

MARIJA PUTICA  
Filozofski fakultet Sveučilišta u Mostaru  
marija.putica@ff.sum.ba

## SEMANTIČKI WEB

### Sažetak

Autorica je u radu predstavila semantički *web* kao novu generacijsku fazu u razvoju *weba*. Nasuprot prethodnim generacijama koje su se temeljile na radu s ljudima, u radu je predstavljena treća generacija *weba*, usmjerena na strojno procesiranje pohranjenih informacija. Informacije se procesiraju na višoj razini jer se semantički *web* temelji na davanju značenja informacijama na *webu*. Definiran je kao plan za postizanje povezanih podatkovnih aplikacija na *webu* u formi kojom oblikuju konzistentnu logičku mrežu podataka. Stoga je u radu predstavljen proces izgradnje semantičkoga *weba* kroz uspostavu nove razine međuoperabilnosti. Zbog važnosti za sintaktičku formu dokumenta i sintaksni sadržaj predstavljeni su jezici semantičkoga *weba*, a zbog važnosti kategorizacije sadržaja *weba* u radu predstavljeni su i rječnici koji obuhvaćaju različita područja. Prethodno su se različite ontologije zasebno razvijale, a semantički *web* donio je potrebu interoperabilnosti te je u radu potrebna pažnja posvećena jezicima ontologija. Govoreći o primjeni semantičkoga *weba*, autorica se posebno usmjerila na upravljanje znanjem te je predstavila alate za održavanje slabo strukturiranih izvora informacija.

*Ključne riječi:* semantički *web*, procesiranje informacija, povezivanje podatkovnih aplikacija, jezici semantičkoga *weba*, rječnici, ontologije, primjena semantičkoga *weba*

## Uvod

*World Wide Web* najpoznatiji je i najčešće korišten internetski servis. Osmišljen je kao alat za razmjenu i povezivanje dokumenata unutar jednoga istraživačkog centra. Utemeljen je na tri standarda (*Transmission Control Protocol/Internet Protocol-TCP/IP*, *Hypertekst Transfer Protocol-HTTP* i *Hypertext Markup language-HTML*). Doživio je dramatičan rast te je pribavljanje relevantnih informacija postalo problem. Iz obilja informacija na *webu* teško je izdvojiti one koje su doista potrebne. Problem predstavlja i provjera njihove vjerodostojnosti kao i kombiniranje različitih izvora. Stoga je pozornost usmjerena na razvoj specijaliziranih *siteova* za pretraživanje podataka. Tehnologija internetskih pretraživača utemeljena je na indeksiranju i pretraživanju ključnih riječi. Takva je tehnologija mogla zadovoljiti prvu razvojnu fazu *weba*, gdje su se uporabom *HTML*-a prikazivale statičke informacije. Sljedeća razvojna faza poznata je po uvođenju dinamičkih komponenti te se uz *HTML* počinju koristiti skriptni jezici. Dinamičke stranice generirane su računalom unaprijed iskodirane nekim od prigodnih jezika (*CGI, PHP, ASP...*). Web-prezentacije zamijenjene su web-aplikacijama zahvaljujući razvoju tehnologije pisanja skripti za izvršavanje na serveru. Obje faze karakterizira usmjerenost na rad s ljudima, a ne s drugim strojevima.

Tolerancija je jedan od temeljnih principa *weba*, a podrazumijeva evolutivan razvoj gdje nove tehnologije ne onemogućavaju korištenje starijih, već se iz njih razvijaju. Treća razvojna faza *weba* usmjerena je na strojeve koji bi na višoj razini mogli procesirati informacije s kojima raspolažu. Zasnovana je na ideji iskorištavanja znanja, principa i tehnologija prethodnih razvojnih faza *weba* za nove faze, a sve u cilju kako bi postao univerzalnim medijem za razmjenu podataka, informacija i znanja. Semantički *web* predstavlja treću razvojnu fazu svjetske računalne mreže *World Wide Web* jer podrazumijeva pridjeljivanje značenja informacijama na *webu*. Sintaktičke i strukturalne metode obrade informacija zamijenjene su semantičkim kako bi se proširila primjena *weba*. Tako bi se od komunikacijskoga medija stvorio globalni informacijski sustav. Informacije pohranjene na *webu*, bez obzira na *metatagove*, nemaju

značenje bez ljudske interpretacije. Cilj semantičkoga *weba* je učiniti ih razumljivim strojevima te u novome okružju biti popunjen sadržajima s formalno izraženom semantikom koja omogućuje potpunu automatizaciju u organizaciji i pronalaženju informacija. Njegova konačna definicija nije utvrđena, ali je razumijevanje i povezivanje informacija te izvođenje zaključaka njegova temeljna karakteristika. Polazi od pretpostavke da internet razumije informacije koje čuva te da je osposobljen uspostavljati logičke veze među njima. Semantika podrazumijeva formalno iskazivanje veze među dokumentima i njihovim značenjem, a organizacija informacija u cijelosti može funkcionirati preko računalne komunikacije (*m2m*). Praćenje povijesti korisnikova pretraživanja rezultira nuđenjem sadržaja prilagođena korisniku. Usvajanjem koncepta nuđenja informacija dograđen je dotadašnji koncept prezentiranja informacija. Navedena sposobnost proistječe iz činjenice da je semantički *web* zapravo sinergija programa koji iz različitih internetskih izvora prikupljaju sadržaj, nakon čega procesiraju informaciju i na globalnoj razini rezultate razmjenjuju s drugim programima. Njegov razvoj zasigurno će dovesti do bitne izmjene u dohvaćanju, sortiranju i klasificiranju informacija.

## 1. Ideja semantičkoga weba

Izvori informacija na *webu* kakve danas koristimo namijenjeni su ljudskom korisniku. Čovjek im pristupa i manipulira putem linkova. Nasuprot tomu, semantički *web* ne zanima struktura poveznica među internetskim dokumentima, već odnosi među elementima i njihovim svojstvima. Semantika je grana lingvistike i bavi se proučavanjem značenja jezičnih znakova, za razliku od informacijske znanosti koju ne zanima značenje informacije nego je promatra po strukturalističkome principu. U novije vrijeme umjetna inteligencija je, kao područje informacijskih znanosti, postala iznimno važno područje pa su obrada prirodnoga jezika, strojno učenje i strojno prevođenje postali važni za obradu prirodnoga jezika na svim razinama, pa i na semantičkoj.

Semantički *web* temelji se na ideji manipuliranja podacima i programima kreiranim neovisno o podacima. *World Wide Web Consortium* (*W3C*) zadužen je za donošenje standarda i razvoj *weba*. Koncem protekloga tisućljeća započeli su s promoviranjem semantičkoga *weba*. Godine 1996. Tim Berners Lee, osnivač *weba* i čelna osoba njegova razvoja, naglasio je činjenicu otežanoga pronalaženja podataka koji nisu izraženi u obliku konkretna semantičkoga navoda. Lee je 1998. godine semantički *web* predstavio kao plan za postizanje povezanih podatkovnih aplikacija na *webu* u formi sposobnoj oblikovati konzistentnu logičku mrežu podataka. Svrha semantičkoga *weba* je povezivanje temeljnih područja ljudskoga znanja i njihova strojna obrada putem kontekstualizirane i stroju razumljive terminologije. Kreiranje inteligentnih programskih agenata sposobnih za samostalno izvršavanje iznimno složenih zadataka pomoću *m2m* (stroj sa strojem) komunikacije cilj je koji je Lee postavio pred razvojem semantičkoga *weba*.<sup>1</sup> Stoga *W3C Semantic Web Activity* punu pozornost pruža standardiziranju ključnih tehnologija koje omogućavaju necentralizirani razvoj, pri čemu svi dijelovi trebaju biti kompatibilni.

*W3C Semantic Web Activity* organizacija utemeljena je s ciljem rukovođenja aktivnostima na izradi otvorenih kolaborativnih tehnologija. Njezina uloga može se svesti na četiri temeljna zadatka. To su: uvođenje standarda, edukacija, suradnja i koordinacija te napredni razvoj. *W3C Semantic Web Activity* ima dvije radne grupe. To su *RDF Core Working Group* (*RDFCore*) i *Web Ontology Working Group* (*WebOnt*) koje rade na zadacima definiranja standarda i tehnologija. *Resource Description Framework* (*RDF*) danas pruža obično okružje za reprezentaciju metapodataka u brojnim aplikacijama te je temeljni zadatak *RDF Core Working Groupa* završetak *RDF* opisnoga rječnika u *RDF Schema Candidate Recommendationu* (*RDFS*). Radna grupa *Web Ontology Working Group* radi na izradi jezika za definiranje strukturiranih ontologija baziranih na *webu* koji će osigurati bogatiju integraciju i interoperabilnost

1 Usp. TIM BERNERS-LEE – JAMES HENDLER – ORA LASSILA, „The semantic web”, *Scientific American*, svibanj 2001., <<http://sciam.com/2001/0501issue/0501berners-lee.html>>, (12. VII. 2017.).

podataka među različitim jedinicama. Polazna osnova za tu aktivnost je *RDF Core*.

Programi za obradu podataka trebaju strojno čitljive navode o izvorima informacija i njihovim međusobnim odnosima kako bi mogli obavljati svoju ulogu. Današnji stupanj razvoja opterećen je brojnim nedostacima. Potrebni podatci nisu uneseni ili nisu uneseni u strojno čitljivom obliku, nisu izloženi u obliku u kojem ih računala mogu procesuirati, ne postoji agent u kojem se problemi mogu sročiti ili ne postoji jezik definiran za takve upite – samo su neki od problema koje tek treba otkloniti. Nužno je izraditi jedan standardizirani model, izraditi vezu među pojmovima u rječniku i donijeti njihove jedinstvene definicije te ih učiniti dostupnima programima. Za razvoj semantičkoga *weba*, poglavito za klasifikaciju struktura znanja, osobito su važni standardi i programska rješenja za terminološku razmjenu te ontologijski rječnici. Tako će se *web* sastojati od opisa i ontologija, a njima će programski agenti upravljati samostalno obrađujući prikupljene podatke i komunicirati djelomičnim razumijevanjem.

Podatci predstavljeni na dosadašnjem *webu* razumljivi su za ljude, dok za strojnu interpretaciju predstavljaju nerješiv problem. Stoga je nužno uvođenje novoga načina predstavljanja podataka na *webu* koji će biti strojno čitljivi i iz kojih će različiti programi izdvajati značenja. Semantički *web* podrazumijeva i pretraživanje uporabom značenja, a ne samo riječima, što je karakteristika dosadašnjega *weba*. Taj cilj zahtijeva semantičku i sintaktičku interoperabilnost predmetnoga rječnika.

Predviđa se razvoj dva oblika semantičkoga *weba*. To su:

1. Klasičan *web* u kojem će dokumenti, pored oblika prilagođena čovjeku, sadržavati i strojno čitljive informacije na način da će se u obične *HTML/XHTML* stranice upisivati podatci koji omogućuju programsko izvlačenje zaključaka iz njih. Tako će se semantički *web* integrirati u postojeći, a isti dokumenti koristit će se za prikaz korisnicima i za strojnu obradu.

2. Semantički *web* u kojem će postojati reference iz semantičkoga *weba* na dokumente običnoga *weba* i obrnuto.<sup>2</sup>

Za lakši prijelaz s dosadašnjega na semantički *web* razvijaju se tehnologije koje bi osigurale automatsko izdvajanje značenja iz nestrukturirana teksta. Utvrđivanje odnosa među elementima i njihovim svojstvima uporabom metapodataka (podataka o podacima) omogućava strukturiranje nestrukturiranih ili polustrukturiranih podataka. Semantički *web* koristi se metapodacima kako bi predstavio informacije. U početnoj fazi svoga razvoja koristi se za *inteligentnije* pretraživanje informacija te za razvoj dijaloških sustava. Stoga se razvijaju specijalizirane tražilice koje se koriste tehnologijama semantičkoga *weba* i metodama za obradu prirodnoga jezika.

## 2. Umjetna i kolektivna inteligencija

Suvremeno društvo vjeruje u pojedince, a za brojne umne ljude koji su obilježili čovječanstvo masa je sinonim za glupost. Tako je Thomas Carlyle tvrdio kako ne vjeruje u kolektivnu mudrost sastavljenu od gluposti pojedinaca, dok je Henry David Thoreau tvrdio kako mase nikada neće doseći standard svojih najboljih pripadnika, već da će se degradirati do razine onih najgorih. Grupe često donose loše odluke, ali mali organizmi predstavljaju suprotne primjere. Opće pravilnosti kolektivne inteligencije možemo vidjeti u brojnim sustavima kao što su mravi, pčele, ptice, bakterije, ribe i sl. Tek je Basil Creighton 2002. godine u noveli *Žrtva* pisao o roju umjetnih nanorobota koji nalikuju insektima. Njima nije upravljao pojedinac već kolektivni um koji im je omogućavao donošenje složenih odluka, pa i anticipiranje budućih događaja.

Termin inteligencije roja 1989. godine ustvrdili su Wang Jing i Gerardo Beni s MIT-a kako bi opisali metode za optimizaciju i kontrolu sustava robota. Thomas D. Seeley je 2010. godine objavio knjigu naziva

2 Usp. LJUBIŠA MILIVOJEVIĆ, „Semantički Web – ideja i tehnologije”, *Institut informacijskih znanosti Maribor*, <[http://home.izum.si/cobiss/oz/2009\\_1'2/html/clanek\\_o2.html](http://home.izum.si/cobiss/oz/2009_1'2/html/clanek_o2.html)>.

*Pčelinja demokracija*.<sup>3</sup> U knjizi je, kao i veći broj suvremenih neurobiologa, ponašanje pčelinjih rojeva usporedio s mozgovima primata. Povezuje ih njihovom oblikovanošću procesom prirodne selekcije kao spoznajne jedinice sposobne prihvatiti i procesirati informaciju s ciljem donošenja odluke. Rojeve i mozgove vidi i kao demokratske sustave odlučivanja koji nemaju vođu koji umjesto njih donosi odluke. Shvaćena kao sposobnost da se kolektivno djeluje inteligentnije nego što to čine pojedinci u poslovnome svijetu – kolektivna inteligencija inicirala je brojne promjene u vrijednostima, paradigmama, modelima te u skupovima vještina. Stoga se kolektivna inteligencija naziva i inteligencija roja ili skupna inteligencija, a često se preklapa s konceptom *data mining*. Definira se kao sposobnost skupine u pronalaženju brzih i djelotvornih rješenja za složene probleme djelotvornije nego njezini pojedini članovi. Tehnologije koje doprinose kolektivnoj inteligenciji uključuju kolaborativno oglašavanje i zajedničke baze podataka za dijeljenje znanja.

Dok je umjetna inteligencija temeljena na strojevima koji razmišljaju bolje od ljudi, kolektivna inteligencija temeljena je na strojevima koji razmišljaju umjesto ljudi. Primjer za to je i pretraživač *Google* koji je postao stroj ideologije popularnoga.<sup>4</sup> Posjeduje enciklopedijsku širinu, ali on odlučuje o onome što je nama potrebno znati. Navedeno postiže zbog svojih sposobnosti da na temelju već pretraživanih podataka i termina, koji su najčešće korišteni, naše upite povezuje u cjelinu.<sup>5</sup> Riječ je o tek djelićima tehnologije semantičkoga *weba* spram čijih je učinaka na društvenim mrežama već iskazana stanovita bojazan. Stoga su u popularnoj kulturi nagoviješteni razmjeri katastrofe, u slučaju da se strojevi približe ljudskoj inteligenciji, dosta daleki.

John McCarthy, tvorac termina umjetna inteligencija i tvorac prvoga programskog jezika umjetne inteligencije *Lisp*, težio je programirati

---

3 Usp. THOMAS D. SEELEY, *Honeybee Democracy*, Princeton University Press, Princeton, 2010.

4 Usp. THOMAS W. MALONE – ROBERT LAUBACHER – CHRYSANTHOS DELLAROCAS, *Harnessing Crowds: Mapping the Genome of Collective Intelligence*, MIT, Cambridge, 2009.

5 Usp. FRANCIS HEYLIGHEN, *Collective Intelligence and its Implementation on the Web: algorithms to develop a collective mental map*, Leo Apostel, Free University of Brussels, 1999., str. 253., <clearm11.vub.ac.be/Papers/CollectiveWebIntelligence.pdf>, (6. VII. 2017.).

stroj koji će proći test Alana Turinga iz sredine prošloga stoljeća. Turing je uspostavio temelje umjetne inteligencije, ali ih nitko poslije nije bitnije nadogradio. McCarthy nije ostvario svoje težnje, a preminuo je u listopadu 2011. godine. Ostala je poznata, a čini se i olako shvaćena, njegova rečenica kako o „procesima ljudskog razmišljanja znamo isto koliko riba zna o plivanju“. Njome je upozorio kako ljudi još nisu naučili razumijevati, ne poznaju vlastiti mehanizam razumijevanja već razumijevaju instinktivno. Stoga taj mehanizam ne mogu kopirati i primijeniti u strojeve. U knjigama *Mind Children* i *Robot Mere Machine to Transcendent Mind*<sup>6</sup> Hans Moravec, profesor robotike i umjetne inteligencije, predvidio je kako bi strojna inteligencija potkraj prve polovine ovoga stoljeća mogla doseći ljudsku. Njegova predviđanja utemeljena su na teoriji eksponencijalnoga porasta proračunskih mogućnosti strojeva. Naime, dosadašnja saznanja o ljudskome mozgu za polazište su imala „računanje“ te je dodatna procesorska snaga smatrana poželjnom. Međutim, složenije strukture mozga moderna računala ne mogu simulirati. Kolektivna inteligencija odnosi se na svaki sustav koji koristi ekspertizu grupe, a ne pojedinca koji bi trebao donositi odluke. Kolektivni projekt umjetne inteligencije *Subcultron*<sup>7</sup> pokrenut je s ciljem stvaranja najvećega jata riba-robotu u svijetu. Među novijim projektima u razvoju umjetne inteligencije je *Chatbot*, robot koji je osposobljen za vođenje dijaloga. Za *Google* su dizajnirali sustav sposoban analizirati postojeće razgovore na temelju filmskih dijaloga. Projekt su vodili istraživači Oriol Vinyals i Quoc V. Le. *Chatbot* dijaloge formira temeljem primjera iz „trening dijaloga“<sup>8</sup>.

Tehnika strojnoga učenja govora zamijenila je pravila za izgradnju razgovorne motorike, a pravila ručnoga kodiranja zamijenjena su strojem koji uči. U sustav su ugrađene neurološke mreže koje oponašaju

6 Usp. HANS MORAVEC, *Robot, Mere Machine to Transcendent Mind*, Oxford University Press, Oxford, 1999.

7 Usp. NICK LEIBER, „Swarming Robots That Keep an Eye on Waterways“, *Bloomberg*, <<https://www.bloomberg.com/news/articles/2016-01-28/subcultron-swarming>>, (3. XII. 2017.).

8 „Chatbot“, *Investopedia*, <<https://www.investopedia.com/terms/c/chatbot.asp>>, (6. II. 2018.).



mrežu neurona ljudskoga mozga. Ideja neuronskih mreža odavno je poznata, ali su je velike IT korporacije nedavno aktualizirale. Skora aktualizacija takvih mreža rezultirala je velikim pomacima u istraživanjima. Riječ je o korporacijama kojima su društvene mreže u središtu interesa pa neuronske mreže na računalima raspoznaju osobe i predmete s fotografija na njima objavljenim. Osim raspoznavanja vizualnoga, neuronske mreže identificiraju izgovorene riječi na androidima te *online* pozive prevode s jednoga jezika na drugi.

### 3. Arhitektura semantičkoga weba

Izgradnja semantičkoga *weba* uvjetovana je uspostavom nove razine međuoperabilnosti temeljene na otvorenim standardima koji, osim kvalitetne definicije za sintaktičku formu dokumenta, trebaju pružiti podlogu za sintaksni sadržaj. Semantički *web* zasnovan je na ideji korištenja znanja, principa i tehnologija, koje su u osnovi običnoga *weba*, za novi *web* koji bi bio univerzalan medij za razmjenu podataka, informacija i znanja.<sup>9</sup> Stoga je, sukladno prethodno razvijenim informacijskim tehnologijama, semantički *web* dizajniran slojevitom arhitekturom. Na mrežne protokole kao postojeću mrežnu infrastrukturu dograđeni su: sintaksa podataka (*XML/XML Schema*), model podataka semantičkoga *weba* (*RDF*), definiranje *RDF* rječnika (*RDF Schema*), ontologije (*OWL*) te logike i mehanizmi za osiguravanje povjerenja (*Logic, Proof* i *Trust*). Logički sloj omogućuje pisanje pravila, *Proof* sloj ih izvršava i evaluira te zajedno sa *Trust* slojem preispituje uvjerenje u točnost iskazana podatka.

*eXtensible Markup Language – XML, Resource Description Framework – RDF* i *Ontologije* čine arhitekturu semantičkoga *weba*.

Od samoga nastanka *weba* 1992. podatci se na internetskim stranicama prikazuju pomoću skupa oznaka jezika *HTML – Hyper Tekst Markup Language*. Način prikaza podataka određuje se *tagovima* (specifičnim oznakama). Skroman obujam semantičkih osobina sadržaja,

---

9 Usp. CHRISTOPHER D. WALTON, *Agency and the semantic Web*, Oxford University Press, Oxford, 2007.

uglavnom u obliku *metatagova*, ugrađuje se u zaglavlje dokumenta. Novi sustav oznaka *XML* strogo odvajja sadržaj od samoga prikaza. *eXtensible Markup Language* – *XML* tehnologija je razvijena u dosadašnjem razvoju *weba* i čini osnovu za sve tehnologije semantičkoga *weba*. Među zaslužnima za razvoj *XML*-a je Jon Bosak, utemeljitelj radne grupe *W3C* s ciljem skaliranja *SGML*-a (*Standard Generalized Markup Language*) u jednostavniju *XML* formu. Tako je *World Wide Web* konzorcij standardizirao *XML*, a koristi se za kreiranje standardnih oznaka (labela) ili polja unutar internetskih stranica koje se razmjenjuju između različitih računalnih aplikacija i organizacija.<sup>10</sup> Glavni element ovoga mehanizma je jezik za definiranje sheme dokumenata – *XML Schema Definition Language (XSDL)*.

Kako bi dokument mogao biti konverziran u *XML* format, treba ispuniti sljedeće uvjete: treba koristiti *DTD* (ili *XML* shemu), sve vrijednosti atributa trebaju biti navedene unutar navodnika, svi *tagovi* trebaju imati otvarajući i zatvarajući element, osim ako konkretni *tag* nije prazan element, ako je *tag* samostojan – treba posjedovati zatvarajući simbol prije kraja *taga*, svi *tagovi* trebaju biti pravilno ugniježđeni, svaki *XML* dokument bez *DTD*-a treba posjedovati attribute tipa *CDATA*, određeni se elementi trebaju referencirati na drugi način kako ne bi razbili strukturu dokumenta te može postojati samo jedan izvorišni element.<sup>11</sup>

*XML* je dizajniran kao univerzalni *markup* (opisni jezik) koji opisuje i stvara strukturu dokumenta neovisno o njegovu krajnjem prikazu. Njegova karakteristika je u mogućnosti prikaza istih informacija na različite načine, s formalnim ili suštinskim razlikama. I kod dogovorenih struktura podataka aplikacije se mogu susretati s dokumentima koji koriste različite verzije strukture, pa *XML* putem svojih mehanizama eksplicitno specificira strukturu unutar dokumenta. *XML* dokument sastoji se od niza *tagova* unutar izvorišnoga, s tim da svaki od *tagova* (ugniježđenih elemenata) može imati proizvoljan broj svojstava. *XML* dokument može predstaviti raznolike strukture, i to na raznolike načine. Izgled

10 Usp. MARIJA PUTICA, *Informatika*, Filozofski fakultet Sveučilišta u Mostaru, Mostar, 2018., str. 148.

11 Usp. *w3schools.com*, [http://www.w3schools.com/xml\\_whatIs.asp#top](http://www.w3schools.com/xml_whatIs.asp#top) (5. I. 2018.).

dokumenta s *.xml* ekstenzijom možemo opisati kao etiketirano stablo, pri čemu je svaki *tag* u vezi s nekim imenovanim čvorom u datotečnome modelu, a svaki ugniježđeni *tag* predstavlja nasljednike toga *taga*<sup>12</sup>. Pogodan je za definiranje strukture dokumenta, ali ne i njegova značenja. Najčešće se koristi za izradu novih specijaliziranih jezika za označavanje, kao semantički *markup* jezik za internetske stranice, kao jedinstveni format za izmjenu podataka i za odvajanje podataka od prezentacije.<sup>13</sup> *XML* je kao format isključivo namijenjen definiranju strukture nekoga resursa, ali ne i njegova značenja. Nemogućnost podržavanja bilo kakve interpretacije podataka sadržanih u dokumentu njegov je glavni nedostatak.

*XML Schema* mehanizam je za definiranje strukture *XML* dokumenta pa predstavlja gramatiku za *XML*. Ona određuje koje elemente smije sadržavati *XML* dokument te redoslijed i broj tih elemenata. Program može odrediti formalnu ispravnost dokumenta zahvaljujući njoj. Omogućuje i postavljanje ograničenja po pitanju tipova dokumenata koji se mogu koristiti u nekome kontekstu i njihove kardinalnosti u okviru nekoga drugog dokumenta. To je nova specifikacija koja je zamijenila *DTD (Document Type Definition)* jer *DTD* nema mogućnost definiranja *XML* struktura te nije moguće definirati tipove podataka unutar *tagova* ili atributa.

*Resource Description Framework – RDF* je u svojim počecima predstavljao jedinstven model za predstavljanje informacija o internetskim resursima. Te informacije uključuju podatke o nastanku resursa, autoru i ostale podatke upisane u *metatagovima HTML* stranica. Kasnije je pojam *web resurs* proširen na sve što se na *webu* može identificirati, bez mogućnosti dohvaćanja. Kada se razvio *XML* kao standard za serijsku reprezentaciju i predstavljanje informacija, prihvaćen je i za predstavljanje i razmjenu *RDF* opisa. *RDF* opisuje semantičke veze među elektroničkim izvorima. Definira se kao sintaksno neovisan, apstraktni model koji određuje standard o metapodacima koji služe za opis resursa na *webu*.

---

12 Usp. *isto*.

13 Usp. „XML and Semantic Transparency”, *Cover Pages*, <<http://xml.coverpages.org/xmlAndSemantics.html>>, (4. II. 2018.).

Razvio ga je *W<sub>3</sub>C* kao osnovu semantičkoga *weba* koji služi kao podloga za razvoj ostalih jezika. Namijenjena mu je uloga u semantičkome *webu* kakvu *HTML* ima u klasičnome *webu*, s razlikom što je riječ o proizvoljno odabranim podacima, a ne međusobno povezanim dokumentima koji zajedno čine zatvoren sustav. To nije programski jezik nego model koji opisuje semantičke veze među elektroničkim izvorima, odnosno on služi za modeliranje metapodataka o resursima na *webu*. Sačinjen je kao niz specifikacija za modeliranje metapodataka, pa je u *W<sub>3</sub>C-u* razvijen kao opća metoda za modeliranje informacija u različitim sintaktičkim formatima. *DRF*-ov model metapodataka izražen je u obliku izraza subjekt-predikat-objekt, koji se u terminologiji *RDF*-a naziva *trojke*. Pri tome subjekt označava izvor, predikat svojstva i izgled izvora te izražava odnos između subjekta i objekta. Navedeni sastavni dijelovi identificiraju se pomoću *URI*-ja (*Uniform Resource Identifier*). Mehanizam za opisivanje izvora glavna je komponenta semantičkoga *weba*, zbog čega je mogućnost *RDF*-a u modeliranju disparatnih, apstraktnih koncepata uvelike ušla i u aplikacije za upravljanje znanjem.<sup>14</sup> Svrha *RDF*-a je omogućiti mehanizmima za interpretaciju informacija da izvore opišu na način na koji ih određeni programi mogu razumjeti, kako bi im pristupili i koristili se informacijama kojima inače ne bi mogli. *RDF* se bavi isključivo modelom podataka i ne govori ništa o jeziku u kojem će se oni čuvati i razmjenjivati, za razliku od *XML*-a koji je jezik za predstavljanje podataka.

Postoji više *RDF* sintaksi, a najvažnije među njima su:

1. Sintaksa uređenih *trojki* kao najjednostavniji način za tekstualno predstavljanje *RDF* podataka, koji leži u osnovi druga dva *RDF* jezika.
2. *RDF/XML* je *XML* sintaksa za *RDF* podatke i definira stroga pravila (shemu) predstavljanja podataka *RDF* grafa.

14 Usp. IVANA OGRIZEK BIŠKUPIĆ – MIHAELA BANEK ZORICA, *Web tehnologije*, Visoka škola za poslovanje i upravljanje s pravom javnosti Baltazar Adam Krčelić, Zaprešić, 2014., str. 73.

3.  $N_3$  je proširenje sintakse uređenih *trojki* kako bi bila pogodnija za praktičnu uporabu pri čemu je *RDF/XML* pogodniji za programsku obradu, dok je  $N_3$  prilagođen korisniku.<sup>15</sup>

Za pohranu *RDF trojki* razvijene su posebne baze podataka nazvane *triplestore*. Kao i u relacijskim bazama podataka, i u *triplestore* baze podataka pohranjuju se podatci do kojih se može doći i korištenjem određenih upita. Za razliku od njih, *triplestore* je baza podataka optimizirana za pohranu i povrat kratkih izjava (*RDF trojke*: subjekt-predikat-objekt). Ovi sastavni dijelovi zajednički se nazivaju *Resource*, a identificiraju se pomoću *URI*-ja (*Uniform Resource Identifier*).<sup>16</sup> Funkcioniraju na način da je subjekt (resurs) povezan s drugim resursom (objektom) preko veze određene trećim resursom (predikatom). Definiranje *Propertiesa* za odgovarajuće domene i njihove semantike zahtijeva dodatne mehanizme koji se ostvaruju u vidu *RDF Sheme*, jednostavna modela kojim se definiraju tipovi podataka za *RDF*.

Za modeliranje domena koriste se osnovni primitivi:<sup>17</sup> *Class* i *SubClassOf* iskazi za definiranje hijerarhija osobina, *Domain* i *Range* iskazi za ograničavanje mogućih kombinacija osobina i klasa te *Type* iskaz za definiranje resursa kao instance određene klase. Uporabom *RDF Shema* interpretacijske mogućnosti *RDF*-a nisu u dovoljnoj mjeri proširene pa se traže dodatna rješenja. Tako su za razvijanje novih primitiva semantici *RDF Shema* razvijeni *OIL (Ontology Inference Layer)* i *DAML (DARPA Agent Markup Language)*, jezici zasnovani na ontologijama.

Ontologije i *OWL (Web Ontology Language)* omogućuju semantički bogatiji opis područja koje je u središtu interesa. Omogućuju da se u strojno čitljivome obliku predstavi znatno potpunija slika, a sačinjavaju je rječnik i određena pravila te ograničenja u uporabi termina ovoga rječnika. Sukladno izvornomu filozofskom značenju znanosti o biću, o onome što postoji, učenju o općim fundamentalnim i konstruktivnim

---

15 Usp. LJ. MILIVOJEVIĆ, *n. dj.*, (3. II. 2018.).

16 Usp. AARON SWARTZ, „The Semantic Web In Breadth”, *Logicerror.com*, <[www.logicerror.com/semanticWeb-long](http://www.logicerror.com/semanticWeb-long)>, (3. II. 2018.).

17 Usp. „The Semantic Web: An Introduction”, *Infomesh*, [www.infomesh.net/2001/swintro](http://www.infomesh.net/2001/swintro) (2. II. 2018.).

određenjima bitka u informatici – ontologija označava formalno definiran sustav pojmova i/ili koncepta i relacija između tih pojmova. *Web*-inženjeri služe se ontologijama poglavito pri registriranju relacija ili karakteristika proučavanih resursa ili objekata. Kao osnovni razlozi rastućega interesa za ontologiju u informacijskim znanostima navodi se sve veća količina informacija koju treba obraditi, pristup relativno besplatnu računalnom prostoru te rastuće zanimanje unutar organizacija za povezivanjem eksternih i internih informacija.<sup>18</sup> Međusobno se razlikuju prema širini pristupa, razini formalnosti, specifičnosti i s obzirom na namjenu. Grupiraju se prema usmjerenju na znanstvena područja ili na znanje u najširem smislu. Kako bi ontologija imala svrhu tijekom svoga stvaranja, nužno je imati u vidu: područje koje će ontologija pokrivati, za što će se koristiti, koje vrste pitanja bi mogle biti postavljene sustavu koji ih održava, tko će se njome koristiti i tko će ju održavati.<sup>19</sup>

Ontologije predstavljaju najvažniju komponentu semantičkoga *weba* i jednu od najvažnijih komponenti u informativno-poslovnim sustavima jer sadržavaju opis programa, koncepata i relacija koje mogu postojati unutar neke zadane hijerarhije u sustavu pojmova.<sup>20</sup> *Ontology Definition Language* (OWL) u semantičkome *webu* definira ontologije. Kako bi ontologije mogle biti strojno čitljive, bilo je potrebno postaviti neke standarde i pronaći programska rješenja koja pružaju sustave kodiranja pogodne za strojnu obradu. Takva programska rješenja obično se nazivaju jezicima za prikaz ontologija i dijele se u tri kategorije: jezike temeljene na logici prvoga reda (npr. *Knowledge interchange format*), jezike temeljene na logici sustava okvira (npr. *Knowledge Language One* i *Frame Representational Language*) i mrežne jezike (*HTML*, *XML*, *RDF*...).<sup>21</sup> Kako bi se principi ontologije primijenili na instituciju interneta, koriste se jezici kao što su *OIL* i *DAML*. *OIL* služi opisu ontologije, a najčešće polazi od *RDF Schema*. Utemeljen je na *XML*-u s utjecajem *RDF* karakteristika (*RDF Schema* proširena je dodatnim znakovima i

18 Usp. JENNIFER ROWLEY – RICHARD HARTLEY, *Organizing knowledge: An introduction to managing access to information*, 4. izdanje, Ashgate, Burlington, 2016.

19 Usp. isto.

20 MICHAEL K. SMITH i dr. (ur.) *OWL Web Ontology Language-Guide*, 2004.

21 I. OGRIZEK BIŠKUPIC – M. BANEK ZORICA, *n. dj.*, str. 78.

simbolima) pa *OIL* ontologije predstavljaju i *RDF* dokumente. *OIL* je organiziran kao serija slojeva koji se zasebno razvijaju po uzoru na arhitekturu semantičkoga *weba*. Pri tome se podrazumijeva da viši slojevi u potpunosti podržavaju niže. U podvrste *OIL*-a spadaju: *CORE OIL* i *Standard OIL Laxers*, *Instance OIL* i *Heavy OIL*. *DAML* je od svih ontologijskih mrežnih jezika napravio najveći pomak. Razvijen je kao proširenje *XML*-a i *RDF*-a. Usklađen je s *RDF Schemom*. Za opis značenja sadržava preciznu semantiku.

Osim navedenih jezika, tehnologije semantičkoga *weba* čine i upitni jezici (*SPARQL*), mehanizmi logičkoga zaključivanja i mehanizmi za osiguravanje povjerenja.

## Zaključak

Iako se pod njim podrazumijeva ogroman informacijski sustav, semantički *web* ponaša se sasvim suprotno. Naime, informacijski sustav počiva na pretpostavci o zatvorenome svijetu jer se informacija koja nije unesena smatra neistinitom. S druge strane, semantički *web* počiva na pretpostavci o otvorenome svijetu jer, ako ne može doći do neke informacije – korisnik uviđa kako o njezinoj istinitosti ne može donositi zaključke. Njihove razlike proistječu iz ideje decentralizacije semantičkoga *weba* i ciljeva koji su uvjetovali odabir tehnologija koje se u njemu koriste.

Razvoj semantičkoga *weba* odvija se na dva plana. Pod pokroviteljstvom *W3C* radne grupe razvijaju se novi standardi, a istodobno se vrši provedba semantičkoga *weba* u postojeći. Bogatstvo današnjega *weba*, osim obilja dostupnih informacija kao pozitivne strane, ima i svoje negativne strane. One se ponajviše ogledaju u posljedicama proisteklim iz izostanka standardizacije i zajedničkoga rječnika, što rezultira heterogenošću koja priječi razmjenu informacija i komunikacija. Internet obiluje bogatstvom neselektirana sadržaja te ogromne količine znanja ostaju skrivene. Cilj semantičkoga *weba* je omogućiti i strojevima da aktivno sudjeluju u obradi i publiciranju sadržaja. Stoga su tehnologije semantičkoga *weba* usmjerene na pomaganje strojevima u razumijevanju

informacija, što će rezultirati kvalitetnijim pretraživanjem, podatkovnom integracijom te automatiziranim obavljanjem tih zadataka. Semantički *web* automatiziranim uslugama pomaže povezivanju informacija iz različitih izvora. Tu se dodavanjem metapodataka svaki resurs može asocirati sa semantički bogatom opisujućom informacijom. Kako pretraživanje kroz uporabu značenja zahtijeva semantičku i sintaktičku interoperabilnost predmetnoga rječnika, nužno je predmete opisati i pozicijskom logikom. Stoga su tradicionalni sustavi organizacije znanja, poglavito tezaursi i klasifikacijski sustavi, prepoznati kao najpogodniji. Razvoj semantičkoga *weba* ovisi o infrastrukturi metapodataka kako bi ih strojevi mogli povezivati i razmjenjivati te je izrada standardiziranih formata za razmjenu rječnika aktivnost koja je u središtu pozornosti.

Na temelje postojećega *weba* nadograđuju se nove tehnologije. Tako svaka tehnologija koja čini semantički *web* koristi prethodno razvijene standarde. Ti su standardi dio ostvarivanja semantičkoga *weba*, dok drugi dio čine ontologije. Razvoj semantičkoga *weba* je, osim vlastitoga napretka, bitan i po tome što je potaknuo i razvoj brojnih drugih disciplina softverskoga inženjeringa, poglavito iz područja konceptualnoga oblikovanja i predstavljanja znanja. Pred razvojem semantičkoga *weba* stoji otvoren problem izrade rječnika pa se rješenje, kao i u brojnim prethodnim slučajevima, vidi u aktivnu angažmanu *open source* zajednica. Razvoj klasifikacijskih shema i tezaurusa uvjet je za dostupnost i razmjenjivanje putem *weba* te je nužan razvoj s njima usklađenih rječnika i mrežnih ontologijskih rječnika. Stoga svaki pojam iz rječnika treba posjedovati jedinstven identifikator kojim bi se povezivalo, semantički tumačilo i višestruko koristilo u različitim shemama metapodataka. Najveći izazov koji stoji pred novom internetskom generacijom je u obradi znanja. Razlog je to zbog kojega se u njegovu provedbu ulazi malim koracima te se u svoje informacijske sustave unutar vlastitih mreža korporacije ugrađuje semantika.



## Literatura

- BERNERS-LEE, TIM – HENDLER, JAMES – LASSILA, ORA, „The semantic web”, *Scientific American*, svibanj 2001., <<http://sciam.com/2001/0501issue/0501berners-lee.html>>, (12. VII. 2017.).
- „Chatbot”, *Investopedia*, <<https://www.investopedia.com/terms/c/chatbot.asp>>, (6. II. 2018.).
- HEYLIGHEN, FRANCIS, *Collective Intelligence and its Implementation on the Web: algorithms to develop a collective mental map*, Leo Apostel, Free University of Brussels, 1999., <[cleamc11.vub.ac.be/Papers/CollectiveWebIntelligence.pdf](http://cleamc11.vub.ac.be/Papers/CollectiveWebIntelligence.pdf)>, (6. VII. 2017.).
- LEIBER, NICK, „Swarming Robots That Keep an Eye on Waterways”, *Bloomberg*, <<https://www.bloomberg.com/news/articles/2016-01-28/subcultron-swarming>>, (3. XII. 2017.).
- MALONE, THOMAS W. – LAUBACHER, ROBERT – DELLAROCAS, CHRYSANTHOS, *Harnessing Crowds: Mapping the Genome of Collective Intelligence*, MIT, Cambridge, 2009.
- MILIVOJEVIĆ, LJUBIŠA, „Semantički Web – ideja i tehnologije”, *Institut informacijskih znanosti Maribor*, <[http://home.izum.si/cobiss/oz/2009\\_1'2/html/clanek\\_02.html](http://home.izum.si/cobiss/oz/2009_1'2/html/clanek_02.html)>.
- MORAVEC, HANS, *Robot, Mere Machine to Transcendent Mind*, Oxford University Press, Oxford, 1999.
- OGRIZEK BIŠKUPIĆ, IVANA – BANEK ZORICA, MIHAELA, *Web tehnologije*, Visoka škola za poslovanje i upravljanje s pravom javnosti Baltazar Adam Krčelić, Zaporešič, 2014.
- PUTICA, MARIJA, *Informatika*, Filozofski fakultet Sveučilišta u Mostaru, Mostar, 2018.
- ROWLEY, JENNIFER – HARTLEY, RICHARD, *Organizing knowledge: An introduction to managing access to information*, 4. izdanje, Ashgate, Burlington, 2016.
- SEELEY, THOMAS D., *Honeybee Democracy*, Princeton University Press, Princeton, 2010.

- SMITH, MICHAEL K. i dr. (ur.) *OWL Web Ontology Language-Guide*, 2004.
- SWARTZ, AARON, „The Semantic Web In Breadth”, *Logicerror.com*, <[www.logicerror.com/semanticWeb-long](http://www.logicerror.com/semanticWeb-long)>, (3. II. 2018.).
- „The Semantic Web: An Introduction”, *Infomesh*, [www.infomesh.net/2001/swintro](http://www.infomesh.net/2001/swintro) (2. II. 2018.).
- *w3schools.com*, [http://www.w3schools.com/xml\\_whatIs.asp#top](http://www.w3schools.com/xml_whatIs.asp#top) (5. I. 2018.).
- WALTON, CHRISTOPHER D., *Agency and the semantic Web*, Oxford University Press, Oxford, 2007.
- „XML and Semantic Transparency”, *Cover Pages*, <<http://xml.coverpages.org/xmlAndSemantics.html>>, (4. II. 2018.).