Semantički web i njegova primjena

Petrović, Tomislav

Undergraduate thesis / Završni rad

2016

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Faculty of Electrical Engineering, Computer Science and Information Technology Osijek / Sveučilište Josipa Jurja Strossmayera u Osijeku, Fakultet elektrotehnike, računarstva i informacijskih tehnologija Osijek**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:200:952451

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-28**

Repository / Repozitorij:

Faculty of Electrical Engineering, Computer Science and Information Technology Osijek

zir.nsk.hr

Image not found or type unknown

SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU
ELEKTROTEHNIČKI FAKULTET

Stručni studij, smjer Informatika

SEMANTIČKI WEB I NJEGOVA PRIMJENA

Završni rad

Tomislav Petrović

Osijek, 2016.

FERIT
FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA I INFORMACIJSKIH TEHNOLOGIJA OSIJEK

Obrazac Z1S: Obrazac za imenovanje Povjerenstva za obranu završnog rada na preddiplomskom stručnom studiju

Osijek, 26.09.2016.

Odboru za završne i diplomske ispite

Imenovanje Povjerenstva za obranu završnog rada na preddiplomskom stručnom studiju

Ime i prezime studenta:	Tomislav Petrović
Studij, smjer:	Preddiplomski stručni studij Elektrotehnika, smjer Informatika
Mat. br. studenta, godina upisa:	AI4261, 01.10.2012.
OIB studenta:	91889200520
Mentor:	Marina Peko
Sumentor:	
Predsjednik Povjerenstva:	Doc.dr.sc. Krešimir Nenadić
Član Povjerenstva:	Krešimir Vdovjak
Naslov završnog rada:	Semantički web i njegova primjena
Znanstvena grana rada:	**Procesno računarstvo (zn. polje računarstvo)**
Zadatak završnog rada	Obraditi temu semantičkog weba, pojasniti pojam ontologije te na primjeru u dogovoru s mentorom pokazati konkretnu primjenu.
Prijedlog ocjene pismenog dijela ispita (završnog rada):	Vrlo dobar (4)
Kratko obrazloženje ocjene prema Kriterijima za ocjenjivanje završnih i diplomskih radova:	Primjena znanja stečenih na fakultetu: 3 Postignuti rezultati u odnosu na složenost zadatka: 2 Jasnoća pismenog izražavanja: 2 Razina samostalnosti: 2
Datum prijedloga ocjene mentora:	26.09.2016.

Potpis mentora za predaju konačne verzije rada u Studentsku službu pri završetku studija:	Potpis:
	Datum:

FERIT
FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA
I INFORMACIJSKIH TEHNOLOGIJA **OSIJEK**

IZJAVA O ORIGINALNOSTI RADA

Osijek, 10.10.2016.

Ime i prezime studenta:	Tomislav Petrović
Studij:	Preddiplomski stručni studij Elektrotehnika, smjer Informatika
Mat. br. studenta, godina upisa:	AI4261, 01.10.2012.
Ephorus podudaranje [%]:	5%

Ovom izjavom izjavljujem da je rad pod nazivom**: Semantički web i njegova primjena**

izrađen pod vodstvom mentora Marina Peko

i sumentora

moj vlastiti rad i prema mom najboljem znanju ne sadrži prethodno objavljene ili neobjavljene pisane materijale drugih osoba, osim onih koji su izričito priznati navođenjem literature i drugih izvora informacija.
Izjavljujem da je intelektualni sadržaj navedenog rada proizvod mog vlastitog rada, osim u onom dijelu za koji mi je bila potrebna pomoć mentora, sumentora i drugih osoba, a što je izričito navedeno u radu.

Potpis studenta:

Sadržaj

1. UVOD 1

1.1. Zadatak završnog rada 1

2. POTREBA ZA SEMANTIKOM 2

2.1. Principi weba 3

2.2. Ideja iza semantičkog weba 4

3. IZVEDBA SEMANTIČKOG WEBA 6

3.1. Standard 6

3.2. Principi 7

3.3. Arhitektura i tehnologije 11

3.4. Ontologija 21

3.5. UML 24

4. PRIMJENA 26

5. ZAKLJUČAK 32

LITERATURA 33

SAŽETAK ..... **Error! Bookmark not defined.**

ŽIVOTOPIS 36

1. UVOD

Sa ubrzanim razvojem Interneta, količina dostupnih informacija na webu se povećava iz dana u dan, konstanto i eksponencijalno. Ono što je web učinilo popularnim, sad ga pomalo guši, nedostatak moderno orijentiranog standarda kao i zajedničkog rječnika i dalje stvara heterogenost u pretraživanju informacija na webu, a ona zauzvrat, sprječava razmjenu informacija i komunikacija.

Ideja semantičkog web-a je ono što je je proizašlo od tvorca njezina prethodnika, Tim Berners-Lee je pokrenuo ostvarenje standarda i jezika koji podržavaju osobni razvitak, te su dizajnirane da pomognu računalima pri boljem razumijevanju "ljudskih" informacija na web-u, samim time profitirali bi korisnici zbog puno preciznijih rezultata u pretraživanju, ali i pouzdanije podatkovne integracije.

Ova tehnologija ne bi samo donijela točnije rezultate u tražilicama, već bi također znali kad bi mogli povezati neke informacije bez obzira na lokaciju izvora, znali bi, isto tako, koji podaci se mogu uspoređivati te bi mogli pružiti razne automatizirane usluge još većem broju domena, počevši, recimo, od svoga budućega stana, ili pak digitaliziranih radnji kao što je državna uprava pa sve do elektronski pokretanih zdravstvenih sustava.

Upravo to je vizija semantičkog weba je vizija, a ovaj završni rad će obraditi, povijest njezina zagovaranja i pisanje standarda i principa. Nakratko se osvrnuti i na radne grupe koje su zadužene za taj posao. I nakon postavljanja teorijskog temelja krenuti objašnjavati zamišljenu arhitekturu ove tehnologije kako je propisuje standard, i nabrojati i opisati detaljno sve tehnologije koje se primjenjuju u razvoju semantičkog weba.

1.1. Zadatak završnog rada

Obraditi temu semantičkog weba, pojasniti pojam ontologije, te na primjeru u dogovoru s mentorom pokazati konkretnu primjenu

2. POTREBA ZA SEMANTIKOM

Godine 1996. otac weba, Tim Berners-Lee, je na pitanje o principu pronalaženja informacija na Internetu istaknuo činjenicu da je izuzetno komplicirano pronaći traženi podatak ako on nije pravilno semantički izražen, odnosno ako traženi pojam nije izražen u obliku u kojem je znanje prikazano kroz neki jezični navod.

U nadolazećim simpozijima, kroz idućih par godina (WWW 7 - Brisbane, 1997 i WWW 8 - Toronto, 1998), Berners-Lee je predstavio svoju ideju globalnog weba sa vlastitom, nezavisnom logikom i službeno najavio idući korak u razvoju weba, ono što je danas poznato kao semantički web. Ovom revolucijom, u svijetu Interneta, pronalaženje informacija na njemu također ulazi u novu eru, budući da je sama srž i cilj semantičkog weba povezati fundamentalna područja ljudske inteligencije i omogućiti njihovu transakciju u računalnu logiku. [1]

Centralni dio ove ideje nalazi se u semantici metapodataka potraživanog objekta (*eng. subject metadata*) i njihovom prikazu kroz računalima razumljivu sintaksu. Osnovni smisao ove ideje, kako je izražava otac weba, je u tome da će računalno obradivi metapodaci biti temelj za nadolazeći naraštaj servisa za pronalaženje informacija koje će i ljudima i računalima olakšati pristup pretraživanje podataka i informacija. Na primjer, bit će moguće dizajnirati inteligentnije botove (bot - skraćeni naziv za robota, u značenju računalnog programa koji se izvršava samostalno) koji će čak i vrlo kompleksne zadatke moći izvršavati totalno samostalno pomoću "m2m" (*eng. machine to machine*) iliti međustrojne komunikacije. [2]

Kako se danas pronalazak i skupljanje informacija na Internetu isključivo oslanjanju na ljudski faktor koji njima upravlja preko jednostavnih linkova, u skoroj budućnosti će takve poslove pretraživanja i skupljanja podataka odrađivati programi kreirani neovisno od traženom pojmu. Za takve programe bit će neophodni računalno čitljivi navodi o izvorima informacija (njihovim referencama) i njihovim međusobnim relacijama koji će se temeljiti na:

1. postojanju jednog standardiziranog modela,

[1] Harry Halpin, Social Semantics: The Search for Meaning on the Web, Springer, 2012.
[2] http://en.wikipedia.org/wiki/Semantic_Web

2. vezi između pojmova u rječniku
3. njihovim jedinstvenim definicijama
4. dostupnosti tih definicija programima.

Ovime se osigurava da će botovi upravljati webom koji će se sastojati od opisa i ontologija, te da će neovisno i slobodno moći obrađivati prikupljene podatke i komunicirati putem djelomičnog razumijevanja (*eng. partial understanding*). [3]

2.1. Principi weba

Razvoj weba, kako ga je zamislio Tim Berners-Lee, zasnovan je na pet tehnoloških i organizacijskih principa:

- jednostavnost,
- modularni dizajn,
- decentralizacija,
- tolerancija,
- distribucija

Prva dva se odnose, prije svega, na softverski inženjering odnosno fizičku stranu internet dizajna, te kao takvi nemaju previše veze s ovom temom. Od ostalih principa, iduća dva, se direktno vežu uz ideju semantičkog weba te objašnjavaju njegov zamišljeni pravac razvoja. Spomena radi, zadnji princip je vezan za posluživanje weba odnosno opisuje posao servera.[4]

Decentralizacija

Princip decentralizacije se ne odnosi isključivo na činjenicu da se informacije nalaze na različitim mjestima na webu, već i na to da je svaki korisnik odgovoran za informacije koje on sam postavi na Internet – kako za sadržaj, tako i za oblik u kojem se one nalaze.

[3] Berners-Lee ,The Semantic Web – LCS
[4] Berners-Lee, Hendlers & Lassila, 2001; Veltman, 2001, 2004

Na Internetu ne postoji nikakva zakonodavna institucija koja izriče što se može, a što ne postaviti na Internet (dok god sam sadržaj nije van okvira zakona). Dapače, decentralizacija se ne odnosi samo na sadržaj, već i na tehnološki razvoj weba – proizvođači raznih programa i vlasnici različitih usluga i servisa na Internetu mogu se koristiti vlastitim programskim rješenjima kako bi poboljšali svoju uslugu. [5]

Tolerancija

Idući princip na jest tolerancija, pod kojom se podrazumijeva da nove tehnologije ne smiju onemogućiti korištenje starih. Time se osigurava postupan, evolucijski, razvoj weba.[6]

2.2. Ideja iza semantičkog weba

Potreba koja je dovela do ideje zaslužne za razvoj semantičkog weba je pitanje: kako iskoristiti sadašnja znanja i principe, te postojeće tehnologije, koje su temelj klasičnog weba, za novi, web koji bi bio univerzalni usmjerivač (eng. router) za razmjenu podataka, informacija i pretraživanje istih. Analizom navedenog problema da se uočiti da je najveći nedostatak postojećeg weba u tome što je oblikovan tako da njime operira čovjek, a ne računalo.

Očito je da za stvari koje su za ljudske korisnike veoma jednostavne, npr. prepoznavanje oglasa neke turističke agencije ili da u tom oglasu prepoznamo i protumačimo adresu i radno vrijeme njihova ureda, za računalne botove, bez obzira na to koliko izvrsno oni bili programirani, predstavljaju nepremostivu zapreku.

Stoga je bilo potrebno uvesti potpuno nov način definiranja informacija i podataka na Internetu, koji će biti razumljiv računalima i omogućiti različitim programima da izdvoje značenje iz njih, a da pri tome osiguraju dovoljno fleksibilnosti da pokriju različite primjene i omoguće daljnji razvoj.

Pitanje koje se nameće jest u kakvoj vezi bi taj semantički web i srodni podaci bili u vezi za trenutnim, svakidašnjim, webom. Po nekim analizama pretpostavlja se da će postojati dva oblika semantičkog weba, a to su:

[5] Berners-Lee, Hendlers & Lassila, 2001; Veltman, 2001, 2004
[6] Berners-Lee, Hendlers & Lassila, 2001; Veltman, 2001, 2004

1. Klasičan web u kojem će dokumenti, pored oblika prilagođenog čovjeku, sadržati i računalno čitljive informacije, odnosno u obične HTML/XHTML stranice upisivat će se podaci koji bi omogućavali da se iz njih programski izdvoji značenje
2. Semantički web kao posebna mreža podataka, gdje , također, mogu postojati reference i navodi iz semantičkog na dokumente običnog weba i *vice versa*.[7]

Ova dva oblika semantičkog weba imaju potencijal za koegzistenciju kako niti jedan od njih ne isključuje onaj drugi, iako, ovaj drugi (semantički), oblik ima neosporne prednosti, te je za vjerovati da će, barem u početku, postojati oba oblika.

Kod pitanja o apsolutnoj migraciji sa običnog na semantički web, treba napomenuti da se ulaže ogroman trud u razvoj tehnologija za automatsko izdvajanje značenja iz ne semantičkog teksta. Ove tehnologije, ako se uspiju razviti, trebale bi znatno olakšati migraciju.

[7] http://www.w3.org/2002/Talks/09-lcs-sweb-tbl/

3. IZVEDBA SEMANTIČKOG WEBA

Razvijanje ideje, u konačnici, je dovelo dovoljno informacija za glavne internet arhitekte da krenu sa popisivanjem principima i definiranjem standarda za ono što će uskoro postati semantički web, kako bi njegovi inženjeri mogli odrediti tehnologije odnosno jezike pomoću kojih će se objavljeni podaci manipulirati.

3.1. Standard

W3C Semantic Web Activity grupa je osnovana da bi preuzela ulogu lidera na području definiranja standarda te pripomogla u izradi otvorenih prototipskih tehnologija. Ona je nasljednik W3C Metadata Activity grupe. Sljedeći zadaci su joj dodijeljeni:

1. Uvođenje standarda
2. Edukacija
3. Suradnja i koordinacija
4. Napredni razvoj [8]

Cilj grupe je definiranje standarda tehnologija i smjernica koje omogućuju decentralizirani razvoj a da pri tome imaju na umu kako svaki od dijelova odgovara jedan drugome. Glavni im je fokus, u istraživanju, informacija, koja bi trebala biti dizajnirana na način da je mogu razumjeti računala. Trenutačno ova grupa ima dvije radne pod-grupe, od kojih svaka ima zadaću u definiciji standarda i tehnologija. One su:

- RDF Core radna podgrupa (RDFCore)
- Web Ontology radna podgrupa (WebOnt). [9]

Trenutačni RDF (*eng*. *Resource Description Framework*) standard preporuča obično okruženje *(eng. framework*) koje reprezentira metapodatke. RDF Core radna podgrupa ima

[8] http://www.w3.org/2002/Talks/09-lcs-sweb-tbl/
[9] Berners-Lee ,The Semantic Web – LCS

primarnu zadaću dovršiti RDF rječnik u RDFS 2000 (*eng. RDF Schema Candidate Recommendation*). Da bi postigla svoj cilj ta podgrupa je izdala je niz dokumenata:

1. RDF/XML Syntax Specification
2. Resource Description Framework Concepts and Abstract Data Model
3. RDF Vocabulary Description Language 1.0: RDF Schema
4. RDF Primer
5. RDF Model Theory
6. RDF Test Cases [10]

Web Ontology radna podgrupa pak, ima zadatak da na temelju rada prve podgrupe dizajnira tehnologiju (jezik) za definiranje arhitekture ontologijskih rješenja baziranih na webu. Ta bi tehnologija trebala omogućiti bolju integraciju te međuoperabilnost podataka između zajednica. U svrhu postizanja svog cilja, radna podgrupa je izdala je niz dokumenata:

1. Requirements for a Web Ontology Language
2. Feature Synopsis for OWL Lite and OWL
3. OWL Web Ontology Language 1.0 Abstract Syntax
4. OWL Web Ontology Language 1.0 Reference
5. Web Ontology Language (OWL) Test Cases[11]

3.2. Principi

Sve može biti identificirano URI-ovima

Prvi princip prati zamisao da se ljudi, mjesta i stvari u fizičkom svijetu mogu referencirati u Semantičkom web-u koristeći niz identifikatora. Onaj tko kontrolira neki dio prostora na Internetu može tamo kreirati URI i objasniti kako on zapravo identificira nešto u fizičkom svijetu. Dio upućenih ljudi je zahtijevalo da samo mali dio web URI prostora

[10] http://www.w3.org/2002/Talks/09-lcs-sweb-tbl/
[11] http://www.w3.org/2002/Talks/09-lcs-sweb-tbl/

bude dostupan za tu namjenu. Semantički web, sam po sebi, ne potražuje, ni ne podržava takva ograničenja. [12]

Ovaj princip još omogućuje opciju da se fizičkim entitetima obraća indirektno. Npr., osobi po imenu «Zrinka Cvitešić» možemo se obraćati, poznajući URI njezinog e-mail sandučića, pa tako možemo služiti upit tipa: «Osoba čija e-mail adresa glasi «mailto:zrinka.cvitesic©gmail.com» i čije je ime «Zrinka Cvitešić». Možemo nastaviti s specificiranjem i puno dublje, te navesti više stvari o toj osobi bez da ikad dodamo novi identifikator za tu traženu osobu. [13]

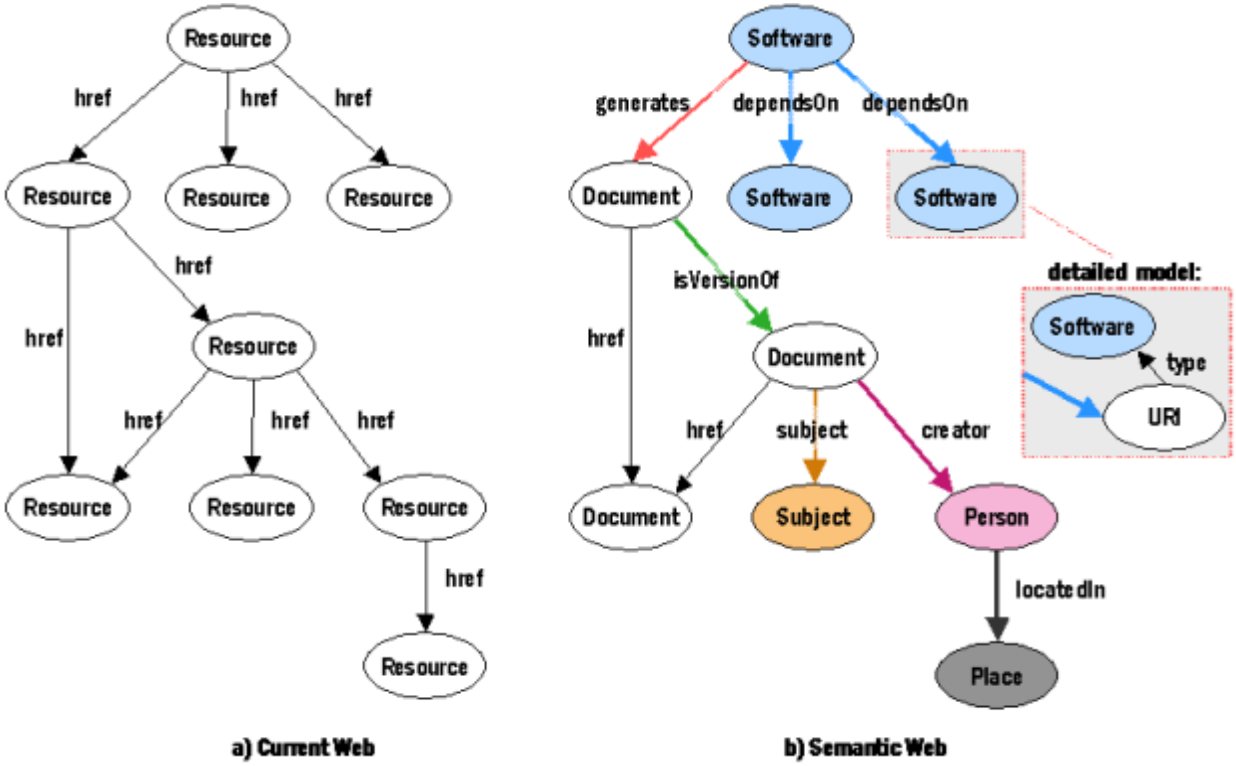
Resursi i linkovi mogu imati tipove

Trenutno, web se sastoji se od ogromne količine resursa i linkova (slika 1). Resursi su u ovom slučaju su zapravo dokumenti spremljeni na mreži, te namijenjeni za ljudsku uporabu i u rijetkim slučajevima sadrže meta podatke koji opisuje čemu služe i definiraju njihove relacije prema ostalim tamošnjim dokumentima. I iako čovjeku ne predstavlja problem zaključivanje vrste dokumenta po njegovu sadržaju npr. jedan je resurs pismo, drugi novela, a treći znanstveni rad, za strojeve je ova informacija često nedostatna za precizno zaključivanje. [14]

Nadalje stroj bi puno više sreće u tome pogledu imao na osnovu analize metatgova pomoću kojih bi stvarao informativne veze npr. «ovisi o», «je verzija od», «ima temu», «autor». [15]

Semantički web se također sastoji od resursa i linkova (slika 1) . No u ovom slučaju, resursi i linkovi imaju tipove koji definiraju koncept pomoću kojega botovi izvlače dodatne informacije npr., neki linkovi mogu reći da je resurs verzija nekog drugog resursa, da je napisan od resursa koji opisuje osobu, ili da resurs sadrži programsku podršku koji ovisi na nekim drugim programskim podrškama.

[12] Harry Halpin, Social Semantics: The Search for Meaning on the Web, Springer, 2012.
[13] Harry Halpin, Social Semantics: The Search for Meaning on the Web, Springer, 2012.
[14] Harry Halpin, Social Semantics: The Search for Meaning on the Web, Springer, 2012.
[15] Harry Halpin, Social Semantics: The Search for Meaning on the Web, Springer, 2012.



Sl. 3.2.1. Usporedba arhitekture klasičnog i semantičkog weba

Tolerira se polu-informacija

Web, kakvog ga poznajemo, je neograničen tj. žrtvuje integritet linka za mogućnost brzog i nenajavljenog rasta. Autori mogu lako povezati vlastiti stranicu sa drugom, bez obzira prema povezivanju na vlastitu stranicu. Bez načina da se o promjeni resursa informira referent linka, on običnim povezivanjem (eng. linking) pasivno prihvaća mogućnost da posjetitelji njegove stranice dobiju *404 error* stranicu koji informira posjetitelje da kliknuti link više ne vodi na oglašavani resurs.

Na sličan način je neograničen i Semantički web , naime bilo tko može reći bilo što o bilo čemu i stvarati različite linkove prema dokumentima. Stvari se nikada neće prestati ažurirati. Neki od povezanih dokumenata mogu prestati postojati ili se njihove adrese mogu biti promijeniti, ostavljajući stare sa zadaćom posluživanja (eng. hosting) novih resursa. [16]

[16] Harry Halpin, Social Semantics: The Search for Meaning on the Web, Springer, 2012.

Alati koji se razvijaju u svrhu dizajniranja i migracije na Semantički web-u trebale bi postaviti određenu toleranciju na taj raspad podataka (*eng. data decay*) koji bi tako trebali biti u mogućnosti funkcionirati unatoč svom raspadu.

Nema apsolutne istine

Jedno od najstarijih pravila Interneta je da: sve što je otkriveno na njemu nije nužno i istinito. Semantički web ne krši to pravilo ni na koji način. Vjerodostojnost se ocjenjuje u svakoj aplikaciji koja obrađuje neku informaciju na mreži. Aplikacije su te koje odlučuju u što da njezini korisnici vjeruju koristeći kontekstualne izjave (tko je rekao što..). [17]

Podržana evolucija

U našem društvu u ovom dobu nije ništa neobično što su koncepti slične tematike vrlo često definirani prema jako različitim grupama ljudi na veoma različitim mjestima na mreži ili čak, istom grupom ljudi ali u sasvim različitim vremenima. Gledajući u retrospektivi, za društvo, i mrežu, bilo bi dosta korisno kombinirati dostupnost podataka na Internetu koji koristi sve te koncepte. [18]

Ideja da semantički web koristi opisni MO (*lat. modus opernadi*) čiji će se opisi kasnije moći bezrezervno širiti kao što se i ljudsko razumijevanje proporcionalno ekspandira. Dapače, ova konvencija dopuštala bi veoma efektivne kombinacije raznih, nepovezanih poslova raznolikih zajednica; pa čak i onda kad bi one same koristile različite rječnike.

Minimalistički dizajn

Kako je već nebrojeno puta spomenuto, ideja semantičkog weba jest da jednostavne stvari ostavi jednostavnim, a složene učini dohvatljivima. Cilj W3C Semantic Web Activity grupe jest da se standardizira samo ono što treba biti standardizirano. Ovakav pristup omogućuje da se postupna implementacija započne već sada, krenuvši najprije od jednostavnih aplikacija koje se baziraju na već standardiziranim tehnologijama.[19]

[17] Harry Halpin, Social Semantics: The Search for Meaning on the Web, Springer, 2012.
[18] Harry Halpin, Social Semantics: The Search for Meaning on the Web, Springer, 2012.
[19] Harry Halpin, Social Semantics: The Search for Meaning on the Web, Springer, 2012.

3.3. Arhitektura i tehnologije

Ovi, ranije opisani, principi primijenjeni su na slojeve web tehnologija i standarda (slika 2). Temelj arhitekture čine Unicode i URI slojevi koji služe da nam omoguće korištenje, u slučaju prvoga, internacionalnog skupa tipkovničkih znakova, a drugi da bi korisnicima pružili način za identifikaciju objekata na semantičkoj mreži. [20]

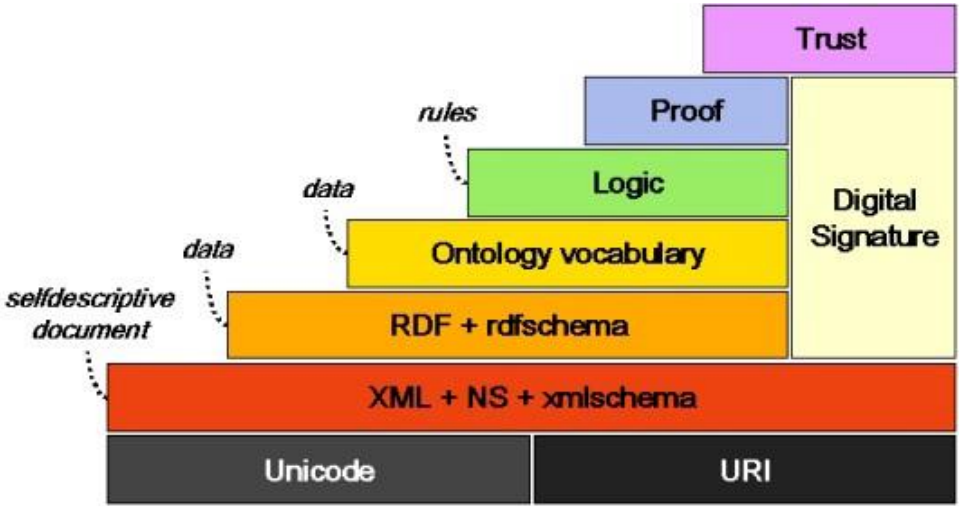
Dalje slijedi XML sloj sa pripadajućom shemom te je zaslužan za osiguravanje integracije definicija ove tehnologije zajedno sa ostalim standardima baziranima na XML standardu. [21]

RDF sloj, zajedno sa svojom shemom (RDFS) daje opciju za kreiranjem izjava o objektima sa URI-ovima i definiranje rječnika. Ovaj sloj je, također, zaslužan za mogućnost tipiziranja resursa i linkova, odnosno davanja imena tim tipovima. [22]

Ontološki sloj služi za omogućavanje daljnjeg unaprjeđivanja i korištenja rječnika, te također, može označavati relacije između odabranih koncepata. [23]

Svi ti slojevi, zajedno sa onim zaduženim za detekciju promjena kod dokumenata (*Digital Signature – desno na slici*), trenutno su još uvijek u postupku standardizacije pod radnom grupom W3C.

[20] Berners-Lee, Hendlers & Lassila, 2001; Veltman, 2001, 2004
[21] Berners-Lee, Hendlers & Lassila, 2001; Veltman, 2001, 2004
[22] Berners-Lee, Hendlers & Lassila, 2001; Veltman, 2001, 2004
[23] Berners-Lee, Hendlers & Lassila, 2001; Veltman, 2001, 2004



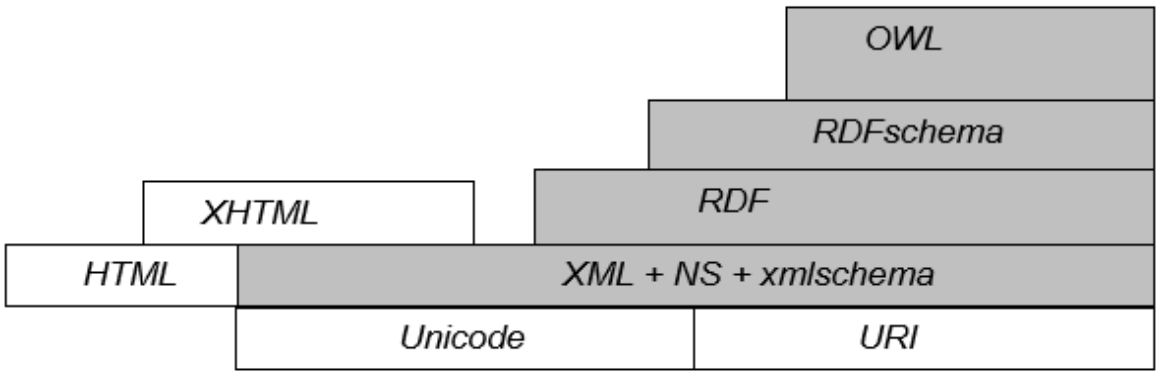
Sl. 3.3.1. Slojna arhitektura semantičkog weba

Preostala tri zadnja sloja (Logic, Proof, Trust) se u ovom trenutku još uvijek razvijaju te se na njima grade isključivo jednostavni demonstrativni primjeri (*eng. proof of concept*). Od ta tri, logički je sloj onaj koji je zadužen za pisanje, dok ih dokazni (*eng. proof*) sloj izvršava i, zajedno sa slojem povjerenja (*eng. trust*), odlučuje da li se danom dokazu vjeruje ili ne.

Izgradnja ovog projekta nije jednostavan proces i kao takav on se može postići samo i jedino ako se moć međuoperabilnosti odvede na višu razinu, tako da bude temeljena na novodefiniranim standardima. Standardi bi trebali prije svega osigurati definiciju za sintaktični oblik resursa, nego i za njegov sintaktički sadržaj. Ovakva međuoperabilnost je primarni cilj radne grupe koja ju izrađuje. Radna grupa bi se, također, trebala oslanjati na već postojeće definicije, te nastaviti sa postupnim predstavljanjem novih jezika namijenjenih webu (slika 3).

- Svaki od tih jezika ima određenu i strogo definiranu primarnu funkciju
- Jezik višeg sloja koristi funkcionalnost jezika nižeg sloja[24]

[24] Berners-Lee ,The Semantic Web – LCS



Sl. 3.3.2. Prikaz i odnos jezika u domeni semantičkog weba

XML

XML (*eng. eXtensible Markup Language*), je jezik za označavanje podataka koji je napravljen s idejom, da bude univerzalni opisni jezik, na umu. Primarna svrha mu je opisivanje strukture i tako tvori arhitekturu bilo kakvog dokumenta nebitno o krajnjem prikazu tog resursa. [25]

Dokument napisan u ovom formatu se sastoji od niza ugniježđenih (*eng. nested*) tagova, od kojih su svi unutar jednog izvorišnog. Svaki od tih tagova ima proizvoljan broj parametara (slika 4).

[25] http://www.w3schools.com/xml/xml_whatis.asp#top

```
<SampleXML>
  <Colors>
    <Color1>White</Color1>
    <Color2>Blue</Color2>
    <Color3>Black</Color3>
    <Color4 Special="Light">Green</Color4>
    <Color5>Red</Color5>
  </Colors>
  <Fruits>
    <Fruits1>Apple</Fruits1>
    <Fruits2>Pineapple</Fruits2>
    <Fruits3>Grapes</Fruits3>
    <Fruits4>Melon</Fruits4>
  </Fruits>
</SampleXML>
```

Sl. 3.3.3. Primjer sintakse neke XML datoteke

Uvjeti koje dokument mora ispuniti da bi bio valjan za konverziju u XML format su slijedeći:

- mora koristiti DTD (ili XML shemu)
- sve vrijednosti atributa moraju biti navedeni unutar navodnika
- svi tagovi moraju imati otvarajući i zatvarajući element osim ako konkretni tag nije prazan element
- ako je tag samostojan onda mora posjedovati zatvarajući simbol '/' (*eng. slash*) prije kraja taga
- svi tagovi moraju biti pravilno ugniježđeni
- svaki XML dokument bez DTD-a moraju posjedovati atribute tipa CDATA
- određeni elementi moraju se referencirati na drugi način kako ne bi razbili strukturu dokumenta
- postoji jedan izvorišni element [26]

[26] http://www.w3schools.com/xml/xml_whatis.asp#top

Izgled dokumenata sa .xml ekstenzijom možemo opisati kao etiketirano stablo, pri čemu svaki tag je u vezi s nekim imenovanim čvorom u datotečnom modelu i svaki ugniježđeni tag predstavlja nasljednike tog taga (*eng. offsprings*). [27]

Ovim formatom se može na više načina predstaviti bilo kakva podatkovna struktura, no pri tome valja upozoriti da ne postoji jedan unificirani način predstavljanja te strukture ili preciznije, moguće je napraviti dokument u .xml formatu na bezbroj načina, od kojih će svaki predstavljaju jednu te istu stvar, a da oni sintaktički nisu isti.

Uz to postoji opcija da se za svaki .xml dokument definira njegova vlastita struktura. Za takvu definiciju služi DTD (*eng. Document Type Definition*) sistem, no on ima veoma siromašne mogućnosti za definiranje struktura unutar .xml-a što se najbolje vidi u njegovoj nemogućnosti definiranja tipova podataka unutar parametara taga. [28]

Zbog tog razloga, definirana je nova specifikacijska shema koja je nadomjestila DTD (sadrži u suštini iste stvari kao i DTD ali samo još proširenije i formalnije). Pa se tako može reći da je XML shema zapravo gramatički dodatak XML-u.

U praksi, to dakle znači, da se XML datoteke mogu koristiti u svrhu:

- serijalizacije drugih markup jezika
- semantičkog markup jezika za web
- jedinstven format za izmjenu podataka [29]

Sad se postavlja pitanje, zašto ne koristiti XML kao osnovni ili čak jedini jezik za izradu semantičkog weba ili preciznije, zbog čega gubiti vrijeme na pronalaženje drugih takvih jezika, ako nam je XML kao markup format sasvim dostatan. Za dobivanje odgovora na to pitanje valja se poslužiti zahtjevima standarda za jezik kojim bi se definirao semantički web, a to su:

1. unificirana ekspresivnost

[27] http://www.w3schools.com/xml/xml_whatis.asp#top
[28] http://xml.coverpages.org/xmlAndSemantics.html
[29] http://xml.coverpages.org/xmlAndSemantics.html

a. pošto ni u teoriji nije moguće predvidjeti ama baš sve potencijalne zahtjeve i upotrebe formata jezika, on mora biti dovoljno fleksibilan da može obraditi bilo koji oblik podataka

2. sintaktična međuoperabilnost

    a. lakoća kojom se iščitavaju podaci iz dokumenta te koliko jednostavno bilo koja aplikacija može iskoriste te podatke

3. semantička međuoperabilnost

    a. Odnosi se na teškoće kod razumijevanja podataka tj. gubi se njihov smisao ili se pak krivo tumači, odnosno općenitije bi bilo mapiranje podataka između nepoznatih i poznatih pojmova. [30]

Ako bi recimo na XML standard gledali kao na opću formu za definiranje gramatike, onda je prvi zahtjev u potpunosti ispunjen jer se apsolutno sve se može prikazati u .xml formatu. Za njega čak i postoji cijeli niz parsera (neki od kojih bi bili: DOM, SAX..) pri čemu i oni sami vrlo fleksibilne, višestruko iskoristive komponente. S postojanjem parsera moglo bi se reći i da je ispunjen drugi zahtjev.

No tu dolazi do problema; jer treći zahtjev (i najvažniji) je nemoguće u potpunosti ispuniti koristeći XML i tu dolazimo do zida.

XML kao format je namijenjen isključivo definiranju strukture nekog resursa, ali ne i njegovog značenja, samim time postaje jasno da je nepodržavanja bilo kakve interpretacije informacija unutar zadanog resursa najveća mana XML-a kao unificiranog jezika na semantičkom webu. I to će se najlakše vidjeti na slijedećem scenariju. [31]

Uzmimo da su isprogramirana dva programa koji pokušavaju međusobno komunicirati. Dalje, u izgradnji takve komunikacije trebalo bi se prebaciti model domene u neki objektni (možda preko UML-a), te onda taj i taj novi model ponovo se prebacuje u neki dobar DTD (odnosno XMLS) dokument.

Kako to prebacivanje iz modela u model pa u DTD nije jednoznačno tako da je vrlo izgledno da će svaki primatelj u toj komunikaciji imati svoj vlastiti DTD.

---

[30] Robin Cover: XML and Semantic Transparency
[31] Robin Cover: XML and Semantic Transparency

Ali, nastavljajući pod pretpostavkom da su sudionici razgovora unaprijed definirali o kojem obliku DTD-a će se raditi, u tom slučaju komunikacija je moguća. No to je tek manji problem, oni pravi nastaju kada se u tu komunikaciju umiješa još jedan sudionik koji ima svoj vlastiti model domene i svoj vlastiti DTD.

U tom slučaju više nije moguće automatski ugurati naknadnog sudionika, već je prije toga nužan detaljan i opsežan rad na usuglašavanju modela, objekata, XML shema i njihovog razmještanja, i to za svakog korisnika ponaosob. U ovom slučaju, koji je usput rečeno, savršeno učestao i uobičajen u svakodnevnici, XML kao univerzalni jezik bio i kao autocesta za pakao.

RDF

Najlakši način za postizanje fleksibilnosti neke tehnologije jest njezino pojednostavljivanje, to je pogotovo istinito za informacijske tehnologije. Sam Internet svoj uspjeh duguje tome što je, barem u svojem začetku, bio dosta jednostavan. Web kao usluga je postojala tako što je postojao neki tekstualni dokument sa jedva najosnovnijim naredbama i opcijama da se u njega ubacuje multimedijski sadržaj ili možda čak i linkovi na druge takve dokumente. Sam sustav je jednostavno bio samo to - jednostavan – jednostavan za upotrebu, dokumente je bilo jednostavno objavljivati i pomoću običnih editora teksta, a postojale su verzije takvih programa koje su služile isključivo za jednostavno pregledavanje web sadržaja i koje su se mogle koristiti čak i na alfanumeričkim terminalima. Prolaskom vremenom te tehnologije su postajale sve složenije, što je znatno proširilo domenu njihove primjene, ali i problemi koji su dolazili s time bili su proporcionalno veći. [32]

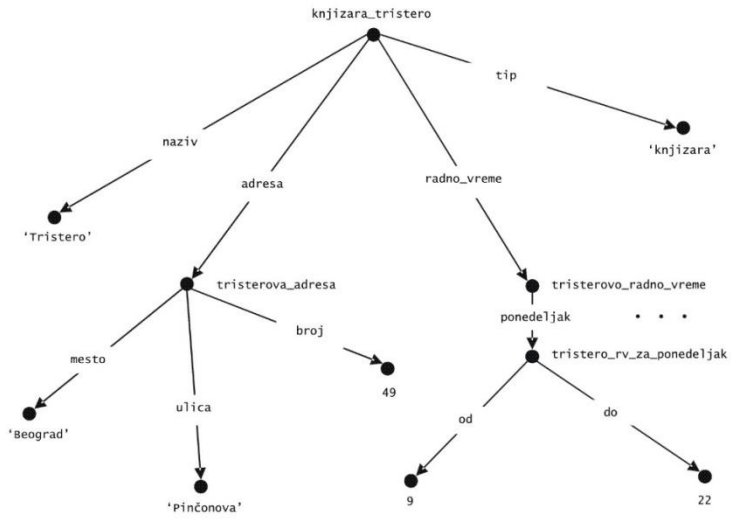
RDF *(eng. Resource Definition Framework)* je u semantičkom webu trebao preuzeti ulogu kakvu je u klasičnome webu obavljao HTML. Osnovna razlika između tog para je u obliku sadržaja koji se postavlja na mrežu – naime, to više nisu samo međusobno povezani dokumenti koji tvore neki zatvoreni sustav, već i proizvoljno odabrani podaci i informacije. Pitanje koje se nameće, je kako definirati strukturu podataka neke datoteke na webu. Pa na

[32] Antoniou, Grigoris; Frank van Harmelen (2008). A semantic Web primer. 2nd ed. London, MIT Press.

primjeru scenarija neke knjižare (slika 5) vidimo informacije koje su za tu knjižaru i posjetitelje te stranice bitne. [33]

Središnje mjesto u tom grafu ima čvor pod nekim imenom , za koji smo rekli da je tipa *knjižara* a naziv *"Tristero"* i na koncu joj dodali još neke dodatne informacije od neke važnosti (radno vrijeme, adresa).

Struktura podataka koju prikazuje ovaj graf mogla bi da se koristi u nekoj centraliziranoj aplikaciji, no totalno je neupotrebljiva za web. Kao osnovni objekt u ovom grafu je stoji središnji čvor etiketiran svojim nazivom i vrstom, međutim, nitko ne može garantirati da negdje na webu ne postoji još barem jedan objekt sa ovim nazivom.



Sl. 3.3.4. Osnovni podaci o knjižari u oblik grafa

Graf nam isto tako govori da naša *knjižara* ima svojstva *tip, naziv, adresa i radno_vrijeme* i da i ta svojstva imaju svoje vrijednosti. Međutim, opet se javlja isti problem kao i maloprije, kad bi se preko pristupilo knjižari nekim programom preko Interneta, korišteni program ne bi imao pojma o je svojstvima zapravo riječ, jer termine sa etiketama: *tip, naziv, adresa i radno_vrijeme,* različite stranice na Internetu mogu da koriste za označavanje potpuno drugačijih pojmova. Potrebno je, naime, pronaći neki način koji bi

[33] Harry Halpin, Social Semantics: The Search for Meaning on the Web, Springer, 2012

putem semantičkog webu izbjegao upotrebu homonima. Tu na scenu stupa URI (*eng*. *Uniform Resource Identifier*). [34]

Kao što mu i samo ime kaže, URI se bavi unikatnim etiketiranjem resursa. Pod resursom se ovdje podrazumijevaju svi objekti koji mogu, ali i ne moraju, biti povezan sa nečim što već postoji na internetu. Adrese dokumenata na webu - URLovi - su zapravo URI s kojima se najčešće susrećemo. [35]

Za razumijevanje ovog primjera nije potrebno opće znanje URI sintakse, jedino što nam je bitno jest to da negdje u njemu mora postojati neki identifikator autoriteta koji je zadužen za dato ime (ili možda neke organizacije koja je zadužena da se ovo ime koristi samo u jednom, točno određenom značenju). Ovaj primjer dobit će rješenje preko URI-a u obliku URL-a koji sadrži naziv domene čiji je kupac autoritet zadužen za dano ime. U prethodnom primjeru, recimo da postoje dva autoriteta koja su zadužena za davanje imena resursima.

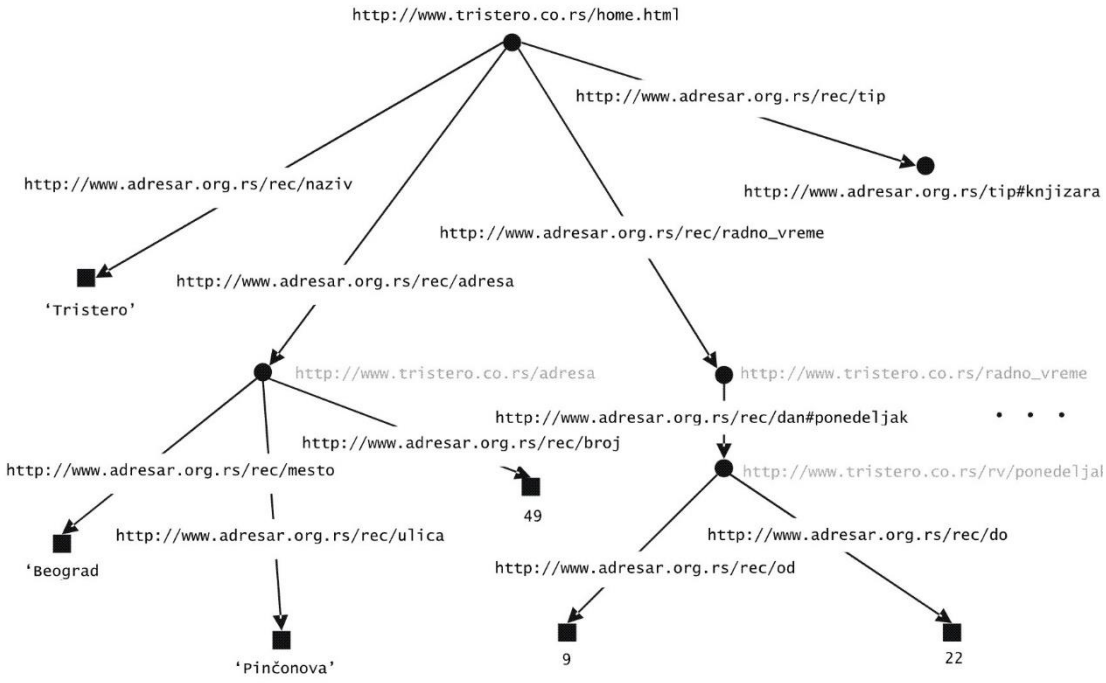
1. sama knjižara (*tristero.co.rs*)

2. neka hipotetska organizacija koja u skuplja sve informacije o firmama na Balkanu (*adresar.org.rs*)

S obzirom na to da je to njezino pravo, knjižara za identifikator središnjeg objekta u grafu, koji je predstavlja, vjerojatno će izabrati adresu naslovne stranice svog web portala (*http://www.tristero*). S druge strane, ta neka hipotetska organizacija, koja je zadužena za sakupljanje podataka o Balkanskim firmama, vjerojatno će URI pod svojom kontrolom identificirati tako da započinje sa *http://www.adresar*

Kad bi sada uzeli graf sa slike 5 i zamijenili nazive svih svojstava sa URI vrijednostima definiranima maloprije dobili RDF graf (slika 6).

[34] Antoniou, Grigoris; Frank van Harmelen (2008). A semantic Web primer. 2nd ed. London, MIT Press.
[35] Antoniou, Grigoris; Frank van Harmelen (2008). A semantic Web primer. 2nd ed. London, MIT Press.



Sl. 3.3.5. Osnovni podaci o knjižari u obliku RDF grafa

Na grafu je vidljivo da su oznake svih grana i većine čvorova na slici zapravo URI vrijednosti. Također je očito da na grafu postoje neki čvorovi čija su imena zadana kao neposredne vrijednosti – nizovi znakova ili brojevi. Takvi čvorovi su na slici prikazani kao kvadrati umjesto kružića, a za njihove oznake se kaže da su imenovane po vrijednosti neposredno danih vrijednosti, tj. to su *literali* - ne mogu imati svojstva.

Kod opisivanja tipa knjižare, koristili smo oznaku koja se sastoji iz URI-a (*http://www.adresar.hr/tip*) i nekog dodatnog identifikatora (*knjižara*) međusobno razdvojenih simbolom "#". Ovaj dodatni identifikator naziva se *fragment.* Slično je i sa radnim danima čije su oznake formirane URI-a "*http://www.adresar.hr/rec/dan*" i fragmenata "*ponedjeljak*", "*utorak*", ... Ovaj općeniti oblik URI-a naziva se URI referenca.

RDF format je unio mogućnost da se na jednostavan način predstave podaci neke organizacije na Internetu. Proces funkcije RDF-a može se opisati kroz još jedan scenarij. Recimo da postoji neki hipotetski servis za pretraživanje semantičkog weba, i recimo da mu korisnik zada da nađe sve knjižare u Zagrebu koje rade nedjeljom prije podne. Izvršavanje takvog upita vjerojatno bi se moglo odviti u slijedećim koracima:

1. U svim RDF grafovima dobivenima kroz rezultat upita izdvojimo one koji imaju neki čvor X koji je povezan sa čvorom http://www.adresar.hr/tip/knjizara i njezinom granom "http://www.adresar.hr/tip". Na ovaj način izdvojili smo sve knjižare.

2. Da bi suzili rezultate pretrage iz prethodnog koraka na one grafove koji imaju vezu sa, iz još ovog koraka izdvojenog, čvora X do nekog čvora preko grane "http://www.adresar.hr/adresa" i od tog čvora do čvora označenog literalom "Zagreb" granom "http://www.adresar.hr/mjesto". Izdvojili smo sve knjižare u Zagrebu.

3. Kako bi suzili pretragu na još specijaliziranije rezultate, recimo na one grafove koji prelaze put od našeg čvora X, pa sve do nekog čvora Y granama "http://www.adresar.hr/radnovrijeme" i "http://www.adresar.hr/dani/nedelja" (čvor preko kojeg ovaj put prelazi nije bitan za pretragu). Te na taj način pretraga se još finije suzila na sve zagrebačke knjižare koje imaju podatak o nedjeljnom radnom vremenu.

4. Za zadnji korak pretrage trebamo pretpostaviti da će naš fiktivni pretraživač, frazu 'prije podne' razumjeti kao 'rade između 07:00 i 12:00' u ovom koraku zadnji puta sužujemo rezultat pretraživanja koji smo do sada dobili na one grafove koji imaju vezu između čvora Y i nekog čvora označenog literalom manjim ili jednakim 12, granom "http://www.adresar.hr/rec/d".

3.4. Ontologija

Ontologija, predstavlja nauku o biću općenito, o onome što postoji. U informatici njezino se značenje malo fokusira tako da joj je definicija: "formalno definirani sistem od pojmova i/ili koncepta i relacija između tih pojmova". Ideja ontologije je bila da ona može povećati funkcionalnost Interneta na nekoliko načina. Ona može biti iskorištena na način da poboljša preciznost u pretraživanju web stranica (umjesto pretrage po ključnim riječima, pretraživački programi mogu tražiti i one stranice koje se odnose na određeni koncept). [36]

[36] Smith, Michael K. i dr. [ur.] (2004). OWL Web Ontology Language – Guide

Napredniji programi mogli bi koristiti princip ontologije za povezivanje podataka na traženim stranicama sa strukturama standarda i pravilima zaključivanja. Web inženjeri se također služe ontologijom, ponajviše u registraciji relacija ili karakteristika proučavanih resursa ili objekata.

Sustav ontologije definira termine, pojmove i njihove međusobne veze. Ontologija praktički ovdje služi kao svojevrsni opis (specifikacija) programa, koncepata i relacija koje mogu postojati unutar neke zadane hijerarhije u tom nekom sustavu pojmova te one predstavljaju najvažniju komponentu semantičkog web-a kao i jednu od najvažnijih komponenti u informativno-poslovnim sistemima. [37]

Ontologija se sastoji od konačne liste termina i veza između tih termina. Termini označavaju klase ili objekte iz neke domene. Veze se obično sastoje isključivo od hijerarhije klasa. Pored veza iz te hijerarhije, ontologija još može sadržavati i informacije tipa:

- svojstva
- ograničenja vrijednosti
- isključive izraze
- specifikacije logičkih relacija među objektima[38]

Izvjesna jezična podrška je, očito, u ovom slučaju nužna, kako bi se omogućilo implementiranje principa ontologije na instituciju Interneta. Postoji nekoliko jezičnih proširenja, koji omogućuju ispis semantičkih informacija preko pravila ili principa ontologije, takvi su recimo OIL *(eng. Ontology Inference Layer)* ili DAML (*eng. DARPA Agent Markup Language*).[39]

OIL je jezik za opis ontologija. OIL za polaznu točku u najvećoj mjeri uzima RDF shemu. OIL je, također, u potpunosti utemeljen XML standardu, tako da mu je sintaksa identična XML-u, mada je i sa RDF-om u isprepleten do određene mjere, tako da OIL ontologija predstavljaju i valjane RDF dokumente. [40]

Ta valjanost se postiže proširivanjem RDF sheme dodatnim znakovima i simbolima. Po mišljenju njegovih autora, samo jedan ontološki jezik ne može zadovoljiti sve potrebe

---

[37] Smith, Michael K. i dr. [ur.] (2004). OWL Web Ontology Language – Guide
[38] Berners-Lee ,The Semantic Web – LCS
[39] Fensel, D., Ontologies: Dynamic networks or formally represented meaning, Stanford University, 2001.
[40] Smith, Michael K. i dr. [ur.] (2004). OWL Web Ontology Language – Guide

korisnika semantičkog web-a. Tako da su se autori opredijelili da i sam OIL organiziraju kao seriju slojeva i imitira arhitekturu semantičkog weba u strukturi i ideji daće se svaki od tih slojeva sam za sebe stalno razvijati. Pod vrste OIL-a spadaju : Core OIL i Standard OIL layers, Instance OIL i Heavy OIL.[41]

Najveći pomak na području ontologijskih mrežnih jezika napravio je DAML. On sam je pažljivo osmišljen kao jezik i alat za ostvarenje ideje semantičkog web-a. U svom početku razvijan je kao proširenje XML-u i RDF-u. On ima karakterističan primjer standarda koji se sastoji od web-kompatibilnog jezika i logičke paradigme posuđene iz područja umjetne inteligencije.

Najnovija njegova verzija (DAML u suradnji sa OIL-om) sastoji se od niza korisnih elemenata, pogodnih za primjenu ontologije i označavanje podataka tako da budu razumljivi botovima. Taktika mu je da u jednu cjelinu spoji mrežni jezik, deskriptivnu logika i sustav rasuđivanja kojeg preporuča OIL.

DAML je isto kao i njegov prethodnik uskladiv sa RDF shemom i sadržava preciznu semantiku za opisivanje značenja.

Semantički web kao ideja se u svojoj srži zapravo i vrti oko definicije ontologije. Semantika bi smislenom web sadržaju dala smisao, a pravila koje propovijeda ontologija u informatici se nameću kao primarni mehanizam iza optimizma ljudi koji su ovi ideju i predložili. Ontologija kao alat inženjerima pruža mogućnost kreiranja novih činjenica u na osnovu činjenica koje su eksplicitno dane, u hodu. Ona omogućavaju postizanje nove razine funkcionalnosti web-a, kroz sljedeće funkcije koje su razlog sve češće primjene ontologijskih principa na Internetu:

- povećana preciznosti u pretraživanju web stranica
- povezivanje informacija na stranicama sa pridruženim im strukturama znanja i pravilima zaključivanja
- razmjenu znanja između ljudi i softvera
- dijeljeno razumijevanje domena

[41] Smith, Michael K. i dr. [ur.] (2004). OWL Web Ontology Language – Guide

- pružanje strukture za prijenos informacija u Semantičkom web-u ili u domenu povezanih podataka
- mogućnost nadograđivanja i mijenjanja u skladu s potrebama[42]

Također, postoje prednosti i nedostatci ontologija, koje ćemo također izdvojiti.

Prednosti ontologija:

1. koherentna navigacija, omogućava kretanje od koncepta do koncepta u strukturi ontologije
2. fleksibilne ulazne točke, jer svaka specifična perspektiva u ontologiji može se pratiti i dovesti u vezu sa svim njegovim povezanim konceptima
3. veze koje ističu relevantne informacije bez zahtjeva prethodnog znanja o domenu ili njegovoj terminologiji
4. sposobnost da predstavi bilo koji oblik informacija, uključujući i nestrukturirane, polu-strukturirane i strukturirane podatke povezivanje koncepta
5. integriranje sadržaja pravilnim povezivanjem i mapiranjem koncepata [43]

Nedostatci ontologija:

1. koncept koji je nov za mnoge korisnike tako da se često javljaju greške
2. problem mapiranja dobivene ontologije sa postojećim domenskim ontologijama,
3. problem gubitka semantike nakon transformacije[44]

3.5. UML

U zadnje vrijeme se jako puno resursa troši na razvoj jezika koji bi predstavljao ontološke principe te koji bi bio kompatibilan sa WWW (*eng. World Wide Web*) standardom. Do tada postoji nelo̸š, i jako fleksibilan i standardiziran jezik za modeliranje, - UML (*eng.*

[42] Harry Halpin, Social Semantics: The Search for Meaning on the Web, Springer, 2012.
[43] Harry Halpin, Social Semantics: The Search for Meaning on the Web, Springer, 2012.
[44] Harry Halpin, Social Semantics: The Search for Meaning on the Web, Springer, 2012.

Unified Modeling Language) koji sadrži kako grafički tako i XML bazirani format. Još neke od prednosti su mu velika zajednica korisnika, visoki nivo podrške. UML je jezik koji služi isključivo za specifikaciju i vizualizaciju sistema [UML, 2005]. [45]

On omogućuje konstrukciju nacrta koje služe za modeliranje sistema opisujući usput koncepte i svojstva tog sustava. Ovaj jezik, isto tako nije ograničen samo na područje informatike ili softverskog inženjerstva, već se može koristiti i u drugim granama industrije i poslovnog svijeta (bankarstvo, zdravstvena zaštita, poslovna inteligencija itd). Iako je isprva zamišljen kao svojevrsna analizi i dizajnu u inženjerskoj branši, UML se vrlo skoro počinje širiti i na druge sfere privatnog i javnog sektora.

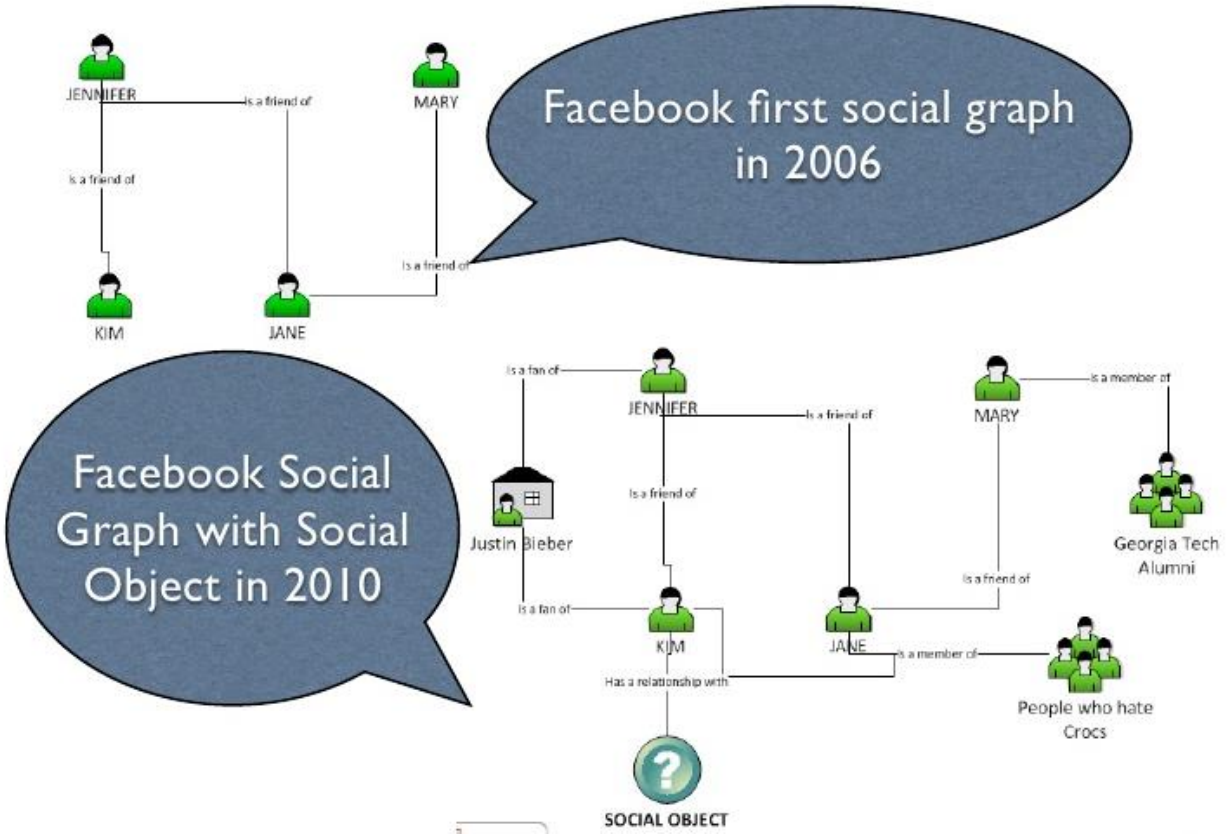
Prema OMG (*eng. Object Management Group*) standardu: UML jest grafički jezik za vizualizaciju, specifikaciju, konstrukciju i dokumentaciju rukotvorina u softverskom sistemu.

UML nudi klasični način zapisivanja shematskih prikaza, uključujući konceptualne prikaze u koje spadaju poslovni procesi i funkcije, te čak i neke konkretnije stvari kao što su izjave programskih jezika ili čak DBS (eng. Database Scheme) i dinamičkih programskih komponenata. UML ubrzo postaje glavni standard za izgradnju objektno orijentiranih programa, te je upotrebljiv za rješavanje OP problema.[46]

[45] Harry Halpin, Social Semantics: The Search for Meaning on the Web, Springer, 2012.
[46] Harry Halpin, Social Semantics: The Search for Meaning on the Web, Springer, 2012.

4. PRIMJENA

Kako je koncept semantike usko vezan za jezik, veći dio populacije vjerojatno će se pitati koliko bi to uopće bilo korisno krajnjem korisniku. U primjeru knjižare pokazano je da semantički web, ako ga se zbilja želi u potpunosti iskoristiti, zahtjeva da se pri zapisivanju informacija koristi unificirana struktura kako bi se dolazilo do precizno filtriranih rezultata. Isti ljudi, vjerojatno, nisu ni svjesni da se web aplikacije kojom se služe već godinama koriste tom metodom; tu su prije svega Googleovi servisi, BBC-ov medijski portal, knjižničarski i trgovinski portali, ali i možda najbolji primjer predstavlja upravo društvena mreža Facebook.



Sl. 4.1. Koncept korištenja semantike na društvenoj mreži facebook.com

Još rane 2010. godine, Facebook je počeo uvoditi svoju, tada još uvijek eksperimentalnu mogućnost, kodnog imena Graph Search, koja je praktički jedan od pionira integracije semantičkog weba na ne-apstraktan, već nadasve, koristan način. Ta integracija ponudila je korisnicima facebooka da potražuju stvari na način koji, zapravo, ima smisla. Korisnik

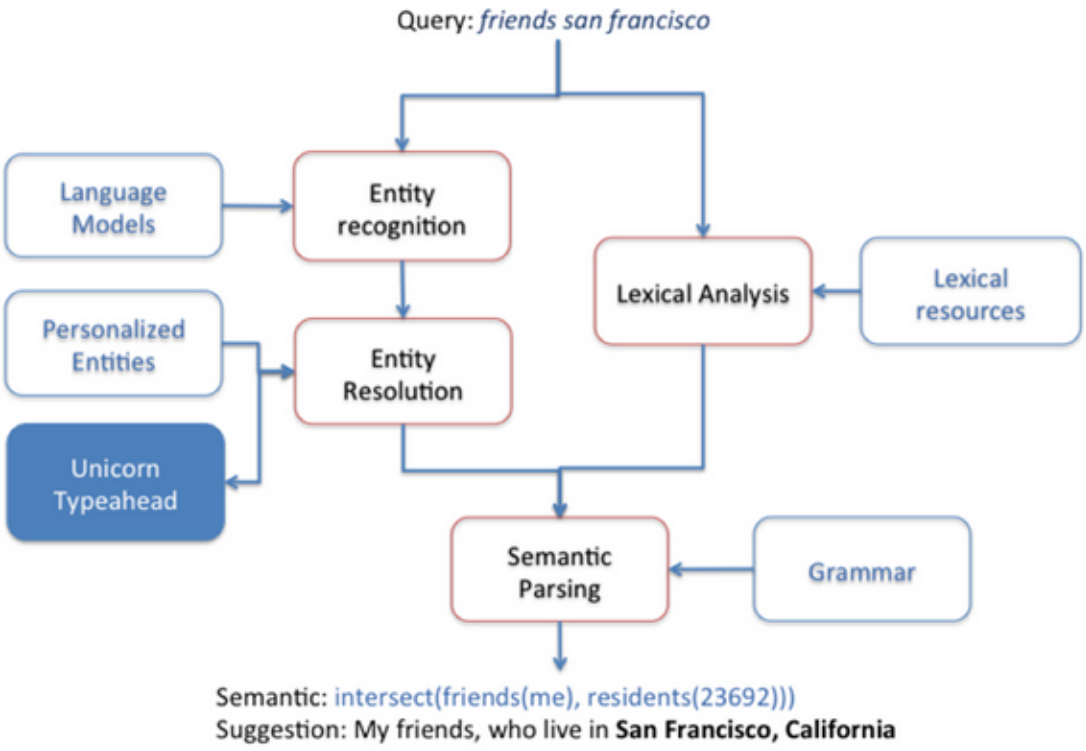
jednostavno u tražilicu treba utipkati: 'coworkers who like sushi', i tražilica će jednostavno izlistati sve ljude zaposlene u korisnikovoj tvrtki koji dijele ljubav prema orijentalnoj kuhinji. Upiše li možda 'photos of weddings uploaded by my friends', tražilica će u tom slučaju, prikazati slike facebook prijatelja koje su negdje u svojim podacima imaju navedeno vjenčanje, bez obzira da li je ono njihovo ili čak sakriveno s naslovnice, što za sobom automatski povlači pitanje narušavanja privatnosti. [47]

Graph Search je stvoren na temelju ogromne količine podataka i vektora: to su podaci o korisnicima, stranicama, grupama, mjestima, slikama, videima, aplikacijama, ali i vezama koje ih spajaju. Pa tako ženska korisnica imena Petra, pohađa srednju umjetničku školu u Osijeku iako joj adresa navodi da ona živi u Belišću, te njezine slike nalažu da se često može zateći na APP-ovom kolodvoru. [48]

Kako je za pretpostaviti da ogromnu bazu podataka kakvu ima Facebook nije jednostavno iskoristiti u semantičke svrhe, njegov programerski tim se potrudio pa prođe kroz sve moguće kombinacije potraživanja i veza između podataka da bi mogli napraviti motor (eng. engine)koji bi pokretao semantičku tražilicu. Komponente arhitekture tog sučelja prikazane su na slici 8. a sadrže:

- Prepoznavanje i rješavanje entiteta
    - Obrađivanje upita i pronalaženje mogućih entiteta i njihovih kategorija i preko tih podataka pretraživanje baze
- Leksičku analizu
    - Analiziranje morfoloških, sintaktičkih i semantičkih informacija u riječima danog upita
- Semantičko raščlanjivanje
    - Filtriranje i prikazivanje N broja rezultata koji zadovoljavaju najviše parametara (odnosno imaju najviše zajedničkih točaka s danim upitom)[49]

[47]http://sodaspeaks.com/2013/02/facebooks-graph-search-the-semantic-webs-first-steps-into-the-mainstream/
[48] https://www.facebook.com/notes/facebook-engineering/under-the-hood-the-natural-language-interface-of-graph-search/10151432733048920/
[49]https://www.facebook.com/notes/facebook-engineering/under-the-hood-the-natural-language-interface-of-graph-search/10151432733048920/



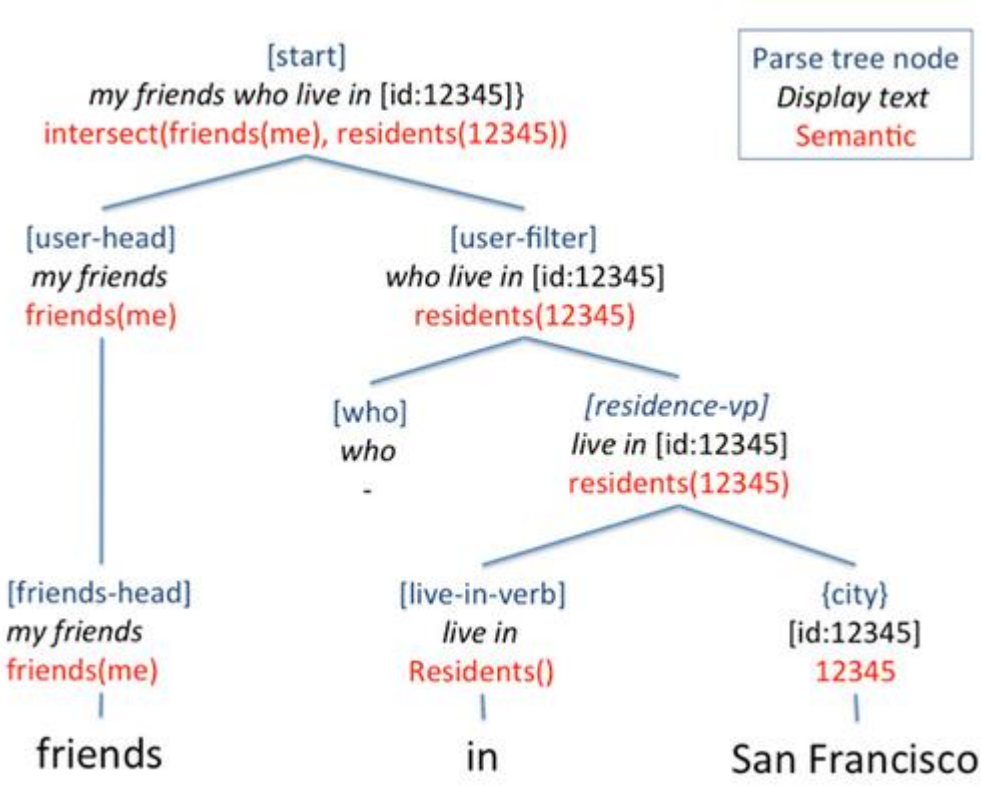
Sl. 4.2. Komponente arhitekture FB Graph Searcha

Graph Search koristi sistem gramatike kako bi upiti bili razumljivi engineu. Za to se koriste parovima simbola i pravila:

- [start] => [users] $1
- [users] => my friend friends(me)
- [users] => friends of [users] friends($1)
- [users] => {user} $1
- [start] => [photos] $1
- [photos] => photos of [users] photos($1)

Simbol [start] je korijen raščlambe pa je tako lijeva strana okupirana ne-terminalnim simbolima, dok sadrži simbole koji upućuju na objekte {korisnik}, {grad}, {grupa}... desna strana također može sadržavati i riječi ili fraze npr. Prijatelji, živi u, radi u, član je, itd..[50]

Drvo raščlambe se generira tako što mu je u korijenu [start] i po sistemu petlje proširuje svoj doseg po metapodacima dok ne dođe do terminalnih simbola. (grafički prikazanu na slici 9.)



Sl. 4.3. Drvo raščlambe za semantički upit

Kako bi otkrili entitete u upitu, Facebookovi programeri su kreirali detektor koji može raščlaniti upit na segmente koji bi mogli biti entiteti te ih klasificirati u kategoriju entiteta npr. u upitu „people who live in san francisco“ 'san francisco' je najvjerojatnije entitet {grad} kategorije. Pa tako postoji preko 20 kategorija ({korisnik}, {grad}, {stranica}, {spol}, {grupa}..). [51]

[50] https://www.facebook.com/notes/facebook-engineering/under-the-hood-the-natural-language-interface-of-graph-search/10151432733048920/
[51] https://www.facebook.com/notes/facebook-engineering/under-the-hood-the-natural-language-interface-of-graph-search/10151432733048920/

Kako je detektor konstruiran da prati jezični model, kao takav sadrži vjerojatnost višestrukih entiteta u upitu kao i onih koji se ponavljaju, te zajedno sa postojećim strojnim modelom daju ovom API-ju mogućnost referenciranja sa 2 jezična modela.

```
p(Class(Qi:j)= K | Q1:N),for all {i, j, k}
```

Za svaki segment upita koji je prepoznat kao entitet, šalje se sistemu za rješavanje entiteta odnosno povlače se podaci iz baze na temelju trenutno obrađivanog, upisanog entiteta. Nakon toga sistem rangira rezultate po dobivenim signalima kao što su raspored u tekstu, fizička pozicija, društvena pozicija...[52]

Facebookovi programeri, također su prikupili ogromnu listu sinonima, pa tako korisnik uz upite kao što su: ' people who like surfing ', ' photos with my friends ' ili ' people who works at facebook ' može koristiti i ' besties from my hood ' što će se prevesti kao ' my friends from my hometown'.

Zadnja stavka u kreiranju drveta raščlambe i prikaza rezultata jest utvrditi gdje ono završava odnosno treba naći terminalne simbole upita. Tokom tog procesa API također obitava informacije kao što su:

1. Početna i završna pozicija upita (i, j) koji se koriste za određivanje pravila
2. Trošak povezan sa svakim parom segmenta i pravila (Rk, Qi, j). Trošak je izračunat na temelju uređivanosti kao i na troškovima samih pravila.[53]

I na posljetku, kako ova tehnologija još uvijek nije usavršena ni gramatika koju Facebook koristi neprobojna, postoji i pričuvna metoda za sve upite koje izlaze van okvira ovog sistema, te se njima i dalje prikazuju rezultati bazirani na ključnim riječima.[54]

Iako Facebookova implementacija ove tehnologije još uvijek treba dosta dorađivanja, njegovi programeri su se nametnuli kao predvodnici u revoluciji weba, a njihov Graph API, koliko god kontroverzan olakšava programerima drugih web servisa korištenje

---

[52] https://www.facebook.com/notes/facebook-engineering/under-the-hood-the-natural-language-interface-of-graph-search/10151432733048920/

[53] https://www.facebook.com/notes/facebook-engineering/under-the-hood-the-natural-language-interface-of-graph-search/10151432733048920/

[54] https://www.facebook.com/notes/facebook-engineering/under-the-hood-the-natural-language-interface-of-graph-search/10151432733048920/

facebookovih usluga, a sami korisnici njihove mreže mogu uživati blagodati jednostavnijeg i preciznijeg pretraživanja.

5. ZAKLJUČAK

U ovom trenutku još uvijek ne znamo kakva se budućnost sprema ideji o semantičkom webu. Veliki broj tvrtki i organizacija ulaže ogromne napore i resurse da se osiguraju softverska i hardverska infrastruktura potrebna za razvoj i funkcioniranje aplikacija i servisa zasnovanih na ovom principu.

Doduše, za njegov eventualni uspjeh, od presudnog utjecaja neće biti softver, već korisnici. Pitanje je kad će semantički tehnologija privući dovoljan broj korisnika da bi motivirala i nove korisnike da ga počnu koristiti. Najvedrije prognoze govore da će semantički web za 2 do 4 godina izaći iz prototipske faze i početi se praktično primjenjivati, a da će između 2018. i 2020. godine doživjeti svoju punu ekspanziju.

Međutim, već je sada razvoj semantičkog weba dao poticaj velikom broju drugih disciplina softverske branše, a posebno se to odnosi na one koji se bave konceptualizacijom informacija i podataka. Samo će sutra pokazati kakav će utjecaj razvoj ove tehnologije imati na društvo i da li će se ostvariti vizija Tim Berners-Leea i drugih zagovornika ove ideje.

LITERATURA

[1] Berners-Lee, Hendlers & Lassila, 2001; Veltman, 2001, 2004

[2] Berners-Lee ,The Semantic Web – LCS

[3] http://www.w3.org/2002/Talks/09-lcs-sweb-tbl/ (20.06.2016.)

[4] Fensel, D., Ontologies: Dynamic networks or formally represented meaning, Stanford University, 2001.

[5] Harry Halpin, Social Semantics: The Search for Meaning on the Web, Springer, 2012.

[6] Robin Cover: XML and Semantic Transparency

[7] http://xml.coverpages.org/xmlAndSemantics.html (20.06.2016.)

[8] http://www.w3schools.com/xml/xml_whatis.asp#top (20.06.2016.)

[9] Antoniou, Grigoris; Frank van Harmelen (2008). A semantic Web primer. 2nd ed. London, MIT Press.

[10] Smith, Michael K. i dr. [ur.] (2004). OWL Web Ontology Language – Guide

[11] http://en.wikipedia.org/wiki/Semantic_Web (19.06.2016.)

[12] http://sodaspeaks.com/2013/02/facebooks-graph-search-the-semantic-webs-first-steps-into-the-mainstream/ (18.09.2016.)

[13] https://www.facebook.com/notes/facebook-engineering/under-the-hood-the-natural-language-interface-of-graph-search/10151432733048920/ (18.09.2016.)

SAŽETAK

Zadatak ovog rada jest definirati što bi to bio semantički web, te opisati za što bi ta tehnologija služila i kako bi ona unaprijedila postojeća korisnička iskustva s Internetom i webom općenito. U radu su također obrađene radne grupe koje su sudjelovale na dizajniranju standarda i popisivanju principa ove tehnologije.

Nakon tog uvoda, nastavlja se sa opisivanjem arhitekture semantičkog weba kako je definiran standardom. Potom je ustvrđeno koje tehnologije i jezici su preporučeni za korištenje ove tehnologije u njenoj, još uvijek, eksperimentalnoj verziji.

I na kraju je obrađen sam termin ontologije koja je služila kao glavni pokretač ove ideje, te je dan primjer gdje se semantički web koristi u stvarnom životu.

Ključne riječi: opisivanje**,** semantički, web, standard, tehnologije, ontologija

ABSTRACT

Semantic web and its usage

The goal of this thesis is explanation of the term semantic web, as well as to describe use cases where this new technology would be beneficial to end user experiences. Workgroups that are in charge of standard definition and principles collection of this technology are mentioned and explained.

After introduction chapter, next chapters describe architecture of the semantic web, as is proposed by W3C standard. Later in thesis, technologies and languages are recommended for this, still in experimental phase, technology.

In the end, term ontology is explained, which is main drive behind the idea of semantic web, and finally, the example of real life implementation was presented.

Ključne riječi: describing**,** semantics, web, standard, technology, ontology

ŽIVOTOPIS

Tomislav Petrović, rođen 05.12.1993 u Vinkovcima, Hrvatska. Osnovnu školu „ Matija Antun Reljković“ završava u Cerni s vrlo dobrim uspjehom, a srednju školu Tehničku školu „Ruđera Boškovića“ završava u Vinkovcima s dobrim uspjehom.

Na sveučilištu Josipa Juraja Strossmayera u Osijeku upisuje Elektrotehnički fakultet, stručni studij smjer Informatika 2012. godine. Tijekom studiranja odradio praksu u privatnoj tvrtki Informatika Fortuno u Vinkovcima.

Od stranih jezika koristi se Engleskim jezikom.