

Rezultati mjerenja hrvatskog Web prostora

Stručni tim formiran u Sveučilišnom računskom centru proveo je od 29. ožujka do 7. svibnja 2002. mjerenje koje je imalo za cilj prikupiti informacije o veličini i sadržaju hrvatskog web prostora. Mjerenjem se prije svega željelo ustanoviti:

- veličinu web prostora,
- korištene formate datoteka (tipove) prema MIME standardu,
- omjer teksta, slike, audio i video zapisa,
- obim i sadržaj meta podataka.

Mjerenjem su obuhvaćeni elektronički resursi dostupni HTTP mrežnim protokolom s poslužitelja u .hr vršnoj internetskoj domeni.

Izmjereni uzorak obuhvatio je 6.564 web poslužitelja s kojih je s uspjehom obrađeno 4.667.920 resursa.

Utvrđeno je da se kao programska podrška na 57 posto web sjedišta koristi Apache Web poslužitelj dok se Microsoft Internet Information Server koristi u 29 posto slučajeva.

Za 79 posto obrađenih resursa izmjerena je ukupna veličina od 263,4 GB. Veličina preostalih 21 posto procijenjena je na 55,3 GB što ukupno daje procjenu od preko 300 GB podataka na javno dostupnim web stranicama u .hr vršnoj internetskoj domeni.

Glede tipova i formata podataka najveći broj resursa, čak 67 posto otpada, prema očekivanju na HTML koji što se obima tiče zauzima tek 15 posto uz prosječnu veličinu od 14 KB. Na slikovne formate otpada 23 posto resursa po broju odnosno samo 6 posto po veličini, dok različiti aplikativni formati (kao što su PDF, Microsoft Word, arhive poput TAR ili ZIP) iako po broju zauzimaju nepunih 7 posto, obuhvaćaju svojom ukupnom veličinom 62 posto mjerenog prostora. Ukupni udio video i audio zapisa svodi se na skromnih 1 posto po broju i ispod 10 posto po veličini.

Uporaba meta podataka registrirana je u 31 posto web stranica (HTML resursa). Utvrđeno je da se Dublin Core standard koristi na samo 0,09 posto web stranica. 25 posto web stranica ima meta podatke koje automatski upisuju alati za izradu web stranica, a gotovo 20 posto web stranica ima zapisane meta podatke prema preporukama internetskih tražilica.

Dobiveni rezultati odgovaraju očekivanjima, ali i rezultatima sličnih istraživanja provedenih u svijetu. Ovo se posebno odnosi na procjene veličine i složenosti web prostora. Dobiveni su očekivani rezultati vezani uz tipove resursa koji potvrđuju tezu o relativno malom broju različitih formata koji se koriste na webu. Analiza meta podataka također je ponudila očekivane rezultate, iako brojnost grešaka ukazuje na nebrigu autora za taj segment web tehnologija.

Prilikom mjerenja, posebni su izazov bile inventivne, ali nestandardne uporabe web tehnologija koje onemogućuju rad konvencionalnih mehanizama za pobiranje (download) podataka.

Kao poseban rezultat mjerenja valja istaći procjenu da uz razvoj kvalitetnih alata za ciljano pobiranje web resursa nužno treba prirediti i preporuke za autore i izdavače kako bi njihovi resursi bili spremniji za pobiranje.

Prikupljeni podaci svakako su temelj za daljnje, ozbiljnije analize i istraživanja hrvatskog Web prostora.

Izloženi rezultati nastali su temeljem suradnje Nacionalne i sveučilišne knjižnice i Sveučilišnog računskog centra Sveučilišta u Zagrebu, na projektu Mjerenje hrvatskog Web prostora za potrebe projekta NISKA".