

SEMANTIČKI WEB – ZA I PROTIV

Mladen Stanojević, Sanja Vraneš, Institut Mihajlo Pupin

Sadržaj – *Semantički Web predstavlja nadogradnju postojećeg interneta koja bi trebalo da obezbedi lakše pronalaženje potrebnih informacija, kao i pružanje inteligentnih servisa. Kada se govori o semantičkom web-u, tada se misli na tehnike za predstavljanje znanja, semantičke web servise i pravila, koji se koriste da bi se klasičnom internetu dodao određeni nivo inteligencije i korisnicima obezbedio viši nivo usluga. U ovom radu je dat kratak pregled osnovnih komponenti semantičkog web-a uz komentar postojećih rešenja, kao i mogući pravci daljeg razvoja.*

1. UVOD

Internet je u poslednjih petnaest godina doživio eksplozivni razvoj i postao jedan od najpopularnijih medija za razmenu informacija. Međutim, upravo je taj eksplozivni razvoj doveo do ograničenja u mogućnosti korišćenja, jer je običnom čoveku postalo veoma teško da u ogromnoj količini informacija pronađe onu koja baš njemu treba. Web pretraživači su nastali kao odgovor na taj problem, međutim on još uvek nije rešen na zadovoljavajući način. I sada se često dešava na neki upit dobijete milione pogodaka, pa je i dalje veoma teško odrediti da li neki od njih možda sadrže informaciju koja vas interesuje.

Istraživači koji su se bavili razvojem interneta su došli do zaključka da je jedini način na koji se ovaj problem može rešiti, da se računari oplemene alatom koji bi na inteligentniji način, uz otkrivanje i prepoznavanje semantike u podacima, pomogli čoveku u pretraživanju interneta. Međutim, web stranice u svojoj sadašnjoj formi nisu čitljive za računare, tako da je bilo potrebno koristiti neki drugi način za predstavljanje znanja koji bi bio razumljiv računarima. Semantički web bi trebalo da omogući predstavljanje podataka na internetu tako da računari, pre svega inteligentni agenti, mogu da ih koriste ne samo za prikazivanje, nego i za njihovu obradu, integraciju i korišćenje u različitim aplikacijama.

Semantički web je pokrenut na inicijativu World Wide Web konzorcijuma [1] (W3C) u koji su uključeni mnogi istraživači i industrijski partneri. Rodonačelnik ideje semantičkog web-a je Tim Berners-Lee, koji je poznat kao autor WWW-a, URI-a, HTTP-a i HTML-a. Tim Berners-Lee je semantički web definisao na sledeći način [2]: „Semantički web predstavlja proširenje postojećeg web-a, gde će informacije imati precizno definisano značenje, što će omogućiti bolju saradnju između ljudi i računara“

U okviru W3C konzorcijuma je već dosta urađeno na unapređenju, proširenju i standardizaciji sistema i mnogi jezici i alati su do sada već razvijeni. Međutim, tehnologije semantičkog web-a su još uvek u samom povelju i, iako mu mnogi proriču svetlu budućnost, još uvek ne postoji saglasnost u vezi daljeg pravca razvoja i karakteristika semantičkog web-a.

Da bi programi koji realizuju različite servise na internetu, gde se najčešće koristi tehnologija inteligentnih agenata, mogli u realnom vremenu da pronađu sve relevantne informacije, integrišu ih i obrade na konzistentan način, bilo bi izuzetno korisno da umesto sirovog sadržaja, najčešće u

prirodnom jeziku, barataju sa preprocesiranim informacijama, sa eksplicitno naznačenim značenjem (semantikom).

Dve važne tehnologije za razvoj semantičkog web-a su već u upotrebi: eXtensible Markup Language (XML) i Resource Description Framework (RDF). XML omogućava svakome da kreira skrivene labele za mapiranje web stranica ili delova teksta. Skripte ili programi mogu koristiti ove labele, ali programer mora da zna kako se one koriste. XML omogućava korisnicima da dodaju proizvoljne strukture u svoje dokumente, ali ništa ne govori o značenju tih struktura.

RDF omogućava predstavljanje značenja u vidu tripleta, pri čemu svaki triplet podseća na subjekat, predikat i objekat proste rečenice. Ovi tripleti se mogu napisati korišćenjem XML labela. Ispostavilo se da ovakva struktura pruža mogućnost za opisivanje velike većine podataka koji se računarski obrađuju. Subjekt i objekat su predstavljeni jedinstvenim identifikatorima resursa (Universal Resource Identifier – URI), na isti način na koji se koriste u linkovima na web stranicama. Predikati su takođe predstavljeni pomoću URI-ja, što svakome omogućava da definiše novi koncept, novi glagol, jednostavnim definisanjem odgovarajućeg URI-ja negde na web-u.

Međutim XML i RDF imaju ograničene mogućnosti u predstavljanju semantike, tako da se ubrzo počelo razmišljati o nekim ekspresivnijim tehnikama za predstavljanje znanja, kao što su ontologije, semantički web servisi i pravila za zaključivanje. Kao što je to već često do sada bio slučaj u okviru računarskih nauka, da se iste tehnike otkrivaju u okviru različitih oblasti samo im se daju nova imena, tako su i istraživači u oblasti semantičkog web-a ponovo otkrivali okvire i relacije (ontologije) i produkciona pravila koji su već davno korišćeni u veštačkoj inteligenciji i ekspertnim sistemima [3].

U nastavku rada će više pažnje biti posvećeno ontologijama, semantičkim web servisima i pravilima.

2. ONTOLOGIJE

U filozofiji, ontologija predstavlja teoriju o prirodi postojanja, o tipovima stvari koje postoje. Istraživači u oblasti veštačke inteligencije i kasnije interneta su preuzeli ovaj termin i za njih je ontologija dokument ili datoteka koja formalno definiše relacije između termina. Tipična vrsta ontologija koja se koristi na web-u ima taksonomiju i skup pravila za zaključivanje. Jedina razlika između ontologija u veštačkoj inteligenciji i semantičkom web-u, je da se definicije semantičkih kategorija i relacija nalaze na istom računaru na kom se nalaze i objekti iz domena aplikacije u oblasti veštačke inteligencije, dok se u slučaju semantičkog web-a definicije mogu nalaziti bilo gde na internetu.

Taksonomija definiše klase objekata i relacije između njih. Na primer, slavuj se može definisati kao ptica pevačica, dok se mogućnost lepog pevanja može ograničiti samo na ptice pevačice, itd. Klase, potklase i relacije između entiteta su veoma korisni u radu sa web-om. Pomoću njih se može opisati veliki broj relacija između entiteta tako što se klasama pridružuju osobine i dozvoljava potklasama da naslede ove osobine.

Ontologije mogu poboljšati funkcionisanje web-a na mnoge načine. One se mogu koristiti na jednostavan način, da bi poboljšale preciznost pretraživanja web-a, tako što bi program za pretraživanje tražio samo one stranice koje se odnose na jasno definisani koncept, umesto sve stranice koje sadrže neprecizno definisane ključne reči. Naprednije aplikacije mogu koristiti ontologije da bi povezale informacije na nekoj web stranici sa odgovarajućom strukturom znanja i pravilima za zaključivanje.

U okviru semantičkog web-a se koriste razni ontološki i šematski jezici kao što su XOL, SHOE, OML, RDFS, DAML+OIL [4] i OWL [5]. XOL je jezik za razmenu ontologija baziran na XML-u, koji je prvobitno bio projektovan za razmenu ontologija u bio-informatici, ali se takođe može koristiti i za ontologije u drugim domenima. XOL je semantički jezik koji se koristi u sistemima baziranim na objektno-orijentisanom predstavljanju znanja, pri čemu se koristi XML sintaksa. SHOE (Simple HTML Ontology Extensions) predstavlja proširenje HTML-a tako što omogućava mapiranje web stranica pomoću informacija koje računari mogu lako da čitaju. OML (Ontology Markup Language) se bazira na deskriptivnoj logici i konceptualnim grafovima i omogućava predstavljanje koncepata organizovanih u taksonomije, relacije i logičke aksiome prvog reda. RDFS (RDF Schemas) predstavlja proširenje RDF-a, gde je jezik za predstavljanje deklarativnog znanja razvijen pod uticajem ideja iz veštačke inteligencije (semantičke mreže, okviri, predikatska logika), kao i jezika za definisanje baza podataka i modela podataka baziranih na grafovima. DAML+OIL je jezik za semantičko mapiranje web stranica baziran na ranijim W3C standardima, kao što su RDF i RDF Schema i proširuje ove jezike korišćenjem bogatijih primitiva za modelovanje. DAML+OIL koristi primitivne za modelovanje koje se obično koriste u jezicima baziranim na okvirima. Web Ontology Language (OWL) je jezik za semantičko mapiranje koji se koristi za objavljivanje i deljenje ontologija na web-u. OWL je razvijen kao proširenje RDF-a sa bogatijim rečnikom i izveden je iz jezika DAML+OIL.

Svi ovi ontološki i šematski jezici predstavljaju vrstu meta jezika za eksplicitno predstavljanje semantike i čine platformu za komunikaciju između inženjera znanja i programera. Posao inženjera znanja je da analizira domen aplikacije i definiše sve semantičke kategorije i relacije između njih od interesa za razvoj te aplikacije. Koristeći isti formalizam za predstavljanje znanja, programeri su u stanju da razvijaju aplikacije semantičkog web-a.

Sam proces predstavljanja znanja podrazumeva definisanje neke vrste meta znanja u smislu definicija osnovnih semantičkih kategorija i relacija, a zatim i objekata – instanci tih semantičkih kategorija kao i relacija između njih. Definisanje meta-znanja se ne može izbeći i predstavlja lakši deo posla. Teži deo posla je prevođenje semantike koja je definisana u prirodnom jeziku na jedan od formalizama za predstavljanje znanja. Ukoliko se ima u vidu broj postojećih web stranica, može se naslutiti koliki je obim tog posla. Upravo zbog toga je i teško zamisliti da će kao konačno rešenje problema predstavljanja znanja u semantičkom web-u biti odabran neki od postojećih formalizama.

Na sreću, taj teži deo posla može biti izbegnut ukoliko se meta znanje o semantičkim kategorijama i relacijama iskoristi za automatsko prepoznavanje objekata u okviru teksta u prirodnom jeziku. Da bi se to omogućilo, potrebne su nove tehnike za predstavljanje semantičkog znanja koje će, umesto zamornog procesa prevođenja domenskih objekata od strane inženjera znanja, obezbeđivati automatsko prepoznavanje

domenskih objekata iz teksta u prirodnom jeziku. Jedan od prvih koraka u tom pravcu predstavlja hijerarhijska semantička forma (HSF) [6]. Korišćenjem tih novih tehnika za predstavljanje znanja bi se izbegao zamoran i obiman posao prevođenja postojećih web stranica, otklonila osnovna prepreka za masovniju primenu semantičkog web-a i pospešio razvoj aplikacija baziranih na semantičkom web-u.

3. SEMANTIČKI WEB SERVISI

Internet je dugo vremena korišćen isključivo kao medijum za prezentaciju statičkih podataka. Međutim, u skorije vreme se došlo na ideju da bi internet mogao da se iskoristi i kao osnova za pružanje određenih usluga, pa su tako nastali web servisi kao osnova za razvoj složenih distribuiranih sistema na web-u. U međuvremenu je veliki napor uloženo radi definisanja i razvoja standarda za masovnu primenu web servisa. Tako je jezik za opisivanje web servisa (Web Services Description Language – WSDL) već postao jedan od osnovnih elemenata u razvoju tehnologija web servisa kao rezultat razvoja i standardizacije u okviru W3C Web Services Description Working Group. WSDL u osnovi omogućava definisanje sintakse ulaznih i izlaznih poruka kao i drugih detalja vezanih za pozivanje servisa. Međutim, WSDL ne podržava definisanje načina izvršavanja osnovnih servisa. Za definisanje toka izvršavanja web servisa se najčešće koristi Business Process Execution Language for Web Services (BPEL4WS) razvijen u okviru OASIS grupe koja obuhvata najveće i najpoznatije proizvođače računarske opreme i softvera (IBM, Microsoft, Hewlett-Packard, Sun, SAP, ...). Za registraciju web servisa, za njihovo oglašavanje i pronalaženje se najčešće koristi Universal Description, Discovery and Integration (UDDI).

Međutim, u isto vreme sazreva potreba za bogatijom semantičkom specifikacijom web servisa da bi se omogućilo potpunije i fleksibilnije automatsko korišćenje servisa, podržao razvoj snažnijih alata i metoda i podržalo semantičko zaključivanje. Bogatiji jezici za predstavljanje znanja omogućavaju razumljiviju specifikaciju brojnih aspekata servisa i obezbeđuju bolju osnovu za širok spektar aktivnosti u životnom veku web servisa. Na primer, bogatija semantika bi omogućila veći stepen automatizacije procesa odabiranja i pozivanja servisa, automatsko prevođenje poruka koje međusobno razmenjuju heterogeni servisi, automatsku ili poluautomatsku kompoziciju servisa i jasniji pristup u prihvatanju rezultata servisa i obradi grešaka. Pored toga, bogatija semantika bi bila od pomoći u obezbeđivanju potpunije automatizacije aktivnosti kao što su verifikacija, simulacija, konfiguracija, obezbeđivanje ulaza, ugovaranje i pregovaranje između servisa. Da bi se odgovorilo toj potrebi istraživači razvijaju jezike, arhitekture, metode i alate što bi sve zajedno moglo da se podvede pod pojam infrastructure semantičkih web servisa.

OWL-S [7] (i ranije DAML-S) predstavlja OWL ontologiju sa tri povezane podontologije koje se nazivaju profil, model procesa i informacije za povezivanje (grounding). Profil se koristi da bi opisao “šta servis radi”, model procesa opisuje “kako process radi” i koristi se za oglašavanje, kreiranje zahteva za servisom, omogućava pozivanje, izvršavanje, kompoziciju, prihvatanje rezultata i obradu grešaka, dok se informacije za povezivanje koriste za mapiranje elemenata modela procesa na detaljnu specifikaciju formata poruka, protokola itd. (što se obično izražava korišćenjem WSDL-a).

WSDL omogućava opisivanje operacija kao osnovnih elemenata web servisa. Operacije čine organizacionu

strukturu oko koje se definiše sintaksa i modeli ulaznih/izlaznih poruka. OWL-S koristi analogne, ali više apstraktne elemente koji se nazivaju atomskim procesima koji su određeni ulazima, izlazima, preduslovima i efektima (inputs, outputs, preconditions, and effects – IOPEs).

Proces pronalazanja informacija korišćenjem semantičkih web servisa uz korišćenje standardnih tehnika za predstavljanje znanja se može približno opisati na sledeći način (Sl. 1):

- Korisnik postavlja upit u prirodnom jeziku
- Parser prirodnog jezika parsira upit i prevodi ga na semantičke kategorije korišćenjem nekog od ontoloških jezika.
- Odgovarajući Web servis se pronalazi na osnovu profila (opisa servisa).
- Korišćenjem modela procesa (opisa ponašanja) vrši se dekompozicija web servisa i njegovo izvršavanje.
- Informacije za povezivanje se koriste da bi se pronašle lokacije atomskih web servisa koji sačinjavaju telo složenog web servisa, ti atomski servisi se zatim izvršavaju i prikupljaju izlazni rezultati u zadatoj formi.
- Na kraju se ti rezultati dalje obrađuju da bi se generisao odgovor u prirodnom jeziku.

Ako analiziramo proces pronalazanja informacija korišćenjem semantičkih web servisa, kao i postojeća rešenja, možemo zaključiti da se različite tehnike za predstavljanje znanja mogu koristiti u različitim fazama obrade.

Parser prirodnog jezika koristi: gramatička pravila za parsiranje upita u prirodnom jeziku, semantičke kategorije za opis elemenata upita.

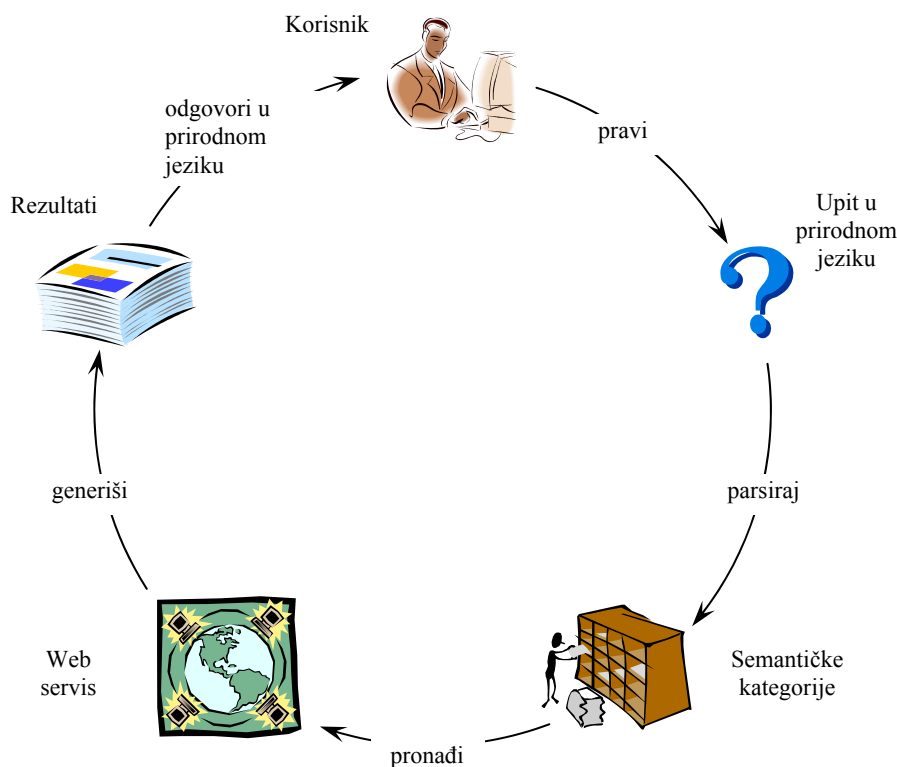
Da bi se odgovarajući web servis pronašao i izvršio, koriste se sledeći podaci: profil servisa - da bi pronašao se potrebni web servis (XOL, OME, RDFS, DAML+OIL, OWL), model procesa (tok izvršavanja) web servisa (DAML-S, OWL-s).

Atomski web servisi se izvršavaju pozivanjem udaljenih procedura (Remote Procedures). U procesu generisanja izlaznih rezultata moraju se predstaviti sledeći podaci: ulazni parametri (korišćeni u pozivima udaljenih procedura), koji moraju biti dostavljeni web servisu u određenoj formi, udaljene procedure (proceduralno znanje), lokalni podaci u okviru udaljenih procedura, izlazni parametri udaljenih procedura, koji moraju biti vraćeni korisniku u zadatom formatu

Korisne informacije, koje u stvari korisnik traži, su najčešće sačuvane u nekoj bazi podataka ili bazi znanja: informacije na nekoj web lokaciji (relaciona baza podataka, XOL, SHOE, DAML+OIL, RDFS, OWL, sistem datoteka)

Na kraju rezultati web servisa se moraju transformisati u odgovor u prirodnom jeziku pomoću generatora prirodnog jezika koji koristi: znanje potrebno za generisanje odgovora u prirodnom jeziku, semantičke kategorije koje sačinjavaju odgovor.

Ovih 11 različitih tipova znanja, neki deklarativni, a neki proceduralni, moraju na neki način biti predstavljeni. U najgorem slučaju će se koristiti 11 različitih tehnika za



Sl 1. Obrada upita u prirodnom jeziku

predstavljanje znanja, a u najboljem – 7 tehnika za predstavljanje:

1. gramatičkih pravila za parser i generator prirodnog jezika
2. internih podataka za parser i generator prirodnog jezika
3. profil servisa
4. model procesa
5. ulaznih i izlaznih parametara udaljenih procedura
6. udaljenih procedura
7. korisnih informacija na web lokacijama

Pored toga što je potrebno uložiti ogroman napor da bi pronašli najpogodnije tehnike za predstavljanje pojedinih tipova znanja i zatim opisali ta znanja upotrebom odabranih tehnika, potrebno je još i prevoditi jedan tip znanja u drugi u toku obrade web servisa.

Nije teško zamisliti koliko je teško definisati sve potrebne interfejse, kako je lako napraviti grešku u toku obrade i koliko mora biti ograničena fleksibilnost sistema koji obavlja tako mnogo transformacija.

Mnogi od ovih problema i ograničenja bi se mogli izbjeći ako bi istu tehniku predstavljanja znanja mogli da koristimo za sve različite tipove podataka koji su gore opisani. Prirodni jezici su idealni kandidati za opisivanje deklarativnog i proceduralnog znanja i potrebno je obezbediti odgovarajuću tehniku za predstavljanje različitih tipova znanja.

4. PRAVILA

Semantika je u okviru semantičkog web-a definisana korišćenjem dva komplementarna koncepta: dodavanjem semantike web sadržajima korišćenjem ontologija i izvršavanjem transformacija i izvođenjem zaključaka nad elementima web-a.

Programska paradigma bazirana na pravilima koristi dva osnovna tipa pravila: produkciona pravila i pravila zamene. Produkciona pravila kao i pravila koja su iz njih izvedena (npr. poslovna pravila ili ograničavajuća pravila), imaju formu „ako su uslovi zadovoljeni onda izvrši neku akciju“ i definisana su za neki skup podataka. Veoma često se koriste u veštačkoj inteligenciji za pravljenje ekspertnih sistema, opisivanje ponašanja robota, rešavanje problema pretraživanja, opisivanje poslovnih pravila, itd.

Pravila zamene imaju formu „zameni upareni entitet nekim drugim“ i prilikom izvršavanja koriste termine ili grafove. Često se koriste u dokazivanju teorema i za algebarske definicije, jer predstavljaju prirodan način za izvođenje operacije jednakosti.

Napori u okviru W3C konzorcijuma su rezultirali usvajanjem standarda za jezik ontologija na web-u, OWL. OWL se bazira na korišćenju deskriptivne logike koja omogućava aksiomatsko opisivanje domena aplikacije kao i zaključivanje nad tim domenom.

U okviru W3C vizije slojevitog semantičkog web-a predviđeno je dodavanje sloja pravila na sloj ontologija. Postoji opšta saglasnost da u aplikacijama semantičkog web-a postoji potreba za pravilima sa dobro definisanom sintaksom i da ona moraju biti dobro integrisana sa slojem ontologija. Iako je oblast razvoja jezika pravila trenutno

veoma živa, postojeća rešenja koja su prilagođena semantičkom web-u su još uvek nedorečena (RuleML [8], Semantic Web Rule Language – SWRL, Datalog, Xcerpt).

5. ZAKLJUČAK

Imajući u vidu ograničenja običnog web-a, jasno je da semantički web, koji bi trebalo da prevaziđe ta ograničenja, ima veliku perspektivu. Međutim, kako se radi o sasvim novoj oblasti, još uvek nije sasvim jasno u kom pravcu će ovaj razvoj zaista teći.

Od svih elemenata semantičkog web-a (predstavljanje znanja, semantički web servisi, pravila), predstavljanje znanja predstavlja najslabiju kariku. Nedostatak postojećih ontoloških jezika se ogleda u potrebi za prevodnjem sadržaja postojećih web stranica korišćenjem definicija semantičkih kategorija i relacija u formu definisanu tim jezicima.

Pravo rešenje za semantički web ne bi trebalo da išta menja u postojećem web-u, već bi trebalo da omogući računarima da razumeju sadržaj postojećih stranica. Da bi se to postiglo, neophodno je razviti odgovarajuće metode za predstavljanje semantike na web-u i zatim razviti meta znanje (definicije semantičkih kategorija i relacija), čijim korišćenjem će aplikacije semantičkog web-a biti u stanju da razumeju sadržaje običnih web stranica.

LITERATURA

- [1] World Wide Web Consortium – W3C, <http://www.w3.org/>.
- [2] T. Berners-Lee, J. Hendler, O. Lassila, “The Semantic Web”, *Scientific American*, Vol. 284, No. 5, 2001.
- [3] S. Vraneš and M. Stanojević, "Prolog/Rex - A Way to Extend Prolog for Better Knowledge Representation", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 6, No. 1, 1994, pp. 22-37.
- [4] DAML+OIL, <http://www.w3.org/TR/daml+oil-reference>.
- [5] OWL Web Ontology Language, <http://www.w3.org/TR/owl-features/>.
- [6] M. Stanojević, S. Vraneš, “Semantic Web with SOUL”, *Proceedings of the IADIS International Conference e-commerce 2004*, Lisbon, Portugal, 2004, pp.123-130.
- [7] OWL-S, <http://www.daml.org/services/owl-s/>.
- [8] The Rule Markup Initiative, <http://www.ruleml.org/>.

Abstract – Semantic Web represents an extension of the existing Web, which should provide easier information retrieval and offer intelligent services. Roughly speaking the main components of Semantic Web are Knowledge Representation techniques, Semantic Web Services and rules. In this paper a brief overview of basic components of Semantic Web has been presented, together with a discussion about existing techniques and hints about possible future developments.

SEMANTIC WEB – PRO ET CONTRA

Mladen Stanojević, Sanja Vraneš