

Semantičke tehnologije – istraživanje i analiza

Ivana Teodor

Abstract – Semantičke tehnologije su, prema mnogima, budućnost web-a. Smatraju se okosnicom web-a 3.0 kao koncepta koji već uzima ogromnu pažnju kako stručne, tako i laičke javnosti. Semantički web, grubo rečeno, funkcioniše zahvaljujući ontologijama (u konceptualnom i strukturalnom smislu) i novih obećavajućih tehnologija poput RDF-a, OWL-a, SPARQL-a itd. (u smislu realizacije).

U ovom radu pokušaću da predstavim razne pristupe kako u kreiranju idejnog koncipiranja tako i u smislu konkretne implementacije semantičkih tehnologija na web-u. Pored pomenutog, pokušaću da prikazem neke od najbitnijih projekata koji se bave ovom tematikom.¹

I. UVOD O SEMANTIČKIM TEHNOLOGIJAMA

Termin *semantika* (grčki *Semantikos*) odnosi se na onog ili ono koje recipijentu daje određene znakove izražene na neki određeni način, u korišćenju jezika, koda ili bilo kako drugačije. U skladu sa temom ovog rada i sferom istraživanja, kao posebno bitna ističe se aksiomska semantika (u skladu sa prirodnom semantikom) koja podrazumeva formalizaciju specifikacije opšteg oblika semantike pri čemu konkretno značenje koda nije dato kao takvo, već se dalje definiše uz pomoć logike simbola, postavljenih aksioma itd, a sve u okvirima i u skladu sa jezičkim konstrukcijama. Semantika, u svom originalnom smislu, je jako usko povezana sa sintaksom. Uzmimo, na primer, rečenicu “*Student voli da radi*”. U ovoj rečenici, sintaksa je forma rečenice, slova, puntuacija itd, dok semantika podrazumeva značenje, tj. to da student ima određeno zadovoljstvo dok obavlja svoje obaveze. U slučaju da promenimo sintaksu, i umesto reči “*voli*” stavimo simbol ♥, sintaksa će promeniti, ali će semantičko značenje ostati istovetno.

¹ Autor: Ivana Teodor (ivana.teodor.259@fit.edu.rs) sa Fakulteta Informatičkih tehnologija u Beogradu.

U vezi sa semantikom, a u smislu istraživanja ovog rada, kao posebno bitna kategorija javlja se komunikacija. Ljudi najčešće koriste glas kako bi komunicirali jedni sa drugima. Ipak, relativno prosta naprava poput diktafona može iste glasove reprodukovati bez znanja značenja glasova tj. reči koje bivaju reprodukovane. Sličan problem javlja se kod komunikacije između računara. Internet omogućava laku, brzu i jeftinu komunikaciju između računara. Ipak, Internet nije dizajniran na taj način da računarima omogući suštinsko razumevanje predmeta komunikacije.

Navedeno je, po mnogim novijim istraživanjima, jako bitna problematika u korišćenju Internet komunikacije. U pogledu Interneta, HTML predstavlja sintaksu komunikacije, dok se WWW može sagledati kao masivan sistem koji korisnicima omogućava smeštanje, pronalaženje i pretragu određenih resursa uz pomoć HTML sintakse. Korišćenjem navedenih mogućnosti, korisnik može, dobiti određenu Web stranu ili informaciju koja mu je potrebna. Ipak, u ovom procesu, sa strane računara značenje se apsolutno gubi. Računar će po zahtevu prikazati Web stranu koju korisnik zahteva, ali sam neće “razumeti” značenje i značaj iste. Na ovaj način, računar na pasivan pomaže korisniku. Postavlja se pitanje “Kako je moguće poboljšati korišćenje Interneta?”, tj. “Kako doći do preciznijih i potpunijih informacija?”. Novije tendencije upravo ukazuju na primenu semantičkih tehnologija. U slučaju pravilne upotrebe istih, računar može “razumeti” stranu koju prikazuje, naučiti puno toga o korisniku i ubuduće ga opskbljivati preciznijim informacijama.

II. PREGLED ISTRAŽIVANJA O KORIŠĆENJU SEMANTIČKIH TEHNOLOGIJA U E-PORTALIMA

E-portal (ili *Enterprise Information Portal*) može se sagledati kao gateway preko koga se ostvaruje pristup podacima, informacijama, sistemu i procesima koji stejkholderi koriste u okviru određene organizacije. Ono što naročito karakteriše e-portale i

odvaja ih od sličnih formi je činjenica da se e-portal može pristupiti kako putem intraneta (od strane zaposlenih u organizaciji), ekstraneta ili Interneta. Pritom, e-portal se kreira tako kako bi se u što većoj meri prilagodio zahtevima i „ukusima“ korisnika koji ga koristi².

Na prvom mestu, a u pogledu istraživanja ove materije, nalazi se W3C (World Wide Web Consortium - <http://www.w3.org>). Na ovom mestu mogu se pronaći brojne publikacije i prezentacije u vezi s semantičkim tehnologijama u opštem smislu, kao i organizacija radnih grupa u vezi s ovom materijom, kako prošlih tako i sadašnjih (primera radi, Web Ontology Working Group ili OWL Working Group). Zahvaljujući koordinaciji i aktivnostima ovih grupa, došlo se do značajnog razvoja tehnologija kao što su npr. OWL, RDF, SPARQL o kojima će biti više reči u nastavku ovog rada.

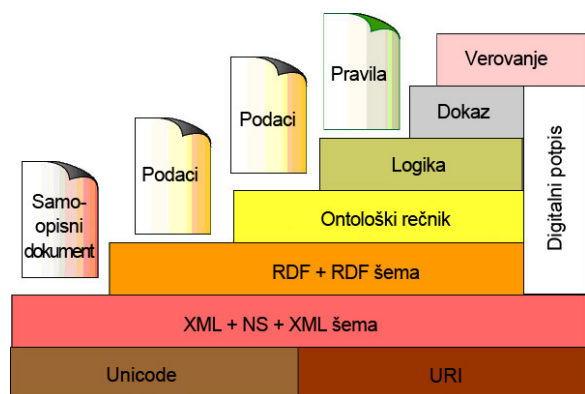
Takođe, bitno je napomenuti da se mnogi industrijski giganti u velikoj meri interesuju za razvoj semantičkih tehnologija u e-portalima. Primera radi, [Microsoft](#)-ovo R&D odeljenje *Microsoft Research* radi na razvijanju i integraciji semantičkih tehnologija u već postojeće portale i servise. Primer i plod ovog istraživanja bio bi *Zentity*, alat koji predstavlja svojevrsan repozitorijum alata, materijala, resursa i blokova informacija koji su namenjeni korišćenju u akademskim krugovima a u cilju stvaranja i održavanja digitalne biblioteke određene organizacije koristeći se ontološkim/semantičkim alatima i principima.

Primera radi, kao posebno interesantno napominjem istraživanje o *semantičkoj realnosti* i *semantičkom svetu* sprovedeno u okvirima Microsoft-ove SamGrail radionice 2007. godine. Ovo istraživanje ispituje postojeću kao i buduću vezu između aktivnosti koje individua poduzima u okviru svoje svakodnevice. U principu, radi se o maksimalnoj integraciji WWW-a i realnog sveta, a uz pomoć semantičkih tehnologija. Na primer, svakodnevni ciljevi koji se postavljaju semantičkoj realnosti su da

² Zvanična definicija e-portala kaže da je e-portal “portal klijentsko-serverske arhitekturom i pripadajućim framework-om gde objektno-orijetisan dizajn razdvaja aplikacioni sloj od korisničkog i razvojnog sloja i tako omogućava fleksibilniju implementaciju sistema i prilagođavanje potrebama korisnika”. API e-portala je napisan u standardnoj Javi i koristi XML interfejs protokol kao dodatak JSP-u”.

vaša lista za šoping zna da nema mleka u prodavnici u koju ste se zaputili, vaš auto zna gde su gužve u saobraćaju itd. Predložene tehnologije koje bi se koristile u ovakvom projektu su P2P, RDF i mikroformati zasnovani na RDF-u, data-strimovi itd. Pored industrijskih giganata, mnoge filantropske, profesionalne i organizacije NGO sektora svojevolljno se uključuju u ovakva istraživanja. Prve konferencije ovog tipa bavile su se osnovnim pitanjima, od toga šta sve semantika podrazumeva i širinom ovog pojma, preciznijim definisanjem pojma semantičke tehnologije, kao i semantičkog Web-a. Na osnovima koje je postavio Tim Berners-Lee sredinom devedesetih godina prošlog veka, istraživan je semantički Web pre svega preko tri stuba na kojima je postavljen od samog početka – RDF, RDFS, OWL.

Počevši od prezentacije Berners-Lee-a održane 2000. godine, ove tehnologije postale su osnova svih daljih istraživanja u ovoj sferi. Utvrđene su okosnice semantičke tehnologije, i mogu se predstaviti sledećim dijagramom:



Slika 1 - Osnovni pojmovi semantičkih tehnologija

Semantičke tehnologije, pre svega zbog svog neumitnog budućeg značaja, postale su jako popularan objekt raznih istraživanja. Mahom, ova istraživanja mogu se grupno podeliti na četiri interesne sfere: Web, Grid i P2P, Usluge, Sadržaj. Sve ove interesne sfere presecaju jedna drugu. U preseku svih ovih interesnih sfera nalazi sama semantika.

Posebna pažnja u ovim istraživanjima posvećena je terminu *Web 3.0*³, koji se smatra naslednikom sada

³ Termin Web 3.0 prvi je upotrebio Džon Markof (*John Markoff*) iz *New York Times*-a 2006. godine tragajući za adekvatnim terminom da opiše treću generaciju web servisa.

već naširoko korišćenom fenomenu Web 2.0. Web 3.0, zahvaljujući drastičnim poboljšanjim tehnike i tehnologije u poslednjem periodu, kao i mnogim istraživanjima i razvoju, predstavlja značajnu šansu za daljim unapređenjem korisničkog doživljaja Web-a, i, sve u svemu, šansu da biznis i media Web portali, laički rečeno, funkcionišu znatno bolje. Web 3.0, smatra se sinonimom za semantički Web, je Web u kome su informacije i podaci povezani korišćenjem semantičkih tehnologija i na taj način pružaju uvid sa značenjem u podatke tj. informacije koje trebaju biti ekstrahirane za korisnika.

Postoji više teorijskih tj. teoretskih prilaza razvoju Web-a 3.0. Ono što je svima zajedničko je činjenica da je Web, kakav jeste, došao do svojih prirodnih granica, i da je potrebna temeljna i sveobuhvatna reorganizacija. Takođe, evidentno je da mnoge od tehnologija koje bi se koristile u punom funkcionisanju Web-a 3.0 nisu još uvek sasvim zrele (TiVo, Pandora...).

Mnogi pod svakodnevnom upotrebom Web 3.0 podrazumevaju kreaciju i korišćenje *internet profila* stvorenih na osnovu browsing istorijata. Korišćenjem ovakvih profila, bilo bi moguće prilagoditi sadržaj korisniku u znatno većoj meri. Primera radi, ako dva korisnika koriste dve iste reči da obave pretragu preko Interneta, čak i preko istog servisa, neće dobiti identične rezultate, zahvaljujući informacijama iz njihovih profila.

Drugi eksperti, pak, smatraju da će fundament Web-a 3.0 ležati na interfejsima aplikacijskog programiranja (API-ima). *API* podrazumeva interfejs koji je dizajniran na taj način da dozvoljava developerima da kreiraju aplikacije na tačno određen i specifičan način koji iskorišćava određene resurse koji su na raspolaganju. Ovaj prilaz već je korišćen u Web-u 2.0; Primera radi, API [Facebook](#)-a omogućava developerima da kreiraju razne aplikacije – igre, media player-e itd. – u okvirima samog portala i na već unapred predviđen i funkcionalan način.

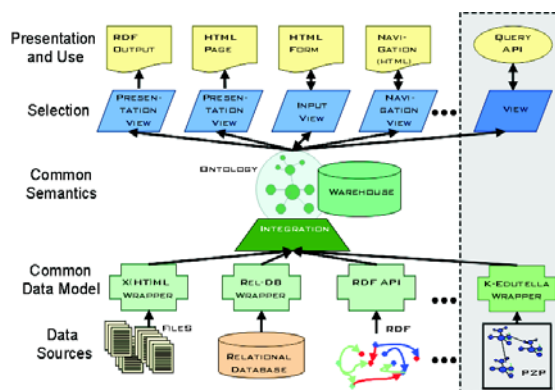
Sledeći pristup razvoja Web 3.0 je *mešanje* ili *mashup*, zasnovan na Web-u 2.0. Mešanje podrazumeva spajanje/kombinaciju dve ili više aplikacija u jednu funkcionalnu aplikaciju. Primera radi, moguće je spojiti portal o restoranima i GPS Web aplikaciju u jednu aplikaciju koja bi dobila na funkcionalnosti.

Naredni pristup podrazumeva razvijanje novog jezika koji bi u potpunosti zamenio HTML. Na ovaj način, Web bi dobio nov početak i mogao da bude od

samog početka razvijan u pogledu maksimalne funkcionalnosti. U okviru daljih napora ove grupe, vršena su istraživanja u vezi s sferom (pandan Web-sfere u smislu semantičkih tehnologija), NeOn-a, SOA4All-a, uslužnoj strani Web-a 3.0.

U vezi s pomenutom semantičkom pretragom, posebnu pažnju raznih istraživača okupirao je jedan od prvih pretraživača zasnovanim na semantičkim tehnologijama – [Wolfram Alpha](#). Cilj Wolfram Alpha je da, po određenom semantičkom algoritmu, svo sakupljeno znanje učini dostupnim svakome po sistematskom principu, tj. da omogući da svaki podatak koji može biti na bilo koji način izračunat, vizuelizovan i prikazan to i bude. Radi se o dugotrajnom projektu koji polako ali sigurno uzima maha. Trenutno, WolframAlpha sadrži više od deset triliona blokova podataka, više od 50000 matematičkih algoritama i modela i jezičko-semantičke alate za više od 1000 domena. Razvijan je na temeljima *Mathematica*-e poslednjih dvadeset godina i sadrži više od 5 miliona linija Mathematica koda. Wolfram Alpha koristi svoj *engine* baziran oko Web 3.0 paradigme kao okosnice i dostavlja informacije korisniku na potpuno prirodan i organistički način. Za davanje odgovora na “nezgodnija i teža” pitanja, Wolfram Alpha koristi *widget*-e i *wiki*-e stvorene u duhu saradnje i kolaboracije od strane samih korisnika ali i developera. U planu je objavljivanje API-a ovog servisa. Ovaj servis je pušten u rad 15.5.2009. godine, ali predstavlja samo početak istraživanja [Wolfram](#)-a na ovom polju.

SEAL (*Semantic PortAL*) je zanimljiv koncept koji predstavlja framework za menadžment web portala na osnovu ontologija. Ove ontologije podržavaju upite ka više izvora i to na taj način da omogućavaju portalu da automatski generiše navigacione prikaze i iste organizuju u odnosu na sadržaj.



Slika 2 - Šema SEAL-a

Jedan od ciljeva SEAL-a je integracija heterogenog sadržaja formatiranog u HTML-u. Ovo se postiže SEAL-ovim *HTML Wrapperom* poluautomatskog tipa razvijenim upravo za ove potrebe. Isti HTML-e mapira kao XML i vrši dalje operacije sa njima koristeći se relacionim bazama podataka, tj. šemama i ontologijama iz istih. Naime, RDF šeme se generišu direktno i automatski iz relacione baze podataka uključene u SEAL arhitekturu. Kolaboracija se vrši na osnovu P2P protokola, specijalno [Edutella](#)-e.

U pogledu menadžmenta web portala, SEAL nudi četiri osnovna prikaza:

- Presentation view – prezentacioni prikaz;
- Input view – prikaz za unos podataka;
- Navigation view – prikaz za navigaciju i
- General view – opšti, uobičajeni prikaz.

U pogledu arhitekture, SEAL se može podeliti na četiri sloja:

- Sloj podataka i udaljenih servisa – opcioni, esterni servisi, najviši sloj. Komunicira sa P2P servisima, nudi *storage* mehanizme itd.
- Srednji (*middleware*) sloj – API visokog nivoa koji manipulira ontologijama i usmerava i sinhronizuje komunikaciju.
- Aplikativno-uslužni sloj – grupiše aplikacije koje koriste servise nižih slojeva.
- Jezgro – osnovna (*KAON*) ontologija

Kao takođe bitne projekte navela bih i [DBPedia](#)-u, projekat koji ekstrahuje informaciju iz Wikipedia-ine baze i prilagođava je i strukturu po osnovu znatno kompleksnijih upita koji ne bi imali svog smisla u Web 2.0 kontekstu, [FoaF](#) – projekat Web 3.0 socijalne mreže, koji kreira strane koje opisuju korisnike (ljude) na mašinski-razumljivom jeziku kao i veze između ljudi, njihove aktivnosti, itd.

Za potrebe ovog rada kreirala sam celokupnu ontologiju izraženu u RDF i OWL formatima, sa temom "Filmovi". U nastavku prilažem deo navedenog kôda:

RDF:

```
<!-- http://www.ivanateodor.com/Film.owl#igraU -->
<owl:ObjectProperty rdf:about="#igraU">
  <rdfs:range rdf:resource="#Film"/>
```

```
<rdfs:domain rdf:resource="#Glumac"/>
</owl:ObjectProperty>
<!-- http://www.ivanateodor.com/Film.owl#imaKoordinator -->
<owl:ObjectProperty rdf:about="#imaKoordinator">
  <rdfs:domain rdf:resource="#Film"/>
</owl:ObjectProperty>
...
<!-- http://www.ivanateodor.com/Film.owl#ZaSvuPubliku -->
<owl:Class rdf:about="#ZaSvuPubliku">
  <owl:equivalentClass>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#imaNivoNasilja"/>
      <owl:allValuesFrom>
        <owl:Class>
          <owl:unionOf rdf:parseType="Collection">
            <rdf:Description rdf:about="#Nema_nasilja"/>
            <rdf:Description rdf:about="#Nizak"/>
          </owl:unionOf>
        </owl:Class>...
      </owl:Restriction>
    </owl:equivalentClass>
```

OWL:

```
xmlns:Film="&www;Film.owl#"
URI="&www;Film.owl">
<EquivalentClasses>
  <Class URI="&www;Film.owl#12+"/>
  <ObjectAllValuesFrom>
    <ObjectProperty
      URI="&www;Film.owl#imaNivoNasilja"/>
    <Class URI="&www;Film.owl#Srednji"/>
  </ObjectAllValuesFrom>
</EquivalentClasses>
<SubClassOf>
  <Class URI="&www;Film.owl#12+"/>
  <Class
    URI="&www;Film.owl#Kategorizovano_Prema_Sertifikaciji"/>
</SubClassOf>
<EquivalentClasses>
  <Class URI="&www;Film.owl#18+"/>
  <ObjectUnionOf>
    <ObjectSomeValuesFrom>
      <ObjectProperty
        URI="&www;Film.owl#imaScenario"/>
      <Class URI="&www;Film.owl#TrilerScenario"/>
    </ObjectSomeValuesFrom>
    <ObjectAllValuesFrom>
      <ObjectProperty
        URI="&www;Film.owl#imaNivoNasilja"/>
      <Class URI="&www;Film.owl#Visok"/>
    </ObjectAllValuesFrom>
  </ObjectUnionOf>
</EquivalentClasses>
<SubClassOf>...
```

III. PREGLED POSTOJEĆIH E-PORTALA ZASNOVANIH NA KORIŠĆENJU SEMANTIČKIH TEHNOLOGIJA

Da bi se određeni e-portal realizovao u smislu semantike, isto podrazumeva korišćenje nekoliko standarda:

- XML-a – pruža osnovnu sintaksu tj. strukturu u okviru dokumenta. Ne pruža semantičko značenje sadržaja;

- XML šema – obeležava granice, tj. postavlja restrikcije XML dokumentima;
- RDF-a – jezik koji opisuje modele podataka i veze između istih, takođe može biti izražen XML sintaksom;
- RDF šema – svojevrsan rečnik za definisanje klasa RDF resursa kao i hijerarhije;
- i OWL-a. – Proširuje pomenuti rečnik, dajući mu nove dimenzije kako u pogledu opisivanja sadržaja, tako i relacija i kardinalnosti.

Kao koristan alat može se pomenuti i SPARQL – jezik upita za semantičke izvore podataka.

Jedan od portala ovog tipa bio bi portal www.suomi.fi na kojem su na dnevnoj bazi objavljivane informacije u vezi javne uprave, o njenom radu i bilo šta drugo što ima veze s svakodnevicom građana Finske. Istraživanje je obuhvatalo i remodeliranje celog portala u smislu dodavanja anotacije (značenja) i semantičkog konteksta uopšte. Pristupljeno je analizi taksonomije tj. bazičnoj organizaciji portala i utvrđeno je da je portal, u smislu sadržaja, podeljen na 12 celina (Kultura, Zdravstvo...). Sadržaj je remodelovan uz pomoć principa semantike tako što je, za početak, uzeto u obzir da svi korisnici nemaju isti *mindset*, tj. poled na svet u celini. Stoga, pristupljeno je ontološkoj kategorizaciji, tj. svakoj kategoriji je dodeljen set objekata za pretragu tj. linkova do određene strane, ali sa pridodatim značenjem (anotacioni linkovi). Ovakvi linkovi nazivaju se *projekcijom*. Dalji sadržaj kategorije je koncipiran kao spoj pomenute projekcije i unija ekstenzija podkategorija. Na ovaj način omogućeno je pretraživanje koje obuhvata više od jednog *mindset*-a, naime, odabirom više ontoloških celina kao granicama pretrage. Takođe, ontologije (definisane u OWL-u) imaju mogućnost da opišu i predstave više pogleda portala:

1. *Teme* – zasnovano na taksonomiji portala u svom prvobitnom obliku
2. *Lokacija* – u svom klasičnom smislu, u skladu s sferom interesovanja
3. *Ciljana populacija* – ova ontologija deli korisnike portala u različite grupe – *studenti*, *zaposleni*, *poslodavci* itd...

Kombinacijom ovakve ontološke podele i stvaranjem svojevrsne hijerarhije u okviru portala omogućena je, pre svega, znatno svrsishodnija pretraga informacija i kompleksan povrat informacija korisniku, tj. moguće je odabrati jednostavan povrat informacija, tj. povrat prostog hiperlinka, ali je moguće i odabrati kompleksan povrat, tj. povrat skupa traženih informacija spojenih i prikazanih kao jedan resurs. Na ovaj način, količina i kvalitet traženih informacija kao i njihova relevantnost za korisnika se agregira i znatno povećava.

Još jedan značajan projekat predstavlja [Semandeks](#). Radi se o Web 3.0 portalu čija je osnovna svrha „pojednostavnjenje“ celog Web 3.0 koncepta i semantičkih tehnologija u smislu upotrebe. Semandeks simulira ponašanje objekata kakvo bi ovi imali u stvarnom svetu i to putem povezivanja „ideja“ tj. entiteta. Korisnici sami biraju entitete kao i relacije koje ih povezuju. Semantički značaj dodaje se preko „Imena“ i/ili „Opisa“ koje se dodeljuju entitetu. Primer ovakvog modusa razmišljanja bila bi rečenica „FIT je fakultet“, u kojem entiteti „FIT“ i „Fakultet“ bivaju povezani relacijom „je“. Entitetu „Fakultet“ je zatim moguće dodati više imena, primera radi, „obrazovna institucija“, „deo školstva“, „deo univerziteta“ itd. Na ovaj način, masovnim unosima od strane korisnika i administratora stvara se ogromna baza pojmova povezanih semantikom. Cilj ovog projekta je stvaranje neke vrste semantičke Wikipedia-e u formi web indeksa. Bitno je i napomenuti da se ovaj portal oslanja na Java tehnologiju.

IV. ULOGA ONTOLOGIJA U TRAŽENJU E-SERVISA

Ontologija u traženju e-servisa je, po mom mišljenju, jedna od najočiglednijih, ako ne i najočiglednija primena ontoloških metoda na web-u.

Primarni zadatak semantičkog web-a je, kao što je napred rečeno, anotacija u funkciji davanja znanju *markup*-u podataka. Ovo se postiže ubacivanjem semantičkih meta-podataka koji se definišu u skladu sa jednom ili više ontologija, što dodaje semantički „sloj“, tj. „sloj značenja“ koji omogućava znatno precizniju (kao i automatizovaniju) pretragu podataka. U principu, ontologija se na web-u prikazuje uz vidudokumenata sastavljenih preko određenih reprezentacionih jezika (OWL, RDFS...). Ovi dokumenti, pored brojnih razlika, imaju i brojne

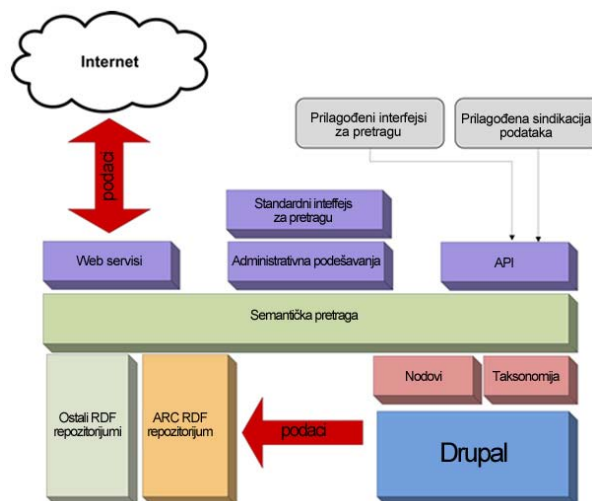
sličnosti sa, primera radi HTML dokumentima, što znači da se i ovi mogu indeksirati, čuvati i prikazivati. Razlika leži u činjenici da ovi dokumenti poseduju znatno kompleksniju i napredniju strukturu, nekada i sa semantičkim vezama sa drugim ontologijama. Neki od semantičkih web servisa za pretraživanje oslanjaju se upravo na ove veze u formiranju rang-liste rezultata pretrage, mada najčešće u kombinaciji sa drugim tehnikama (primer bi bio [OntoSelect](#)). Metričke mere koje prikazuju intezivitet ove pojave mogu se opisati kao *Poklapanje klase*, *Gustina* (oba pojma ukazuju na strukturalna poklapanja rezultata pretrage dvaju tehnika), *Semantička sličnost*, *Razmak* (ovi termini ukazuju na semantičku „težinu“ rezultata pretrage). Većina web servisa koji (načelno) prate ova pravila, nalik postojećim servisima za pretragu web-a, koriste algoritme koji određuju tzv. *Page rank*. Najčešće, algoritam pre svega uzima u obzir koliko ontologija importuju traženu ontologiju, a zatim i koliko je ontologija tražena ontologija importavala. Ukupni skor se računa prema zadatim pravilima:

$$\begin{aligned}
 cIO(O) &= \text{Broj importovanih ontologija za } O \\
 cIRO(O) &= \text{Broj importovanih ontologija koje mogu biti parsovane (za } O) \\
 cFO(O) &= \text{Broj ontologija koje importuju } O \\
 IO(O) &= \{x|x \text{ importuje ontologiju } O\} \\
 iS(O, \textit{nivo}) &= \frac{cFO(O)}{2^{\textit{nivo}}} + \sum_{O' \in IO(O)} iS(O', \textit{nivo} + 1) \\
 \textit{povezanost}(O) &= \begin{cases} cIO(O) > 0 : \frac{iS(O,0) * cIO(O)}{iS(O,0) * cIO(O)} \times \frac{\textit{broj } IRO(O)}{\textit{broj } IO(O)} \\ \textit{else} : 0 \end{cases}
 \end{aligned}$$

Danas postoji više semantičkih servisa za pretragu web-a. Jedan od njih je [Swoogle](#), koji uz pomoć *REST* interfejsa (*Representational State Transfer* – korišćen za distribuirane hipermedia sisteme) kako bi rezultat prikazao kao dinamičku web-stranu (uz pomoć RDF-a i XML-a). Upotreba nekoliko ontologija osigurala je semantičko formiranje meta-podataka i kvalitetan opis rezultata upita.

Naravno, u ovaj tip servisa mogao bi se ubrojati i već famozni WolframAlpha, ali je o istom već bilo reči u prethodnom delu ovog rada.

Kao posebno zanimljiv portal za semantičku pretragu izdvojila bih [Semantic Search](#). Cilj ovog servisa je omogućavanje semantičke pretrage u okviru [Drupal](#) CMS-a. Ovo se postiže korišćenjem RDF-a i veze između RDF-a i Drupal-ovog RDBMS-a (putem modula za Drupal). Za sada, Semantic Search podržava RDF, ARC, YARS i Sesame standarde. Arhitektura Semantic Search-a može se videti na sledećem dijagramu:



Slika 3 - Arhitektura Semantic Search-a

Ovaj servis se već nalazi u primeni kod nekih portala, primera radi <http://newahirc.org/>.

V. ANALIZA PREDNOSTI I NEDOSTATAKA KORIŠĆENJA SEMANTIČKIH TEHNOLOGIJA U E-PORTALIMA

Semantičke tehnologije bi mogle dodati potpuno novu dimenziju web-u, samim tim i e-portalima, kao što je napred rečeno, dodavanjem značenja određenim entitetima tj. njihovim opisima i to na način čitljiv/razumljiv ne samo ljudima, već i servisima/računarima.

Sama srž semantičkih tehnologija je razvoj semantičkih ontologija u smislu određenih standarda, kako bi se postigla formalizacija i jasna struktura koja bi dalje olakšala razumevanje ne-ljudskim entitetima. Do sada je definisano 12 prednosti semantičkog u odnosu na klasičan web:

1. Pretraga – Semantički web ima ogroman potencijal da poboljša mehanizme pretrage weba, koristeći se anotacijama i opisima servisa, posebno u odnosu na sadašnji način pretrage („sve ili ništa“ rezultati pretrage po ključnim rečima). Ova prednost vidljiva je i u primeni u sferi *e-commerce*-a, gde semantički web omogućava pronalaženje pravih proizvoda, kupaca, prodavaca itd. Trenutno najsavršenija tehnika pri *e-commerce*-u je *ebXML* (*e-business XML*, prilagođen upravo ovoj sferi upotrebe) koji i dalje ne poznaje značenja i ne može

pobezati ključne komponente sa poslovnim procesima.

2. Agenti – programi koji delaju u ime i za račun druge osobe, entiteta ili procesa i karakteristični su isključivo za semantički web.
3. Upravljanje znanjem – hvatanje, ekstrahiranje, procesuiranje i čuvanje znanja. Ovaj proces je, očigledno, praktično nemoguć u konceptu klasičnog web-a. Upravljanje znanjem ima poseban značaj u B2B portalima jer je preko semantičkog web-a moguće prikazati značajne, povezane podatke (bili oni strukturirani ili ne) i na osnovu njih raditi dalju analizu.
4. Integracija – Ontologija se može koristiti kako bi se konkretizovala terminologija heterogenih sistema. Ontološke mape mogu rešiti „nesporazume“ između sistema. Ova stavka posebno je bitna u sferi B2B, tj. automatizacije B2B. Ontologija bi eliminisala opasnost da poslovni partneri ne mogu razumeti sintaksu neke materije, kao i da njihov sistem neće moći da ispravno funkcioniše u skladu sa sistemom date organizacije.
5. Kreiranje kompleksnih sistema – Uz pomoć DAML-S i BPEL4WS (kompozicioni jezici) moguće je, koristeći se ontološkim i semantičkim sredstvima, kreirati kompleksan sistem od više prostijih sistema.
6. Multimedija – Znatno je lakše pretraživati i raditi sa određenim A/V sadržajem ako je korisniku poznato njegovo značenje. Ogroman broj organizacija, bez obzira na njihov tip, svakodnevno radi sa velikom količinom multimedijalnih formata.
7. Filtriranje informacija – pandan filtriranju preko ključne reči. Odvajanje korisne od beskorisne informacije, ponovo, znatno olakšano sa pridodatim značenjem.
8. Omogućen mašinski dijalog – značajno u smislu dalje automatizacije kao i komunikacije na relaciji čovek-mašina.
9. Virtuelna zajednica – Ontološki servisi mogu naći primenu u formiranju, organizaciji i komunikaciji u pogledu

virtuelnih zajednica koje mogu obuhvatati jednu ili više organizacija.

10. Online marketing – Anotacija omogućava znatno bolji plasman marketinških kampanja, plasirajući marketinške poruke na prava mesta, pravim klijentima.
11. Rečnik, fleksibilnost i standardizacija (mapiranje) – Korisnik portala ne mora poznavati korišćene standarde, dovoljno je poznavati značenje određenog pojma i znati šta želi dobiti od pretrage.

Naravno, obzirom da semantički web još uvek nije zaživeo u punom smislu reči, ovo su samo pretpostavke raznih istraživača. Isto važi i za nedostatke. Njih, u smislu brojnosti, ima manje, pre svega iz razloga jer se, mahom nedostaci otkrivaju tokom upotrebe. Očigledni nedostaci, ako se tako mogu nazvati, su:

1. Visoki troškovi pri reorganizaciji portala na web-u
2. Potpuna promena modusa rada SEO-a
3. Potreban dalji razvoj
4. Potrebna dalja ulaganja
5. Potrebno dalje testiranje
6. Potencijalna „glomaznost“ zbog kompleksnije strukture web dokumenata

Na kraju, pravi odgovor na pitanje doći će tek kasnije, kada semantički web bude uzeo maha. Najkritičniji je segment semantičke pretrage, čija efikasnost ne može biti na pravi način upoređena sa klasičnim servisima ovog tipa dok prava web 3.0 arhitektura i resursi ne budu bili raspoloživi u dovoljnom broju.

VI. ZAKLJUČAK

Većina tekućeg rada na semantičkoj anotaciji zasnovanoj na ontologiji pretpostavlja ontologije koje su tipično razvijene za konkretan zadatak. Umesto toga, realističniji pristup bi bio da se pristupa ontološkoj biblioteci i da se selektuje jedan ili više odgovarajućih ontologija. Iako je razvoj velikih projekata i objavljivanje ontologija tek u začetku, mnoge su već dostupne. Odabir odgovarajuće ontologije (ili kombinacije komplementarnih) će biti

veoma bitan podproces Semantičkih Web aplikacija. Do skoro, rešenje ovog problema je ležalo u biblioteci ontologija. U poslednje vreme, dinamičke web-bazirane ontološke biblioteke i ontološki pretraživači kao što su OntoKhoj, OntoSelect, SWOOGLE, i Watson su osposobljeni za ontološku pretragu i prikazivanje na bazi podataka. Web 2.0 neće trajati zauvek. Web je entitet koji konstantno evoluira i kreće se svojim prirodnim putem. Web 3.0 bi se mogao definisati kao ono što dolazi posle minjavanja Web 2.0 fenomena. Ono što je Web 2.0 doneo je veću integraciju i međupovezanost kao i socijalnu konotaciju Web-a, kompleksne AJAX web aplikacije, ali ne i značenje i inteligentni pristup terminima, tako da je očigledno da budućnost najbitnijeg svetskog komunikacionog kanala ne može ležati na manje-više banalnim temeljima i nišama.

Sve u svemu, Web 3.0, teoretski, poseduje dosta potencijala i znatno bolju početnu tačku u odnosu na nekadašnji Web 2.0, oslanjajući se na semantiku, mikroformate, prirodne forme pretrage, napredni *data mining*, mašinsko učenje, tehnologije veštačke inteligencije itd, a u cilju kreacije čitave nove generacije web servisa koji bi svojim korisnicima, ali i medijatorima (mašinama) pružile potpuno novi uvid u materiju i omogućile *modus operandi* do sada nezamislivog tipa. Semantičke tehnologije spremaju teren za potpuno novi nivo korisničkog doživljaja, pre svega osetiv u sferi biznisa, gde se, u današnje doba, ni jedan poslovni proces ne može ni zamisliti bez adekvatnog IT fundamenta i ogromnog broja entiteta koji u njemu učestvuju. Relativno prostim procesom, kao što je uvođenje semantičkog sloja, očekuje se masivno poboljšanje web-a kao komunikacionog kanala, znatno uprošćavanje poslovnih procesa u (IT smislu), automatizaciju poslovnih procesa, lakdu pretragu, personalizovan pristup i prilagođavanje servisa e-portala korisniku, i u krajnjoj liniji niže cene usluga, što je cilj svakog biznisa.

LITERATURA

“From Chaos to Order - Winning the Information Management Game” – Grupa autora, CBSi, 2009.

“Ontologies: Silver bullet for knowledge management and electronic commerce (2nd Edition)” - Fensel, D., Springer-Verlag 2004

“Semantic e-Government Portals – A Case Study” - Teemu Sidoroff, Eero Hyvönen, HIIT 2008.

“Semantic Reality – Connecting the Real and the Virtual World” - Manfred Hauswirth, Stefan Decker, DERI 2007.

“Semantic Analysis of E-Business Operations” - Mehmet Sayal, Akhil Sahai, VijayMachiraju, Fabio Casati, HP Laboratories 2002.

“An Organization Ontology for Enterprise Modelling” - Mark S. Fox, Mihai Barbuceanu, Michael Gruninger, Jinxin Lin, Enterprise Integration Laboratory 2002.

“Ontology-Based Personalized Search and Browsing” - Susan Gauch, Jason Chaffee, Alexander Pretschner, NSF 2007.

“Personalized Access to Web Services in the Semantic Web” - Nicola Henze and Daniel Krause, IVS - Semantic Web Group, University of Hannover 2007.

“Architectural Styles and the Design of Network-based Software Architectures” – Rob Thomas Fielding, 2000.

“Evaluating Ontology Search” - Paul Buitelaar and Thomas Eigner, German Research Center for Artificial Intelligence (DFKI GmbH) Language Technology Lab & Competence Center Semantic Web, 2007

SAKE - Semantic-enabled Agile Knowledgebased e-Government - Anna Siemek, 2008

www.wikipedia.org

searchcio.techtargget.com

www.e-businessguide.gov.au

whitepapers.silicon.com

<http://www.informit.com>

<http://www.semantic-conference.com>

<http://www.w3.org>

<http://www.semanticgrid.org>

<http://computer.howstuffworks.com>

<http://blog-en.saltlux.com>

<http://sw-portal.deri.at>

<http://www.yso.fi>

<http://www.internetnews.com>