

Web2.0 - a gdje su podaci?

Dobrica Pavlinušić

CARNet

Dobrica.Pavlinusic@CARNet.hr

098-304-971

Sažetak

U današnje doba Interneta problem dostupnosti informacija napokon više nije zanimljiva tema svakodnevnog razgovora. Baš naprotiv, tolika dostupnost stvorila je sasvim novi problem viška informacija koje nas napadaju sa svih strana i zahtijevaju od nas dodatan napor u odabiru samo onih koje su nam stvarno korisne.

Zbog toga ćemo u ovom radu pokušati pogledati koje sve tipove informacija susrećemo i prikazati jedan od mogućih načina za organiziranje različitih izvora informacija u korisnu cjelinu koristeći IRC, RSS, tagove i mnogo mašte.

1. Da li je Web 2.0 džungla podataka?

Da bi lakše pogledali sa kojim se sve Web2.0 podacima svakodnevno susrećemo, podijelit ćemo ih prema načinu primanja: oni mogu dolaziti do nas bez naše intervencije (*push*), ili moramo po podatke otići na neku adresu na mreži (*pull*). Da bi bilo jasnije, pogledajmo nekoliko najpopularnijih mrežnih servisa prema ovoj podjeli:

- *pull* (korisnik "odlazi" po podatke na mrežu)
 - Web (HTTP)
 - RSS (Atom, podcasts...)
- *push* (dolaze bez intervencije korisnika)
 - e-mail (SMTP)
 - Instant Messaging (IRC, IRQ, Jabber...)

Ako pogledamo ovu podjelu od vrha prema dnu, primijetit ćemo da podaci imaju različitu "svježinu" u zavisnosti od toga kojim načinom dolaze do nas. Štoviše, mi zapravo koristimo različite načine komunikacije s istim ljudima ovisno o tome koliko nam brzo treba njihov odgovor.

Većina spomenutih načina komunikacije je asinkrona i ne zahtijeva da oba korisnika budu spojena na mrežu u isto vrijeme. Iznimka je instant messaging za koji je potrebno da obje strane budu istovremeno spojene na mrežu.

Zanimljivo je primijetiti da mnogi korisnici interneta koriste više različitih davatelja usluga istog servisa. Uobičajeno je imati nekoliko različitih e-mail adresa u zavisnosti sa kojim krugom ljudi komuniciramo (npr. poslovnu i privatnu). Taj problem davno je riješen mogućnošću prosljeđivanja e-mail poruka na jedno centralno mjesto, a sličan pristup koristimo i za RSS izvore koje zapravo sakupljamo na jednom mjestu da bi ih pročitali.

Upravo to okupljanje podataka na jednom mjestu i njihovo filtriranje je princip koji pokušavamo iskoristiti ne bi li na jednostavan način mogli zavladatai tom gomilom podataka kojom nas mreža zatrpava.

Sa jedne strane, standardni protokoli za razmjenu podataka (HTTP, SMTP, Jabber, IRC...), i načina zapisa (HTML, XML, e-mail poruke) omogućavaju nam da takav *mashup* napravimo, ali sa druge strane naći alat koji točno zadovoljava naše potrebe (i ima podršku za sve tipove izvora podataka koji su nam zanimljivi) nije lako.

2. Primjer integracije različitih izvora podataka

Naš primjer je grupa računarskih zanesenjaka koja se okuplja u stvarnom svijetu i virtualno na IRC kanalu. Osim IRC kanala koji je centralni komunikacijski medij, koristi se također i wiki te mailing lista, ali većina komunikacije (i najveći izazov) je upravo bio IRC kanal.

Prva logična ideja je bila napraviti web arhivu koja bi omogućavala da na jednostavan način pogledamo što se dešavalo na irc kanalu kada nismo bili spojeni.

Tako je nastao projekt irc-logger (pomalo očekivanog imena) koji je prerastao u zanimljivu kombinaciju ulazno-izlaznih protokola, a sve u svrhu objedinjavanja zanimljivih informacija na jednom mjestu.

3. IRC arhiva

Da bi stvorili arhivu poruka na IRC kanalu, irc-logger je spojen kao običan korisnik (bot) na IRC server. Na taj način on vidi sve poruke na kanalu i može stvoriti arhivu, te je prikazati na web-u (po datumu) ili pretražiti po bilo kojoj riječi iz sadržaja.

Zanimljiva je mogućnost pretraživanja arhive korištenjem IRC klijenta. Slanjem poruke

```
/msg irc-logger search proba
```

dobiti ćemo (u obliku poruka) sve zapise u arhivi koji su sadržavali riječ proba. Na taj način zapravo imamo mogućnost korištenja arhive i pretraživanja na dva načina: preko web-a i preko IRC poruka.

4. Kako organizirati poruke? Tagovi!

Da bi omogućili jednostavnije organiziranje poruka, odlučili smo uvesti dogovor za tagove u poruke na IRC-u.

Bilo koja riječ koja završava sa dva znaka kroz (//) na taj način postaje tag. Tagovi se prikazuju u obliku oblaka korištenja u web arhivi, ali i omogućavaju jednostavno pozivanje svih poruka koje koriste isti tag.

5. A svi ostali izvori podataka?

Mnogo potencijalno zanimljivih podataka za članove naše grupe već se nalazi na nekim servisima na mreži. Najbolji primjer su možda linkovi na del.icio.us-u ili filtrirane poruke iz raznih RSS čitača. Na sreću, većina tih servisa omogućava generiranje RSS feeda sa tim podacima.

Zbog toga, irc-logger može pratiti proizvoljan broj RSS feedova i slati promjene u njima na sam kanal. Prikazivanje promjena stranica na wiki-ju (preko RSS feeda promjena) na IRC kanalu pokazao se kao zgodan mehanizam međusobnog pritiska kada treba napisati neku najavu: netko napiše početak, svi dobiju obavijest na IRC kanal i ubrzo je cijela obavijest gotova!

6. Svaka prava komunikacija je dvosmjerna!

Osim mogućnost da prima podatke preko RSS-a, irc-logger može i stvoriti RSS feedove sa oblakom korištenih tagova ili zadnjim porukama koje su u sebi imate neki od tagova.

Taj RSS feed je onda koristan za uključivanje u druge servise, a mi ga koristimo unutar wiki-ja koji na taj način postaje uvijek živ, čak i kada ga nitko ne editira.

7. Sinergijski efekt povezivanja

Većina mogućnosti irc-loggera na prvi pogled izgleda prilično uobičajeno i ne naročito zanimljivo. Ali, kada pogledamo njihovu kombinaciju dolazimo do zaključka da kombinacijom standardnih protokola i malo mašte možemo napraviti zanimljiv *mashup* koji se u našem slučaju pokazao kao vrlo koristan članovima naše grupe.