

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Ivo Prelc

**TEHNOLOGIJE SEMANTIČNEGA SPLETA IN
NJIHOVA UPORABA V PODJETJU ZA
PROIZVODNJO VINA**

DIPLOMSKO DELO
NA VISOKOŠOLSKEM STROKOVNEM ŠTUDIJU

Mentor: viš. pred. dr. Damjan Vavpotič

Ljubljana, 2011

Št. naloge: 00110/2011

Datum: 05.04.2011



Univerza v Ljubljani, Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Kandidat: **IVO PRELC**

Naslov: **TEHNOLOGIJE SEMANTIČNEGA SPLETA IN NJIHOVA UPORABA V
PODJETJU ZA PROIZVODNJO VINA**
**USE OF SEMANTIC WEB TECHNOLOGIES IN A WINE PRODUCTION
COMPANY**

Vrsta naloge: Diplomsko delo visokošolskega strokovnega študija prve stopnje

Tematika naloge:

V okviru diplomske naloge predstavite področje semantičnega spleta in tehnologije, ki se pri tem uporabljajo. V prvem delu naloge predstavite prednosti, ki jih semantični splet prinaša, pristope k uvajanju semantičnega spleta in težave s katerimi se pri uvajanju semantičnega spleta srečujemo. V drugem delu naloge preučite možnosti za uporabo tehnologij semantičnega spleta za potrebe opisa izdelkov podjetju za proizvodnjo vina. Predstavite primer opisa takšnega izdelka ter na kratko opišite prednosti, ki jih takšen opis prinaša v konkretnem primer.

Mentor:

viš. pred. dr. Damjan Vavpotič



Dekan:

prof. dr. Nikolaj Zimic

IZJAVA O AVTORSTVU

diplomskega dela

Spodaj podpisani Ivo Prelec,
z vpisno številko 63070436,

sem avtor diplomskega dela z naslovom:

Tehnologije semantičnega spleta in njihova uporaba v podjetju za proizvodnjo vina.

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal samostojno pod mentorstvom (naziv, ime in priimek) viš. pred. dr. Damjana Vavpotiča;
- so elektronska oblika diplomskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko diplomskega dela;
- soglašam z javno objavo elektronske oblike diplomskega dela v zbirki »Dela FRI«.

V Ljubljani, dne _____ Podpis avtorja: _____

ZAHVALA

Rad bi se zahvalil mentorju prof. viš. pred. dr. Damjanu Vavpotiču za predlagano temo in strokovne napotke, ki so pripomogli k nastajanju diplomske naloge.

Prav tako se zahvaljujem odgovornim v podjetju Vipava 1894 d.d., da so mi omogočili študij ob delu in sofinanciranje šolnine.

Posebej se zahvaljujem domačim, prijateljem, sodelavcem in vsem, ki so me v času študija spodbujali, motivirali in potrpeali z mano.

KAZALO VSEBINE

1 UVOD.....	3
2 SEMANTIČNI SPLET.....	4
2.1 Semantični splet kot znanost.....	7
2.2 Pristopi k semantičnemu spletu.....	9
2.2.1 Pet glavnih pristopov pri dodeljevanju pomena podatkom.....	9
2.2.2 Dilema, kako se bo semantični splet uvajal.....	10
2.3 Težave pri uvajanju in kritike.....	12
2.3.1 Zakaj semantični splet ne bo uspel.....	12
2.3.2 Pomeni semantični splet manj dela za spletne razvijalce.....	12
2.3.3 Lažni metapodatki.....	12
2.3.4 Problem integracije različnih ontologij.....	13
3 TEHNOLOGIJE SEMANTIČNEGA SPLETA.....	14
3.1 Konzorcij svetovnega spleta W3C in njegov pristop.....	14
3.2 Semantični splet kot skupek standardov.....	15
3.2.1 Elementi semantičnega spleta.....	17
3.2.2 Ontologije in OWL.....	18
3.2.3 Metapodatki in RDF.....	19
3.2.4 Inteligentni agenti, SPARQL in Linked Data.....	21
3.3 Mikroformati (uF).....	23
3.3.1 Kako okolica sprejema mikroformate.....	24
3.4 RDFa.....	26
4 UPORABA SEMANTIČNEGA SPLETA V PODJETJU ZA OPIS VIN.....	27
4.1 O podjetju.....	27
4.2 Spletna stran podjetja.....	27
4.3 Prednosti, ki jih prinaša uporaba semantike na spletni strani.....	28
4.4 Izbira tehnologije.....	30
4.5 Kako začeti z vnosom semantike.....	31
4.6 Katere vsebine označevati.....	33
4.7 Semantično označevanje izdelkov.....	34
4.7.1 Zasnova povezav med elementi.....	35
4.7.2 Izbira ontologije za opis vina.....	36
4.7.3 Na poti k lastni ontologiji.....	39
5 SKLEP.....	42
6 LITERATURA IN VIRI.....	44

KAZALO SLIK

Slika 1: Razvoj spleta po dekadah – dogaja se Web 3.0.	4
Slika 2: Intenzivnost poizvedovanja po različicah spleta na Googlu	5
Slika 3: Aplikacije posameznih dekad spleta	6
Slika 4: Oblak povezanih odprtih podatkov	6
Slika 5: Kaj o semantičnem spletu misli Google – http://www.googlism.com	7
Slika 6: Aktivna področja pri obdelavi besedila.....	8
Slika 7: Razporeditev pristopov k Semantičnemu spletu glede na zahteve	10
Slika 8: Različna pristopa glede na način vnosa semantike	11
Slika 9: Plasti povezanih tehnologij semantičnega spleta	16
Slika 10: Elementi semantičnega spleta	17
Slika 11: Opis domene, izbira ontologije	19
Slika 12: Trojček RDF.....	20
Slika 13: Sklepanje iz podatkov	20
Slika 14: SNORQL – vmesnik za poizvedovanje SPARQL	22
Slika 15: Dodatka za Firefox: Operator in Tails.....	25
Slika 16: Sklad tehnologij Semantičnega spleta.....	26
Slika 17: CMS aplikacija Moj Portal.....	28
Slika 18: SEO analiza spletne strani 'vipava1894.si'	29
Slika 19: Uporaba Mikroformatov in RDFa v spletu (% vseh spletni strani)	31
Slika 20: Spletna aplikacija za generiranje zapisa hCard	32
Slika 21: Graf RDF.....	35
Slika 22: Rezultat, kot ga prikaže Google, ko uporabimo GoodRelations.....	38
Slika 23: Primer razrednega opisa vina	40
Slika 24: Delovno okolje orodja za kreiranje ontologij Protégé	41

KAZALO TABEL

Tabela 1: Semantični splet z veliko in malo začetnico.....	14
Tabela 2: Glavna načela, ki vodijo delo v W3C.....	15
Tabela 3: Dvoumen zapis z XML	19
Tabela 4: Nedvoumen zapis povezave z RDF.....	19
Tabela 5: Delujoče končne točke SPARQL	22
Tabela 6: Kaj mikroformati so in kaj niso	23
Tabela 7: Najbolj uporabljeni mikroformati.....	24
Tabela 8: Kdo koristi/vgrajuje uF in kakšna orodja uporablja	25
Tabela 9: Nekateri izmed možnosti za strukturirano označevanje	32
Tabela 10: Ontologije v uporabi.....	33
Tabela 11: Zapisa z mikroformatoma hCard in geo	33
Tabela 12: Kontakt, zapisan z RDFa	34
Tabela 13: Zapis v RDF/XML.....	36
Tabela 14: Zapis v N3	36
Tabela 15: Lastnosti, ki opisujejo vino – Barbera Ventus.....	37
Tabela 16: Opis vina kot izdelka, z ontologijo GoodRelations.....	38

KRATICE, OKRAJŠAVE, SIMBOLI

API	(Application Programming Interface) – vmesnik uporabniškega programa
CMS	(Content Management System) – sistem za upravljanje vsebin
DAML	(DARPA Agent Markup Language) – označevalni jezik DARPA agentov
DARPA	(Defense Advanced Research Projects Agency) – agencija za napredne raziskave amerškega obrambnega ministrstva
GRDDL	(Gleaning Resource Descriptions from Dialects of Languages) – paberkovanje opisa virov iz narečij jezikov
HTML	(Hypertext Markup Language) – označevalni jezik za oblikovanje in povezovanje večpredstavnostnih dokumentov
LSI	(Latent Semantic Indexing) – prikrito semantično indeksiranje
OIL	(Ontology Interchange Language) – jezik za izmenjavo ontologij
OWL	(Web Ontology Language) – standard za izražanje ontologije
RDFa	(Resource Description Framework - in - attributes) – konceptualni model za opisovanje virov
SEO	(Search engine optimization) – optimizacija spletnih strani v skladu z iskalniki
SKOS	(Simple Knowledge Organization System) – sistem za organizacijo znanja
SPARQL	(Simple Protocol and RDF Query Language) – protokol in poizvedovalni jezik za RDF
SQL	(Structured Query Language) – strukturirani povpraševalni jezik
SS	Semantični splet
W3C	(World Wide Web Consortium) – konzorcij svetovnega spleta
XSLT	(Extensible Stylesheet Language Transformations) – pretvorbe razširljivega slogovnega jezika

POVZETEK

V podjetju lahko izkoristimo semantične tehnologije za spletno prodajo in oglaševanje izdelkov ali storitev. V te namene zadostujejo v spletu že deklarirani objekti za semantični opis poslovnih učinkov, za opis njihovih specifičnih lastnosti pa je običajno potreben razvoj lastne ontologije. To ontologijo lahko izkoristi katerakoli aplikacija v spletu za poizvedovanje po naših semantično opisanih primerih.

Prvi sklop naloge zajema razvoj Semantičnega spleta. Začne se z vizijo ustanovitelja spleta Tima Berners-Lee-ja o povezanih podatkih v spletu, in nadaljuje z iskanjem načina za predstavitev spletnih vsebin računalnikom na bolj enostaven način. Opisani so pristopi k dodeljevanju pomena podatkom ter težave in vprašanja, ki se odpirajo pri razvoju semantičnih tehnologij. Semantični splet temelji na nadgrajevanju obstoječih spletnih tehnologij in spreminja splet v podatkovno bazo. Kot njegova zaključena celota so predstavljene spletne ontologije, okvir za opisovanje virov in inteligentni agenti.

Zadnje poglavje začne z opisom podjetja in njegove spletne strani. Zajema prednosti, ki jih tu prinaša uporaba spletnih tehnologij. Tehnološki pristopi k semantičnemu označevanju podatkov na spletni strani obravnavajo Mikroformate, RDFa in Microdata. Srečamo se z vprašanjem vnosa semantike preko sistema CMS. Ciljne skupine objektov za semantični opis so vina. Njihove specifične lastnosti potrjuje izbor tehnologije RDFa. Nadaljnje raziskovanje napeljuje v razvoj lastne ontologije za opis domene vina in prikaže možnost njene uporabe na spletni strani.

Ključne besede: semantični splet, mikroformati, RDFa, ontologije, vino

ABSTRACT

In a company, one can take advantage of semantic technologies for online sales and advertising of products or services. Objects for semantic descriptions of business effects are already declared on the web and suffice for such purpose. On the other hand, describing specific characteristics of the business effects generally requires developing our own ontology. Afterwards, this ontology can be used by any application on the web for querying our semantically described cases.

The first part describes the Semantic Web development. It starts with a vision of the web founder, Tim Berners-Lee, about web linked data, and continues to search a way to present web contents to computers in a simpler manner. This part also describes approaches to assigning the meaning to the data, and issues which arise in developing semantic technologies. Semantic Web works on the already existing web technologies, and changes the web into a database. Web ontologies, resource description framework and intelligent agents are presented as its complete unit.

The final chapter begins with a description of the company and its website. It captures the benefits brought by the use of the semantics. Technological approaches to the semantic marking of data on the website address Microformats, RDFa and Microdata. We encounter the issue of semantics input via the CMS. Wines are the target groups of objects for semantic description. Their specific characteristics confirm choosing RDFa technology. Further research leads to the development of wine domain ontology and displays its potential use on the website.

Keywords: semantic web, microformats, RDFa, ontologies, wine

1 UVOD

Tim Berners-Lee, so-izumitelj svetovnega spleta, direktor W3C konzorcija in oče semantičnega spleta, pravi: »*Sanjam o spletu, v katerem bodo računalniki sposobni analizirati vse podatke, ki krožijo po internetu: vsebino, povezave in interakcije med ljudmi in računalniki. Semantični splet, ki bi to omogočal, še mora nastati, a ko bo prišel na dan, bodo v vsakodnevnem življenju, prometu, birokraciji, stroji komunicirali s stroji.*« [12]

Prvotna Berners-Lee-jeva vizija o prihodnosti svetovnega spleta je bila zelo abstraktna. Še zdaj posamezniki povezujejo semantični splet z umetno inteligenco, drugi pa menijo, da je njegovo bistvo opisovanje podatkov. Analizirajmo torej besedo semantika. To je nauk o pomenu, zanima jo pomen besed. Semantični splet je torej splet s pomenom. Tudi današnji splet, vsebina spleta, ima svoj pomen, vendar je to še vedno bolj splet, razumljiv ljudem, prihodnji splet pa naj bi bil razumljiv tudi računalnikom, strojem. Gre torej za to, da skušamo računalnikom pomagati, da bi se znali pogovarjati med seboj v omrežju. [9]

Komunikacija računalnik – računalnik ni mogoča brez vnaprejšnjega dogovora o strukturi in vsebini podatkov, ki se izmenjujejo. Izvirna ideja je namreč ustvariti metapodatke za opis podatkov, kar bo omogočilo računalnikom procesiranje pomena stvari. Poznamo XML, vendar še vedno se moramo pozanimati, kaj pomeni posamezna oznaka podatka, da jo lahko uporabimo. Nimamo skupnih slovarjev za opis pomena podatkov. Na primer, ali 'poštna koda' in 'zip code' pomenita isto? Potreben je torej standard, dogovor o skupnem jeziku, ki bi dal pomen podatkom. Doslej smo imeli dogovorjene predvsem jezike za opis oblike¹. Stvari, ki imajo različno formo, vendar isti pomen, je cilj poenotiti. V prvih letih razvoja se semantični splet ukvarja s pomenom podatkov tipično v besedilih, kasneje pa se predvideva njegovo širitev tudi na ostale multimedijske vsebine: slike, zvočne zapise, videe itd. [9]

Cilj Semantičnega spleta je izpostaviti glavnino virov informacij na svetovnem spletu kot podatke, ki si jih bodo znali računalniki samodejno razložiti. Z dodajanjem pomena podatkom v spletnih dokumentih bo splet postal univerzalni medij za izmenjavo podatkov. Aplikacije bodo znale uporabiti vsebino spleta kot bazo podatkov. Na primer: Pogledamo na stran borze in aplikacija nam svetuje, katere delnice naj kupimo. Opišemo zdravstvene težave, pa nam ponudi kontakt zdravnika, izpiše recept tablet itd. V končni fazi bi radi našli boljši način za iskanje po ključnih besedah in prvi korak v tej smeri je označevanje podatkov z njihovo vsebino in dodeljevanje povezav mednje. Seveda je pri uporabi terminov poimenovanja treba paziti, da so stvari standardne, da se ujemajo z nečem, kar je že uveljavljeno.

Številne tehnologije, ki jih predlaga konzorcij W3C, že obstajajo in se uporabljajo v različnih kontekstih, še posebej tistih, ki se ukvarjajo z informacijami, ter kjer je izmenjava podatkov skupna potreba, kot so znanstvene raziskave in izmenjava podatkov med podjetji. Ker pa se prvotna zamisel o SS kot sistemu, ki omogoča strojem razumevanje kompleksnih človeških zahtevkov še ni realizirala, so se s podobnimi cilji pojavile še druge tehnologije, kot npr. Mikroformati in Microdata. Tako lahko za uporabo semantike na spletni strani izbiramo med temi tehnologijami, kot prvo glede na to, ali tehnologija zadovoljuje naše potrebe, nato pa še glede na njeno priljubljenost oziroma podprtost v spletu.

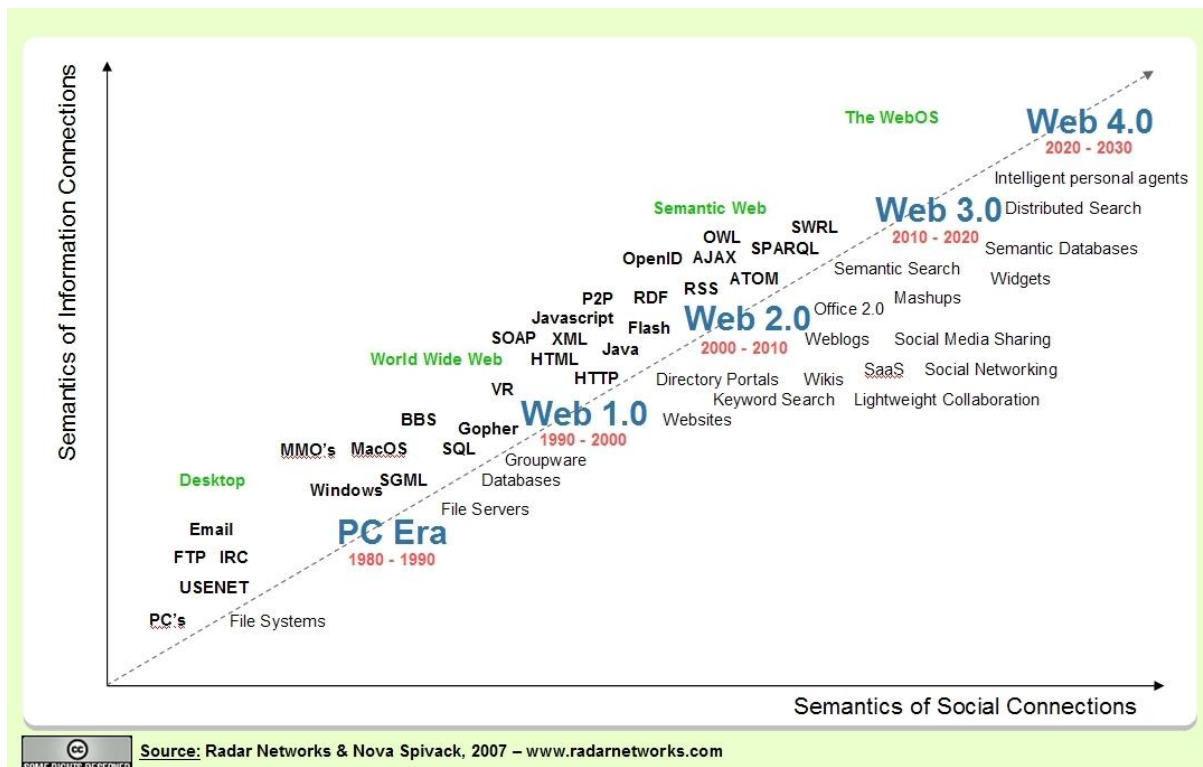
¹ HTML in CSS opisujeta, kako se bodo podatki na spletni strani prikazali, ne pa kaj pomenijo.

2 SEMANTIČNI SPLET

Splet je bil prvotno le splet dokumentov. S preprostim klikom na povezavo v spletnem brskalniku je brskalnik sprožil zahtevo spletnemu strežniku, da nam pošlje dokument, ki nam je bil nato prikazan. Brskalnika ni skrbelo vsebina dokumenta, ampak je samo sledil svojim notranjim pravilom za prikazovanje strani. Podatke, ki smo jih po zahtevi prejeli, smo morali razumeti in povezati sami. S semantičnim spletom bodo za povezovanje rezultatov iskanja in prikaz smiselnega odgovora skrbele aplikacije – spletni agenti.

Tim Berners-Lee pojasnjuje, da pri SS ne gre za iskanje neke odmevne aplikacije, ki bi postavila mejnik za novo obdobje spleta. Ni ravno navdušen nad razdelitvijo razvoja spleta v dekada, in poimenovanja označevanja ter povezovanja podatkov z 'Web 3.0'. Poudarja, da gre za skupno delo, množico aplikacij, ki jih spletni mojstri postopoma dodajajo, prehod pa je počasen. Razvija se skupno okolje, ki omogoča izmenjavo podatkov in njihovo ponovno uporabo čez različne aplikacije, bodisi v podjetjih ali drugih skupnostih. [9]

Semantični splet je prva začela uporabljati vojska z namenom, da bodo lahko sklepali iz svojih podatkov, na primer ugotavljali dejstva, katera mesta bodo najbolj verjetno tarča terorističnih napadov. Ameriška agencija DARPA je ustvarila označevalni jezik DAML. V sodelovanju z Evropsko unijo je kasneje razvila DAML+OIL, ki se šteje za enega semantično najbolj izraznih jezikov za opis dokumentov v svetovnem spletu. Na konceptih DAML+OIL je bil leta 2003, pod okriljem konzorcija W3C, zgrajen jezik za kreiranje ontologij OWL. Spodnja slika prikazuje glavne valove sprememb, ki se dogajajo nekako v dekadah. S prehodi med dekadami lahko opazimo izmenjavo med pomembnostjo izgleda (front end) in vsebine (back end) spletnih dokumentov.



Slika 1: Razvoj spleta po dekadah – dogaja se Web 3.0.

WEB 1.0 je vse o izgradnji infrastrukture. Je množica datotek, ne podatkov. Srečamo se s protokoli: HTML, HTTP, JPG, GIF in URL. Dokumenti so distribuirani po spletu, podatki pa še vedno centralizirani. Gre za splet z mnogimi uporabniki in le nekaj ponudniki vsebin. Oglaševanje ni kontekstualno, večina vsebin HTML je ročno vnesenih.

WEB 2.0 je družbeni splet. Kot pojem je deležen nešteti marketinških zlorab. Gre za splet, v katerem so podatki centralizirani, ponudniki vsebin pa so lahko tudi uporabniki. Dostop do podatkov je olajšan s pomočjo tehnologij RSS in API. Veliko vsebin HTML je avtomatsko generiranih. Pomanjkljivost je, da je Web 2.0 še vedno zgolj spletišče posameznih dokumentov in aplikacij, ki centralizirane podatke hranijo predvsem zase.

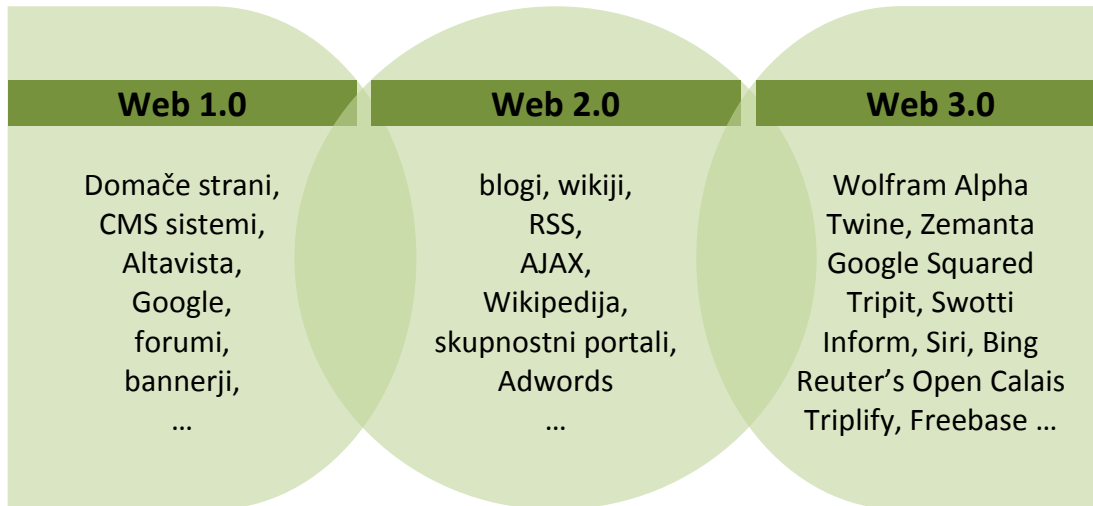
WEB 3.0 je baza podatkov oziroma podatkovni svetovni splet. Podatki so distribuirani, razdrobljeni in vseprisotni na spletu. To je splet za ljudi in končno tudi inteligentne stroje (agente), ki znajo te podatke razumeti, analizirati in ponuditi ljudem. Googlov direktor, dr. Eric Schmidt, je podal tako definicijo: *Web 3.0 bodo povezane aplikacije, majhne, hitre, zmožne izvajanja na različnih platformah, distribuirane viralno (socialna omrežja, elektronska pošta itd.), podatki pa raztreseni po medmrežju.*

WEB 4.0 bi lahko bil splet umetne inteligence, superiorni splet, ki reši vse probleme.

Web 3.0 je tako imenovani tretji glavni val, tretja dekada spleta, ki prinaša semantični splet. Omenili smo, da glavni izumitelj spleta samega, Tim Berners-Lee, ideje o različicah spleta ne mara preveč in si predstavlja semantični splet bolj usklajeno s svojo izvirno vizijo – kar pomeni, da smo dejansko priča razvoju spleta, o kakršnem je on razmišljal pred skoraj 20 leti. Semantični splet je veliko starejša ideja kot Web 3.0, kar lahko vidimo tudi na Google Trends, ki prikazuje intenzivnost poizvedb po novem spletu (Slika 2).

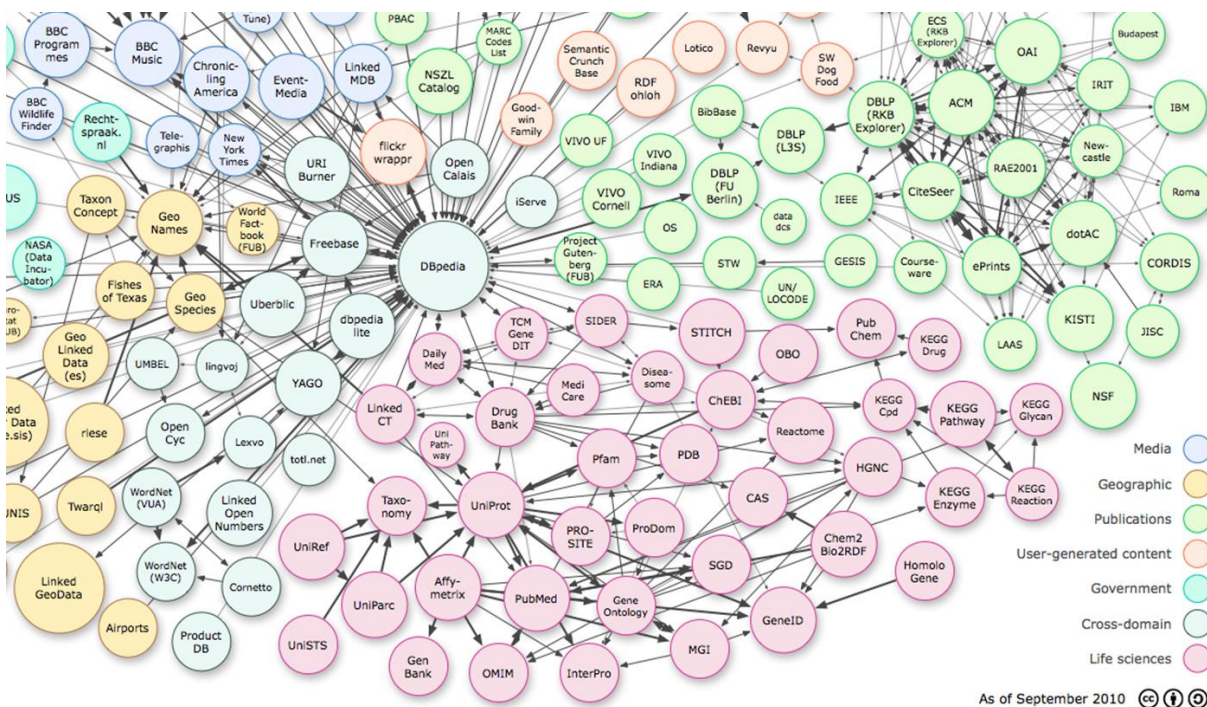


Slika 2: Intenzivnost poizvedovanja po različicah spleta na Googlu



Slika 3: Aplikacije posameznih dekad spleta

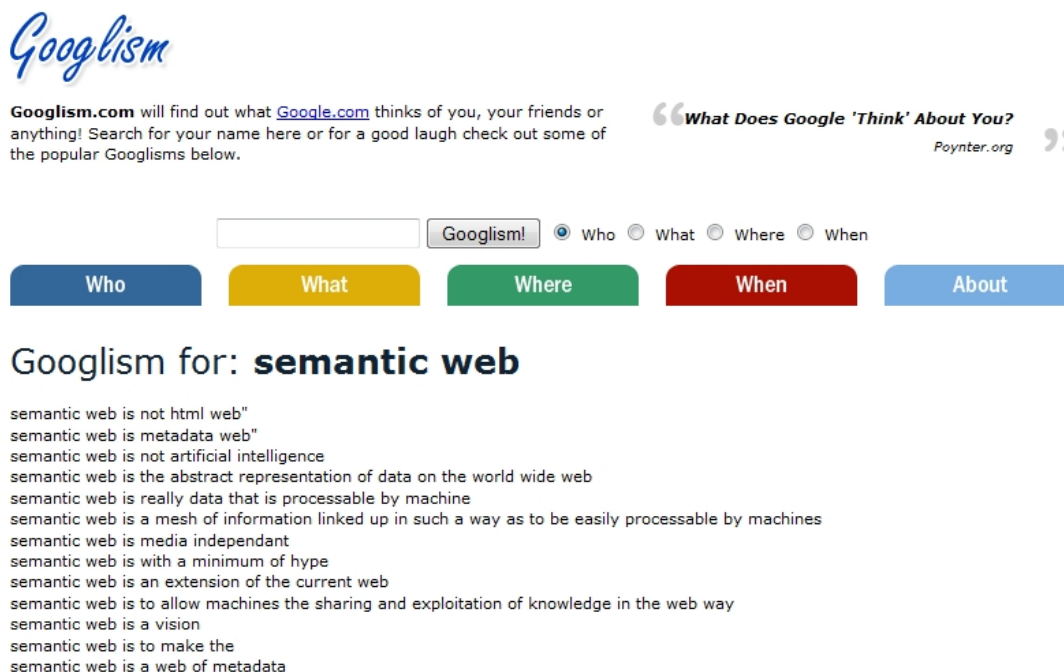
SS prinaša tudi poslovne prednosti. Z izpostavljanjem javnih poslovnih podatkov, kot povezanih podatkov, je iskanje izdelkov in storitev po spletu enostavnejše, za ljudi in spletne agente. SS je postal zanimiva marketinška tehnika.



Slika 4: Oblak povezanih odprtih podatkov

2.1 Semantični splet kot znanost

Poglejmo si nekaj sprejemljivih pomenov semantičnega spleta, ki nam jih vrne Googlism (Slika 5):



Slika 5: Kaj o semantičnem spletu misli Google – <http://www.googlism.com>

- semantični splet ni html splet – *html je opis oblike besedila, ne pa pomena*
- semantični splet so metapodatki – *podatek, ki govori o podatku, to je pravzaprav semantika, toda s kakšnimi jeziki naj opišemo podatke?*
- semantični splet ni umetna inteligenca
- semantični splet je abstraktna predstavitev podatkov na svetovnem spletu – *abstraktna v tem smislu, da so to podatki, ki govorijo o podatkih*
- semantični splet je medijsko neodvisen – *pogovarjamo se o pomenu, ki je lahko kjerkoli: v računalniku, telefonu, na papirju ...*
- semantični splet je razširitev trenutnega spleta – *prelivanje tehnologije*
- semantični splet je W3C-jeva velika vizija bolj strukturiranega in strojnega ... *W3C konzorcij – neprofitna ustanova, ki skrbi za standarde, ki so na spletu*
- semantični splet napravi stvari uporabne ... *tudi za računalnike; grafično dodelana stran s flash dodatki je za človeka pomensko razumljiva, za računalnik pa ne*
- semantični splet temelji na sistemih, ki bolj podrobno opisujejo podatke

Srce semantičnega spleta predstavlja tehnologija, ki uporabnikom omogoča iskanje zelenih informacij ter korelacij med njimi, ne glede na to, kje na spletu se iskani podatek nahaja. Gre za pomensko, ne sintaksno, označevanje in povezovanje podatkov, s čimer lahko le-ti postanejo uporabni za strojno obdelavo.

Eden izmed problemov pri obdelavi besedila je, kako iz obširnega besedila narediti povzetek, ki ohrani bistvo. Sklepamo lahko, da moramo za pripravo dobrega povzetka poznati vsebino besedila. Najti bistvene točke in mejo², v kakšnem obsegu naj se te zajamejo v povzetek, je za aplikacijo zahtevna naloga. Kot prvo je tu potrebna analiza besedila. Kako analizirati besedilo in za kakšne namene, pa je izziv posameznih znanstvenih področij, ki se ukvarjajo z obdelavo besedila (Slika 6). [9]



Slika 6: Aktivna področja pri obdelavi besedila

Strojno učenje, Odkrivanje zakonitosti v besedilih – Osredotočeno je na način računalniškega učenja, analitske tehnike in statistično obdelavo podatkov (zelo dobri na inštitutu Jožef Stefan).

Obdelava naravnega jezika – Ukvarjanje s pravili slovnice, iskanje slovničnih struktur v besedilu (osebek, povedek, predmet).

Iskanje informacij – Kako poizvedovati, kako iz množice podatkov izvleči prave? Pomembna je struktura podatkov in indeksiranje le teh.

Semantični splet – Poskuša pripeljati še pomen besedila. Odkrivanje zakonitosti v besedilih in Analitika gledata na besedilo kot na neko množico besed in se ukvarjata predvsem z agregacijo in ekstrakcijo 'surovih' podatkov. SS se po drugi strani ukvarja predvsem z integracijo in predstavitvijo podatkov, ki lahko poveča ponovno uporabo danih informacij.

Vsa področja se nekje stapljajo, prekrivajo in govorijo o drugih področjih. Stapljajo v smislu ljudi, idej, problemov, orodij in tehnik, ki jih uporabljajo.

² Na inštitutu Jožef Stefan so razvili model aplikacije za ustvarjanje povzetka, z nastavljenim pragom po pomembnosti vsebine, ki naj se zajame v povzetek.

2.2 Pristopi k semantičnemu spletu

Pri uvajanju semantičnega spleta poznamo več pristopov. Ti se ločijo po tem, da za poudarjanje pomena stvari nekateri potrebujejo pametnejše aplikacije, drugi pa pametnejše podatke. Najpametnejše aplikacije so tiste, ki znajo iz stavka v naravnem jeziku izvleči pomen. Pametnejši podatki jim to delo olajšajo. Dejansko so to metapodatki – opisi podatkov vključeni v podatke, zapisani po standardih tako, da jih vsaka aplikacija lahko razume. Ideja je, da podatke, ki jih neka aplikacija ustvari, zna druga aplikacija brati – tako vsi ustvarjamo podatke, ki lahko koristijo vsem. Toda, kdo bo vnesel vse te metapodatke? Katera orodja bi lahko uporabil za označevanje?

2.2.1 Pet glavnih pristopov pri dodeljevanju pomena podatkom

Označevalni pristop (Tagging approach) – Podatkom ročno dodajamo značke, ki opisujejo pomen podatka.

Primer: <naslov>Izdelki</naslov>, <artikel><ime>Sauvignon</ime></artikel> ...

Za: To je enostaven pristop in ne potrebuje učenja nove tehnologije.

Proti: Zanesti se moramo na ljudi, da dodajajo značke, in če jih, da jih dodajajo pravilno.

Statistični pristop (Statistics approach) – Tolmačenje velike količine podatkov z uporabo matematičnih in statističnih orodij.

Primer: LSI (Google), odprt kodni iskalnik Lucene ...

Za: Zelo razširljiv, prilagodljiv, enostaven za uporabnike in neodvisen od jezika.

Proti: Ne razume popolnoma podatkov. Algoritmi najdejo povezave, vendar ne razumejo pomena. Problem je napisati pravo poizvedbo z besedami v pravem vrstnem redu. Kako vprašati, da bo sistem razumel, kaj iščemo?

Jezikoslovni pristop (Linguistics approach) – Poskus ustvariti programsko opremo, ki razume jezik, njegovo slovnico, sintakso in enote pomena. Išče pomen stavka, kot bi to naredil človek.

Primer: Podjetje Powerset (Microsoft, 2008), Podjetje Inxight (Business Objects, 2007)

Za: Možnost poizvedb v naravnem jeziku. Črpa znanje iz besedila.

Proti: Ta pristop je računsko zahteven, potrebno ga je programirati za vsak jezik posebej.

Pristop semantičnega spleta (Semantic web approach) – Bistvo je v ustvarjanju smiselne strukture za splet. Meta-splet poskuša graditi spletne baze podatkov, omogoča, da delujejo spletne poizvedbe kot poizvedbe v relacijski bazi podatkov.

Primer: Twine (Radar Networks – Evri, 2010), DBpedia, Metaweb (Google, 2010)

Za: Natančnejše poizvedbe. Ni tako računsko zahteven kot jezikovni pristop.

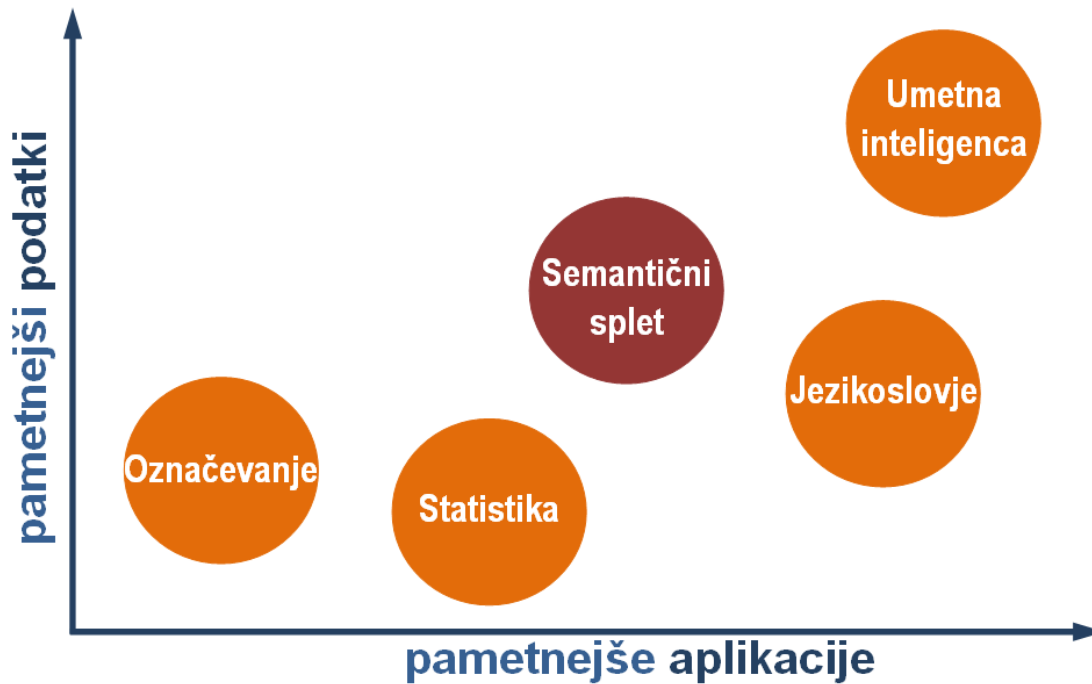
Proti: Potrebuje metapodatke. S strani urednikov spletišč nesprejeta OWL in RDF, čeprav je oba potrdil W3C. Premalo spodbude za ljudi, da bi dodajali semantiko na svoje spletne strani.

Pristop z umetno inteligenco (Artificial Intelligence approach) – Približevanje znanju strokovnjaka. Strategije učenja, ki jih uporablja človek. Konceptualna analiza podatkov.

Primer: WolframAlpha

Za: Uporabniku ne bo potrebno več razmišljati, kako vprašati računalnik. Vprašal bo na svoj način. Kaj je uporabnik vprašal in kaj ponuditi kot odgovor, pa bo razmišljal računalnik.

Proti: Računsko drag pristop, težko dosegljiv s sedanjo infrastrukturo in znanjem. Kako vnašati znanje, če tako hitro narašča?



Slika 7: Razporeditev pristopov k Semantičnemu spletu glede na zahteve

Iz zgornje slike (Slika 7) je razvidno, da lahko pri statističnem pristopu ostanejo podatki nedotaknjeni, uporabiti pa moramo pametne aplikacije, ki bodo znale analizirati podatke. Označevalni pristop zahteva nekaj več dela na podatkih, jezikoslovni pa več na aplikacijah. Semantični splet in umetna inteligenca imata kompromis glede zahtev po poseganju v aplikacije in podatke, s tem da je umetna inteligenca pametnejša v vseh pogledih. [3]

2.2.2 Dilema, kako se bo semantični splet uvajal

Tim Berners-Lee govori o semantičnem spletu že precej časa, ampak zadeva se bolj počasi odvija. Problem pri dodajanju pomena podatkom v spletu je, da bi se morali vsi držati enakih standardov, kar pa je težko izvedljivo, pa tudi malo verjetno.

Temeljni cilji SS so v osnovi:

- avtomatska gradnja/učenje ontologij iz poljubnih (spletnih) virov,
- avtomatsko označevanje dokumentov z metapodatki,
- razvoj storitev semantičnega spleta (B2B, spletni agenti in storitve).

Do teh ciljev pa na splošno vodita dve poti:

Klasični pristop (Bottom-Up) – Vsak prispeva en delček, nekdo poskrbi za zlaganje teh delčkov, in kot celota dobi stvar neko novo vrednost. To pomeni masivno, ročno vnašanje metapodatkov, z upoštevanjem razvitih standardov semantičnega spleta. *Zelo težavno.*

Spustni pristop (Top-Down) – Se uporablja danes. Generiranje RDF dokumentov je avtomatično, s pomočjo aplikacij, ki analizirajo podatke in njihov pomen pripenjajo podatkom. *Veliko bolj praktično.*

KLASIČNI PRISTOP

- vsak nekaj prispeva
- popolna semantika
- mikroformati



SPUSTNI PRISTOP

- na obstoječih podatkih
- LSI – matematični pristop k semantičnemu indeksiranju
- Stvarno, osredotočeno na uporabnika

Slika 8: Različna pristopa glede na način vnosa semantike

Prednost spustnega pristopa k semantičnemu spletu je ohranjanje obstoječe infrastrukture. Žal pa je pot do avtomatskega generiranja RDF precej zahtevnejša, zato lahko tudi dlje traja. [11]

W3C ima okvir RDF/OWL celovito zasnovan, vendar je težko razumljiv ljudem brez ozadja iz matematike in računalništva. Kot zagovarja Spustni pristop, potrebujemo mehanizem, ki samodejno obstoječo vsebino HTML pretvori v metapodatke RDF/OWL. Seveda, če bi imeli tak mehanizem, metapodatkov ne bi niti potrebovali, saj bi že sam mehanizem razumel vsebino. Zlata sredina med izgradnjo zapletenega mehanizma za razumevanje človeškega jezika in ročnim vnosom metapodatkov je torej orodje, ki bi opravilo vsaj večji del pretvorbe, česar ne bi zmoglo, pa bi interaktivno vnesli sami. [8]

2.3 Težave pri uvajanju in kritike

2.3.1 Zakaj semantični splet ne bo uspel

Kot vsaka nova tehnologija je tudi SS prejel mnoge kritike. Razlogov, zakaj opustiti projekt semantičnega spleta, se je kar vrstilo. Če povzamemo, je največ zgražanja požela kompleksnost SS. Kritiki so navajali, da so projekti SS ogromni, zahtevajo veliko denarja, časa in truda, preden bi sploh dospeli do uporabnikov. In to v letih, ko je splet poln revolucij, majhnih in poceni projektov (socialne vsebine, mobilni splet), bolj in manj uspešnih, ki potiskajo splet naprej. Nadaljevali so, da je za uporabo tehnologij SS potrebno najeti jezikoslovca, taksonomista, ontologa in programerja, ki se bo znašel med RDF okvirji ter izdelal aplikacijo. Sporno je bilo tudi vprašanje, kako bodo ti ljudje nadomestljivi.

Sodobni splet teži k zmanjševanju zapletenosti njegove uporabe, perfekcionizem, ki se je skušal doseči s SS, pa je to zapletenost samo še povečeval. Nekateri so obsojali, da je to maslo preveč pametnih inženirjev, ki povečujejo obseg SS, ker jim je bolj zanimivo postavljanje vprašanj, kot pa iskanje rešitev. Očitali so, da je SS izgubil stik s stvarnostjo. To sklepanje se danes izkazuje za napačno, že zaradi idej SS, ki jih uporabljamo pri vsakdanjem 'spletnem' delu. Je pa res, da semantične tehnologije omogočajo rešitve, ki segajo daleč preko današnje spletne prakse.

SS je odpiral tudi vprašanja neposredno poslovne narave. Med najbolj znanimi nasprotniki SS je bil Kanadčan Stephen Downes, ki pravi, da ne bomo nikoli dosegli, da bodo podjetja med sabo sodelovala, zlasti konkurenčna. Sodelovanje pa je pogoj, da se najde skupen slovar in da se SS zgodi. Kot odgovor lahko rečemo samo, da se SS že dogaja. Podjetja se prilagajajo standardom, ker hočejo ostati konkurenčna in ker v prihodnosti lahko računamo na, v največji meri, spletno prodajo. Tekma pa bo res huda. Nekonkurenčne cene je bolje skrivati, kot jih objavljati, če le niso upravičene s kakovostjo. [13]

2.3.2 Pomeni semantični splet manj dela za spletne razvijalce

Semantični splet ne poskuša nadomestiti ljudi s stroji, ampak uporabiti stroje, da delajo stvari, za katere ljudje izgubimo preveč časa: npr. filtriranje množice zadetkov, ki nam jo vrne poizvedba na Googlu, ukvarjanje s Spami, označevanje podatkov ... Na tak način bi nam omogočili, da lahko delamo, kar znamo najbolje, kar je naš namen, in smo s tem bolj produktivni. SS lahko pomeni prihodnjo evolucijo spletne izkušnje, ki lahko korenito spremeni predvsem področji spletne prodaje in oglaševanja. Način dela spletnih razvijalcev se za zdaj ne bo bistveno spremenil, oziroma se bo spreminjal konstantno kot doslej. V spletu se stalno razvijajo nove tehnologije in vsaka prinaša nove rešitve, pa tudi nove izzive.

2.3.3 Lažni metapodatki

Dejstvo je, da so metapodatki, ki jih ljudje pripenjamo na splet, odvisni od naše usposobljenosti, motivacije in cilja. Potem pa pomislimo na semantični splet, poln Spama. Takorekoč, nikoli ne moremo zaupati spletnim mestom, da zagotavljajo poštene metapodatke.

Toda tako kot SS omogoča preprosto vnašanje Spama, tako s pomočjo povezanih podatkov omogoča tudi enostavno zaznavanje le-tega. Stran, ki bo ponujala goljufive vsebine skupaj s pravimi, bo označena kot goljufiva in uporabniku ne bo ponujena med rezultati.

2.3.4 Problem integracije različnih ontologij

OWL jezik ter orodja za ontologije so poenostavili projektiranje ontologij. Nastal pa je problem, kako povezati različne ontologije. Izkušnje kažejo, da je pogosto lažje ustvariti ontologije po lastni meri za različne koncepte, kot pa poskušati integrirati obstoječe ontologije teh konceptov.

Vsaka ontologija ima svoja poimenovanja, filozofske usmeritve, poglede na domeno, ki jih vpeljejo ljudje iz lastnih potreb, ko ustvarjajo. Povezovanje teh različnih svetovnih nazorov je težek problem. Preprosto navajanje, da so si razredi ali lastnosti enakovredni, ni rešitev, saj njihovo dedovanje ni nujno enakovredno in se lahko dejansko po funkciji pomensko precej razlikujejo. OWL bo moral veliko bolj izrazito opredeliti preslikavo med ontologijami za rešitev teh težav.

Cilj je ustvariti ontologije, ki so elegantne, enostavne za vzdrževanje, razširitev, razumevanje in uporabo. Uvoz ontologij v druge ontologije se ne zdi način za doseganje tega. Različne ontologije se običajno ne ujemajo dobro, v nekaterih primerih pa sploh ne. Zato je včasih lažje ontologijo napisati na novo kot predelati in povezati obstoječe. Najti torej preprosto in učinkovito pot za povezavo ontologij ali prepričati vse, naj uporabljajo isto ontologijo? Po drugi strani si lahko mislimo, da bolj kot bo celovita in učinkovita ta ontologija, manj bo ljudi, ki se bodo strinjali z njo, kaj šele razumeli dovolj dobro za uporabo zaradi njene zapletenosti. Število ontologij v spletu vse bolj raste in odgovor na to vprašanje je treba poiskati čim prej.

Semantični splet je v vzponu in se še naprej razvija, tudi če semantično povezovanje ne bo tako enostavno izvedljivo. Originalna vizija semantičnega spleta je decentraliziran sistem, le da je manjkajoči člen dobro orodje, metodologija za povezovanje ontologij. Take metodologije še ne poznamo in Nova Spivack, spletni futurist, kot se je poimenoval, je predvideval, da je bolj verjetno, da bo na koncu prevladalo nekaj velikih ontologij in ponudnikov storitev, s katerimi se bodo ostali povezali. Kot bomo videli v nadaljnjem opisu razvoja tehnologij, se to dejansko tudi dogaja. [10]

3 TEHNOLOGIJE SEMANTIČNEGA SPLETA

V svetu ima najvidnejšo vlogo, pri pospeševanju uporabe in razumevanja semantičnega spleta, gibanje v okviru konzorcija W3C. Ta je leta 2004 izdal model za opisovanje virov RDF in jezik za spletne ontologije OWL kot W3C standarda. V prvih dneh semantičnega spleta se je predvidevalo, da bodo vse vsebine predelane v RDF in bomo kmalu imeli na voljo množico semantično opisanih podatkov. Prehajanje pa se je počasi izvajalo, morda zato, ker je glavna XML prezentacija RDF-ja izgledala preveč zapleteno. Z namenom poenostavitve semantičnega označevanja so bili predstavljeni bolj strnjeni in bralni RDF zapisi, kot sta Notation3 (N3) in njegova podskupina Terse RDF Triple Language (Turtle), vendar so se izkazali kot še vedno prezahtevni. Neke vrste rešitev k enostavnejšemu ustvarjanju metapodatkov so prinesli **Mikroformati**, imenovani tudi 'semantični splet z malo začetnico', ki pa niso imeli veliko skupnega z veliko vizijo o SS. Kot prikazuje Tabela 1, je prišlo do dveh ločenih pristopov k semantičnemu spletu, oziroma razmejitev na dolgo pričakovani »Semantični splet« iz akademskih krogov in »semantični splet«, katerega pobudniki so razvijalci spletnih strani, ki jim je bilo dovolj čakanja na uresničitev izvirne ideje o SS.

Semantični splet (The Semantic Web)	Zbirka tehnologij, ki izvira iz W3C: RDF, OWL, N3, Turtle, RDFa, GRDDL, SPARQL ...
semantični splet (lower-case semantic web)	Realna semantika. Označbe in tehnologije, sprejete s strani realnih ljudi – izven akademskih in teoretičnih razprav. <ul style="list-style-type: none"> • Preprosta semantika z mikroformati <i>Ni potrebno definirati sveta. Enostavno združevanje delčkov.</i> • Evolucijsko, ne revolucionarno <i>Semantika v današnjemu spletu, ne ustvarjanje spleta prihodnosti.</i> • Razvoj, usmerjen za uporabnika <i>Ljudje prvi, stroji drugi. Ljudje pomagamo ustvarjati metapodatke</i>

Tabela 1: Semantični splet z veliko in malo začetnico

3.1 Konzorcij svetovnega spleta W3C in njegov pristop

W3C je mednarodna skupnost, kjer organizacije članice, zaposleni v W3C-ju in javnost sodelujejo pri razvoju spletnih standardov. Z izumiteljem spleta Berners-Lee-jem in generalnim direktorjem Jeffreyem Jaffe-jem je poslanstvo W3C-ja vodenje spleta do njegove polne zmogljivosti. Poslanstvo uresničujejo z razvojem protokolov in smernic, ki zagotavljajo dolgoročno rast spleta. Nekaj pomembnih načel tega poslanstva prikazuje Tabela 2.

Splet za vse	Družbeni pomen spleta je, da omogoča človeške komunikacije, trgovino in priložnosti za izmenjavo znanja. Eden od osnovnih ciljev W3C je, da bi bile te ugodnosti na voljo vsem ljudem, ne glede na strojno opremo, programsko opremo, omrežne infrastrukture, materni jezik, kulturo, geografsko lokacijo ali fizične/psihične sposobnosti.
Splet na vsem	Število različnih vrst naprav, ki lahko dostopajo do spleta, je izjemno naraslo. Mobilni telefoni, pametni telefoni, dlančniki, interaktivni televizijski sistemi, odzivniki in celo nekateri gospodinjski aparati so lahko povezani v splet.
Vizija	Vizija W3C vključuje sodelovanje, izmenjavo znanja in krepitev zaupanja na globalni ravni – torej splet z bogato interakcijo. Splet je bil izumljen kot komunikacijsko orodje, namenjeno komurkoli in kjerkoli za izmenjavo informacij. Vrsto let je bil splet za mnoge samo orodje za branje. Blogi in Wikiji so pritegnili več avtorjev v splet in socialne mreže so postale cvetoči trg osebnih vsebin in spletnih izkušenj. W3C standardi so podprli ta razvoj zaradi močne arhitekture in oblikovnih načel.
Splet podatkovnih storitev	Nekateri ljudje vidijo splet kot veliko zbirko povezanih podatkov, drugi pa kot velik nabor storitev za izmenjavo sporočil. Oba pogleda se dopolnjujeta, katerega pa se uporabi, je odvisno od aplikacije.
Splet zaupanja	Splet je spremenil način komuniciranja med seboj. S tem je spremenil tudi značaj naših družbenih odnosov. Ljudje sestankujejo na spletu, opravljajo poslovne in osebne odnose, ne da bi se kdaj srečali v živo. W3C priznava, da je zaupanje družbeni pojav, vendar tehnološka načrtovanja lahko ojačajo zaupanje. Več kot se dejavnosti dogaja 'on-line', pomembnejša postaja podpora kompleksnih interakcij med skupinami po svetu.

Tabela 2: Glavna načela, ki vodijo delo v W3C

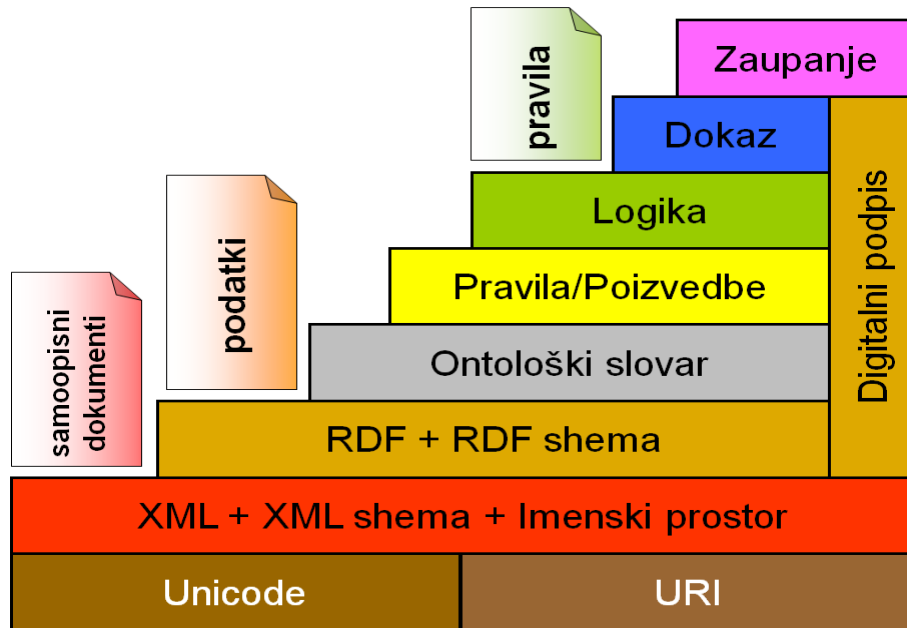
Poleg klasičnih 'spletnih dokumentov' W3C pomaga graditi sklad tehnologij za podporo 'spletnim podatkom' – podatkom, kot se jih najde v zbirkah podatkov. Končni cilj spletnih podatkov je omogočanje računalnikom, da naredijo več koristnega dela z njimi, in razvijanje sistemov, ki podpirajo zaupanja vredne interakcije prek omrežja. Tehnologije semantičnega spleta omogočajo ljudem, da ustvarjajo skladišča podatkov na spletu, gradijo slovarje in pišejo pravila za obvladovanje podatkov. Povezani podatki so podprti s tehnologijami, kot so RDF, SPARQL, OWL in SKOS.

3.2 Semantični splet kot skupek standardov

Semantične tehnologije lahko obravnavamo v smislu plasti. Vsaka plast se naslanja na razširitev funkcionalnosti plasti pod njo. To nam da vedeti, da **semantični splet ni ločen splet, ampak nadgradnja obstoječega**. Je razširitev in okrepitev obstoječih spletnih strani in ne njihova zamenjava. Nenazadnje tu ne gre za spletne strani in povezave, temveč za **odnose med podatki**. SS je skupek tehnoloških standardov (Unicode, URI, XML, RDF, RDFS, OWL ...), ki omogoča strojem razumeti semantične dokumente in podatke. Z uporabo

standardov splet postane baza podatkov. Ne ustvarjamo posamezne baze podatkov, ampak **spreminjamo splet v bazo podatkov**. [12]

Zelo znana je predstavitev nivojev povezanih tehnologij, poimenovana 'Semantic Web Layer Cake' ali 'W3C Grand Vision' (Slika 9), ki prikazuje razvoj tehnologij na poti do semantičnega spleta. O nekaterih tehnologijah ni dvoma, da so koristne za današnji splet, o drugih pa še vedno potekajo dogovori.



Slika 9: Plasti povezanih tehnologij semantičnega spleta

Unicode – Najprej se nam je uspelo dogovoriti o predstavitvi znakov. Povsem jasno je že, da potrebujemo črke.

URI (Uniform Resource Identifier) – Enotni označevalnik vira, standard za identifikacijo in lociranje virov. Naslednji korak je bil dogovor o naslavljanju obstoječih stvari, identifikaciji objekta na spletu, v računalniku (predvsem govorimo o digitalnih vsebinah).

XML (eXtended Markup Language) – Po daljšem času smo sprejeli dogovor o zapisu podatkov tako, da jih lahko tudi drugi razberejo. Rodil se je XML, ki je tudi že dobro uveljavljen na spletu. Potekala je linija dogovorov, kako se omejiti pri zapisu podatkov, da lahko z njimi kaj naredimo, jih poiščemo, iz njih sklepamo in izvajamo operacije nad njimi, jih strukturiramo. Podobno kot pri zapisu strukture izročka, se približno ve, kako ga je treba narediti. Sledile so enostavne sheme: XML-ji in DTD-ji, ki ne dovoljujejo več uporabe poljubnih značk in poljubnega vrstnega reda. Te sheme standardizirajo označevanje, da bomo znali uporabljati te podatke.

RDF – Ogrodje za opis virov je dogovor o predstavitvi metapodatkov, ki označevanje še nekoliko bolj omeji, hkrati pa definira model za opis relacij med spletnimi viri. Poleg XML formata se uporablja še druge sintaktične predstavitve ogrodja RDF.

Primer je graf s točkami in povezavami – še vedno svoboden zapis, ki ni vsem razumljiv, vendar računalnikarji in matematiki znajo iz njega razbrati, kaj predstavlja.

ONTOLOGIJA – Filozofska disciplina o obstoju, stvarnosti. Zakaj nekaj je, obstaja? Z informacijskega pogleda je to skupen slovar ali podatkovni model, ki opisuje objekte v domeni in razmerja med njimi. V osnovi je kot XML shema – dokument, ki opisuje strukturo in pravila uporabe podatkov ter da podatkom smisel. Primer ontologije so podatkovni modeli iz relacijske baze. Standardi/jeziki na tem področju se še vedno sprejemajo in večina (XML, RDF, OWL) se jih ukvarja z neko stopnjo ontološke predstavitve znanja.

Pravila – Pravila, akcije, ki se dogajajo nad dogodki. RIFF (Resource Interchange File Format) – različni formati, vpadljivi eden na drugega.

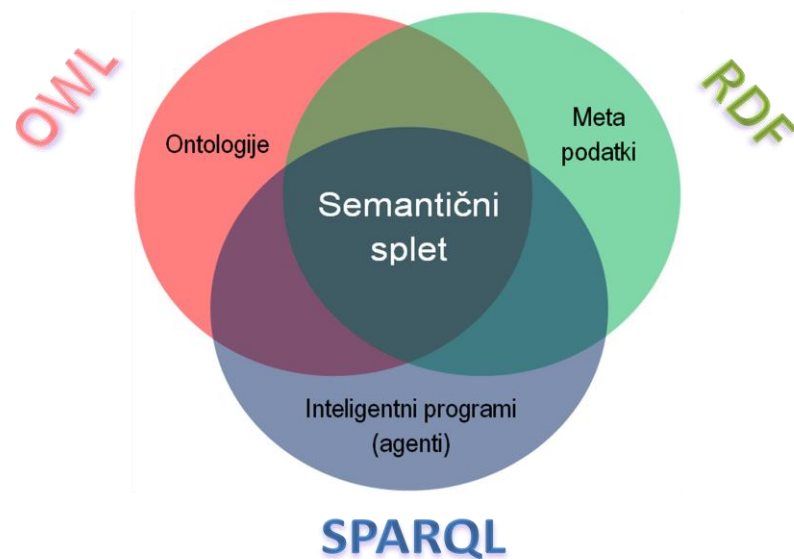
Poizvedbe – Svet se je dogovoril, kako povpraševati po podatkih.

Logika, dokazovanje – Tvorita avtomatičen sistem sklepanja. Ta spletnim agentom omogoča odločitve o tem, ali določen vir zadošča podanim zahtevam.

Zaupanje – Nista dovolj kodiranje in digitalni podpis, ampak so še druge nianse tega zaupanja. Npr.: Ali zaupamo, da je informacija na Wikipediji resnična? Semantični splet lahko podpira zapis zaupanja.

3.2.1 Elementi semantičnega spleta

Elemente semantičnega spleta predstavljajo ontologije, metapodatki in inteligentni programi. Za prehod iz nestrukturiranega spleta v splet, bogat semantične vsebine, je W3C konzorcij predstavil označevalni jezik RDF in ontologijo OWL kot poskus za celoten zajem in opis informacij. Kot poizvedovalni jezik za tako opisane informacije pa je razvil SPARQL. [12]



Slika 10: Elementi semantičnega spleta

3.2.2 Ontologije in OWL

V informatiki so ontologije formalna predstavitev množice konceptov z nekega področja in razmerij med njimi. Takorekoč ontologije združujejo zbirko objektov in pravil, ki so sestavljene tako, da se lahko združujejo. Pomembno je vedeti, da objekti niso zgolj besede, temveč koncepti. Uporabljamo jih za predstavitev človeškega znanja in sklepanje. Ontologije lahko sami ustvarjamo in objavljamo v spletu.

Ontologija opredeljuje način opisovanja stvari. Za sporazumevanje in nedvoumno izražanje je zelo pomembno, da uporabljamo skupno ontologijo, na primer stavek »Jan je avtor knjige.« Kaj če bi namesto »avtor« uporabili »pisatelj«? Pomen ostaja več ali manj isti, za nas, kaj pa za računalnik? Ob taki spremembi se ne znajde več, razen če ima dostop do ontologije, ki bi mu razjasnila analogijo besed.

Sodobne ontologije si delijo mnogo strukturnih podobnosti, ne glede na jezik, v katerem so izražene. Te **skupne komponente ontologij** so:

Posamezniki: primeri ali predmeti (objekti).

Razredi: sklopi, zbirke, koncepti, entitete, vrste predmetov, elementi.

Lastnosti: atributi, vidiki, značilnosti, parametri predmetov.

Odnosi: zveze, razmerja, načini povezav med razredi in posamezniki.

Pogojne funkcije: kompleksne strukture, izpeljane iz odnosov.

Omejitve: formalni opisi pogojev za sprejem izjav ob vnosu.

Pravila: izjave v obliki pogojnih stavkov za opis logičnega sklepanja.

Aksiomi: trditve v logični obliki, ki skupaj tvorijo teorijo, ki jo ontologija opisuje v svoji domeni.

Dogodki: ob spreminjanju lastnosti ali odnosov.

Za pisanje ontologij običajno koristimo posebna orodja. Zapisujemo jih z uporabo jezikov za opis ontologij. Podmnožica teh jezikov so označevalni jeziki, ki kot označevalni sistem za kodiranje znanja najpogosteje uporabljajo XML. Mednje štejemo: DAML + OIL, OIL, OWL, RDF, RDF Schema in SHOE.

Po strukturi ločimo ontologije, ki temeljijo na podlagi: **okvirjev** ([FLogic](#), [OKBC](#), [KM](#)), **opisovalne logike** ([OWL](#), [RACER](#), [KL-ONE](#)) in **logike prvega reda** ([CycL](#), [KIF](#)). Vsebinsko pa ji delimo na domenske ontologije in vrhnje ontologije. **Domenska ontologija** opisuje specifično področje. Predstavlja pomen izrazov, ki se uporabljajo v tem področju. **Vrhnja ontologija** (ali temeljna) je model skupnih objektov, ki so splošno uporabljeni v širokem razponu domenskih ontologij. Primeri standardiziranih vrhnjih ontologij so: [Dublin Core](#), [GFO](#), [OpenCyc/ResearchCyc](#), [SUMO](#) in [DOLCE](#).

OWL je W3C-jev standard za izražanje ontologij. Omogoča semantični opis poljubne povedi. Ontologije OWL uporabljajo v večini RDF in sintakso XML. Obstajajo trije nivoji OWL: OWL Lite, OWL DL in OWL Full. Vsak nivo zavzema svoje mesto v kompromisu med izrazitostjo in izračunljivostjo.

Razred IZDELEK vsebuje lastnosti	Razred PODJETJE vsebuje lastnosti
<ul style="list-style-type: none"> - ime - simbol kakovosti - izdelovalca (tip: Podjetje) 	<ul style="list-style-type: none"> - naziv - naslov - telefon

Slika 11: Opis domene, izbira ontologije

3.2.3 Metapodatki in RDF

Metapodatki so podatki, ki opisujejo druge podatke, so reference do pojmov, ki so definirani v ontologijah ter tako ustvarjajo globalni pomen metapodatkov. Eden od načinov za zapis metapodatkov je XML. Zaradi možne dvoumnosti zapisa se XML ne izkaže kot najboljši, zato je bil v te namene posebej razvit označevalni jezik RDF.

```

<avtor>
  <uri>http://www.vipava1894.si/vina/c21</uri>
  <ime>Špela</ime>
</avtor>

<oseba ime="Špela">
  <delo> http://www.vipava1894.si/vina/c21</delo>
</oseba>

<dokument href="http://www.vipava1894.si/vina/c21" avtor="Špela" />

```

Tabela 3: Dvoumen zapis z XML

Razmerje, izraženo v vseh primerih, prikazanih v Tabela 3, je: Špela je avtor strani »http://www.vipava1894.si/vina/c21«. Dokaj preprosta izjava, vendar izražena na več načinov v XML. Zelo težko bi bilo zgraditi programsko opremo, ki lahko zaključi vse možne načine zapisa razmerja in izrazi njegovo bistvo. Ampak RDF izraža to razmerje le na en način, in tako omogoča generično gradnjo pri opisovanju stvari (Tabela 4).

```
<stran> <avtor> <Špela>
```

Tabela 4: Nedvoumen zapis povezave z RDF

Osnovna celica zapisa RDF je trojček (ang. triple) RDF. Trojček RDF ima tri dele, ki si po dogovoru sledijo v vrstnem redu: osebek, predikat in predmet (Slika 12). Osebek označuje vir, predikat označuje lastnosti ali vidike virov in izraža odnos med osebkom in predmetom. S takim zapisom lahko tvorimo entitete s številnimi lastnostmi. Vsak del trojčka lahko izrazimo kot vir na spletu, ki je povezava URI. Če URI spremenimo, postanejo RDF-ji, ki so se nanj nanašali, neuporabni. Zelo pomembna je torej smiselna, jedrnata in dosledna struktura, o

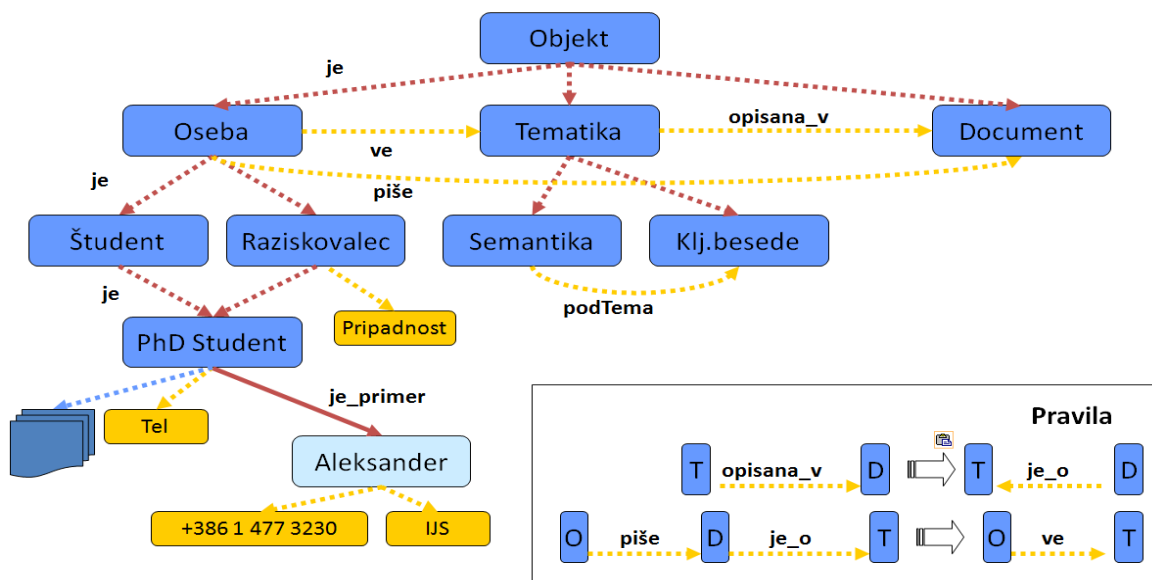
kateri moramo razmišljati vnaprej. Spreminjanje po objavi prinese kvečjemu negativne točke s strani obiskovalcev.



Slika 12: Trojček RDF

Za sam zapis trojčka pa lahko uporabimo RDF/XML zapis (Tabela 13). Uporabimo lahko tudi ne-XML izpeljanko, kot je N3 (Notation 3) (Tabela 14), ali njegovi podmnžici Turtle (Terse RDF Triple Language) in Ntriple. Lahko pa ga prikažemo z grafom (Slika 21). Graf RDF je množica trojčkov RDF. Množica vozlišč v RDF grafu je množica osebkov in predmetov, povezave med njimi so predikati.

RDF je torej jezik za opis informacij in virov na spletu. Slika 13 prikazuje razširjen graf RDF, ki prinaša bogatejši spletni pomen: objekti in razmerja imajo pomen. Aplikacije lahko iščejo, poizvedujejo skozi vse te podatke spleta enako, kot prehajamo iz ene spletne strani na drugo, le da se to dogaja na drugem nivoju. In ker se podatki sami opisujejo, nosijo smisel, lahko uporabljamo zelo preproste aplikacije, ki delajo koristne stvari. Inteligenca je namreč v podatkih.



Slika 13: Sklepanje iz podatkov

Na tako zapisanih podatkih lahko logično sklepamo, oblikujemo pravila, ki jih bodo znali Inteligentni agenti prebrati, razumeti in tudi primerjati z drugimi podatki na spletu.

3.2.4 Inteligentni agenti, SPARQL in Linked Data

Številni iskalniki preiskujejo splet in uporabniku posredujejo podatke in dokumente, da bi zadovoljili njegove informacijske potrebe. Rezultati pa velikokrat niso v skladu z uporabnikovimi zahtevami. Da bi uporabnik vedno dobil tisto, kar želi, so bili za preiskovanje spleta in iskanje pravih, omejenih in učinkovitih informacij razviti Inteligentni agenti (Intelligent Web searching Agents).

Inteligentni agent je zaprt računalniški sistem, sestavljen iz arhitekture in programa. Ta računalniški sistem se nahaja v nekem okolju, ki ga razume in je v njem sposoben delovati samostojno. V semantičnem spletu je njegova naloga analizirati metapodatke in uporabiti ontologije, ki so dostopne na spletu, za razumevanje teh metapodatkov. Po zaslugi povezanih podatkov (Linked Data) so danes agenti sposobni slediti iz enega vozlišča do drugega po ogromnem globalnem grafu.

V semantičnem spletu lahko Inteligentne agente uporabimo za več stvari, na primer kot:

- spletni-indeksirni agent, ki spletnim dokumentom dodaja semantiko;
- osebni agent, ki uporablja bazo znanja SS za rezervacijo počitnic, sestankov ...
- večuporabniški agent, ki je sposoben gradnje in vzdrževanja dodatnih zbirk povezanih podatkov.

Agenti pri komunikaciji z omrežjem koristijo spletne povezovalne tehnologije (Web Services). Spletne povezovalne tehnologije so standardizirani načini integracije spletnih aplikacij z uporabo standardiziranih tehnologij, kot so XML, SOAP, WSDL in UDDI z namenom doseči povezljivost aplikacij. Npr. XML pri tem skrbi za označevanje, SOAP za prenos podatkov, WSDL opisuje storitev, ki je na voljo, UDDI pa skrbi za seznam storitev, ki so na voljo.

Ko imamo enkrat vire podatkov v RDF in imamo ontologije, ki določajo razmerja med njimi, potrebujemo še način za iskanje podatkov po teh virih. **SPARQL** je poizvedbeni jezik za RDF. Za izmenjavo rezultatov teh poizvedb med aplikacijami, storitvami in drugimi agenti pa SPARQL ponuja še svoj protokol. Z uporabo WSDL 2.0 ta protokol opisuje enostavno spletno povezovalno tehnologijo z eno operacijo: poizvedba.

SPARQL sintaksa za izražanje poizvedb v RDF-ju je podobna SQL-u, pa tudi poizvedbe same so videti in delujejo kot podatki RDF. Poizvedovanje temelji na ujemanju vzorcev, je prilagodljivo, namenjeno za delo v spletu na podatkih, sestavljenih iz različnih virov. Na primer: ujemanje je mogoče izbrati kot opcijo; ta lastnost ga naredi boljšega kot SQL pri poizvedovanju po nestrukturiranih podatkih. Takšni tudi so podatki, združeni iz različnih virov, nepredvidljivi in nestrukturirani, in niso urejeni kot pri običajni zbirki podatkov SQL.

Viri povezanih podatkov običajno ponujajo končne točke SPARQL za dostop do njihove baze podatkov. Končna točka SPARQL (SPARQL endpoint) je protokolarna spletna storitev, ki uporabnikom (ljude, strojem) omogoča poizvedovanje po bazi znanja z jezikom SPARQL (Slika 14). Rezultati poizvedbe se po navadi vrnejo v enem ali več strojno bralnih formatih

(XML, JSON, navaden tekst; RDF/XML, Ntriple, Turtle, N3), zato se končne točke SPARQL smatra kot stroju prijazen vmesnik do baze znanja. Tako oblikovanje poizvedb kot prikaz rezultatov sta mišljena kot delo za programsko opremo, in ne za ročno vnašanje s strani uporabnikov. Danes obsegajo največje končne točke SPARQL tudi po več sto milijonov trojčkov. Nekaj teh delujočih točk prikazuje Tabela 5.

SPARQL Explorer for <http://dbpedia.org/sparql>

SPARQL:

```

PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX : <http://dbpedia.org/resource/>
PREFIX dbpedia2: <http://dbpedia.org/property/>
PREFIX dbpedia: <http://dbpedia.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dbo: <http://dbpedia.org/ontology/>

SELECT ?name ?birth ?person WHERE {
  ?person dbo:birthDate ?birth .
  ?person foaf:name ?name .
  FILTER (?birth >= "2011-01-01"^^xsd:date) .
}

```

Results:

SPARQL results:

name	birth	person
"Nicolaas Petrus van Wyk Louw"@en	"2011-06-06"^^xsd:date	:Nicolaas_Petrus_van_Wyk_Louw ↗
"Nimā Yushij"@en	"2012-11-12"^^xsd:date	:Nima_Yooshij ↗

Slika 14: SNORQL – vmesnik za poizvedovanje SPARQL

Končne točke SPARQL	stanje	Spletni obrazec
BBC Programmes and Music	2010-06-29	Ajax based Visual Query Builder
BBC playcount data iz DBTune	2010-01-07	n/a
data.gov	2010-05-22	spletni obrazec
DBLP Bibliography Database (Freie Universität Berlin)	2010-01-07	spletni obrazec (Firefox)
DBpedia (Freie Universität Berlin, Universität Leipzig, OpenLink Software)	2010-01-07	SNORQL (Firefox/Safari/Opera); Ajax based Visual Query Builder
DBpedia-live (Freie Universität Berlin, Universität Leipzig, OpenLink Software)	2010-01-07	spletni obrazec

Tabela 5: Delujoče končne točke SPARQL

Poizvedbe SPARQL h končnim točkam SPARQL so v osnovi poizvedbe 'HTTP GET' s parametrom *query* (GET /sparql?query=...). Najbolj prikladen način za klicanje poizvedb pa

je uporaba knjižnic za programske jezike, kot so: SPARQL JavaScript Library, ARC for PHP, RAP (PHP), Jena/ARQ (Java), SPARQL Wrapper (Python) ...

Glavnina semantičnega spleta se nahaja v skladiščih trojčkov, ki so dostopni le prek končnih točk SPARQL. Povezava teh skladišč, med seboj in z drugimi viri, je določena z načinom objave in medsebojnega povezovanja strukturiranih podatkov na spletu, ki ga je Tim Berners-Lee dopolnil z arhitekturo **Linked Data**. Osnovni načeli te arhitekture sta:

- uporaba modela RDF za objavo strukturiranih podatkov v spletu in
- uporaba povezav RDF za medsebojno povezavo podatkov iz različnih virov.

Ob upoštevanju obeh načel postane splet podatkov dostopen široki paleti obstoječih brskalnikov RDF (npr. [Disco](#), [Tabulator](#), [OpenLink Browser](#)), pajkov RDF (npr. [SWSE](#), [Swoogle](#)), in poizvedovalnih agentov (npr. [SemWeb Client Library](#), [SWIC](#)).

3.3 Mikroformati (uF)

Zavedajoč se kompleksnosti RDF in OWL se skupina ljudi trudi z drugačnim pristopom, imenovanim Mikroformati (okrajšano μF ali uF). Mikroformati vključujejo semantiko neposredno na obstoječe strani (X)HTML, z dodajanjem enostavnih označb na berljive elemente za človeka, kot so dogodki, kontakti, lokacije ... Njihov cilj je s takimi označbami napraviti te elemente berljive programski opremi ter enostavne za: indeksiranje, iskanje, shranjevanje, navzkrižno sklicevanje in združevanje. Povzeto jih lahko opišemo tudi takole:

Mikroformati SO

- način razmišljanja o podatkih (začeti kar se da preprosto).
- načela načrtovanja za formate.
- namenjeni prvotno ljudem in šele nato računalnikom.
- prilagojeni trenutnim pojavom in primerom uporabe (npr. socialnim omrežjem).
- povezani z vsebinami, kot so: semantični XHTML, semantika realnega sveta, semantični splet z malo začetnico, XHTML brez izgub ...
- zbirka preprostih odprtih formatov, ki jih mnogi aktivno razvijajo za bolj strukturirano bloganje in spletno objavljjanje mikrovsebin na splošno.

Mikroformati NISO

- nov jezik.
- razširljivi in odprti (kot RDFa).
- poskus, ki bi zahteval spremembo obnašanja in nova orodja.
- popolnoma nov pristop, ki bi delujočo tehniko postavil v zgodovino.
- rešitev za vse taksonomije, ontologije in druge podobne abstrakcije.
- kompleksni ali namenjeni opisu celotnega sveta.

Tabela 6: Kaj mikroformati so in kaj niso

Načela uporabe mikroformatov so modularnost, spodbujanje decentraliziranega razvoja vsebin/storitev, uporaba že sprejetih standardov in prosta uporaba. Tehnično je semantična vrednost dodana obstoječim (X)HTML vsebinam s pomočjo množice skupnih elementov, navedenih kot vrednosti atributov *class* in *rel/rev*.

Mikroformat	Za predstavitev
hCard	ljudi, podjetij in organizacij
geo	geografskih koordinat
xFolk ali hAtom	zbirk zaznamkov, blogov
rel-tag	označb s povezavami
XFN	razmerij med ljudmi s hiperpovezavami
XOXO	opisa strukture podatkov
hResume	povzetka življenjepisa CV
hReview	ocen izdelkov, storitev, podjetij
VoteLinks	glasovanj
rel-payment	podpore strani
hCalendar	dogodkov

Tabela 7: Najbolj uporabljeni mikroformati

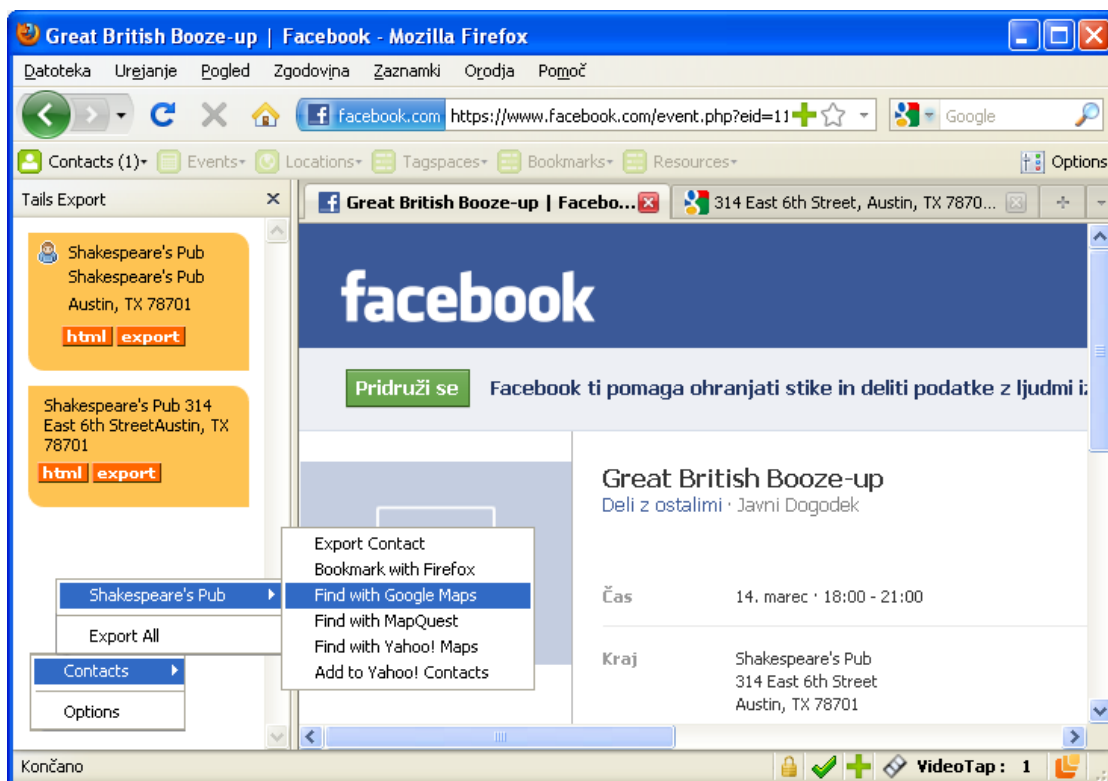
Mikroformati obstajajo za ozko skupino podatkov, kot so informacije o kontaktih, koledarji, opisi (izdelkov, storitev, dogodkov). Nekaj jih temelji na obstoječih podatkovnih formatih, kot sta *vcard* in *ical*. Vzemimo kot primer mikroformat za opisovanje kontaktnih podatkov, imenovan hCard (Tabela 11). Z uporabo hCard je mogoče opremiti HTML, tako da lahko brskalnik ali iskalnik, ki zazna mikroformate, razbere podatke o osebi, kot so ime in priimek, podjetje ali telefonska številka. Drugi zelo uporabljen mikroformat, imenovan hCalendar, omogoča avtorjem strani opis dogodkov. Tabela 7 navaja nekaj najbolj uporabljenih mikroformatov.

3.3.1 Kako okolica sprejema mikroformate

Nekateri ljudje menijo, da je pristop z uF zgrešil bistvo Semantičnega spleta, kar naj bi bil enotni model metapodatkov, ki omogoča shemam razširitve, dostopnost in deklarativno semantiko. Vsak na novo predstavljen uF ima namreč določeno opisno kodo, ki ne ponuja nobenih razširitev, kar za uvedbo novih lastnosti pomeni predstavitev zopet novih uF. Z uF ne bomo nikoli mogli 'opisati sveta', kot se tega oklepa vizija SS. Obstoječi sklop uF opisuje malo stvari, ki jih srečujemo na spletu, in še pri teh redkih stvareh moramo pogosto dejanski opis prilagajati tistemu, kar uF ponujajo, namesto da izrazimo tisto, kar želimo. Poleg tega pa uF prinašajo še vse težave klasičnega pristopa k SS: strani moramo opremljati ljudje. Dobra novica je le, da si zaradi preprostejšega zapisa lahko več pomagamo z obratnim inženirstvom in avtomatizacijo.

Z uF se ne da izražati vsega, kar zmoreta RDF in OWL, nedvomno pa je pristop z njimi precej preprostejši. Striktnost in enostavnost njihove uporabe je odpravila možnost dvoumnega

zapisa in ponudila natančnejše rezultate iskanja. Tako so uF postali zelo priljubljeni. Google je začel z razpoznavanjem hCard, hReview in hProduct uF, predstavil je iskanje receptov, podprto z uF. Facebook je dodal mikroformata hCard in hCalendar milijonom dogodkov.



Slika 15: Dodatka za Firefox: Operator in Tails

Vse več je dodatkov za spletne brskalnike, ki omogočajo pregled in iskanje po podatkih, označenih z uF. Prav tako lahko označene podatke izvažamo v formate, primerne za uvoz v druge aplikacije; npr. izvoz kontaktov/dogodkov s spletne strani v Microsoft Outlookov imenik/koledar. Z dodatkom za brskalnik Firefox, Operator (v menijski vrstici) in Tails (stranski stolpec), lahko izvažamo z mikroformati označene podatke ali pa dalje iščemo po njihovi vsebini (Slika 15).

Koriščenje uF	Vgrajevanje uF	Orodja za uF
Yahoo SearchMonkey, Google Rich Snippets, Technorati, Google Buzz, Uporabniki z uF orodji.	Twitter, Facebook, Digg, Flickr, Yahoo, Google, Technorati, MovableType, Wordpress.	Operator, Oomph 2, Tails, H2VX, Optimus, CMS plugins (MovableType, WordPress), Orodja za kreiranje uF (dodatki za Dreamweaver).

Tabela 8: Kdo koristi/vgrajuje uF in kakšna orodja uporablja

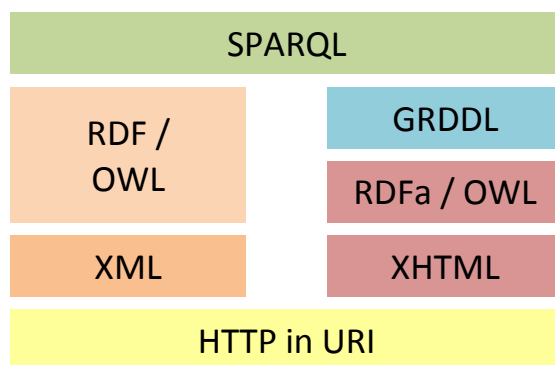
Pojavljajo se tudi prvi primeri iskalnikov, ki temeljijo na uF, na primer Technorati, ki v realnem času preiskuje uporabniško generirane strani po oznakah in ključnih besedah ter

hkrati zagotavlja indeksiranje rezultatov po priljubljenosti. Podobna orodja, ki vključujejo uporabo ali generiranje uF, nam našteva Tabela 8.

3.4 RDFa

Neposredno vnašanje semantike na obstoječe (X)HTML strani je bil izziv tudi za W3C. Razvili so svojo izpeljanko označevanja podatkov, ki so jo poimenovali RDFa (predhodnik je bil eRDF, ki je zastarel). Implementacija je nekoliko bolj zapletena kot pri Mikroformatih, vendar veliko bolj generična. Dejansko gre za RDF opis, izražen znotraj XHTML. Čeprav je RDFa podmnožica RDF, je prav tako dobro zasnovan in omogoča uporabo lastnih slovarjev ali razširitev obstoječih. Vsebina strani je berljiva tako za človeka kot tudi za stroj. RDFa namreč niso več trojčki, zapisani v ločenih datotekah, ampak so opisi podatkov, tesno povezani s podatki samimi znotraj XHTML.

RDF, vdolan v dokumente XHTML kot RDFa, ni več domač vsem semantičnim orodjem, ki sicer obvladajo RDF. Zato je bilo potrebno najti način, ki bo predhodno razpoznaval prisotnost vsebin RDFa in jih pretvoril v RDF. Rešitev W3C-ja za ta manjkajoči člen je GRDDL (izgovorjeno »gridl«). Osnovna ideja GRDDL je transformacija obstoječega dokumenta XHTML v RDF, skozi XSLT, ki je določen s povezavo v glavi dokumenta. [RDFa Parser](#) je primer spletne aplikacije, ki iz RDFa zapisa izlušči RDF/XML zapis.



Slika 16: Sklad tehnologij Semantičnega spleta

Kot je prikazano na zgornji sliki (Slika 16), je osnovna plast Semantičnega spleta HTTP in URI, ki sicer spada še pod prejšnji splet, vendar je temelj semantičnemu spletu. Na levi strani imamo tehnologiji XML in RDF/OWL, ki predstavljata prvo zasnovo SS. Na desni je RDFa omogočil uporabo RDF znotraj XHTML, hkrati pa zahteval uveljavo transformacije GRDDL. Obe smeri nam na koncu ponudita podatke, pripravljene za poizvedovalno plast SPARQL. [5]

4 UPORABA SEMANTIČNEGA SPLETA V PODJETJU ZA OPIS VIN

4.1 O podjetju

Vipava 1894 d.d.³ je družba, ki je heterogena po svoji dejavnosti, saj se ukvarja s predelavo vina in mleka. Sestavljata jo dve strateški enoti, vinska klet in mlekarna, ter štirje sektorji (trženje, finance in računovodstvo, splošni sektor in izvoz). Ustanovljena je bila leta 1990, se lastninsko preoblikovala leta 1998, vendar njene korenine pridelovanja grozdja in vina segajo še veliko dlje v preteklost. Leta 1894 je v Vipavi nastala prva vinogradniška zadruga na Slovenskem, katere izročilo je še danes veliko bogastvo, saj je tradicija pridelave vina izjemno pomembna za ime vina in za vinorodno deželo, v kateri ga pridelujejo.

Vinska klet družbe Agroind spada po številnih kriterijih na prvo mesto med slovenskimi vinskimi kletmi: po velikosti, po letni količini prevzetega grozdja in po prodanih litrih vina. Vendar fizični kriteriji za poslovno filozofijo že dolgo niso najpomembnejši. Že v preteklih letih in desetletjih so vina iz vipavske kleti prejemale številna priznanja za vrhunsko kakovost, zato se velika pozornost namenja kakovosti vin in njihovi promociji na domačih in tujih trgih. Klet je prepoznavna tudi po posebnih sortah vina, in sicer: avtohtoni sorti Zelen in Pinela pri belih vinih, od rdečih pa je posebnost Barbera, ki je pred 40 leti svojo drugo domovino našla v Vipavski dolini. Posodobljena tehnologija je spodbudila k oblikovanju nove, ustrežnejše blagovne znamke. Zdaj je bogat izbor vin razdeljen v štiri blagovne znamke: Vipava, Ventus, Lanthieri in Storia.

Vipavska mlekarna spada med manjše v Sloveniji. Mlečni izdelki se tržijo pod blagovno znamko Mleko Vipava. Mlekarna ima izvozno številko za EU, še posebej pa so zapriseženi pridelavi izdelkov z znakom varovalnega živila, ki ga podeljuje Društvo za zdravje srca in ožilja Slovenije. Na slovenskem trgu so poznani tudi po pridelavi prvega probiotičnega izdelka – Bifidus, ki še posebej ugodno vpliva na zdravje ljudi. Svoje mlečne izdelke tržijo predvsem na domačem trgu.

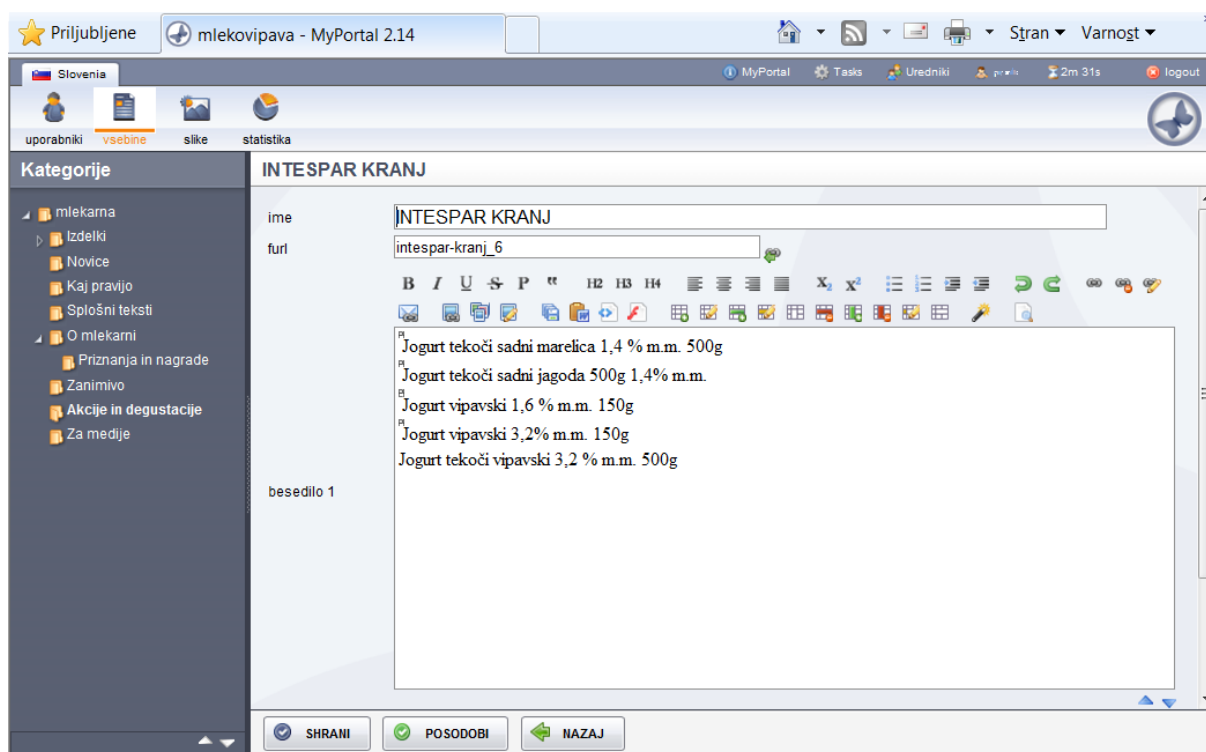
4.2 Spletna stran podjetja

V družbi Vipava 1894 d.d. so se odločili za zunanje razvijalce in oblikovalce spletne strani, in sicer za podjetje Editor d.o.o. Slednje je za upravljanje spletnih vsebin razvilo sistem MojPortal, ki je tudi dostopna točka in urejevalnik spletnih strani Vipava 1894 d.d.

MojPortal spada v panogo CMS aplikacij, ki omogočajo dinamično izdelavo in upravljanje spletnih strani ter enostavno dodajanje in spreminjanje vsebin. Jedro sistema portala in vsi dodatni moduli so izdelani v programskem jeziku PHP in se povezujejo s podatkovno zbirko MySQL.

Preko sistema MojPortal vnašajo vsebine, kot so novice, kontakti in podatki o izdelkih, običajno zaposleni iz oddelka marketinga in informatike. Za izgled strani, način vnosa vsebin, možnosti uvoza in izvoza podatkov pa poskrbi podjetju dodeljena franšizna enota podjetja Editor d.o.o.

³ Daljši naziv družbe je 'Agroind Vipava 1894 Vipava d.d.', zato je podjetje ponekod poimenovano 'Agroind'.



Slika 17: CMS aplikacija Moj Portal

Podjetje Editor d.o.o. je vložilo precej truda v razvoj nove različice sistema MojPortal. Zaradi temeljnih sprememb sistema je bila onemogočena enostavna nadgradnja in potreben je bil prenos vsebin v novo različico, kar je predstavljalo dodaten strošek za Agroind. V Agroindu so se zaradi spremembe blagovne znamke mlečnih izdelkov in varčevanja pri stroških tako odločili za prehod na novo različico le za mlečni del. Vinske vsebine so ostale dostopne preko starejše različice portala, ki temelji še na HTML dokumentih.

Kljub novemu izgledu, poenostavitvi uporabe in prehodu spletnih strani iz HTML na dokumente XHTML v novi različici sistema MojPortal nisem zasledil sprememb, ki bi kaj bolj nakazovale na uporabo semantičnega spleta. Semantična posodobitev spletnih strani je namreč mišljena kot delo razvijalcev in ne končnih uporabnikov. Raje kot začeti z ročnim označevanjem, ponudniki spletnih aplikacij, kot je sistem CMS, globalno razmišljajo o integraciji semantičnega označevanja že v sam sistem. Smiselno opisati in povezati podatke na celotni strukturi strani znotraj sistema CMS pa je lahko velik projekt, ki utegne biti izziv tudi razvijalcem na Editor d.o.o.

4.3 Prednosti, ki jih prinaša uporaba semantike na spletni strani

Z vidika oglaševanja in trženja je nedvomno zelo pomembna ocena optimiziranosti strani za spletne iskalnike (SEO). Boljšo kot imaš oceno, višje si uvrščen med rezultati iskanja. V spletu so številna orodja, za oceno SEO na splošno, pa tudi bolj usmerjena na oceno SEO za določene iskalnike. Firefox-ov dodatek SEO WooRank upošteva mnoge kriterije optimiziranosti. Izrezek ocen, ki opisujejo spletno stran podjetja Vipava 1894 d.d. (Slika 18), nakazuje, kako semantika pridobiva pomen pri optimiziranosti strani za iskalnike.

In-Site SEO						
Home analysis						
URL	http://vipava1894.si Domain length: 10 characters					
Title	Missing					●
Meta description	Missing					●
Meta keywords	Missing					
Headings	H1	H2	H3	H4	H5	H6
	0	0	0	0	0	0
Images	No images found on this website.					●
Text/HTML ratio	0% (view text)					●
Frames	No					●
Flash	No					●
Inside analysis						
Inside pages analysis	It looks like your titles are the same on the most pages!					●
Website compliance						
www resolve	Be careful! Your website without www doesn't redirect to www (or the opposite). It's duplicate content!					●
robots.txt	http://www.vipava1894.si/robots.txt					●
XML Sitemaps	Missing					●
Language	Declared: Missing Detected: en					●
Doctype	Missing					
Encoding	No					●
Google Analytics	No					●
W3C validity	Invalid					●
Microformats	Missing					●
Dublin Core	Missing					●
Geo Meta Tags	Missing					●

Slika 18: SEO analiza spletne strani 'vipava1894.si'

Neposredna prednost uporabe semantike je sama vizija o sematičnem spletu, ki vidi splet kot bazo podatkov. S spletnimi agenti lahko poljubno poizvedujemo po semantično označenih podatkih. Agenti lahko uporabimo znotraj svoje strani, enako pa lahko z našimi podatki razpolagajo tudi drugi agentje (aplikacije). Ustvarjamo lahko prepletene storitve (mashups), podobno kot to v Web 2.0 omogočajo API-ji, le da tu zaradi uporabe semantike ni potrebno

ustvarjati ločenih baz in programirati API-jev, ki bodo posredovali podatke iz teh baz. Podatki so že sami opisani in strukturirani, tako da si naključen agent lahko kadarkoli postreže z njimi. Poglobiti se mora le v semantično tehnologijo, s katero so podatki opisani.

4.4 Izbira tehnologije

Odločiti se moramo, ali bomo naše podatke za spletne agente ločeno opisovali, na primer v datotekah RDF, oziroma bomo opis podatkov vključili neposredno v datoteke HTML. Načeloma je razvijalcem spletnih strani urejanje dokumentov HTML bolj domače. Poleg tega tehnologije semantičnega spleta omogočajo, da v html vključimo semantično opisane podatke, ki bodo vidni samo spletnim agentom, ali pa semantično opišemo podatke, da bodo vidni tudi človeku, tako da se prikažejo na spletni strani. Danes najbolj uveljavljene tehnologije, ki omogočajo neposreden vnos semantike na spletne strani, so: Mikroformati, RDFa in Microdata.

Microdata je najnovejši način pomenskega označevanja spletnih strani, ki bo do izraza prišel šele, ko bo na polno zagnan HTML5. V duhu je to razširljiva različica Mikroformatov, ki jo je predlagal Ian Hickson kot odgovor na RDFa. Nima nikakršne povezave z razvojem RDF, je pa kompatibilen z njim kot tudi z JSON in drugimi podatkovnimi formati. Z njim lahko opisujemo trojčke RDF, vendar ne takih, ki uporabljajo podatkovne tipe, literale XML ali prazna vozlišča. Microdata nekateri iskalniki (Google, Bing) že podpirajo in ima dobre možnosti, da postane zelo uporabljen označevalni jezik za semantiko, vendar je še vedno premalo dokumentiran, zato je tu samo grobo opisan.

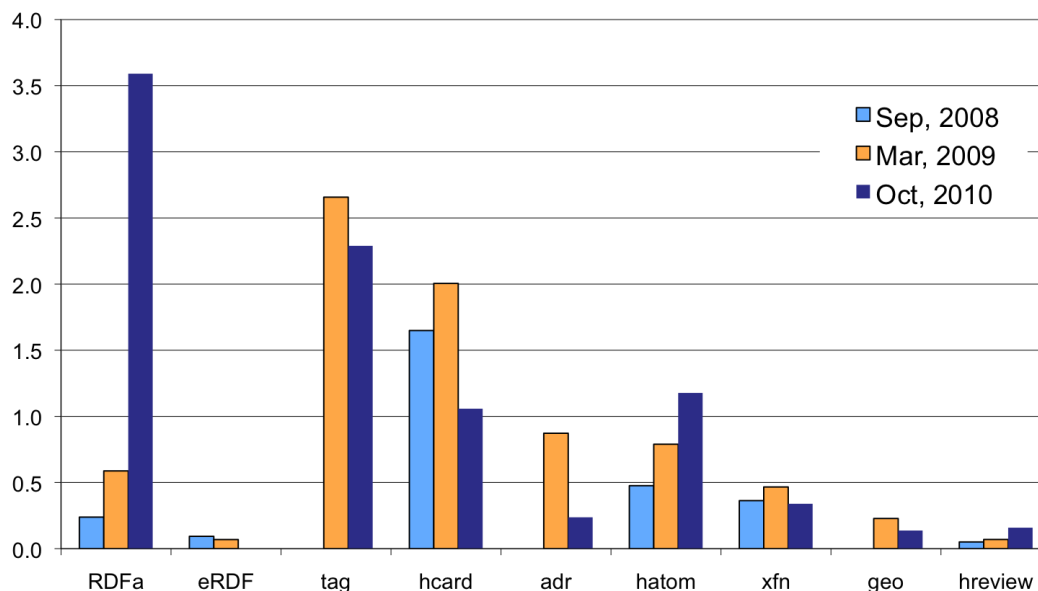
Mikroformati so striktno določen jezik, nerazširljiv, vendar preprost. **RDFa** je še vedno kompleksen, zahteva XML ali XHTML podlago, je pa dobro zasnovan, razširljiv jezik, ki omogoča implementacijo lastnih slovarjev (glej poglavje 3).

Razdvojenost pristopa k semantičnemu spletu, s prihodom Mikroformatov, je bila za mnoge, predvsem pa konzorcij W3C, zaskrbljujoča. Mikroformati so postajali vse bolj priljubljeni. S ciljem boljšega spleta se je W3C odločil, da bo z mehanizmom GRDDL v transformacijo poleg RDFa zajel še Mikroformate. GRDDL je tako z integracijo Mikroformatov, ki običajno pokrivajo pomensko reven jezik HTML, postal močan in prilagodljiv mehanizem, ki je povezal obe tehnologiji in omogočil, da [RDFa in Mikroformate uporabljamo skupaj](#). Z njim so postali Mikroformati dostopni tudi orodjem SS.

Čeprav so bili med razvijalci Mikroformati bolj priljubljeni, nam nove raziskave Yahoo! kažejo, da je RDFa dosegel eksplozivno rast v letu 2010. RDFa je najhitreje rastoči označevalni jezik podatkov na spletu in se uporablja na več kot 430 milijonih spletnih straneh, kar predstavlja približno 3,6 % vseh spletnih strani na internetu. Zahvala gre prizadevanjem zaposlenih na Yahoo! ([SearchMonkey](#)), Google ([Rich Snippets](#)), Facebook ([Open Graph](#)) in vsem, ki so priporočali uporabo RDFa.

Spodnji graf (Slika 19) prikazuje primerjavo uporabe nekaterih Mikroformatov in označevanja RDFa na spletu kot odstotek vseh spletnih strani, ki temelji na analizi 12 milijard spletnih strani, indeksiranih pri Yahoo! Search. Enaka analiza je bila opravljena v treh različnih časovnih obdobjih, tako grafikon prikazuje tudi razvoj uporabe. [4]

Ker sta obe tehnologiji, Mikroformati in RDFa, uveljavljeni med spletnimi iskalniki, poleg tega pa ju združuje še GRDDL, bomo v nadaljevanju izbirali med njima glede na enostavnost uporabe in primernost za opis posameznih objektov. Kot bomo videli, sta obe tehnologiji v spletu podprti z mnogimi orodji, ki poenostavijo dodajanje pomena vsebini.



Slika 19: Uporaba Mikroformatov in RDFa v spletu (% vseh spletni strani)

4.5 Kako začeti z vnosom semantike

Ena izmed možnosti za vnos semantike je neposredno vnašanje metapodatkov. Druga je generiranje semantičnega opisa s pomočjo orodij ali spletnih aplikacij ter kopiranje ustvarjene kode na naše strani. Da ne bi bilo odvečnih korakov ali potrebnega dodatnega izobraževanja za moderatorje vsebin na spletni strani, se kot najboljša izbira izkaže vnosni obrazci znotraj sistema CMS, ki z aplikacijo v ozadju vnešeni vsebini samodejno pripnejo ustrezen pomen.

Primer vnosnega obrazca za tehnologijo mikroformatov je spletna aplikacija [hCard Creator](#) (Slika 20). Seveda CMS potrebuje prilagoditev, ki bo omogočala vnos semantike na stran. Za odprtokodne CMS, kot so [Wordpress](#), [Joomla](#) in [Drupal](#) ([Drupal 7](#) ima že vgrajene module s podporo RDFa), se na spletu najde že mnogo dodatkov za semantično označevanje. Uporabniki sistemov CMS, ki še nimajo podpore za semantiko, smo tako prepuščeni razvijalcem, toda ne povsem. Večina sistemov za urejanje vsebin omogoča tudi neposreden vnos kode znotraj strani HTML. Za začetek si lahko pomagamo s tem, vendar je na mestu opozoriti, da obstaja nevarnost, da sistem CMS s shranjevanjem preko lastnega generatorja html vsebine preoblikuje ali celo prekrije naše vnesene značke.

hCard Creator

hCard-o-matic

name Ivo Prelc

company Vipava 1894 d.d.

street Vinarska cesta 5

city Vipava **state** **zip** 5271

phone 05 3671 000

url www.vipava1894.si

photo url

code

```
<div class="vcard">
<a class="url fn" href="www.vipava1894.si">Ivo Prelc</a>
<div class="org">Vipava 1894 d.d.</div>
<div class="adr">
<div class="street-address">Vinarska cesta 5</div>
<span class="locality">Vipava</span>
<span class="postal-code">5271</span>
</div>
<div class="tel">05 3671 000</div>
</div>
```

preview

[Ivo Prelc](#)
 Vipava 1894 d.d.
 Vinarska cesta 5
 Vipava 5271
 05 3671 000

Slika 20: Spletna aplikacija za generiranje zapisa hCard

Naj bo prenovljen CMS, izdelava obrazcev ali ročni vnos, v vsakem primeru moramo najprej določiti kodo, ki bo opisala izbrane elemente in jih primerno prikazala. Glede na vrsto elementa, ki ga opisujemo, in priljubljenost v iskalnikih lahko izbiramo med paletto označevalnih struktur iz tehnologije mikroformatov ali RDFa (Tabela 9).

Opisovani element	Strukturirano označevanje
Ljudje in organizacije	hCard, RDF vCard
Koledarji in dogodki	hCalendar, RDF Calendar
Mnenja, bonitete, ocene	VoteLinks, hReview
Socialna omrežja	XFN, FOAF
Licence	rel-license

Tabela 9: Nekatere izmed možnosti za strukturirano označevanje

Pri uporabi mikroformatov enostavno izberemo označevalno strukturo za opisovani element, medtem ko se pri RDFa osredotočamo na izbiro ontologije ali skupine ontologij, ki bodo opisovale naš element, oziroma njegovo domeno. Načrtujemo lahko svojo ontologijo. Zaradi boljše podpore v spletu s strani iskalnikov pa je bolje poiskati kakšno od obstoječih ontologij, ki bi se najbolj prilegala podatkom, ki jih želimo predstaviti. Tabela 10 navaja nekaj že obstoječih ontologij za opisovanje stvari, ki se uporabljajo v spletu. Zraven teh je še veliko ontologij domensko specifičnih za področja, kot so tehnologija, okolje, kemija in jezikoslovje. Te niso tako poznane med iskalniki. Če zadovoljijo naše potrebe, se odločimo zanje, sicer pa vseeno razmislimo o razvoju lastnih slovarjev, ker bodo bolj prilagojeni našim podatkom.

Kontakte bi lahko opisali tudi z uporabo tehnologije RDF/OWL. Spodnji primer nam lepo prikazuje zapis kartice zaposlenega, ki se bo prikazal na spletni strani; berljiv bo torej za človeka in hkrati razumljiv za spletne agente (Tabela 12). Vsebino med značkami lahko skrijemo, oziroma prikažemo le spletnim agentom, z uporabo atributa '*content*', ki mu priredimo to vsebino. Velikokrat ta atribut uporabimo za prikaz podatkov agentom v drugačni obliki, kot se prikažejo na zaslonu, na primer zapis datuma pri podatkovnih tipih:

```
<span property="dc:date" content="2011-05-01">maj 2011</span>.
```

```
<div xmlns:contact="http://www.w3.org/2001/vcard-rdf/3.0#"
class="contactinfo" about="http://primer.si/ivop">
  <span property="contact:fn">Ivo Prelc</span>,
  <span property="contact:title">programer</span> pri
  <a rel="contact:org" href="http://vipava1894.si">Vipava 1894 d.d.</a>.
  Kontaktirate me lahko po
  <a rel="contact:email" href="mailto:ivo@vipava1894.si">e-pošti</a>
  ali pokličete na moj
  <span property="contact:tel">
    <span property="contact:type">službeni</span>
    telefon
    <span property="contact:value">05 3671 200</span>.
  </span>
</div>
```

Tabela 12: Kontakt, zapisan z RDFa

Opis zapisa kontakta z RDFa:

- Primer se začne z uporabo imenskega prostora *xmlns*. Ta določa, kje se nahaja slovar (ontologija) s seznamom lastnosti kontakta, kot ga določa *vcard*: *contact="http://www.w3.org/2001/vcard-rdf/3.0#"*.
- V naslednjih vrsticah lahko namesto celotne povezave uporabimo predpono, kot na primer: *property="contact:fn"*, kar določa ime osebe.
- Podobno nadaljujemo z opisom ostalih podatkov kontakta. Pri telefonu opazimo, da ima posamezna lastnost lahko tudi podlastnosti.

4.7 Semantično označevanje izdelkov

Podjetje Vipava 1894 d.d. ima na svoji spletni strani mnogo opisanih izdelkov v človeško razumljivem jeziku, ki predstavljajo izziv, da jih še semantično opišemo za računalnike. Med izdelke sodijo vina, penine, vinjak, mleko, jogurti, siri, puding, zdrob ... To so raznovrstni izdelki, ki zahtevajo raznovrstne opise. Ta raznovrstnost je ključna za korak k izbiri semantične tehnologije.

Dokler je naš namen označevanje elementov, ki so široko v uporabi, je zadeva preprosta. Z Mikroformati lahko opišemo te elemente dokaj enostavno in zelo natančno, žal pa smo z njimi omejeni na točno določene elemente s točno določenimi karakteristikami. Ne ponujajo

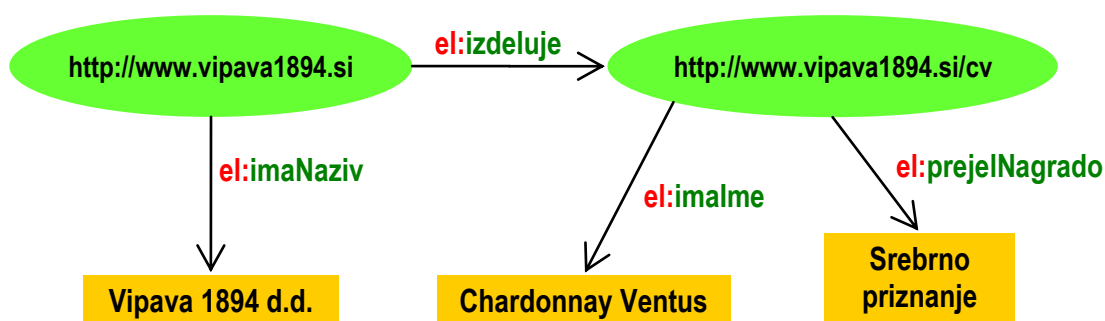
razširitev ločenega opisa lastnosti in ne izdelave lastnih slovarjev. Za opis izdelka Mikroformati sicer ponujajo tri strukture:

- **hListing** – namenjena oglaševanju izdelka za prodajo,
- **hReview** – namenjena oceni izdelka s strani uporabnikov izdelka in
- **hProduct** – prinaša nekaj dodatnih lastnosti za opis izdelka.

Z razvojem v RDFa smo veliko bolj svobodni, kar se tiče razširljivosti opisa elementa in ustvarjanja povezav med elementi. Poleg tega pridobiva RDFa tudi vse večjo podporo v spletu s strani označevalnih orodij, razvijalcev in iskalnikov. Za izdelke, ki jih običajno opisujejo posebne lastnosti, je zato bolje, da se odločimo za tehnologijo, ki bo omogočala opis njihovih posebnosti in hkrati zanje pomenila dodano vrednost v spletu; v našem primeru bo to RDFa.

4.7.1 Zasnova povezav med elementi

Pri načrtovanju spletne strani SS s tehnologijo RDFa je zelo pomembna izbira ontologije. Kakšne možnosti nudi za opis izbranih elementov in kakšen je njen koncept povezav med temi elementi?



Slika 21: Graf RDF

Slika 21 opisuje, da na strani <http://www.vipava1894.si> podjetje Vipava 1894 d.d. predstavlja izdelek Chardonnay Ventus, ki je prejel srebrno priznanje. Zaradi poenostavitve grafičnega prikaza povezav so izpuščene podrobne lastnosti podjetja in vina. Pojem `prejelNagrado` bi lahko še razširili na npr. lokacijo prejema, leto prejema in skupino izbora.

Z namenom poenostavitve RDF zapisa je W3C-jeva skupina SWIG (Semantic Web Interest Group) razvila jezik Notation 3 (N3) (Tabela 14), ki je kompaktna in berljiva alternativa sintaksi RDF/XML. Jezik N3 je še celo bolj razširjen, da omogoča večjo izraznost kot RDF/XML. Spodnja primera v RDF/XML in N3 predstavljata grafu RDF enakovreden zapis.

Ontologije se sicer ne ukvarjajo s posameznimi primeri, vendar so posamezni primeri tisti, ki nas usmerjajo pri razvoju ontologije. V ontologiji se posamezni primeri pretvorijo v razrede,

ki imajo svojo hierarhijo. Tudi povezave med razredi imajo lahko hierarhično strukturo in vsaki povezavi imamo možnost določiti karakteristike⁴ in omejitve (npr. števnost).

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:el="http://www.primer.si/elementi/">

  <rdf:Description rdf:about="http://www.vipava1894.si/">
    <el:imaNaziv>Vipava 1894 d.d.</el:imaNaziv>
    <el:izdeluje rdf:resource="http://www.vipava1894.si/cv"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://www.vipava1894.si/cv">
    <el:imaIme>Chardonnay Ventus</el:imaIme>
    <el:prejelNagrado>Srebrno priznanje</el:prejelNagrado>
  </rdf:Description>

</rdf:RDF>
```

Tabela 13: Zapis v RDF/XML⁵

Prva vrstica RDF dokumenta je deklaracija XML. Sledi ji korenski element RDF dokumentov: `<rdf:RDF>`. Imenski prostor `xmlns:rdf` označuje, da so elementi s predpono `rdf` iz imenskega prostora »`http://www.w3.org/1999/02/22-rdf-syntax-ns#`«. Imenski prostor `xmlns:el` označuje, da so elementi s predpono `el` iz imenskega prostora »`http://www.primer.si/elementi`«. Element `<rdf:Description>` opisuje vir, ki ga določa atribut `rdf:about`. Elementi `<el:imaNaziv>`, `<el:izdeluje>`, `<el:imaIme>` in `<el:prejelNagrado>` so lastnosti posameznega vira.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix el: <http://www.primer.si/elementi/>.
@prefix vp: <http://www.vipava1894.si/>.

vp: el:imaNaziv "Vipava 1894 d.d."; el:izdeluje vp:cv.
vp:cv el:imaIme "Chardonnay Ventus"; el:prejelNagrado "Srebrno priznanje".
```

Tabela 14: Zapis v N3

4.7.2 Izbira ontologije za opis vina

Čprav je RDFa razširljiv in omogoča uporabo lastnih ontologij, je zaradi razpoznavnosti opisa v iskalnikih še vedno boljše, da uporabimo obstoječe ontologije. Težava pa se pojavi, ko imamo element, v našem primeru vino, ki mu želimo opisati lastnosti, in ga obstoječe ontologije še nimajo v svojih slovarjih – ali pač, to je dobro prej raziskati.

⁴ Povezava je lahko funkcionalna, inverzna, tranzitivna, simetrična, antisimetrična, refleksivna ali irefleksivna.

⁵ Pravilnost RDF/XML zapisa lahko preverimo s spletno storitvijo na strani <http://www.w3.org/RDF/Validator/>.

```

<div>
  <div>Barbera Ventus</div>
  
  <div>
    <div>Vinska sorta, ki v Sloveniji uspeva samo v Vipavski dolini. Na
    glinasto peščenih, s soncem obsijanih terasah nam da živahno, rdeče
    rubinasto suho vino z vonjem sljive in drobnih gozdnih sadežev.</div>
    <div>Letnik: 2009</div>
    <div>Sorta: 100% barbera</div>
    <div>Predelava vina:</div>
    <div>
      <ul>
        <li>Vinograd: Terasasti vinogradi</li>
        <li>Trgatev: Ročna, konec septembra</li>
        <li>Maceracija: 6 dni pri kontrolirani temperaturi do 25 oC.</li>
        <li>Fermentacija: V inox posodah pri temperaturi 20-25 oC.</li>
        <li>Zorenje: 8 mesecev</li>
      </ul>
    </div>
    <div>Opis vina:</div>
    <div>
      <ul>
        <li>Barva: Intenzivno temno rdeča barva z vijoličnimi odtenki</li>
        <li>Vonj: Svež, sortno saden, prevladuje vonj ribeza, sliv.</li>
        <li>Aroma: Vino je bogato, polno. Svežino daje nekoliko višja
        kislina, tanini so mehki, prijetni in se lepo vkomponirajo v vino.</li>
        <li>Harmonija: Vino je harmonično, svežina izhaja iz nekoliko
        višje kisline.</li>
        <li>Sposobnost staranja: 3-5 let</li>
      </ul>
    </div>
    <div>Tabela:</div>
    <div>
      <ul>
        <li>Alkohol: 11,7% vol.</li>
        <li>Vsebnost sladkorja: 1,9 g/l</li>
        <li>Skupna kislina: 5,9 g/l</li>
        <li>Temperatura serviranja:14 - 16 oC.</li>
        <li>Poroka vina in hrane: Testenine, rižote z gobami, glavne jedi
        s svinjskim mesom.</li>
        <li>Nagrada:2010,zlata medalja,Vino Slovenija GORNJA RADGONA</li>
      </ul>
    </div>
  </div>
</div>

```

Tabela 15: Lastnosti, ki opisujejo vino – Barbera Ventus

Večina iskalnikov podpira listo razširjenih ontologij, vključno z [Dublin Core](#) (standardi), [FOAF](#) (ljudje), [hReview](#) (ocene) in [GoodRelations](#) (poslovni učinki). Google je celo razvil svojo ontologijo [data-vocabulary.org](#). Paleta s strani iskalnikov podprtih ontologij se bliskovito širi.

Za opis vina, kot prodajnega izdelka, bi lahko uporabili Googlovo ontologijo, ampak zaradi širše podpore raje izberemo GoodRelations (podpora pri [Google Rich Snippets](#) in [Yahoo! SearchMonkey](#)). Ta ontologija ponuja besednjak za zapisovanje poslovnih ponudb, prodaje,

najema, popravil, vzdrževanja blaga in izdelkov ter blagovnih storitev. Z njo poudarimo posebnosti naših izdelkov in storitev ter jih s tem napravimo bolj dostopne za najnovejšo generacijo iskalnikov. Iskalniki zaznajo dodane metapodatke (Tabela 16) in uvrstijo strani višje v lestvici zadetkov, ceno izdelka ali ocene obiskovalcev pa lahko prikažejo že kar v rezultatih (Slika 22).

Barbera Ventus

★★★★☆

St. ocen: 99

The excerpt from the page will show up here. The reason we can't show text from your webpage is because the text depends on the query the user types.

freeweb.t-2.net/v1894/SW/barberaVentus.html - [Cached](#) - [Similar](#)

Slika 22: Rezultat, kot ga prikaže Google, ko uporabimo GoodRelations

```
<div xmlns="http://www.w3.org/1999/xhtml"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:gr="http://purl.org/goodrelations/v1#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">

  <div typeof="gr:Offering" about="#offering">
    <div rev="gr:offers" resource="http://fv/vipava.html#company"></div>
    <div property="gr:name" xml:lang="en">Barbera Ventus</div>
    <div property="gr:description">Vinska sorta, ki...</div>
    <div rel="foaf:depiction"
      resource="http://fv/barberaVentus.jpg"></div>
    <div rel="gr:hasBusinessFunction"
      resource="http://purl.org/goodrelations/v1#Sell"></div>
    <span rel="gr:hasReview">
      <span typeof="gr:Review-aggregate" about="#review_data">
        Ocena: <span property="gr:rating"
          datatype="xsd:string">4</span>,
        na podlagi <span property="gr:count"
          datatype="xsd:string">99</span> ocen.
      </span>
    </span>
    <div rel="foaf:page" resource="http://fv/barberaVentus.html"></div>
  </div>
</div>
```

Tabela 16: Opis vina kot izdelka, z ontologijo GoodRelations

Zgornji RDFa opis vina (Tabela 16) je ustvarjen s pomočjo orodja [Rich Snippet Generator](#). Zgenerirano kodo lahko enostavno kopiramo na zelene strani. Poleg izdelka (storitve) nam generator ponudi tudi izdelavo kartice podjetja, ki izdelek ponuja, in kartice trgovine, kjer ga prodaja. Ali se podatki izdelka, podjetja in trgovine nahajajo v eni datoteki ali ločenih datotekah, je vseeno, saj je povezljivost med njimi določena z atributoma *rel* (relation) in *rev* (reverse relation). Povezave so zelo pomembne za končno poizvedovanje SPARQL.

Z uporabo ontologije GoodRelations smo naše izdelke in podjetje napravili razpoznavne v spletu. Še vedno pa nismo rešili problema opisa vina. S to ontologijo bi vse attribute navedli kvečjemu kot skupen opis, ki spletnim agentom ne bi bil razumljiv. Če želimo imeti 'spletnega sommelier-a' potrebujemo ontologijo, ki bo omogočila poznavanje vrste vina, njegove značilnosti in razmerja. Za domeno [vina in hrane](#) ter povezav med njima je W3C izdal različico ontologije že leta 2004. Predstavil jo je kot ontologijo, ki bo omogočala odgovor na naslednjo poizvedbo: »Povej mi, katera vina naj ponudim k jedilnim menijem; mimogrede, ne maram Sauterna (francosko desertno vino).« Ampak kot pove ta stavek, se v ontologiji bolj osredotočajo na povezavo med hrano in vinom. Definirane imajo že nekatere sorte vina in pokrajine, kjer jih predelujejo. Se pa redki opisi vin ujemajo z našimi vini in ni potrebno ugibati, da tudi slovenski vinorodni okoliši niso vneseni. Ker torej vino ni glavna domena, bi se morali sprijazniti z zelo okrnjenim opisom vina, z nekaj lastnostmi, kot so: `locatedIn`, `hasMaker`, `hasSugar`, `hasFlavor` in `hasBody`.

Izkaže se, da v spletu ni najbolj primerne ontologije, ki bi opisala naš primer. Tako ne kaže drugega, kot začeti razmišljati o razvoju lastne ontologije. Oziroma, odkar vemo, da so ontologije razširljive, lahko to izkoristimo in splošno uporabljeno in podprto ontologijo GoodRelations dopolnimo z razredom vino, njegovimi lastnostmi in ustreznimi povezavami.

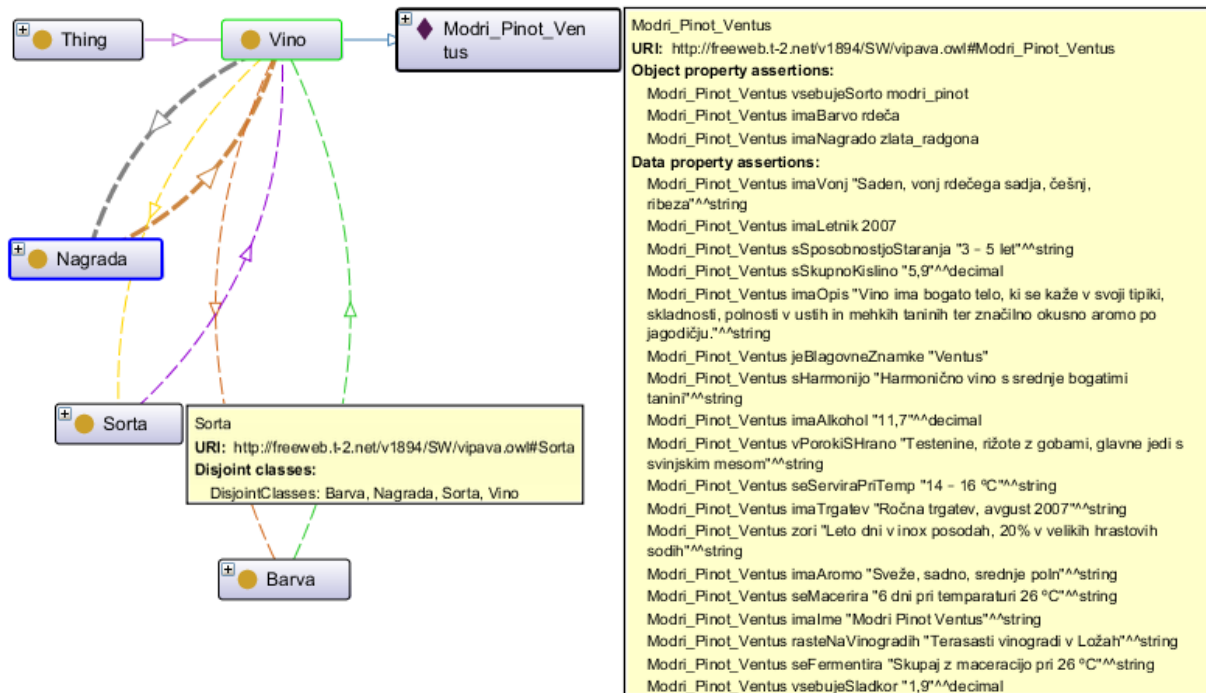
4.7.3 Na poti k lastni ontologiji

Omenili smo že, da razvoj ontologije ni preprost, primerjamo ga lahko s kreiranjem relacijske podatkovne baze. Zastaviti si moramo konceptualni model, ki bo opisoval vino in ga povezoval s proizvajalcem, prodajo in še kakšnimi entitetami iz drugih ontologij. Lastnostim razreda vino moramo določiti podatkovne tipe. Najbolje, da jih uporabimo že definirane iz W3C ontologije [XMLSchema](#). Kljub tem podobnostim iz podatkovnih baz ima [OWL](#) svoje specifike, ki zahtevajo poglobitev v strukturiranje ontologij in poznavanje izpeljank RDF-ja, vendar to še vedno ni znanstvena fantastika, kot opisujejo mnogi članki. Poleg tega obstajajo v spletu v pomoč številni urejevalniki ontologij, kot so: [Protégé](#), [NeOn Toolkit](#), [TopBraid Composer](#), [Altova SemanticWorks](#) ... Seveda bi postopek razvoja ontologije vseeno precej razširil vsebino našega konteksta, zato so ti odstavki predvsem kazalo, kam lahko pelje naš naslednji korak.

Za zajetje znanja domene vina v ontologijo je torej potreben dogovor o lastnostih vina, ki jih želimo izraziti. Vino kot razred lahko neposredno opišemo z njegovimi podatkovnimi lastnostmi, lahko pa tudi oblikujemo podrazrede, ki podrobneje opišejo vino. Taki podrazredi za vino bi lahko bili:

- Sorta z lastnostmi: `imaImeSorte`, `imaProcentSorte` ...
- Barva z lastnostmi: `vOsnovi` (belo vino, rdeče vino), `imaOdtенок` ...
- Nagrada z lastnostmi: `imaLetoNagrade`, `imaNazivOcenjevanja`, `imaVrstoNagrade` ...

Lastnosti, ki povezujejo razred z literalom oz. podatkovnim tipom, imenujemo *podatkovne lastnosti*. Lastnosti, ki povezujejo razrede (objekte) med seboj, pa so *objektne lastnosti*. Razred Vina lahko tako povežemo z zgornjimi podrazredi z objektnimi lastnostmi: `vsebujeSorto`, `imaBarvo` in `imaNagrado`. Tako hierarhično strukturiranje razredov imenujemo taksonomija (Slika 23).

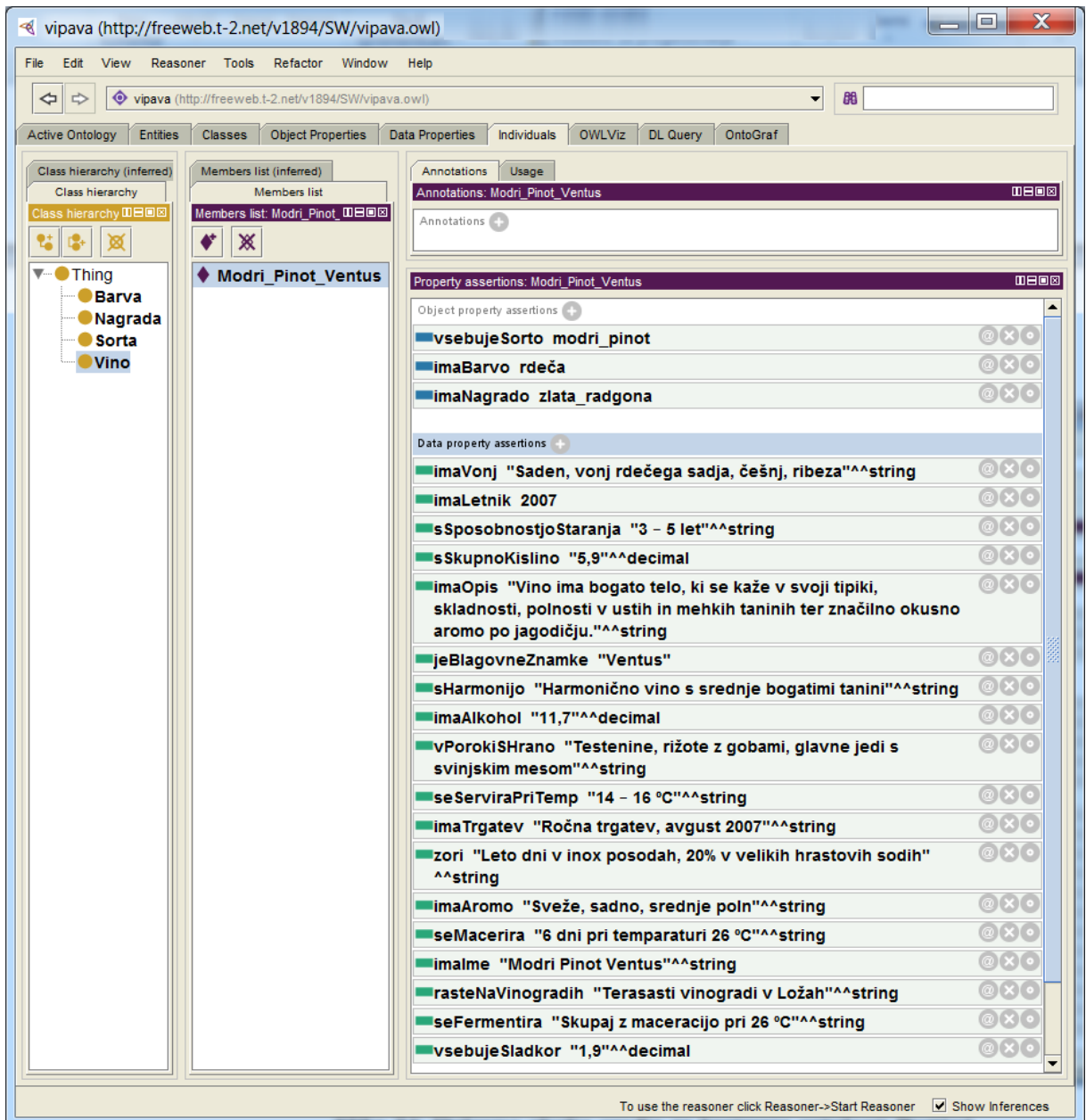


Slika 23: Primer razrednega opisa vina

Eden izmed dobro dokumentiranih programov za urejanje ontologij je odprtokodni program Protégé. Razvit je bil na univerzi Stanford University. Protégé omogoča kreiranje, vizualizacijo in manipuliranje z ontologijami v različnih predstavitvenih načinih. Arhitektura orodja Protégé je zgrajena modularno na osnovi programskega jezika Java. Modularnost omogoča razvijalcem razširljivost ter dodajanje novih funkcionalnosti. Poleg tega ga je možno prilagajati glede na različne zahteve za uporabnikom prijaznejšo podporo pri kreiranju modelov znanja in vnašanju podatkov. Protégé lahko uporabimo za ustvarjanje Protégé-Frames ontologij in OWL ontologij [2]. Slika 24 prikazuje delovno okolje za ustvarjanje OWL ontologije v orodju Protégé; ustvarja se posamezen primer (Modri Pinot Ventus).

Če povzamemo, podprto ontologijo nameravamo razširiti z entiteto vina, ki spletnim iskalnikom ne bo veliko pomenila; kaj nam to koristi? Z uporabo knjižic SPARQL lahko v Javi, PHP-ju ali drugih jezikih, naredimo na naši spletni strani lastnega agenta. Oblikujemo obrazec za poizvedbo, ki omogoči obiskovalcu vnos povpraševanja po lastnostih vin ali celo priporočenih porokah z jedjo, če vino povežemo z entiteto hrane. Aplikacija bo v povezani strukturi s SPARQL poizvedbo poiskala natančne odgovore ter jih ponudila obiskovalcu v branje ali izvoz. Poizvedovanje po naših podatkih, s prosto dostopno ontologijo, bo lahko izkoristila tudi poljubna aplikacija v spletu.

S preprosto zastavljeno strukturo vina (Slika 23) lahko z aplikacijo na spletni strani ponudimo možnost izpisa vin po barvi ali vin, ki so prejela nagrado v nekem obdobju. Izpišemo lahko vsa vina, ki vsebujejo določeno sorto grozdja, poizvedemo po sladkorju ali alkoholu, ki ga vsebujejo, blagovni znamki itd. S SPARQL lahko sestavljamo tudi kombinirane poizvedbe, ki vključujejo več lastnosti, če utegnejo biti zanimive za obiskovalca spletne strani.



Slika 24: Delovno okolje orodja za kreiranje ontologij Protégé

5 SKLEP

Pri spletnem poizvedovanju nam semantični splet obljublja konec brskanja skozi kupe nepomembnih informacij. Odslej naj bi bilo povezovanje zadetkov, ki jih vrnejo iskalniki v smiseln odgovor, delo računalnikov. Čeprav smo že v dekadi Web 3.0, ki je sinonim za obdobje semantičnega spleta, se ta obljuba oziroma sanje, še niso povsem uresničile. Za uresničitev namreč ni dovolj imeti semantiko in poznati vse odnose med podatki. Računalniki bi si morali zapisovati naše vedenje, ko brskamo po spletu, in vzpostaviti interakcijo z nami, ko bi potrebovali dodatno razlago o podatkih, ki jih iščemo. Skratka, semantika se mora prepresti še z mnogimi informacijskimi področji, da se izrazi njena koristnost pri obdelavi spletnih podatkov. Semantične tehnologije in orodja, ki sem jih zajel v tej nalogi, pa so že prvi sadovi dolgoletnega razvoja SS. Označevanje podatkov, ki je bilo vseskozi pereč problem, izgublja svojo zapletenost z uporabo novih orodij za urejanje semantike.

V duhu iskanja semantičnih tehnologij, za uporabo na spletni strani podjetja, sem v diplomski nalogi predstavil, kako tehnologije SS obravnavajo potrebo po strukturiranih podatkih dosledno po standardih; kako dodajajo pomensko vrednost, neposredno za stroje in posredno za človeka, obstoječim plastem spleta, od HTTP in URI naprej; kako opisovalni okvir RDF omogoča nedvoumno izražanje odnosov med objekti samimi in v povezavi le-teh s podatkovnimi tipi oz. literali. Naštel sem glavne vrste in lastnosti ontologij ter predstavil tiste, ki temeljijo na standardu OWL. Krog SS sem zaključil z opisom inteligentnih agentov in poizvedovalnega jezika SPARQL, ki ga uporabljajo pri iskanju podatkov v spletu opisanega znanja. Nadalje sem predstavil Mikroformate, ki so ubrali svojo pot k semantičnemu spletu, in mehanizem GRDDL, ki je pristop z Mikroformati povezal s pristopom spletnega konzorcija W3C. Omenil sem tudi Microdata, tehnologijo, za katero menim, da bo nasledila Mikroformate, ko bo HTML5 polno funkcionalen, in ki zaradi enostavnosti uporabe utegne biti tekmelec tudi za RDFa.

Semantika preplavlja splet, po drugi strani pa ugotavljam, da nikakor nisem edini, ki je v zadnjih dveh letih prvič slišal za SS. In mogoče še vedno ne bi, pa sem izbral to temo za seminarsko nalogo, ravno iz radovednosti, kaj je to SS. Resda SS niso tehnologije, s katerimi bi se moral globlje seznaniti vsak uporabnik spleta, menim pa, da vseeno primanjkuje člankov na temo njihove praktične uporabe, zlasti v slovenskem jeziku. Podjetja se danes zavedajo prednosti spletnega oglaševanja, ne vedo pa, kakšno prednost jim pri trženju izdelkov lahko prinese SS. Slednje so že okusila podjetja v svetu, kot na primer Best Buy, ki je z uporabo RDFa dosegel 30% povišanje prometa na njihove spletne strani. Implementacija informacijskih tehnologij v podjetjih je velikokrat samoiniciativa razvijalcev, iz njihovega zanimanja za novo tehnologijo. Upravo prepričati, da podpre uvedbo nove tehnologije, je velikokrat težja naloga kot implementacija sama, zlasti ko nimaš zagotovila, da bo ta prinesla dodano vrednost podjetju. Z vidika konkurenčnosti v poslovnem svetu pa predstavlja tehnologija SS ugodnosti, ki jih lahko žanješ le, dokler si med prvimi, ki jo uporabljajo. Iskalniki v prikazu dajejo prednost semantično obogatenim stranem, toda slej kot prej se bo semantika znašla na večini spletnih strani in iskalniki jih bodo zopet obravnavali 'enakovredno'. Seveda to ne pomeni, da poslej SS ne bo več koristen. SS je odprta rešitev, ki je razvita v skupno dobro in ne samo v poslovne namene. Če jo posamezniki ali podjetja izkoristijo v svoje dobro, je to kvečjemu prednost za tehnologijo samo, kajti nihče noče zaostajati za drugimi in tako se lahko ta hitreje uveljavi.

Za cilj sem v tej nalogi izbral iskanje načina, oziroma tehnologije SS, za semantični opis objektov na spletni strani podjetja Vipava 1894 d.d., tako da bodo razumljivi spletnim

agentom. Ob tej raziskavi so me navdušile prednosti, ki jih prinaša SS pri oglaševanju izdelkov. Danes se sicer vedno bolj uveljavlja oglaševanje in trženje preko socialnih omrežij, ki že koristijo semantično označevanje. Za podjetje pa se še vedno spodobi, da ima tudi svojo spletno stran. Z ozirom na izvedbo obstoječe spletne strani podjetja sem si postavil vprašanje: Kako vnašati semantiko na spletno stran preko sistema za upravljanje vsebin, ki še nima podpore za ustvarjanje metapodatkov? Odgovor mi je bil jasen: z vnašanjem oznak neposredno v html. Ni pa bila to rešitev. CMS namreč ob shranjevanju vnesene vsebine ponovno generira html zapis, ki ni nujno enak vnesenemu. Poleg tega tudi glava trenutnih spletnih strani, do katere nimamo dostopa preko CMS, ni v skladu z RDFa, ki zahteva za osnovo dokument 'XHTML+RDFa' in ne navaden dokument 'HTML'. Izkazalo se je, da je za semantično uporabo, dolgoročno potrebna nadgradnja CMS.

Pri opisovanju izdelkov sem si zastavil, da bi v semantični opis zajel vse lastnosti, kot jih je doslej opisoval tekst na spletnih straneh. Izbral sem tehnologijo RDFa. Mikroformati so sicer preprostejši, toda neprilagodljivi. RDFa je močnejša in razširljiva tehnologija, ki temelji na dolgoletnem razvoju in spletnih praksah. Naslednji korak je bilo iskanje ontologije. Pri izbiranju ontologij se zaradi boljše razpoznavnosti v spletu odločamo za tiste, ki so podprte s strani večjih iskalnikov. Za opisni objekt sem si izbral vino, ki ima mnogo specifičnih lastnosti, zato v spletu nisem našel ontologije, ki bi najbolj ustrezala opisu teh lastnosti. Pot me je tako napeljala na razvoj lastne ontologije, z namenom združene uporabe z obstoječimi v spletu bolj uveljavljenimi ontologijami. Toda zakaj uporabiti lastno ontologijo, nepoznano spletnim iskalnikom? Trženje izdelkov z uporabo semantike me je nekoliko oddaljilo od osnovne zamisli o SS, ki daje spletnim agentom možnost razumevanja vsebine spletnih strani in uporabe spleta kot baze podatkov. Jasno je, da SS ni samo splet za iskalnike, ki se osredotočajo na širše uporabljene ontologije. Tudi naše aplikacije lahko izkoristijo strukturiranost podatkov z novimi tehnologijami. Zakaj torej ne ustvariti svojega spletnega agenta, ki bo obiskovalcem naše spletne strani postregel s podatki o vinih na enostaven način, brez brskanja po vseh straneh, kjer so sicer vina opisana. To mi je postalo izziv, vendar hkrati preobširna tema, da bi jo zajel v tej nalogi. Nakazal sem jo bolj kot smer, ki bi izkoristila tehnologijo SS za enostavnejši dostop do podatkov, razpršenih po spletnih straneh podjetja.

6 LITERATURA IN VIRI

- [1] Michael C. Daconta, Leo J. Obrst, Kevin T. Smith, The Semantic Web: A guide to the Future of XML, Web Services, and Knowledge Management, Indiana: Wiley Publishing Inc., 2003, pogl. 8.
- [2] (2011) A Practical Guide To Building OWL Ontologies Using Protege 4 and CO-ODE Tools. Dostopno na:
http://owl.cs.manchester.ac.uk/tutorials/protegeowltutorial/resources/ProtegeOWLTutorialP4_v1_3.pdf
- [3] (2008) Making Sense of the Semantic Web. Dostopno na:
<http://vimeo.com/684381>
- [4] (2011) Microformats and RDFa deployment across the Web. Dostopno na:
<http://tripletalk.wordpress.com/2011/01/25/rdfa-deployment-across-the-web/>
- [5] (2008) Planning a Semantic Web site. Dostopno na:
<http://www.ibm.com/developerworks/xml/library/x-plansemantic/>
- [6] (2011) RDFa Microformat Tagging For Your Website. Dostopno na:
<http://www.customerparadigm.com/index/503/RDFa-Tagging-Services.php>
- [7] (2011) RDFa: The Inside Story from Best Buy. Dostopno na:
<http://searchnewscentral.com/20110207129/Technical/rdfa-the-inside-story-from-best-buy.html>
- [8] (2008) Semantic Web Patterns: A Guide to Semantic Technologies. Dostopno na:
http://www.readriteweb.com/archives/semantic_web_patterns.php
- [9] (2007) Semantični splet - več kot le hype! Dostopno na:
http://videlectures.net/kp07_grobelnik_spvkh/
- [10] (2006) The Ontology Integration Problem. Dostopno na:
http://novaspivack.typepad.com/nova_spivacks_weblog/2006/08/the_ontology_in.html
- [11] (2007) Top-Down: A New Approach to the Semantic Web. Dostopno na:
http://www.readriteweb.com/archives/the_top-down_semantic_web.php
- [12] (2007) Web 3.0 – Semantični splet. Dostopno na:
<http://blog.morphix.si/2007/uporabniske-izkusnje/web-30-%E2%80%93-semanticni-splet/>
- [13] (2007) Why the Semantic Web Will Fail. Dostopno na:
<http://halfanhour.blogspot.com/2007/03/why-semantic-web-will-fail.html>