

PRESEK

List za mlade matematike, fizike, astronome in računalnikarje

ISSN 0351-6652

Letnik 25 (1997/1998)

Številka 6

Strani 342-347

Aleksander Vesel:

PRIPOMOČKI ZA DESKANJE PO INTERNETU

Ključne besede: računalništvo, Internet, HTTP, WWW, spletno iskalo, hierarhično kazalo, spletni brskalnik.

Elektronska verzija: <http://www.presek.si/25/1354-Vesel.pdf>

© 1998 Društvo matematikov, fizikov in astronomov Slovenije

© 2010 DMFA - založništvo

Vse pravice pridržane. Razmnoževanje ali reproduciranje celote ali posameznih delov brez poprejšnjega dovoljenja založnika ni dovoljeno.

PRIPOMOČKI ZA DESKANJE PO INTERNETU

Nastanek Interneta

Za začetek je treba poseči več kot trideset let nazaj v Združene države Amerike. To je bilo obdobje *hladne vojne* oziroma stalnih napetosti med tedanjima velesilama. To je bil čas, ko računalniki še zdaleč niso bili tako razširjeni kot sedaj. Privoščile so si jih lahko samo večje raziskovalne ustanove, velika podjetja in nekateri državni organi, predvsem vojska. Računalnike so takrat že povezovali v *računalniška omrežja*, v njih pa so prevladovali vojaški računalniki.

Hladna vojna je silila sodelujoče v priprave na morebitni resni izbruh sovražnosti. Mednje je sodilo tudi preučevanje posledic jedrskega napada na ZDA. Pri tem so ugotovili, da bi uničenje nekaj velikih ameriških računalniških središč povsem ohromilo računalniške komunikacije. V želji po rešitvi tega problema so obstoječe centralizirano omrežje postopoma preoblikovali. V novo nastalo omrežje so bili enakopravno vključeni računalniki in manjša omrežja ameriškega obrambnega ministrstva ter nekaterih raziskovalnih ustanov. Taka zasnova je omogočala delovanje omrežja tudi ob nenadnem izpadu večjega števila računalnikov ali računalniških povezav. Raziskovalcem, ki so novo omrežje razvijali, je bilo verjetno bolj ali manj vseeno, kaj se bo zgodilo z omrežjem po jedrskem uničenju, a vojaki očitno razmišljajo drugače in znanstveniki so to uspeli izkoristiti. Za znanstvenike je bilo pomembno, da so povezani računalniki omogočali boljšo izrabo računalniških zmogljivosti, hiter in zanesljiv prenos podatkov med njimi in pošiljanje elektronskih sporočil oziroma *elektronsko pošto*. Nastalo omrežje so leta 1969 poimenovali ARPANET in iz njega je zrasel današnji Internet.

Osemdeseta leta so v takratno Sovjetsko zvezo prinesla perestrojko in s tem tudi dolgo pričakovano otoplitev med velesilama. Leta 1986 je nadzor nad omrežjem ARPANET dobila civilna znanstvena organizacija NSF (National Science Foundation). Za znanstvenike je, v nasprotju z vojaki, izredno zanimiva izmenjava informacij in idej. Zato so začeli podpirati povezovanje raziskovalnih, kasneje pa tudi komercialnih organizacij v omrežje, ki je kmalu dobilo globalne razsežnosti in povežalo večino razvitega sveta.

Uporabniki Interneta so se hitro navadili, da so lahko v nekaj minutah dosegli znanca na drugem koncu sveta z elektronsko pošto. Marsikoga je pritegnila izmenjava mnenj na elektronskih konferencah, spet druge (predvsem mlajše uporabnike) pa elektronski klepet z vrstniki po svetu. Kljub globalni razširjenosti pa so do začetka devetdesetih let Internet uporabljali predvsem raziskovalci in šolajoča se mladina.

Leta 1989 so za pomembno novost poskrbeli Evropejci. Zanimivo je, da ta ni prišla z računalniške ustanove, temveč z Evropskega centra za fiziko delcev (CERN) v Ženevi. Tamkajšnje raziskovalce (anekdota pravi, da predvsem starejše profesorje) je motila množica različnih *protokolov* oziroma načinov sporazumevanja med računalniki. Razvili so protokol HTTP. Novi protokol je bil namenjen prenosu *hiperbesedil*. Hiperbesedilo je datoteka, ki poleg navadnega besedila vsebuje še posebej označene besede. Te predstavljajo povezave, prek katerih se lahko preselimo na druge datoteke poljubne vsebine (hiperbesedilo, slika, zvočni zapis itd.). Za oblikovanje hiperbesedil so razvili poseben jezik in ga poimenovali HTML.

Prednosti novega načina predstavitve in prenosa podatkov so kmalu spoznali tudi na drugih raziskovalnih ustanovah in v nekaterih računalniških podjetjih. Na računalnikih, povezanih v Internet, so uporabnikom začeli ponujati datoteke, napisane v jeziku HTML in dosegljive s protokolom HTTP. Nastal je *Svetovni splet* ali *Splet* (angleški izraz je World Wide Web ali krajše WWW).

Sprehajanje ali *deskanje* po spletnih datotekah oziroma *spletnih straneh* je bilo dokaj okorno, dokler ni leta 1993 študent z univerze v Illinoisu napisal svoj program za deskanje po Spletu ali *spletni brskalnik*. Poimenoval ga je Mosaic, njegova poglavitna prednost v primerjavi s sodobniki pa je bila, da je deloval v grafičnem načinu in je omogočal deskanje s pomočjo miške. Tako je postal Splet dostopen najširšemu krogu ljudi, tudi tistim, ki se z računalništvom poklicno ali kako drugače podrobneje ne ukvarjajo. Ocenjujejo, da je danes v Internet vključenih že več deset milijonov računalnikov, kakšna desetina od teh pa ponuja skupaj okoli 50 milijonov spletnih strani.

Ogromno število spletnih strani seveda pomeni tudi veliko količino bolj ali manj zanimivih informacij. Žal ta razbohotenost pomeni tudi težave pri njihovem iskanju in obvladovanju. Dodatno težavo predstavlja decentralizirana, skoraj anarhična zasnova Interneta. Spletne strani so po obliki zelo heterogene, mnogo je podvajanj in prekrivanj informacij. Splet v marsičem spominja na živ organizem, saj se spletne strani ves čas spreminjajo, vsak dan pa nastane na tisoče novih in ugasne nekaj starih spletnih strani. Brez nedavno razvitih orodij za iskanje in razvrščanje spletnih informacij bi bila zato uporabnost Svetovnega spleta precej omejena.

Orodja za iskanje po Spletu lahko razdelimo v tri skupine:

1. spletna iskala,
2. hierarhična kazala,
3. spletna meta iskala.

Spletna iskala

Namen spletnega iskala je zbrati podatke o spletnih straneh v kazalo in omogočiti uporabnikom Spleta iskanje po teh kazalih. V kazalu so shranjeni deli spletnih strani oziroma *ključne besede*. Za izdelavo kazal se uporabljajo posebni programi, ki noč in dan preiskujejo Splet in z zbranimi podatki zagotavljajo aktualnost kazala. Ti programi se imenujejo *roboti* ali *pajki*. Prvi tak program je bil Lycos (njegov spletni naslov je <http://www.lycos.com>), ki je dobil ime po neki vrsti pajka. Tako kazalo lahko potem uporabniki Interneta ali *deskarji* preiskujejo s posebnimi, v ta namen razvitimi programi. Običajno ali enostavno iskanje poteka tako, da uporabnik v spletno iskalo vnese zaporedje ključnih besed, za katero domneva, da ga bo pripeljalo do željene infomacije oziroma iskane spletne strani. Rezultat iskanja je seznam spletnih strani (oziroma njihovih naslovov), ki vsebujejo katero od vnesenih besed.

Kot primer si pogledjmo iskanje spletnih strani, ki vsebujejo podatke o znanem francoskem matematiku (bil je tudi eden od začetnikov računalništva) Blaiseu Pascalu. Pri enostavnem načinu iskanja bi najverjetneje vnesli zaporedje:

```
blaise pascal .
```

Pozoren bralec je verjetno opazil, da v zaporedju, v nasprotju s prapopisom, nismo uporabili velikih začetnic. S tem smo zahtevali, da iskalo poišče spletne strani, ki vsebujejo vsaj eno od ključnih besed, napisanih v kakršnikoli kombinaciji malih in velikih črk. Nekatera spletna iskala namreč ločijo velike in male črke. Če v vneseni ključni besedi uporabimo vsaj eno veliko črko, bo takšno spletno iskalo poiskalo samo tiste strani, ki vsebujejo besede, ki se popolnoma ujemajo z vneseno. Ker ni nobene pravila o uporabi velikih in malih črk v spletnih straneh, je v večini primerov varneje uporabiti samo male črke.

Zelo znani spletni iskali sta še AltaVista (<http://altavista.digital.com>) in Excite (<http://www.excite.com>). Za svoje delovanje potrebujejo spletna iskala zelo močno aparaturno opremo. AltaVista je v času, ko nastaja ta prispevek, opremljena s triindvajsetimi zelo zmogljivimi računalniki AlphaServer in AlphaStation. Šestnajst od teh je posebej dobro opremljenih: vsak ima dvanajst procesorjev in 8 GB hitrega pomnilnika, za shranjevanje kazal pa uporabljajo več diskov s kapaciteto 300 GB. Za primerjavo dodajmo, da pomnilniške kapacitete enega od teh računalnikov ustrezajo skupnim pomnilniškim kapacitetam 500 srednje zmogljivih osebnih računalnikov.

Hitrost pa ni edino merilo za kvaliteto spletnega iskala. Pomembni dejavniki kvalitete so še:

1. pogostost obnavljanja podatkov v kazalu,
2. število spletnih strani, zajetih v kazalu ter
3. iskalne metode, vgrajene v spletno iskalo.

Bežno si oglejmo le najpomembnejše iskalne metode. Večina spletnih iskal omogoča uporabo predpon $\boxed{+}$ in $\boxed{-}$, s katerima lahko izrecno zahtevamo ali izključimo besede (oziroma besedne zveze). Tako postane iskanje natančnejše. Pri tem je treba paziti, da med predpono in besedo ne vpišemo presledka. Prej omenjeno iskanje podatkov o Blaiseu Pascalu ne bilo posebej uspešno, saj bi poleg iskanih podatkov o slavnem matematiku med drugim našli še množico spletnih strani, na katerih se nadobudni programerji hvalijo, kakšne neverjetne programe v programskem jeziku pascal so napisali. Število nepotrebnih spletnih strani bi lahko zmanjšali, če bi v spletno iskalo vnesli izraz

$\boxed{+blaise+pascal}$,

s katerim zahtevamo, da najdena spletna stran vsebuje obe ključni besedi. Še bolj bi število najdenih spletnih strani skrčili z izrazom

$\boxed{+blaise+pascal-"programming language"}$,

saj bi na ta način izključili strani, ki vsebujejo podatke, povezane s programskim jezikom pascal. Omenimo, da smo z narekovaji zahtevali tako imenovano *blizino* oziroma iskanje besedne zveze. Tako smo v zadnjem opisanem primeru izključili tiste strani, pri katerih besedi **programming** in **language** stojita zaporedoma.

Seveda je lahko takšno omejevanje števila najdenih spletnih strani tudi dvorezno. Hitro se nam namreč lahko zgodi, da je najdenih strani premalo in da iskalo kakšno od zanimivih spletnih strani spregleda. Način iskanja je zato potrebno čim bolj prilagoditi cilju iskanja. Če iščemo čisto določeno informacijo, npr. rojstni datum Blaisea Pascala, je smiselno iskanje čim bolj omejiti. Včasih pa nas zanimajo vse informacije o neki tematiki, npr. življenje in delo Blaisea Pascala. V tem primeru moramo paziti, da iskanje zastavimo širše in da katere od zanimivih strani ne spregledamo.

Mnoga spletna iskala poznajo tudi zahtevnejše načine iskanja. Tako lahko v AltaVisti iskanje omejimo na posamezne dele spletnih dokumentov. Iščemo lahko npr. samo med besedami iz naslovov ali pa med imeni slik v spletnih straneh. Strokovnjaki, ki skrbijo za spletno iskalo Excite, pa so razvili *koncept inteligentnega iskanja* ICE, ki omogoča, da se iskanje razširi tudi na tiste spletne strani, ki sicer ne vsebujejo nobene od vnesenih ključnih besed, vsebujejo pa besede, ki so z njimi pomensko povezane.

Hierarhična kazala

Podobno kot pri spletnih iskalih so tudi tukaj podatki o spletnih straneh shranjeni v kazalu. Od spletnih iskal pa se razlikujejo v dveh pomembnih značilnostih:

- podatki o spletnih straneh so hierarhično organizirani,
- razvijajo in vzdržujejo jih ljudje oziroma posebne skupine strokovnjakov.

Hierarhična organiziranost pomeni, da so vse zajete spletne strani razdeljene v več velikih področij. V priljubljenem hierarhičnem kazalu Yahoo! (<http://www.yahoo.com>) so to: umetnost in humanistika, gospodarstvo, računalniki in Internet, izobraževanje, razvedrilo, država, zdravje, novice in občila, rekreacija in šport, imeniki, regije, znanost, družbene vede ter družba in kultura (slika na III. strani ovitka). Našteta področja se delijo na manjša, ta na še manjša itd. Zaradi take organiziranosti je iskanje informacij ponavadi enostavnejše. Podatke o Blaiseu Pascalu lahko najdemo tako, da na področju **znanost** izberemo podpodročje **matematika**, tam pa podpodročje **matematiki** (v času, ko je nastajal ta prispevek, je ena od dveh spletnih strani, ki ju najdemo na opisani način, že ugasnila, kar dobro ponazarja hitro spreminjanje Svetovnega spleta).

Ker za oblikovanje in vzdrževanje hierarhičnih kazal ne skrbijo računalniški programi, vsebujejo precej manj spletnih strani kot kazala spletnih iskal. Njihova pomembna prednost pa je, da so vsebovane spletne strani skrbno izbrane in razvrščene. Posebne skupine strokovnjakov odločajo o tem, katera spletna stran bo izbrana, in o tem, na katerem mestu v hierarhičnem kazalu bo prisotna. Zaradi tega so hierarhična kazala zelo priljubljena, tako da so jih začeli vgrajevati tudi v nekatera spletna iskala (npr. v Excite in Lycos).

Omenimo še, da je v hierarhično kazalo Yahoo! vgrajena dodatna pomoč uporabniku, program MyYahoo!. Ta spada v novo vrsto programov, pisanih za Internet, ki jih imenujemo *agenti*. Agenti so "inteligentni" vmesniki, ki nudijo deskarjem posebne storitve. Sposobni so se prilagajati določenemu okolju in z njim komunicirati. Gre za področje računalništva, ki se šele razvija. Današnji agenti nudijo pomoč pri delu z elektronsko pošto, pri navezovanju stikov med deskarji, pri pridobivanju in filtriranju informacij s Spleta itd. MyYahoo! omogoča, da uporabnik

vnese svoje osebne podatke in izbere področja, ki ga zanimajo. Na podlagi teh podatkov program ustvari posebno spletno stran, ki se dnevno spreminja glede na tiste novosti s Spleta, za katere program presodi, da bi uporabnika lahko zanimala.

Spletna meta iskala

Kot smo že povedali, se spletna iskala med seboj razlikujejo. Zato ni presenetljivo, da najdeta dve različni spletni iskali za isto zaporedje vnesenih ključnih besed običajno dva različna seznama spletnih strani. Če želimo popolnejšo informacijo o stvari, ki nas zanima, moramo torej uporabiti obe spletni iskali. Več spletnih iskal uporabimo, več spletnih strani najdemo in popolnejša je dobljena informacija.

Takšno večkratno iskanje je seveda dolgočasno, vzame pa lahko tudi zelo veliko časa. Spletna meta iskala so pravzaprav orodja, ki to težaško delo naredijo namesto nas. To so posebni programi, ki za iskanje uporabljajo več "tujih" spletnih iskal in zato nimajo lastnega kazala spletnih strani. Rezultat iskanja je spet seznam najdenih spletnih strani, narejen z združitvijo seznamov, ki jih priskrbijo uporabljena spletna iskala. Primer spletnega meta iskala je MetaCrawler (<http://www.metacrawler.com>), ki si pri delu pomaga z devetimi spletnimi iskali.

Pogled v prihodnost

Napovedovanje je v računalništvu silno neprijetna naloga. Ne bomo pa se dosti zmotili, če napovemo, da bo šel razvoj predvsem v smeri večje "inteligence" iskalnih orodij. Že omenjeni agenti in koncept inteligentnega iskanja so verjetno le predhodnica orodij, ki bodo uporabniku omogočala, da bo lažje izluščil čim več uporabnih informacij.

Drug pristop k boljšemu obvladovanju Spleta pa je večja specializacija iskalnih orodij. Ker morajo specializirana iskala obvladovati le manjši del Spleta, so tako zbrani podatki običajno kvalitetnejši, pa tudi iskanje željene informacije je enostavnejše. Danes že obstajajo specializirana iskala, ki so namenjena iskanju podatkov o programski opremi, kompaktnih ploščah, knjižnih izdajah, revijah itd. Pojavljajo pa se tudi iskala, ki so namenjena iskanju podatkov v posameznih državah in regijah (npr. Mat'Kurja na naslovu <http://www.ijs.si/slo> za iskanje po slovenskem delu Spleta).



Yahoo! Games
chess, bridge, hearts...

win 2 tickets and \$5,000
to tour the globe! theglobe.com click here now!



[Oscar® Winners](#)

Search [options](#)

Yahoo! Online – \$14.95 a month Internet access – [sign up today](#)

[Yellow Pages](#) – [People Search](#) – [Maps](#) – [Classifieds](#) – [Personals](#) – [Chat](#) – [Email Shopping](#) – [My Yahoo!](#) – [News](#) – [Sports](#) – [Weather](#) – [Stock Quotes](#) – [more...](#)

- **Arts and Humanities**
[Architecture](#), [Photography](#), [Literature...](#)
- **Business and Economy [Xtra!]**
[Companies](#), [Finance](#), [Employment...](#)
- **Computers and Internet [Xtra!]**
[Internet](#), [WWW](#), [Software](#), [Multimedia...](#)
- **Education**
[Universities](#), [K-12](#), [College Entrance...](#)
- **Entertainment [Xtra!]**
[Cool Links](#), [Movies](#), [Music](#), [Humor...](#)
- **Government**
[Military](#), [Politics \[Xtra!\]](#), [Law](#), [Taxes...](#)
- **Health [Xtra!]**
[Medicine](#), [Drugs](#), [Diseases](#), [Fitness...](#)
- **News and Media [Xtra!]**
[Current Events](#), [Magazines](#), [TV](#), [Newspapers...](#)
- **Recreation and Sports [Xtra!]**
[Sports](#), [Games](#), [Travel](#), [Autos](#), [Outdoors...](#)
- **Reference**
[Libraries](#), [Dictionaries](#), [Phone Numbers...](#)
- **Regional**
[Countries](#), [Regions](#), [U.S. States...](#)
- **Science**
[CS](#), [Biology](#), [Astronomy](#), [Engineering...](#)
- **Social Science**
[Anthropology](#), [Sociology](#), [Economics...](#)
- **Society and Culture**
[People](#), [Environment](#), [Religion...](#)

[What's New](#) – [Weekly Picks](#) – [Today's Web Events](#) – [Yahoo! Internet Life](#) – [Message Boards](#)
[Yahoo!igans!](#) – [Seniors' Guide](#) – [Apply for a Yahoo! Visa Card](#) – [Yahoo! Gear](#) – [3D Stock Viewer](#)

World Yahoos [Australia & NZ](#) – [Canada](#) – [Denmark](#) – [France](#) – [Germany](#) – [Japan](#) – [Korea](#)
[Norway](#) – [SE Asia](#) – [Sweden](#) – [UK & Ireland](#)

Yahoo! Metros [Atlanta](#) – [Austin](#) – [Boston](#) – [Chicago](#) – [Dallas / Fort Worth](#) – [Los Angeles](#)
Get Local [Miami](#) – [Minneapolis / St. Paul](#) – [New York](#) – [S.F. Bay](#) – [Seattle](#) – [Wash.D.C.](#)

Smart Shopping with

[How to Suggest a Site](#) – [Company Info](#) – [Privacy Policy](#) – [Contributors](#) – [Yahoo! How-To](#)