

# KONFERENCA ONLINE INFORMATION 2009

Konferenca Online Information 2009 z naslovom *Information + Conversation = Collaboration + Innovation* je potekala v Londonu od 1. do 3. 12. 2009. Razdeljena je bila na konferenčni (delegatski) in razstavnih del. Iz Slovenije sva se konferenčnega dela udeležila Nataša Godec iz Osrednje družboslovne knjižnice Ljubljana in jaz, medtem ko so se razstavnega dela udeležili tudi predstavniki NUK-a, Centralne tehniške knjižnice in UKM-a. Skupaj nas je bilo preko 600 delegatov iz 45 držav, medtem ko naj bi se obeh delov udeležilo okrog 9.000 obiskovalcev iz 70 držav.

**Konferenčni del** je bil vsak dan razdeljen na tri vsebinske sklope, osrednja tema konference pa je bila semantični splet.<sup>1</sup> V poročilu povzemam nekaj ključnih poudarkov iz predstavitev semantičnega spleta.

## SEMANTIČNI SPLET

Cilj semantičnega spleta (Web 3.0) je uporabniku dostaviti vse pomembne informacije v strukturiranem formatu, ki ga lahko uporabi katera koli aplikacija. Omogočil naj bi iskanje po kateri koli javni bazi podatkov na svetu z uporabo skupnega jezika in filtriranja nerelevantnih informacij, ki jih današnji iskalniki ne zmorejo. Baze znanja naj bi bile medsebojno povezane (angl. *web of linked data*), kar bi omogočilo avtomatizirano kombiniranje virov in dostavo tiste informacije, ki jo uporabnik potrebuje. Razvoj semantičnega spleta je pospešil socialni splet (angl. *social web* – Web 2.0) in že kaže praktične rezultate tako v akademski, kot tudi v javni in poslovni sferi.

Danes ignoriramo okrog 80 % nestrukturiranih podatkov, kot so spletne strani, e-naslovi, dokumenti, članki, blogi, forumi, novice itd. in se osredotočamo le na 20 % strukturiranih podatkov. Semantična tehnologija lahko pri tem pomaga z ekstrakcijo informacij nestrukturiranih podatkov in njihovim informacijskim zlitjem s strukturiranimi podatkovnimi silosi.

Semantična tehnologija postaja prevladujoči trend. Uveljavil se je format *rdf*. Tudi Google (Rich Snippets), Yahoo (Search Money) in Microsoft (Powerset) so

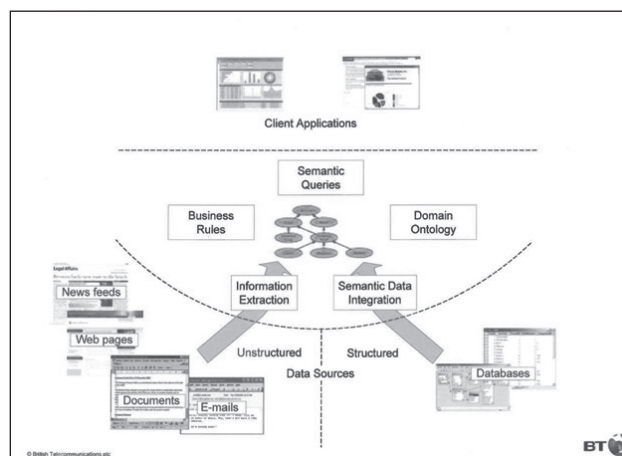
postali aktivni na tem področju. V mnogih sektorjih se semantična tehnologija uporablja že danes (<http://www.w3.org/2001/sw/sweo/public/UseCases/>).

## VIZIJA INTERNETA PRINCIP SEMANTIČNEGA SPLETA

Ker želimo reševati težave čim hitreje, je hitro in kakovostno iskanje informacij zelo pomembno. Zato se je treba usmeriti v razumevanje potreb uporabnikov in njihovega poizvedovanja. Ta trend bo zelo verjetno



Slika 1: Internet danes in jutri



Slika 2: Prikaz delovanja semantičnega spleta

močno vplival na razvoj spletnih iskalnikov, osebnih razvrščanj in uporabniškega vmesnika. Pri tem razvoju igra splet 2.0 pomembno vlogo. V zadnjih letih je namreč splet postal digitalni Gutenberg, saj je postal medij, v katerem objavljajo tako založniki kakor splošno

prebivalstvo (blogi, forumi, tviterji, vikiji).

Ta obseg in kompleksnost objavljenih vsebin (blogi itd.) ter naraščajoče število uporabnikov zahteva posodobitev iskalnih orodij, saj tehnologija iskanja izvira še iz časov šestdesetih in sedemdesetih let in ni več kos kakovostnemu poizvedovanju za vse bolj zahtevne uporabnike. To iskanje ni več le v domeni specializiranih knjižničarjev, temveč ga izvajajo uporabniki spleta. Večina predavateljev vidi kot del rešitve semantični splet, ki bo znal ob ustreznih modifikacijah iskalnikov popolnoma izkoristiti potenciale spleta. Pri tem je pomembno opozoriti, da ustreznost presojujejo ljudje in ne računalnik – ta zna le računati in prikazati rezultate iskanja. V tem smislu bi bilo treba iskalnike usmeriti v podobno vlogo, kot jo imajo referenčni knjižničarji – prvo vprašanje zastavi uporabnik, naslednja vprašanja pa računalnik, cilj pa je priti do želene informacije.

Govorilo se je tudi o spletu v realnem času (angl. *realtime web*), problemu ažurnosti podatkov ter indeksiranja. Prihajajo mladi s svojimi mobilnimi napravami in pritiskajo na informacijska podjetja, ki se trudijo izboljšati ažurnost podatkov na spletu. V nekaterih panogah je realni čas zelo resna in pomembna zadeva (finance, zdravstvo, novice, logistika itd.) in trenutno ta problem vse informacijske družbe jemljejo resno in kot velik izziv. Dotaknili so se tudi problema dezinformacij na spletu in tudi tukaj niso imeli prave rešitve. Omenili so pomembnost kritičnega pogleda na objavljene vsebine, ki jih objavlja vse večje število uporabnikov spleta.

Trenutna evolucija spleta je spodbujena z interaktivnim ter semantičnim spletom in koristnostjo funkcij, ki jih omogočajo vmesniki API-ji (angl. *application programming interface*). Z rastočim omrežjem se pojavljajo problemi identifikacije/avtentikacije uporabnikov spleta in njihove varnosti. Trenutno obstajajo naslednje rešitve:

1. Avtentikacija IP, ki identificira uporabnika na nivoju institucije. Deluje dobro za "on-campus" uporabnike, je pa neuporabna za oddaljene uporabnike.
2. Federativna avtentikacija identificira uporabnika prav tako na nivoju institucije in je uporabna za uporabnike "on-campus" in oddaljene uporabnike. Primer tovrstne avtentikacije predstavljata protokola Athens za Veliko Britanijo in Shibboleth.
3. Druge sheme "uporabniško ime/geslo" individualno identiteto rešujejo zunaj konteksta institucije v povezavi s kreiranjem uporabniških računov znotraj

sistemov založnikov.

4. Druge sheme oddaljenih dostopov, kot je EZproxy, omogočajo oddaljenim uporabnikom, da se identificirajo preko svoje institucije v programskem okolju založnika.
5. Personalizirana prijava omogoča unikatno identifikacijo uporabnikov znotraj institucije in večinoma zahteva dodaten nivo prijave znotraj založnikovega programskega okolja, kar zahteva od uporabnika kreiranje in vzdrževanje dodatne identitete na strani založnika. Teh različnih načinov avtentikacije ni mogoče kombinirati ali jih združevati, kar predstavlja oviro v prihodnosti. Potrebna je nova rešitev.

Problem presonalizacije se pojavlja ob vse rastočem in uporabnem spletu 2.0 (Twitter, Facebook, Google), kjer se mora uporabnik registrirati, če želi sodelovati v blogih, vikijih in v preostalih socialnih spletnih omrežjih. Eno izmed rešitev predstavlja protokol OpenID, ki omogoča enotno identiteto na številnih spletnih servisih.

Srečujemo se tudi s problemom vodenja statistik (standard COUNTER) in tremi nivoji identitete na spletu:

1. Osebni račun za vsebine, ki so namenjene za osebno, privatno uporabo.
2. Vsebine, do katerih lahko dostopa le posamezni oddelek in ne celotna institucija.
3. Institucionalna identiteta in njihove pravice dostopa do posameznih vsebin.

Splošno uveljavljeni standard COUNTER ne uspe rešiti beleženja statistik tovrstnih nivojev identitete, zato semantični splet predstavlja številne izzive na področju statističnih matrik.

Zastavljajo se tudi številna vprašanja na nivoju zaupanja in varnosti:

1. Naj zaupamo svojo identiteto ponudnikom servisov?
2. Kdo nadzoruje osebne podatke spletnih servisov (Facebook, LinkedIn itd.)?

V spletu povezanih podatkov (semantični splet ali splet 3.0) je identiteta osrednjega pomena za determiniranje zaupanja in avtentikacije.

Informacije na spletu so trenutno predvsem v dokumentih *html*. Html in današnji splet ponujata malo možnosti opisovanja vsebine. Semantična spletna tehnologija naj bi premostila tovrstno slabost s pomočjo opisnega jezika RDF in OWL. Te tehnologije, ki so standardi W3C

(World Wide Web Consortium), omogočajo opisovanje vsebin na spletu s pomočjo strojnega procesiranja deskriptorjev, avtomatizirane analize in črpanja spletnih informacij. Semantični splet ne bo nastal čez noč. Verjetno bodo ostala tudi področja nesemantičnega spleta. Semantični splet nekatera področja že podpirajo in so še nekoliko bolj interne narave v poslovnih procesih, se pa počasi širijo tudi v odprto sfero spleta. Na področju semantičnega iskanja imamo praktične primere, kot so Hakia, Powerset, Siderean in Ontotext. Na področju informacij in procesne integracije so primeri Metatomix in Ontoprise, na področju semantične socialne mreže pa se pojavlja Radar Networks. Med velikimi podjetji so tudi Oracle s svojo podporo RDF in Thomson Reuters na področju poslovnih informacij (Calais initiative).

Predstavljenih je bilo nekaj praktičnih izvedb uporabe semantične tehnologije. Ena izmed najudarnejših predstavlja rešitev angleške medijske hiše BBC, ki sodi med največje medijsko podjetje na svetu. Soočajo se z veliko količino dokumentov in gradiva v svojih arhivih, kar je pripeljalo do tega, da uporabnikom niso uspeli ponuditi kakovostnega iskalnika, ki bi omogočal iskanje zelenih vsebin. Pri iskanju rešitve so uporabili semantično tehnologijo. Predvsem na področju multimedijskih vsebin (video, glasba) so vpeljali avtomatizirane postopke klasificiranja, analize in strukturiranega zajemanja metapodatkov ter njihovega medsebojnega povezovanja.

Predstavilo se je tudi založniško podjetje 24 ORE Group iz Italije, ki je bilo zaradi gospodarske krize prisiljeno zmanjševati stroške poslovanja in se je usmerilo predvsem na spletno objavljanje in pri tem uporabilo semantično tehnologijo (uporabljajo iskanje v "naravnem jeziku"). Pomagajo si z orodjem Cogito, ki ga razvija Expert System. Gre za orodje, ki omogoča avtomatizirano razumevanje besed, stavkov, paragrafov in celotnega dokumenta (semantično razumevanje vsebine dokumenta). O tem orodju imajo zelo pozitivno mnenje.

Več o semantičnem spletu, preostalih vsebinskih sklopih in posameznih predstavitev je zapisano v zborniku.

## RAZSTAVNI DEL KONFERENCE

Več kot 230 podjetij je predstavilo svoje rešitve na svojih razstavnih prostorih. Kakor prejšnja leta so bili v spremljajočem sklopu konference organizirani kratki seminarji,<sup>2</sup> ki so bili za udeležence konference brezplačni.

## ZAKLJUČEK KONFERENCE

V zaključnem delu konference se je razpravljalo o štirih temah, ki naj bi v prihodnje predstavljale svetovni trend na področju informacij:

1. *Google in njegov pritisk* na preostale tekmece ponujanja informacij. Dejstvo je, da je Google še vedno največji igralec (Google Ecosystem), ki močno vpliva na mlade in njihove navade. Je pravzaprav ogromen val, na katerem je mogoče deskati, seveda pa obstaja tudi alternativa, ki pa je verjetno manj prijetna. Zastavlja se tudi vprašanje, kako se bodo mladi vključili v delovna okolja, ki predstavljajo bolj rigidni prostor, kakršnega vsaj trenutno niso vajeni.
2. *Ali bo format xml prevladal?* Večina razpravljavcev se je strinjala, da je iluzorno pričakovati, da bi bili vsi dokumenti, objavljeni na spletu, v strukturi *xml* (strukturirani podatki).
3. *Rich Media*. Gre za pričakovani prevladujoči trend objavljanja video vsebin na spletu.
4. *Varnost in zasebnost podatkov na spletu*. Vsi so se strinjali, da popolne varnosti in zasebnosti na spletu ne bo. Podatki so in bodo pod nadzorom državnih varnostnih služb.

## Opombe

- 1 Celoten program je objavljen na spletnem naslovu [http://www.online-information.co.uk/online09/conference\\_2009.html](http://www.online-information.co.uk/online09/conference_2009.html).
- 2 Program teh seminarjev in posamezne predstavitev je mogoče najti na naslovu [www.online-information.co.uk/online09/seminars\\_2009\\_new.html](http://www.online-information.co.uk/online09/seminars_2009_new.html).

## Reference

- [1] Online information 2009: Conference proceedings. London, 2009.
- [2] Online information 2009: Event Guide, London, 2009.

Miran Petek