vsebina so v veliki meri neznanka, ki jo lahko raziskave zaradi hitre rasti ocenijo le delno. Predvsem velja upoštevati, da rezultati zastarajo izredno hitro. Zato se npr. pojavljajo predlogi, da bi morale za kvalitetne primerjave znotraj posameznih raziskovalnih projektov med izvajanjem dveh merjenj preteči manj kot 90 dni (CyberAtlas, 1997). Z drugimi besedami: opazne – ali celo pomembne – spremembe se dogajajo že v treh mesecih.

Zaradi nagle širitve Interneta je razumljivo, da odgovarjajoči koncepti, teorije in konstrukti pogosto zaostajajo za hitro spreminjajočo se realnostjo. Povsem mogoče je tudi, da bo Internet v sedanji tehnološki obliki že v nekaj letih izginil oziroma se preobrazil in zliil z drugimi tehnologijami. V tem okviru je pomembno predvsem povezovanje osebnega računalnika (PC), televizije (WEBTV) in telefona (videofona), kar vse lahko odmeri Internetu povsem novo vlogo.

Z vidika samega raziskovanja Interneta pomeni hiter razvoj predvsem dejstvo, da realnost neprestano prehitava obstoječo metodologijo raziskovanja. Merjenje Interneta zato predstavlja nelahko nalogo, saj smo še daleč od tega, da bi razpolagali z vzpostavljenim, celovitim in zaokroženim instrumentarijem statističnega spremljanja. Problem pomanjkanja standardov je najbolj pereč na področju osnovnih definicij ter na področju mednarodnih primerjav. Posebej kritično je npr. dejstvo, da ni splošno sprejetega dogovora, kdo je uporabnik Interneta in kaj pravzaprav Internet obsega. Zaradi vseh naštetih težav je tudi razumljivo, da vsebin, povezanih z Internetom, ne merijo uradni zavodi za statistiko, ampak predvsem komercialne in akademske raziskave.

Verjetno je v okviru raziskovanja Interneta poleg samega ugotavljanja števila uporabnikov najbolj tipično in tudi najbolj aktualno prav merjenje obiskanosti WWW strani. Kljub številnim nedorečenostim glede terminologije in kljub določenim tehničnim težavam je danes mogoče razmeroma natančno meriti dostope do posameznih WWW strani. Karakteristično pa je, da kljub podrobnim podatkom o obisku, vemo zelo malo o tem, kdo so obiskovalci. Zaenkrat seveda ne kaže, da bi lahko statistike obiskanosti avtomatično vključile tudi demografski profil obiskovalcev. Lastniki WWW strani poskušajo problem delno rešiti tako, da obiskovalce silijo v postopek registracije, kar ima seveda številne omejitve. Anonimnost obiskovalcev, ki je ena pomembnejših lastnosti Interneta, je torej pri statistikah obiskanosti njegova poglobljena pomanjkljivost. Zato ostajajo izjemne tehnološke možnosti merjenja obiskanosti v veliki meri neizkoriščene, razvoj metodologije pa se razvija predvsem v smeri že uveljavljenega proučevanja TV avditorija, ki največkrat temelji na panelnem opazovanju vzorca respondentov.

Kljub nemajhnim težavam pri statističnem spremljanju Interneta, velja dodati, da se terminološki in metodološki standardi, čeprav z določeno zamudo, hitro vzpostavljajo.

Predvsem velja to za fenomene, ki so vezani izključno na Internet (npr. obiskanost, domene, računalniki, povezani v Internet). Seveda posega Internet tudi na številna siceršnja področja našega življenja, kjer pa že obstajajo uveljavljene metodologije spremljanja pojavov (npr. ekonomska statistika, statistika medijev).

Globalni značaj tehnologije Interneta pomembno vzpodbuja tudi oblikovanje standardiziranega in mednarodno primerljivega merjenja. Zaradi nuje po mednarodno primerljivih ter hitro posredovanih rezultatih kot tudi zaradi zahtevnosti takih raziskav in obsežnosti s tem povezanega usklajevanja so pri tovrstnih projektih učinkovite predvsem globalne komercialne organizacije kot npr. International Data Corporation, Nielsen, Gallup, pa tudi novonastale organizacije, specializirane za ožje področje Interneta, npr. NUA, MIDS, Network Wizzard. Navedena podjetja opravijo torej tudi primarno zbiranje podatkov, kar je v primeru »*computer-to-computer* tehnik« – kljub vsem težavam – razmeroma enostavno.

Mednarodno usklajena nacionalna statistična raziskovanja, ki jih spremlja dolgotrajno usklajevanje metodologij - kar je običajno ne samo na področju uradne statistike, ampak tudi pri (mednarodnih) raziskavah rodnosti, AIDS-a, pismenosti, pa tudi socioloških in politoloških vsebin ipd. - so za globalno proučevanje Interneta manj primerna. Pričakovati je torej, da bo globalna narava Interneta pokazala svojo moč tudi v pogledu statističnega spremljanja in kmalu vzpostavila vse potrebne mednarodne statistike. S tem bo presegla običajne omejitve in težave, ki pri tem največkrat nastajajo. Glede na dejstvo, da Internet obstaja kot javni fenomen šele nekaj let, lahko zaključimo, da je doseženi nivo statističnega spremljanja pravzaprav zadovoljiv. Trendi na tem področju pa napovedujejo, da bodo hitro in učinkovito opravljeni tudi nadaljnji koraki statističnega spremljanja.