

SVEUČILIŠTE U ZAGREBU

FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Seminarski rad iz SPVP-a

Semantički web

(Zagreb, lipanj 2005)

Student : Ognjen jović
2401021367

Sadržaj

Uvod	3
Semantički web	4
XML	5
RDF	6
Ontologija	7
OIL	7
DAML	8
Zaključak	8
Objašnjenja	9
Reference	10

UVOD

World Wide Web predstavlja svakako najpopularniji servis Interneta. Načini prikazivanja informacija su bitno evoluirali. U prvoj fazi na webu su prikazivane samo statičke informacije koje su se prezentirale korištenjem HTML-a. Prezentacije sa dinamičkim karakteristikama bio je sljedeći korak. U tu svrhu, kao dodatak HTML-u koriste se skriptni jezici kao npr. JavaScript.

Razvoj tehnologija koje su omogućile pisanje skripti koje se izvršavaju na strani servera (npr. ASP i PHP) donijele su revoluciju i pomak od web prezentacija prema web aplikacijama. Posljednji pomak u razvoju web-a okarakteriziran je napuštanjem koncepta prezentiranja informacija prema konceptu nuđenja informacija preko web servisa.

Kako je web doživio dramatičan rast, sve je veći problem bilo doći do relevantnih informacija. Rješenje se našlo u razvoju specijaliziranih sajtova za pretraživanje podataka tzv. internet pretraživači. Tehnologija koju pretraživači koriste zasnovana je na indeksiranju i pretraživanju ključnih riječi. Ali nedostatak toga je da ovakav pristup ignorira semantiku informacija pa se javlja problem dobivanja irelevantnih informacija.

SEMANTIČKI WEB

Semantički web je nadopuna, ekstenzija World Wide Web-a a nastao je u potrazi za efikasnijim rješenjima za pronalaženje informacija, te kao takav omogućuje bolju suradnju između računala i korisnika. Ideja je da se sve informacije koje se na web-u pojavljuju označe posebnim tagovima tako da računala mogu automatizirano povezivati podatke iz jednoga tipa informacija (primjerice – fotografija s označenim mjestom, datumom i vremenom nastanka) s nekim drugim tipom informacija (primjerice – meteorološko izvješće za period kada je nastala fotografija) [1].

Semantički web kao koncept predstavlja i omogućava da webu dostupni izvori informacija mogu biti organizirani i korišteni semantičkim, a ne sintaktičkim ili strukturalnim metodama, on predstavlja sinergiju programa koji prikupljaju sadržaj sa weba iz različitih izvora, zatim procesira informaciju i razmjenjuje rezultate sa drugim programima, na globalnoj razini [2].

World Wide Web uz pomoć računala za pretraživanje i ogromnog broja raspoloživih meta-podataka omogućava dobivanje informacija koje za sada još ne mogu u potpunosti zadovoljiti informacijske potrebe korisnika. Sve više je istraživačkih napora da se u samom web-u poveća efikasnost pretraživanja kako bi se došlo do što relevantnijih informacija. Kao jedan od najnovijih rezultata takvih napora W3C konzorcijuma je Semantički web.

Semantički web je osmislio Tim Berners-Lee, otac WWW i W3C konzorcijuma. Tim Berners-Lee zajedno s W3C konzorcijem u suradnji s velikim brojem istraživača rade na poboljšanju, proširenju i standarizaciji sistema u različitim jezicima, publikacijama i alatima.

Arhitekturu Semantičkog web-a čine tri važna nova standarda - XML (eXtensible Markup Language), RDF (Resource Description Framework) i Ontologije [3]. Semantički web je u razvoju i još uvek nije u općoj upotrebi, ali obećava radikalno poboljšanje mogućnosti pronalaženja, sortiranja i klasificiranja informacija.

XML

Još od nastanka World Wide Weba 1992. godine podaci se prikazuju na web-stranicama pomoću skupa oznaka nazvanih HTML - "*Hyper Text Markup Language*". Specifičnim oznakama ili "tagovima" određuje se način prikaza podataka, npr. podebljani tekst dolazi između oznaka i , a odlomak se smješta između oznaka <P> i </P>.

U vrlo skromnom opsegu određuju su se i neke semantičke osobine sadržaja, uglavnom u obliku tzv. META-tagova unutar zaglavlja dokumenata.

Za razliku od toga, novi sustav oznaka nazvan XML - "*Extensible Markup Language*" izrazito odvaja sadržaj od samog prikaza. XML specificira strukturu dokumenata, za razliku od HTML-a koji je specificirao samo prikaz [4]. Pomoću tagova je definirano samo sadržajno značenje dokumenta, dok je način prikaza određen posebnim datotekama, tzv. "stylesheets"-ima pomoću kojih se, na primjer, može prilagoditi boja i veličina slova u dokumentu, način prikaza linkova, isticanje pojedinih riječi, podloga dokumenta, i sl.

XML je mehanizam za razmjenu strukturiranih podataka na WWW-u te omogućuje vlasniku definiranje vlastitih oznaka, odnosno jezika specifično za određeno područje primjene [5]. To u određenom smislu predstavlja ugrađivanje semantike na web, međutim ovako ugrađena semantika je implicitna što je čini prepoznatljivom samo ljudima, ali i ne strojevima.

XML-om se dakle približilo jedan korak više semantičkom web-u u kojem su dijelovi web-stranica tagovima definirani po svojem značenju, a ne kao do sada, po načinu prezentacije u web-preglednicima.

Primjer :

HTML prikaz podataka

```
<TABLE>
  <TR>
    <TD> Ime </TD> <TD> Tina </TD>
  </TR>
  <TR>
    <TD> Plaća </TD> <TD> 5000Kn </TD>
  </TR>
  <TR>
    <TD> Grad </TD> <TD> Zagreb </TD>
  </TR>
</TABLE>
```

XML značenje podataka

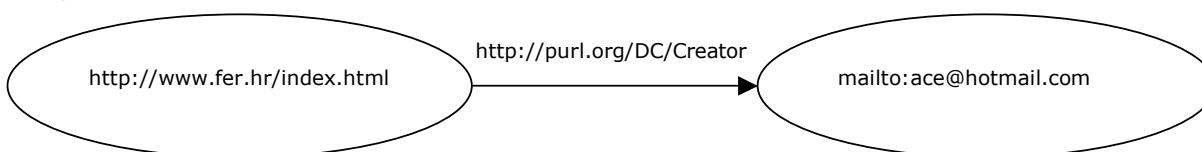
```
<radnik>
  <ime> Tina </ime>
  <plaća> 5000Kn </plaća>
  <grad> Zagreb </grad>
</radnik>
```

RDF

RDF (Resource Description Framework) je predložen od strane W3C konzorcijuma za modeliranje meta-podataka o resursima na web-u [6]. Kao što samo ime govori RDF nije jezik već model koji se može predstavljati na različite načine. Kako je XML postao standard za serijsku reprezentaciju i razmijenu informacija, najšire je prihvaćen i za predstavljanje i razmijenu RDF opisa.

RDF zapravo opisuje semantičke veze između elektronskih izvora. Informacije u RDF-u su tzv. izjave u obliku tripleta, tj. osnovni elementi u RDF modelu su: Subjekt, Predikat i Objekt. Sva tri sastavna dijela se općenito nazivaju Resource a indentificiraju se pomoću URI (Uniform Resource Identifier) [7]. Subjekt (resurs) je povezan s drugim resursom (objektom) preko veze koja je određena trećim resursom (predikatom).

Primjer :



Po ovome možemo reći da je Ace kreirao index.html.

Iz prethodnog primjera vidimo da RDF ne definira sintaksu nego samo model za predstavljanje metapodataka. Definiranje *properties*-a za odgovarajuće domene i njihove semantike zahtjeva dodatne mehanizme koji se implementiraju u vidu RDF shema (*RDF Schemas, RDFS*).

RDF shema je jednostavan model koji definira tipove podataka za RDF. Korištenjem RDF shema možemo reći npr. da je "Fido" tip "Pas", a da je "Pas" podtip životinje. Tako pomoću RDF shema možemo definirati domenski specifične *properties*-e i klase resursa na koje se te osobine odnose.

Osnovni primitivi koji se koriste za modeliranje domena su: *Class* i *SubClassOf* iskazi (za definiranje hijerarhije osobina), *Domain* i *Range* iskazi (za ograničavanje mogućih kombinacija osobina i klasa), *Type* iskaz (za definiranje resursa kao instance određene klase) [2].

RDF model omogućuje standardni opis činjenica o web resursima, što podacima daje određenu interpretaciju. Upotrebom RDF shema te se interpretacijske mogućnosti proširuju. No kako je za realizaciju semantičkog web-a potrebno omogućiti tehnike koje omogućavaju izražavanje semantike na još višem nivou, neophodna su daljnja proširenja. Tako npr. jezici koji se zasnivaju na ontologijama kao OIL ili DAML, dodaju neke nove primitive formalnij semantici RDF shema.

ONTOLOGIJA

Ontologija (grč. *Ontos* biće, stvarnost; grč. *logia* nauka), u svom izvornom filozofskom značenju, predstavlja nauku o biću, o onome što postoji, učenje o općim, fundamentalnim i konstitutivnim određenjima bitka.

U informatici u području "Wissensrepräsentation" (representacija znanja) ontologija znači formalno definirani sistem od pojmova i/ili koncepta i relacija između tih pojmova [7].

Web konstruktori se služe ontologijama za registriranje odnosa, relacija i karakteristika objekata. Sistemi ontologija definiraju termine, pojmove i njihove međusobne veze, ontologija je opis (kao formalna specifikacija programa) konceptata i veza koje mogu postojati unutar hijerarhije u sistemu pojmova.

Da bi se omogućilo implementiranje ontologija na web-u, potrebna je određena jezična podrška. Za sada postoje više jezika odnosno proširenja jezika, koji omogućuju "kodiranje" semantičkih informacija primjenom ontologija, kao što su npr. OIL ili DAML.

OIL

OIL (Ontology Inference Layer) predstavlja jezik za opis ontologija na web-u. U velikoj mjeri OIL se zasniva na, i uzima kao polaznu točku RDF sheme [8]. OIL je u potpunosti zasnovan na XML sintaksi, dok je sa RDF-om u određenoj mjeri i isprepleten. Kako se OIL ontologije zasnivaju na RDF sintaksi, one ujedno predstavljaju i validne RDF dokumente. Ovo se postiže proširivanjem definicije RDFS-a dodatnim jezičnim primitivima.

Po mišljenju autora OIL-a mala je vjerojatnost da samo jedan jezik za ontologije može zadovoljiti sve potrebe korisnika i primjene semantičkog web-a. Zato su se autori opredijelili da OIL organiziraju kao seriju slojeva koji će se stalno razvijati. Podrazumijeva se da viši sloj u potpunosti podržava funkcionalnost nižih slojeva.

Trenutna specifikacija OIL-a predviđa pored *Core OIL* i *Standard OIL* slojeva, i *Instance OIL* i *Heavy OIL*.

DAML

DAML (DARPA Agent Markup Language) je jezik zasnovan na ontologijama, razvijen je kao proširenje XML-a i RDF-a.

DAML u odnosu na RDF Schema ide još jedan korak dalje, pružajući dodatnu dubinu kod properties-a i klasa, omogućavajući time širi spektar izražavanja nego kod RDF Schema [2].

DAML je konceptualno vrlo sličan OIL-u i da je riječ o pravcima koji za sada neće ići u novija istraživanja dovoljno svjedoči i posljednji iz grupe DAML jezika – DAML+OIL. Ovakav pristup, naravno, DAML uvrštava u najozbiljnijeg kandidata za buduće standarde.

ZAKLJUČAK

Semantički web je osmišljen kako bi imali traženi podatak na web-u koji je definiran i povezan, linkovan na način da je razumljiv strojevima, ne samo za svrhu prikazivanja već i za upotrebu u različitim aplikacijama pretraživanja i dostupnosti podataka.

Tehnologija za realizaciju semantičkog web-a postoji i poznat je način kako izgraditi terminologiju i kako koristiti metapodatke. Cijela ideja ovisi o zajedničkom korištenju standarda, nešto što će se koristiti i biti rasprostranjeno svuda.

Semantički web bi trebao da definira strukturu sadržaja stranica na web-u i da kreira uvjete gdje bi software agenti "lutali" od stranice do stranice i spremno odgovarali na upite i zahtjeve korisnika.

Tim Berners-Lee je uvjeren da će se u bliskoj budućnosti semantički web sve više koristiti što se povećava primjena RDF kompatibilnih dokumenata. On semantički web vidi kao jedinstvenu moć koja će razbiti sve barijere, intelektualne i kulturne, koje kroz web danas postoje [9].

Vjeruje da bi ovakvo stvaranje semantičkog web-a koje zahtijeva veliki manualni trud prilikom stvaranja web-stranica i stavljanja podataka na Internet, moglo u budućnosti čak dovesti do stvaranja prave umjetne inteligencije koja bi svoja znanja crpila iz podataka s Internet-a.

OBJAŠNJENJA

Meta-podaci : meta-podaci ili meta-informacije se općenito nazivaju podaci, a informacije dobivaju preko drugih podataka.

META-tagovi: Mogu se koristiti na različite načine kao npr. opis stranice, ključne riječi, autor stranice i td. Kao i obični HTML tagovi i META tagovi su nevidljivi za posjetioce stranice. Za razliku od HTML tagova, meta tagovi se postavljaju između <HEAD> i </HEAD> tagova stranice.

Semantika : je područje lingvistike koja se bavi smislom i značenjem jezika, odnosno jezičnih znakova. Semantika se bavi pitanjem kako se smisao i značenje kompleksnih pojmova može opisati jednostavnije.

URI : Uniform Resource Identifier služi za indentifikaciju abstraktnog ili fizičkog resursa. URI se upotrebljavaju za opis resursa npr. web stranica, različitih fail-ova, kao i e-mail primatelja [7]. On je jednostavan web indentifikator: poput stranica koje počinju s "http:" ili "ftp:" koje često nalazimo na web-u.

W3C : World Wide Consortium je gremij za standarizaciju određenih tehnika World Wide Web-a. Utemeljitelj i predsjednik W3C-a je Tim Berners-Lee, koji je poznat i po tome da je osmoslio i World Wide Web.

REFERENCE

- [1] <http://www.bug.hr/vjesti/index.asp?id=58328>
- [2] <http://infomesh.net/2001/swintro/>
- [3] <http://eprints.rclis.org/archive/00001986/>
- [4] http://www.geografija.hr/novosti.asp?id_novosti=62&id_projekta=0
- [5] <http://staticweb.rasip.fer.hr/staff/gjakovljevic/xml/contents.html>
- [6] <http://www710.univ-lyon1.fr/~champion/rdf-tutorial/rdf-tutorial.pdf>
- [7] <http://www.wikipedia.org>
- [8] Fensel,D., Van Harmelen,F., Horrocks,I., McGuinness,D.L. and Patel-Schneider,P.F., "*OIL: An Ontology Infrastructure for the Semantic Web*", IEEE Inteligent Systems, Vol. 16, Num. 2, March/April 2001.
- [9] <http://smh.com.au/>